Project Report

On

FORECASTING RUBBER PRICES OF KOTTAYAM USING SARIMA AND HOLT-WINTER'S EXPONENTIAL SMOOTHING METHOD

Submitted

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

in

APPLIED STATISTICS AND DATA ANALYTICS

by

K S KRISHNAKRIPA

(RegNo.SM23AS006)

(2023-2025)

Under the Supervision of

VISMAYA VINCENT

DEPARTMENT OF MATHEMATICS AND STATISTICS

ST. TERESA'S COLLEGE (AUTONOMOUS)

ERNAKULAM, KOCHI – 682011

APRIL 2025

ST. TERESA'S COLLEGE (AUTONOMOUS), ERNAKULAM

THE FEAR OF THE LORD IS THE BEGINNING OF WISDOM

ST. TERESA'S COLLEGE
(AUTONOMOUS)
ESTD 1925

CERTIFICATE

This is to certify that the dissertation entitled, FORECASTING RUBBER PRICES OF KOTTAYAM USING SARIMA AND HOLT-WINTER'S EXPONENTIAL SMOOTHING METHOD is a bonafide record of the work done by K S KRISHNAKRIPA under my guidance as partial fulfillment of the award of the degree of Master of Science in Applied Statistics and Data Analytics at St. Teresa's College (Autonomous), Ernakulam affiliated to Mahatma Gandhi University, Kottayam. No part of this work has been submitted for any other degree elsewhere.

Date:

Place: Ernakulam

Vismaya Vincent

Assistant Professor,

Department of Mathematics and Statistics

St. Teresas's College (Autonomous)

Ernakulam

ST. TERESA'S COLLEGE (AUTONOMOUS) DEPARTMENT OF MATHEMATICS AND STATISTICS ERNAKULAM

Nisha Oommen

Assistant Professor & HOD

Department of Mathematics and Statistics

St. Teresas's College (Autonomous)

Ernakulam

External Examiners

1. Sangeetha Chandran 30.04.2025

2. Anju N.B 30.04.25

DECLARATION

I hereby declare that the work presented in this project is based on the original work done by me under the guidance of Vismaya Vincent, Assistant Professor, Department of Mathematics and Statistics, St. Teresa's College (Autonomous), Ernakulam and has not been included in any other project submitted previously for the award of any degree.

Ernakulam

Date: 30/04/2025

K S KRISHNAKRIPA

SM23AS006

ACKNOWLEDGEMENTS

I take this opportunity to thank everyone who has encouraged and supported me to carry out this project.

I am very grateful to my project guide Ms. Vismaya Vincent for her immense help during the period of work.

In addition, I acknowledge with thanks to the Department for all the valuable support and guidance that has significantly contributed to the successful completion of this project.

I would also like to thank the HOD for her valuable suggestions and critical examinations of the project.

Ernakulam:

Date: 30/04/25

K S KRISHNAKRIPA

SM23AS006

ABSTRACT

In this study, monthly rubber price data of two different grades of rubber RSS1 and RSS3 of Kottayam from January 2010 to December 2023 are analyzed and used for forecasting future rubber price value. Two time series model were deployed to forecast future monthly rubber price data for grades RSS1 and RSS3 of Kottayam. SARIMA and Holt-Winter's Exponential Smoothing were used to forecast the prices from January 2024 – December 2025. Comparison of these models for both grades were done using matrices like RMSE and MAE. Holt-Winter's model was selected as the best model for both grades RSS1 and RSS3.

ST.TERESA'S COLLEGE (AUTONOMOUS)
ERNAKULAM

Certificate of Plagiarism Check for Dissertation

Author Name	K S KRISHNAKRIPA
Course of Study	M.Sc. Applied Statistics & Data Analytics
Name of Guide	Ms. Vismaya Vincent
Department	P.G. Dept of Mathematics & Statistics
Acceptable Maximum Limit	20
Submitted By	library@teresas.ac.in
Paper Title	FORECASTING RUBBER PRICES OF KOTTAYAM USING SARIMA AND HOLT-WINTER'S EXPONENTIAL SMOOTHING METHOD
Similarity	7% AI - 6%
Paper ID	3409028
Total Pages	53
Submission Date	2025-03-17 15:52:58

Signature of Student **Signature of Guide**

Checked By
College Librarian

CONTENTS

1. INTRODUCTION 1

1.1 Natural Rubber Rate 1

1.2 Objectives 2

2. LITERATURE REVIEW 3

3. MATERIALS AND METHODS 5

3.1 Data Collection 5

3.2 Methodology 5

3.3 Tools for Forecasting 5

3.4 Tools for Analysis 5

3.5 Python Programming Language 6

4. EXPLORATORY DATA ANALYSIS AND TIME SERIES 7

4.1 EDA 7

4.1.1 Descriptive Statistics 7

4.1.2 Time Series Plot 7

4.1.3 Seasonal Decomposition 7

4.2 Time Series 8

4.2.1 Time Series Analysis 8

4.2.2 Mathematical Model 9

4.2.3 Time Series Modeling: Basic Definitions 10

4.3 Forecasting 13

4.3.1 SARIMA 13

4.3.2 Exponential Smoothing Methods 14

4.3.3 Holt-Winter's Method 14

4.4 Statistical Tests 16

4.4.1 ADF test 16

5. RESULTS AND DISCUSSION 17

5.1 Modelling of Rubber Prices of RSS1 using Holt-Winter's model 19

5.2 Modelling of Rubber Prices of RSS3 using Holt-Winter's model 23

5.3 Modelling and Forecasting of Rubber Prices of RSS1 using

SARIMA model 27
5.4 Modelling and Forecasting of Rubber Prices of RSS3 using SARIMA model 35
5.5 Comparison between SARIMA and Holt-Winter's model 44

CONCLUSION 45

REFERENCE 46

CHAPTER 1

INTRODUCTION

1.1 NATURAL RUBBER RATE

Several plant species form natural rubber, but economic considerations and quality control limit its commercial source to just one species, Hevea brasiliensis. Indigenous to the Amazon region, it has been introduced since the late 19th century to tropical areas of Asia and Africa. It is regarded as one of the greatest and most successful plant introductions in history, resulting in plantations of more than 9.3 million hectares globally, 95% of which are in Asia. The rubber tree has a lifespan of more than a century, but its economic life in plantations is generally 32 years—consisting of a 7-year immature period and a 25-year productive period. Commercial production of rubber in India began in 1902.

India is the fourth largest producer of natural rubber in the world. Kerala is the largest producer of natural rubber in India. Kerala accounts for 78% of the area and 90% of total rubber produced in the country (George & Chandrashekar, 2014). It was introduced to tropical Asia and Africa by the efforts of the British Government during the later part of 19th century (Kohjiya, 2015). An analysis of rubber production and consumption from 2005-06 to 2020-21 shows a decline in production growth, raising concerns for stakeholders. Meanwhile, rubber consumption increased from 801,110 tonnes in 2005-06 to 1,096,410 tonnes in 2020-21 (Vasagan & Chakraborty, 2021). Activities to widen the genetic pool of the rubber tree using breeding and biotechnology, including molecular biology and micropropagation, are meant to improve productivity and sustainability. Alternative sources such as guayule and Russian dandelion are being considered (Venkatachalam et.al., 2013). India's natural rubber production fluctuated despite area expansion, mainly due to declining productivity, especially in Kerala, which dominates cultivation. Tripura and Karnataka saw positive growth, while other states expanded primarily in area. Declining productivity was linked to adverse climate and farmers pausing tapping during price drops. Kerala’s trends significantly impacted national figures, as it accounted for 78% of cultivated land. A minimum support price policy could help stabilize production (Nithin & Mahajanashetti, 2017).

Rubber latex is critical throughout industries because it is elastic, resilient, and water-resistant. It is used everywhere in medical gloves, car tires, industrial adhesives, and household products such as mattresses and balloons. Volatility in rubber latex prices affects production costs and market stability, with industries that depend on it affected. Excessive volatility impacts supply chains and investment, and maintaining price stability is important for long-term growth. Understanding price fluctuations helps businesses and policymakers manage risks, ensuring steady supply and economic stability. Un-stability in the price of rubber is caused by international demand-supply imbalances, hedge funds, currency volatilities, and crude prices. The Cuddy Della Valle Index reported increasing instability at Kottayam and Bangkok markets (11% & 14% for 2005-08 versus 13% & 20% for 2009-12). There is increased volatility in global markets, led by crude oil and synthetic rubber prices, while the Price Stabilization Fund (2003) stabilized domestic prices (Anuja et.al., 2013).

Price forecasting of rubber is important because the market is volatile, and prices are affected by world demand, oil prices, exchange rates, and speculation. Precise price forecasts enable producers and manufacturers to plan production, control costs, and reduce risks. Policymakers use forecasts to formulate stabilizing policies, and investors use them for making informed investment decisions. Better forecasting makes the market more efficient, provides stability to the supply chain, and facilitates sustainable growth in the rubber sector. One of the recommended model for forecasting rubber rates is Holt-Winter's exponential Smoothing method (Chawla & Jha, 2009). Another reliable method to forecast the prices is ARIMA model (Sukiyono et al., 2019) (Mathew & Murugesan, 2020).

'Forecasting Rubber Prices of Kottayam using SARIMA And Holt-Winter's Exponential Smoothing Method' a research on forecasting rubber price for two grades of rubber RSS1 and RSS3 utilized Holt-Winters and SARIMA models. Due to the price volatility of rubber based on global dynamics, precise forecasting benefits producers and traders. Holt-Winters identified short-term trends, while SARIMA described long-term behaviors more effectively. Accurate forecasts facilitate price risk management and market stability. The comparison of the two models is also done. The data was collected from website of Rubber Board Ministry of Commerce, and Industry. The dataset comprises of monthly rubber prices(in Indian Rupees) of two different grades of rubber RSS1 and RSS3 from January 2010 to December 2023.

1.2 OBJECTIVES

The main objectives of this study are:

1. To forecast the rubber rate using Seasonal ARIMA model.
2. To forecast the rubber rate using the Holt-Winter's Exponential Smoothing method.
3. To compare the accuracy between forecast models Seasonal ARIMA and Holt's Exponential Smoothing method

CHAPTER 2

LITERATURE REVIEW

This chapter presents an overview of existing literature on predicting rubber prices, with emphasis on statistical methods. It discusses different methodologies employed for examining past price trends and determining the most critical influencing factors. On the basis of previous studies, this review attempts to determine gaps and present effective methods for enhancing prediction precision.

Chawla and Jha (2009) presented a paper in which an attempt is made to forecast the price of natural rubber in India by using monthly data for the period from January 1991 to December 2005. The forecasts are obtained up to December 2008 using Linear Trend Equation, Semi-log Trend Equation, Holt's Method, Winter's Method and ARIMA Model. The accuracy of forecast obtained through various methods is compared with the monthly actual production data of natural rubber for the period from January 2006 to March 2007, the last month for which actual price data were available. This period is also used as a hold-out sample. It is found that Winter's method gives the best result followed by Holt's method and Semi-log trend equation. The MAPE of Winter's, Holt's method, and Semi-log trend equation are 8.01 per cent, 8.09 per cent and 8.12 per cent respectively. It is therefore suggested that Winter's method could be used for forecasting the natural rubber prices in India. The forecasting method for prices of natural rubber in India, as shown in this paper, can be a very useful tool for the Indian rubber industry professionals and policy makers in India.

Anuja et al. (2013) presented a study about price fluctuations in rubber. Rubber, a perennial crop with a 10-20 year economic lifespan, is highly affected by price fluctuations, influencing production and farm income stability. The Cuddy Della Valle Index indicated significant price instability in domestic and international markets. GARCH (1,1) analysis of RSS 4 rubber prices (2005–2012) confirmed high volatility. Johansen's co-integration test revealed long-run market integration, suggesting price uniformity and improved marketing efficiency.

Venkatachalam et al. (2013) did a study on Hevea brasiliensis, the primary source of natural rubber, though alternatives like guayule and Russian dandelion show potential. Guayule is suitable for medical latex, while dandelion rubber is being developed for tires. Other plants, such as lettuce and fig trees, require further study. An ideal rubber crop would be fast-growing and widely adaptable. This review examines Hevea cultivation and alternative latex sources for future production.

George and Chandrashekar (2014) did a study on production of rubber. India, the fourth-largest natural rubber producer, led global productivity at 1,841 kg/hectare in 2011-12. Kerala accounts for 90% of India's output. Rubber is marketed through traders and dealers, with exports peaking at 75,905 tonnes in 2003-04 before declining due to reduced incentives. Major buyers include China, Malaysia, and Indonesia. This study analyzes Kerala's rubber production and marketing trends using CAGR.

Kohjiya (2015) presented a study on origin and history of rubber plant. The rubber tree (*Hevea brasiliensis*) is native to the Amazon Basin in South America. Indigenous people in the region used its latex for waterproofing. In the 18th century, European explorers introduced rubber to the world. By the late 19th century, British botanists smuggled rubber seeds to Southeast Asia, leading to large-scale plantations in countries like Malaysia, Indonesia, and Thailand.

Nithin and Mahajanashetti (2017) presented a study on trend and growth analysis of area, production and productivity of Natural rubber. Natural rubber is a key industrial material. From 2005-06 to 2014-15, Kerala led cultivation (69.16%), while Tripura saw the highest growth in area (9.27%) and production (11.55%). Despite a 3.36% annual increase in cultivation, production declined (-1.19%) due to climate factors and farmers pausing tapping during price drops.

Sukiyono et al. (2019) presented a study aimed to analyze and select the possible forecasting methods for monthly natural rubber prices in Indonesia and World Markets. The univariate model of Double Exponential Smoothing, Decomposition, and ARIMA models are applied to forecast price data from 2012:1 – 2016:12. The selection of an accurate model is based on the lowest value of MAPE, MSD, and MAD. ARIMA is the possible methods for world rubber price forecasting while Double Exponential Smoothing should be applied for predicting domestic rubber prices because it allows for better predictive performance.

Mathew and Murugesan(2020) presented a study on Indian natural rubber price forecast–An Autoregressive Integrated Moving Average (ARIMA) approach. The objective of this study was to forecast the price of natural rubber in India during April 2019 to March 2020 by employing autoregressive integrated moving average (ARIMA). The monthly pricing data for the period from April 2008 to March 2018 was used for the study. The analysis was carried out during the year 2018–19. The prices of RSS4, Latex(60%DRC) and ISNR 20 were taken for modelling. AIC was used as a selection criterion for the best-fitted model. ARIMA(3,1,2) for RSS 4, ARIMA (3,1,2) for Latex 60% DRC, and ARIMA (4,1,3) for ISNR20were the most suited models to forecast the price. The evaluation metrics were R2 , Adjusted R2 , Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE). These were employed for validating the forecasting model.

Vasagan and Chakraborty (2021) presented a study on demand and supply trend of rubber in India. This study analyzes trends in rubber production and consumption in India. Rubber cultivation expanded to 822,500 hectares, but imports grew faster than production. Data from the Rubber Board indicate a 2.09% annual increase in cultivated land over 15 years, with trappable area rising from 447,015 hectares (2005-06) to 692,900 hectares (2020-21). Despite increasing consumption (801,110 to 1,096,410 tonnes), production declined, leading to a supply-demand gap. Prices rose by 3.3%, and imports surged by 10.57%. To reduce reliance on imports, the study recommends boosting domestic production for farmers' benefit.

CHAPTER 3

MATERIALS AND METHODS

3.1 DATA COLLECTION

The data set used in this study is from website of Rubber Board Ministry of Commerce, and Industry (http://rubberboard.org.in/). The dataset comprises of monthly rubber prices(in Indian Rupees) of two different grades of rubber RSS1 and RSS3 from January 2010 to December 2023. The dataset consist of 'Date' , 'Price of RSS1' (in Indian Rupees) and 'Price of RSS3' (in Indian Rupees)

3.2 METHODOLOGY

The first crucial step in the analysis was a detailed examination of the data. This includes outlining the data collection methods, processing techniques, and analytical models used to achieve the study's objectives. The objective of this study is to examine the fluctuations in the prices of both grades of rubber in Kottayam and to predict future prices using statistical techniques. In the study we analysed the data to understand the characteristics of the dataset then forecast it using SARIMA and Holt-Winter's Exponential Smoothing model. The comparison of the two is done by using MAE and RMSE values.

3.3 TOOLS FOR FORECASTING

1. Seasonal ARIMA (Autoregressive integrated moving average) model.
2. Holt-Winter's Exponential Smoothing method.

3.4 TOOLS FOR COMPARISON

1. **RMSE (**Root Mean Square Error**)**

The formula for RMSE is:

$$RMSE = \sqrt{\frac{\Sigma(y_i - \widehat{y_i})}{N-P}}$$

where,

y_i is the actual value for the i th observation

$\widehat{y_i}$ is the predicted value for the i th observation

N is the number of observations

P is the number of parameter estimates, including the constant

2. **MAE (**Mean Absolute Error**)**

The formula for MAE is:

$$MAE = \frac{1}{n}\sum_{i=0}^{n} |y_i - \widehat{y_i}|$$

where,

n is total number of observations

y_i is the actual value

$\widehat{y_i}$ is the predicted value

| y_i - $\widehat{y_i}$| is the absolute error for each observation

3.5 PYTHON PROGRAMMING LANGUAGE

Python is a versatile and highly sought-after programming language, renowned for its clarity, versatility, and adaptability. Its wide-ranging applications span multiple domains, including data science, machine learning, and web development, cementing its position as a leading language in the industry.

CHAPTER 4

EXPLORATORY DATA ANALYSIS AND TIME SERIES

4.1 EXPLORATORY DATA ANALYSIS

Exploratory Data Analysis (EDA) is a crucial initial step in data science projects. It involves examining and visualizing data to identify key features, detect patterns, and explore relationships between different variables. EDA can also be used to understand the quality of the data ,check null values, missing values etc.

4.1.1 Descriptive Statistics

Descriptive statistics summarize the central tendency, distribution, and variability of a dataset by computing measures like the quartiles, mean, median, mode, standard deviation, etc. Measures such as quartiles and standard deviation identify outliers and extreme values in the dataset. Descriptive statistics generally give a complete summary of the important characteristics of the data, enabling better understanding and enabling further accurate analysis.

4.1.2 Time Series Plot

A time-series plot, or a time plot, is a form of graph used to present data points that were gathered in sequence over time. The x-axis in a time-series plot illustrates the time, while the y-axis illustrates the variable under measurement.

4.1.3 Seasonal Decomposition

The seasonal decomposition method is a powerful method for decomposing time series data into its basic components: trend, season, and residuals. This breakdown allows us to detect seasonality, identify whether the dataset displays regular, recurring patterns that are predictable over a fixed interval and identify the trend. By knowing the trend and seasonality, we are able to choose the appropriate model and increase forecasting accuracy.

The three elements of seasonal decomposition are:

1. Trend: The long-term rise or fall in the data.
2. Seasonal: Routine, predictable patterns which repeat over a fixed interval.

3. Residual: The difference between the actual value and the forecast value, reflecting how well the data conforms to the model.

In total, exploratory data analysis (EDA) is an essential step towards understanding the data, finding missing or null values, and coming to understand the right models for precise forecasting.

4.2 TIME SERIES

A time series is a sequence of data points recorded at successive, equally spaced intervals over time, making it a series of discrete-time observations.

4.2.1 TIME SERIES ANALYSIS

A method of analyzing a series of data points collected over time is referred to as time series analysis. Data points in time series analysis are measured at regular intervals over a fixed time frame, not at random. But conducting this type of research takes more than collecting data over time. The capability of analyzing how factors over time is what separates time series data from any other data. Alternatively, time is a significant element as it appears to influence the way the information alters as it gathers and the end result is attained. It offers a source of additional data and some criteria between the information.

To ensure consistency and reliability, time series analysis most likely requires a

huge number of data points. A huge data set makes your analysis able to filter out irregular data and have a representative sample size.

The underlying pattern and structure of a data is embodied by the components of time series.

The prominent components of time series are:

1) Secular Trend : Trend is the general direction of data over the long term, i.e., whether it rises, falls, or stays constant. A time series with no trend of increasing or decreasing is stationary in the mean.

2) Seasonal Variations : Seasonal fluctuations in a time series are brought about by rhythmic forces that act periodically over a 12-month period, according to a given pattern every year. Climate, weather, local customs, and customary practices are typical items that bring about seasonal fluctuations. Measuring seasonal fluctuations is primarily aimed at isolating them from the trend and assessing their effect.

3) Cyclical Variations : Cyclical variations are persistent rising or falling patterns of a time series that exceed a year and occur with irregular time intervals. Such variations occur very

frequently in economic statistics and have durations between 5 and 12 years, or even longer. Cyclical variations differ from seasonal variations since they are not periodic, but rather the length varies with the business or industry under analysis.

4) Irregular Fluctuations : Irregular variations are unexpected, short-term movements in a time series with no regular pattern. Also referred to as residual variations, they are the leftover random fluctuations once trend, cyclical, and seasonal variations are eliminated. Such variations are frequently the result of unexpected events like natural disasters, wars, or economic shocks.

4.2.2 MATHEMATICAL MODEL

These are the two models we commonly prefer for the decomposition of a time series into its four components. The objective is to estimate the four types of variations and separate them so that their relative effect on the overall behavior of the time series can be demonstrated.

1) Additive Model : Based on the additive model the time series decomposition is performed on the assumption that the impact of different components is additive or in other words,

$$Y_t = T_t + S_t + C_t + I_t$$

Where Y_t is value of time series and T_t, S_t, C_t and I_t represents trend, seasonal variations, cyclical variations and irregular variations respectively. S_t, C_t and I_t are absolute values in this model and may be positive or negative. The model postulates that four components of the time series are independent of one another and none has any influence on the other three components. In real practice, this hypothesis does not hold true as these factors influence one another.

2) Multiplicative Model : The multiplicative model breaks down a time series into four elements (trend, seasonality, cycle, and irregularity) with the underlying assumption that their impacts are dependent on one another. The model is given by:

$$Y_t = T_t \times S_t \times C_t \times I_t$$

where T_t, S_t, C_t and I_t are relative changes, i.e., rates or indices, and not absolute values. The model can be converted with the help of logarithms to facilitate calculations.

$$logY_t = logT_t + \log S_t + logC_t + logI_t$$

4.2.3 TIME SERIES MODELING : BASIC DEFINITIONS

1) Autocorrelation Function (ACF) :

The Autocorrelation function (ACF) of a stationary time series $\{Z_t\}$ quantifies the correlation between Z_t and its k- step ahead value, Z_{t+k}. The ACF at lag k, i.e., ρ(k), is given by:

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)}$$

where $\gamma(k)$ is the covariance of Z_t and Z_{t+k}, and $\gamma(0)$ is the variance of Z_t .

2) Partial Autocorrelation Function (PACF) :

The partial autocorrelation function (PACF) yields information about the structure of dependence in a stationary time series process. In contrast to the autocorrelation function, the PACF controls for the impact of intermediate lags, quantifying the correlation between X_t and X_{t+k} after accounting for the effect of $X_{t-1}, X_{t-2},\ldots, X_{t-k+1}$.

The partial autocorrelation at lag k, φ_{kk}, is the correlation between X_t and X_{t-k} controlling for the influence of prior terms. It is equivalent to the partial regression coefficient in the following equation:

$$X_t = \varphi_{k1}X_{t-1} + \varphi_{k2}X_{t-2} + \cdots + \varphi_{kk}X_{t-k} + \varepsilon_t$$

3) Autoregressive (AR) process

An Autoregressive process of order p, AR(p), is a time series $\{ X_t \}$ that can be represented as:

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \cdots + \varphi_p X_{t-p} + \varepsilon_t$$

where:

$\varphi_1, \varphi_2,\ldots, \varphi_p$ are parameters, ε_t is a purely random process with 0 mean and constant variance σ^2

4) Moving Average (MA) process

A time series $\{ X_t \}$ is said to be a moving average process of order q, abbreviated as MA(q) if it is a weighted linear sum of the last q random shocks,

$$X_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \cdots - \theta_q\varepsilon_{t-q}$$

where $\{\varepsilon_t\}$ denotes a purely random process with mean 0 and constant variance σ^2.

5) Stationarity

The ARIMA process needs a stationary series of data. Stationarity guarantees that AR coefficients satisfy certain conditions, making useful parameter estimates possible.

Conditions for stationarity differ by model:

i)White noise series and pure MA models are always stationary.

ii)AR(1) and ARMA(1,q) models need $|\varphi_1| < 1$.

iii)AR(2) and ARMA(2,q) models need: $|\varphi_2| < 1$, $\varphi_1 - \varphi_2 < 1$, $\varphi_1 + \varphi_2 < 1$

These assumptions guarantee that ARIMA is stationary and can make accurate predictions.

6) Invertibility

ARIMA models also need to meet the condition of invertibility, which provides that the weights given to the past observations fall as the observations get older. Invertibility is required because it is desirable to assign heavier weights to current observations.

The invertibility conditions differ by model:

i)White noise series and pure AR processes are always invertible.

ii)MA(1) and ARMA(p,1) need $|\theta_1| < 1$.

iii)MA(2) and ARMA(p,2) need: $|\theta_2| < 1$, $\theta_1 + \theta_2 < 1$, $\theta_1 - \theta_2 < 1$.

These conditions guarantee that the ARIMA model is invertible and yields meaningful results.

7) ARMA process

Mixed Autoregressive moving average model with p autoregressive terms and q moving average terms is abbreviated as ARMA(p, q) is given by,

$$X_t - \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \cdots + \varphi_p X_{t-p} = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}$$

Using the backward shift operator B, the ARMA(p, q) can be written in the form,

$$X_t - \varphi_1 B X_t + \varphi_2 B^2 X_t + \cdots + \varphi_p B^p X_t = \varepsilon_t - \theta_1 B \varepsilon_t - \theta_2 B^2 \varepsilon_t - \cdots - \theta_q B^q \varepsilon_t$$

$$\varphi(\mathrm{B}) X_t = \theta(\mathrm{B}) \varepsilon_t$$

where $\varphi(\mathrm{B})$ and $\theta(\mathrm{B})$ are the polynomial in B of degree q and p.

8) ARIMA and SARIMA

In reality, most time series are non-stationary so it is not appropriate to use stationary AR, MA or ARMA processes directly. One of the standard ways of dealing with non-stationary series is by differencing the data to make them stationary. The first difference is computed as:

$$Z_t - Z_{t-1} = (1 - B)Z_t$$

Further differencing can be applied to obtain second differences, and so on. The d th difference is expressed as:

$$(1 - B)^d Z_t$$

When a time series is differenced d times before fitting an ARMA(p, q) model, the resulting model for the original (un-differenced) series is called an ARIMA(p, d, q) model, where 'I' stands for Integrated, and d represents the number of differences applied. Algebraically, an ARIMA(p, d, q) model is represented as:

$$\phi(\mathrm{B})(1 - B)^d Z_t = \theta(\mathrm{B})\varepsilon_t$$

where $\{\varepsilon_t\}$ is a purely random process with mean zero and constant variance σ^2. Apart from autoregressive (AR) and moving average (MA) terms, ARIMA models can also have a constant.

Seasonality in a time series means there are patterns which recur every s periods and where s is the number of periods until the pattern recurs. Monthly data would have s=12, for instance.

In a seasonal ARIMA model, seasonal AR and MA terms forecast X_t based on data values and past time-point errors with multiples of s lag. A seasonal ARIMA model combines non-seasonal and seasonal parts in multiplicative structure and is expressed as ARIMA(p, d, q) × (P,D,Q)s, where:

(p, d, q) represents the non- seasonal components and
(P, D, Q)s represents the seasonal components

Mathematically, a seasonal ARIMA model is represented as:

$$\Phi(B^s)\varphi(\mathrm{B})(1 - B^s)D(1 - B)^d X_t = \Theta(B^s)\theta(\mathrm{B})Z_t$$

$$\text{where, } \varphi(\mathrm{B}) = 1 - \varphi_1\mathrm{B} - \varphi_2\mathrm{B}^2 - \ldots - \varphi_p\mathrm{B}^p,$$

$$\theta(\mathrm{B}) = 1 - \theta_1\mathrm{B} - \theta_2\mathrm{B}^2 - \ldots - \theta_q\mathrm{B}^q$$

$$\Phi(\mathrm{B}^S) = 1 - \Phi_1\mathrm{B}^S - \Phi_2\mathrm{B}^{2S} - \ldots - \Phi_q\mathrm{B}^{PS}$$

$$\Theta(\mathrm{B}^S) = 1 - \Theta_1\mathrm{B}^S - \Theta_2\mathrm{B}^{2S} - \ldots\ldots\ldots - \Theta_q\mathrm{B}^{QS}$$

where:

p = order of non-seasonal autoregression (AR)
d = order of non-seasonal differencing
q = order of non-seasonal moving average (MA)

P = order of seasonal autoregression (SAR)
D = order of seasonal differencing
Q = order of seasonal moving average (SMA)
s = number of periods in a season

The term { Z_t } represents a purely random process with zero mean and constant variance σ^2.

9) Akaike information criterion (AIC)

Akaike information criterion (AIC) is a prediction error estimator and, by extension, relative quality of statistical models for the data at hand. For a set of models for the data, AIC estimates each model's quality relative to any one of the other models. Accordingly, AIC offers a model selection method.

$$AIC = 2k - 2ln(L)$$

10) Diagnostic Checking

Time series diagnostic checking involves residual analysis to check that the model is a good fit. Residuals should first be randomly looking and not show any patterns when plotted. The ACF plot must not have any significant correlations, and the Ljung-Box test can verify if residuals are uncorrelated. Then, normality checking is needed; a histogram or Q-Q plot must display a roughly normal distribution. Lastly, model performance can be measured with metrics such as RMSE, MAE and information measures such as AIC or BIC. When diagnostic checks indicate problems, options for solutions include transformations such as log or differencing, inclusion of AR/MA terms or changing over to different models.

4.3 FORECASTING

4.3.1 Seasonal ARIMA (Autoregressive Integrated Moving Average)

In a seasonal ARIMA model, seasonal AR and MA terms forecast X_t based on data values and past time-point errors with multiples of s lag. A seasonal ARIMA model combines non-seasonal and seasonal parts in multiplicative structure and is expressed as ARIMA(p, d, q) × (P,D,Q)s, where:

(p, d, q) represents the non- seasonal components and
(P, D, Q)s represents the seasonal components

Mathematically, a seasonal ARIMA model is represented as:

$$\Phi(B^s)\varphi(\mathrm{B})(1 - B^s)D(1 - B)^d X_t = \Theta(B^s)\theta(\mathrm{B})Z_t$$

$$\text{where, } \varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \ldots - \varphi_p B^p,$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q$$

$$\Phi (B^S) = 1- \Phi_1 B^S - \Phi_2 B^{2S} - \ldots - \Phi_q B^{PS}$$

$$\Theta (B^S) = 1- \Theta_1 B^S - \Theta_2 B^{2S} - \ldots - \Theta_q B^{QS}$$

where:

p = order of non-seasonal autoregression (AR)
d = order of non-seasonal differencing
q = order of non-seasonal moving average (MA)
P = order of seasonal autoregression (SAR)
D = order of seasonal differencing
Q = order of seasonal moving average (SMA)
s = number of periods in a season

The term $\{Z_t\}$ represents a purely random process with zero mean and constant variance σ^2.

4.3.2 Exponential Smoothing methods

Exponential smoothing techniques are mean methods that have only three inputs: the future forecast for the latest period (F1), the observed value for that period (y_t) and the estimate of the smoothing constant (α).

Exponential smoothing can easily be extended to handle time series with trends and seasonal variation. The form to handle a trend with non-seasonal data is most commonly known as Holt's exponential smoothing, while the form that also deals with seasonal variation is commonly known as the Holt-Winters method.

4.3.3 Holt-Winter's Method

Holt's technique can be generalized to handle time series that have both trend and seasonal fluctuations. Holt-Winters technique comes in two forms, additive and multiplicative, the \ application of which is based on the nature of the given time series. Let L_t, T_t, I_t represent the local level, trend and seasonal index, respectively at time t.

Its interpretation will be based on whether seasonality is assumed to be additive or multiplicative. In the additive scenario i, y_t-I_t is the deseasonalized value, whereas in the multiplicative category it is y_t / I_t. The levels of the 3 quantities L_t, T_t, and I_t, all have to be estimated and so we require 3 updating equations with three smoothing parameters, say $\propto$, γ and δ. As with the smoothing parameters typically being in the range (0, 1) beforehand. The shape of the updating equations is again intuitively plausible.

Suppose the seasonal variation is multiplicative. Then the (recurrence form) equations for updating L_t, T_t, I_t, when a new observation y_t becomes available are

$$L_t = \alpha\left(\frac{y_t}{I_{t-s}}\right) + (1-\alpha)(L_t + T_{t-1})$$

$$T_t = \gamma(L_t - L_{t-1}) + (1-\gamma)T_{t-1}$$

$$I_t = \delta(y_t/L_t) + (1-\delta)I_{t-s}$$

and the forecasts from time (t) are then,

$$y_{t+h} = (L_t + hT_t)I_{t-s+h}$$

for h= 1,2, …, s. If the seasonal variation is additive, the equations for updating L_t, T_t, I_t, when a new observation y_t becomes available are

$$L_t = \alpha(y_t - I_{t-s}) + (1-\alpha)(L_{t-1} + T_{t-1})$$

$$T_t = \gamma(L_t - L_{t-1}) + (1-\gamma)T_{t-1}$$

$$I_t = \delta(y_t - L_t) + (1-\delta)I_{t-s}$$

and the forecasts from time (t) are then

$$y_{t+h} = L_t + hT_t + I_{t-s+h}$$

For starting values, it seems sensible to set the level component L0, equal to the average observation in the first year i.e,

$$L_0 = \sum_{t=1}^{s} y_t$$

where s is the number of seasons. The starting values for the slope component can be taken from the average difference per period between the first and second year averages i.e,

$$T_0 = \left(\frac{\sum_{t=s+1}^{2s} y_t}{s} - \frac{\sum_{t=1}^{s} y_t}{s}\right)/s$$

Finally, the seasonal index starting value can be calculated after allowing for a trend adjustment, as follows:

$$I_0 = \frac{y_k - (k-1)T_0}{2}$$

$$I_0 = \frac{y_k - (L_0 + (k-1))T_0}{2}$$

Where k= 1,…,s. This will lead to (s) separate values for I_0, which is what is required to gain the initial seasonal pattern.

4.4 STATISTICAL TESTS

4.4.1 ADF test

The ADF test is a type of test known as 'Unit Root Test', which is the correct way to test the stationarity of a time series. Augmented Dickey-Fuller (ADF) test is a popular statistical test that is employed in order to test if a given time series is stationary or not. It is one of the most popularly employed statistical tests when it comes to testing the stationarity of a series.

It is testing the below two null and alternative hypotheses:

H_0: The time series is considered as non-stationary.

H_1: The time series is considered as stationary.

Now if the p-value of this test is less than a certain value (say $\alpha = 0.05$) then in those situations the null hypothesis is rejected and states that the time series is stationary.

CHAPTER 5

RESULTS AND DISCUSSION

This chapter discusses a comparative study of time series modelling and forecasting of monthly rubber price of grades RSS1 and RSS3 using SARIMA, Holt-Winters Exponential Smoothing forecasting. The data comprises 168 observations from January 2010 to December 2023.

Descriptive Statistics :

Table 5.1: Descriptive Statistics

RSS1	
count	168
mean	16490.7916
std	2967.4198
min	11350.0000
25%	14050.2500
50%	16026.5000
75%	18183.0000
max	25805.0000

RSS3	
count	168
mean	15520.7142
std	2984.3710
min	9570.0000
25%	13204.0000
50%	14997.0000
75%	17447.7500
max	24683.0000

Time Series plot :

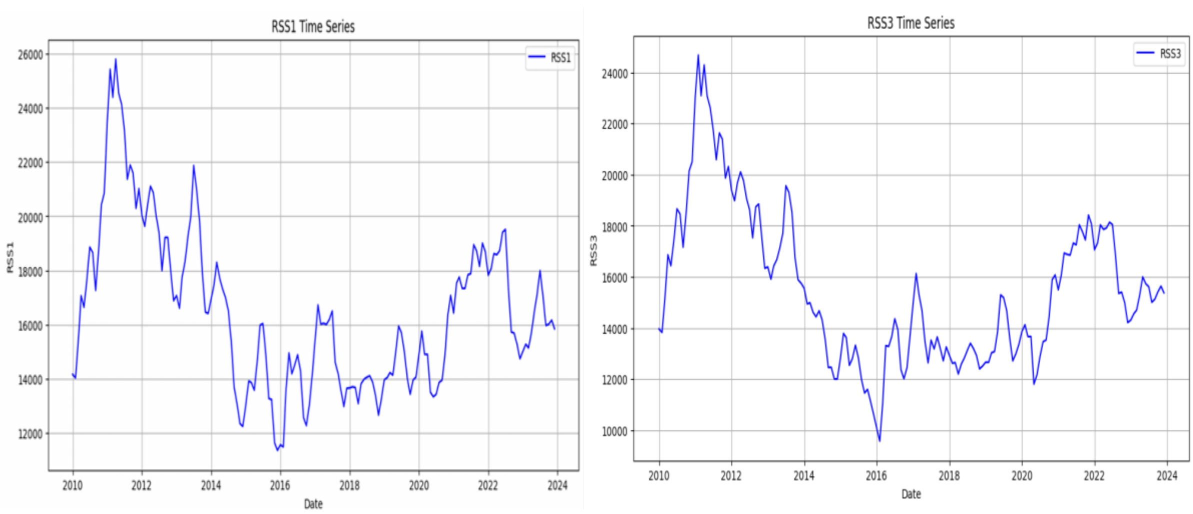

figure 5.1: Time series plot

Time Series Decomposition:

To evaluate the trend, seasonality and random components seasonal decomposition is done, from figure 5.2 and figure 5.3 it is clear that both RSS1 and RSS3 shows seasonality.

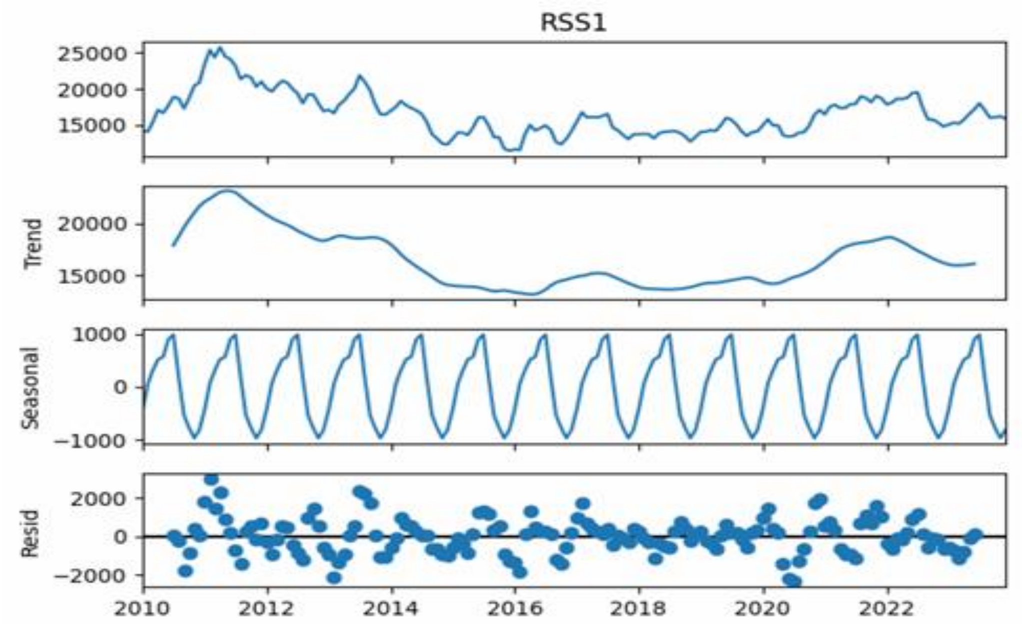

figure 5.2: Time series decomposition of RSS1

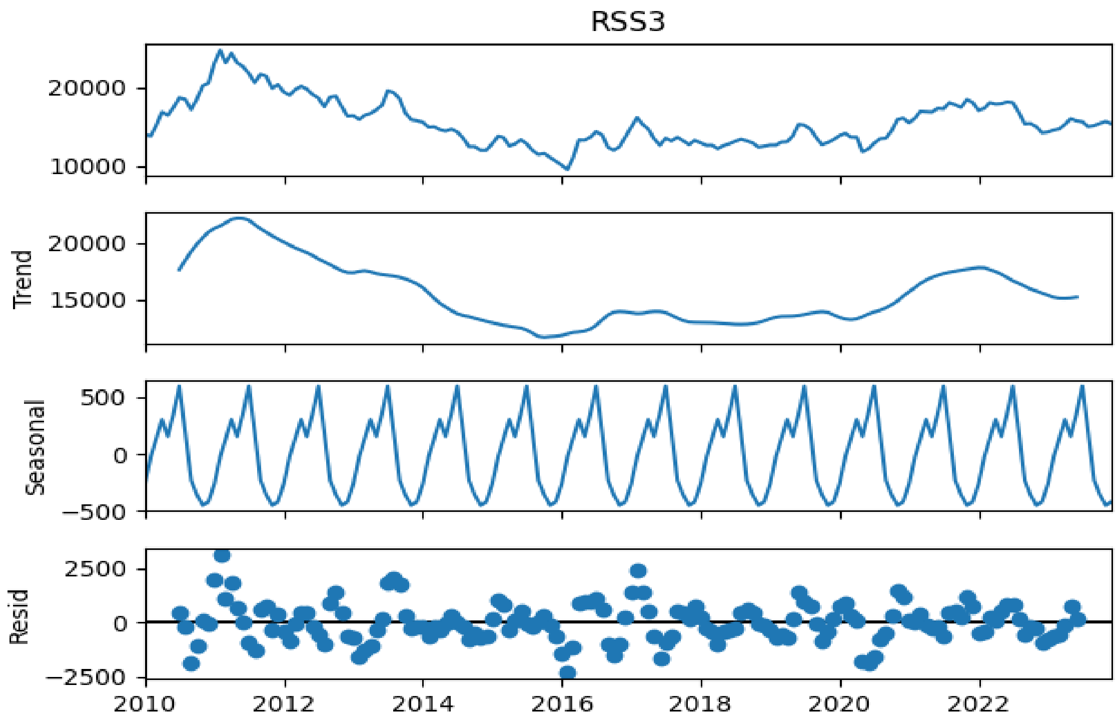

figure 5.3: Time series decomposition of RSS3

5.1 MODELLING OF RUBBER PRICES OF RSS1 USING HOLT-WINTER's MODEL

Holt's Winters Forecasting Procedure is used to forecast the rubber prices. The parameter estimates for the model obtained are listed in the table

Table 5.2: Holt-Winter's model parameters

Parameter	Parameter Estimates
Alpha (Level)	0.895
Beta (Trend)	0.071
Gamma (Season)	0.001

Then the equation for the forecast value is given as:

$$L_t = 0.895(y_t - I_{t-s}) + (1 - 0.895)(L_{t-1} + T_{t-1})$$

$$T_t = 0.071(L_t - L_{t-1}) + (1 - 0.071)T_{t-1}$$

$$I_t = 0.001(y_t - L_t) + (1 - 0.001)I_{t-s}$$

$$\hat{y}_{t+h} = L_t + hT_t + I_{t+h-s}$$

Equation 5.1: Holt- Winter's Formula

Diagnostic Checking :

Diagnostic checking is a vital step in statistical modelling that ensures the model's accuracy, reliability, and effectiveness. It assesses the model's precision and helps refine it to achieve better prediction accuracy.

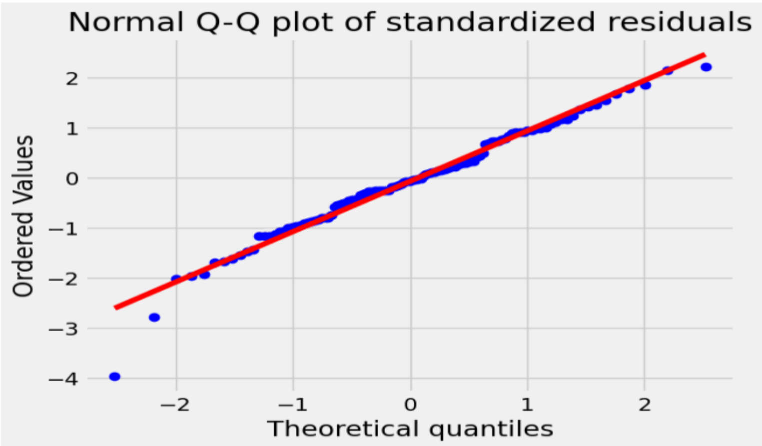

figure 5.4: Normal Q-Q plot

The figure 5.4 depicts the Q-Q plot , it is clear that the most of the residual values lie on the straight line, which indicates that residuals are approximately normally distributed.

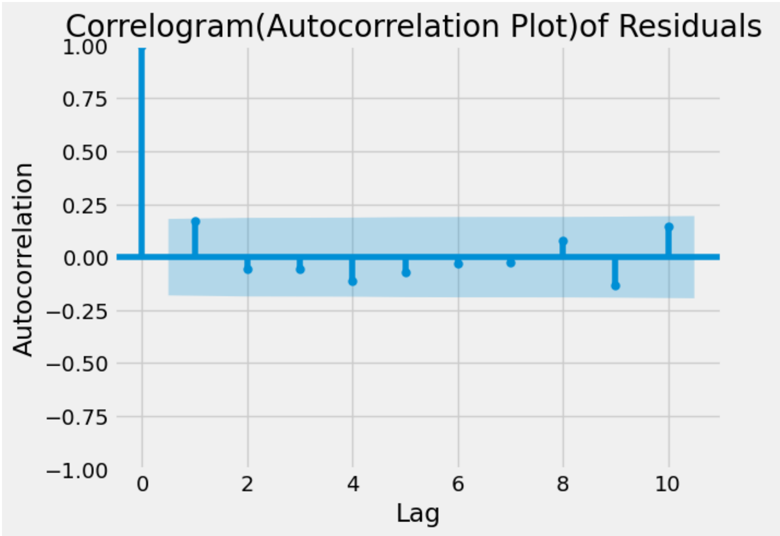

figure 5.5: Autocorrelation plot

The correlogram in figure 5.5 reveals that the autocorrelation at all lags is insignificant and decays to zero. This indicates that there is no substantial autocorrelation present in the data.

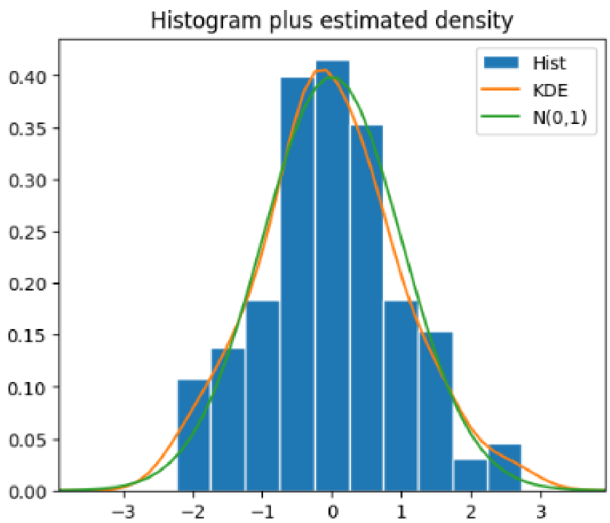

figure 5.6: Histogram of Residuals

The figure 5.6 shows that the histogram matches the normal distribution curve, the residuals are approximately normal.

The diagnostic checking confirms that the fitted Holt-Winters model is statistically sound and adequate. Therefore, the model is suitable for forecasting monthly rubber prices.

In-sample Forecasting:

The fitted model is used to do In-sample forecasting. In-sample Forecasting done for the year 2023 is given in the below table 5.3

Table 5.3: In sample forecasted values

Months	Actual Value	Predicted Value	Months	Actual Value	Predicted Values
January	15008	16090.815202	July	17996	17624.859392
February	15275	17047.570368	August	17040	17403.294683
March	15133	16638.122085	September	15962	16416.538661
April	15695	16544.581736	October	16013	15791.246006
May	16456	16511.163645	November	16162	15521.131111
June	17140	17179.855036	December	15840	15380.115726

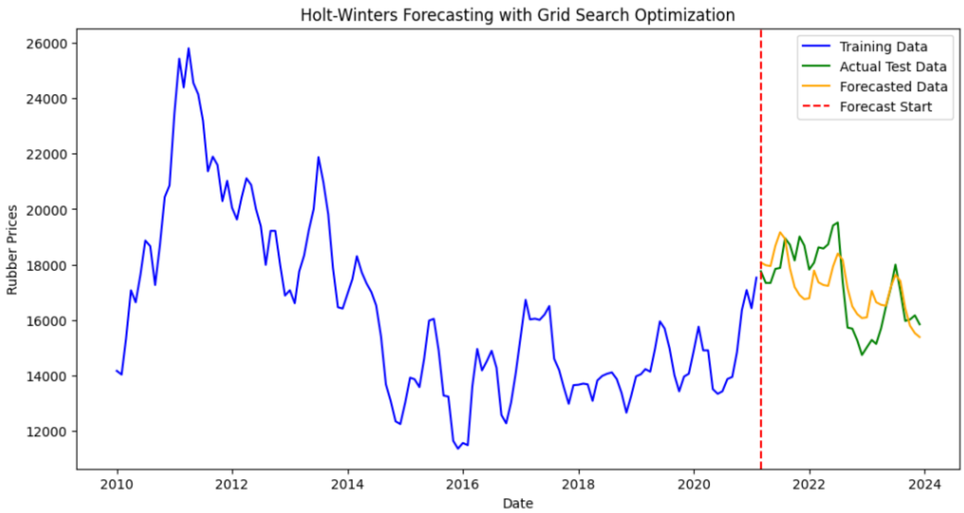

figure 5.7: Actual vs Predicted values plot

In figure 5.7 the blue line are the training data, green line is the actual testing data and orange line is the predicted testing data.

Forecasting of Rubber Prices using Holt-Winters model:

Rubber Prices from January 2024 to December 2025 is forecasted using the model.

Table 5.4: Holt-Winter's forecast of RSS1

Months	Forecasted Values	LCL	UCL	Months	Forecasted Values	LCL	UCL
Jan 2024	18070.0985	12893.7580	22638.5834	Jan 2025	17354.1103	12882.1583	22643.9276
Feb 2024	17973.6322	11194.9068	20939.7341	Feb 2025	17259.1069	11185.7602	20948.6245

Mar 2024	17942.47 89	10584.28 71	19024.54 10	Mar 2025	17226.82 12	10572.39 78	19032.46 73
Apr 2024	18674.53 64	13258.65 84	21753.25 41	Apr 2025	17927.19 58	13267.96 86	21769.10 25
May 2024	19163.83 59	13987.25 46	23510.68 42	May 2025	18394.34 83	14005.98 65	23521.83 77
Jun 2024	18928.47 35	13025.15 94	22931.35 71	Jun 2025	18165.81 31	13011.75 98	22946.17 48
Jul 2024	17860.51 36	11246.25 31	20864.56 42	Jul 2025	17138.52 61	11235.34 18	20877.34 06
Aug 2024	17185.33 03	11147.02 45	20014.35 10	Aug 2025	16488.28 83	11132.51 58	20025.46 18
Sep 2024	16896.42 80	10475.26 53	19584.02 45	Sep 2025	16208.95 93	10463.67 25	19596.54 52
Oct 2024	16747.96 77	9975.698 7	19024.74 10	Oct 2025	16064.04 17	9986.573 2	19039.98 75
Nov 2024	16778.32 02	11783.31 59	18953.65 41	Nov 2025	16090.81 02	11776.41 21	18960.94 23
Dec 2024	17778.55 70	12049.58 46	20147.68 94	Dec 2025	17047.53 68	12534.95 19	20154.82 93

The Table 5.4 is the forecasted values and its LCL and UCL of rubber prices of RSS1 for January 2024 to December 2025.

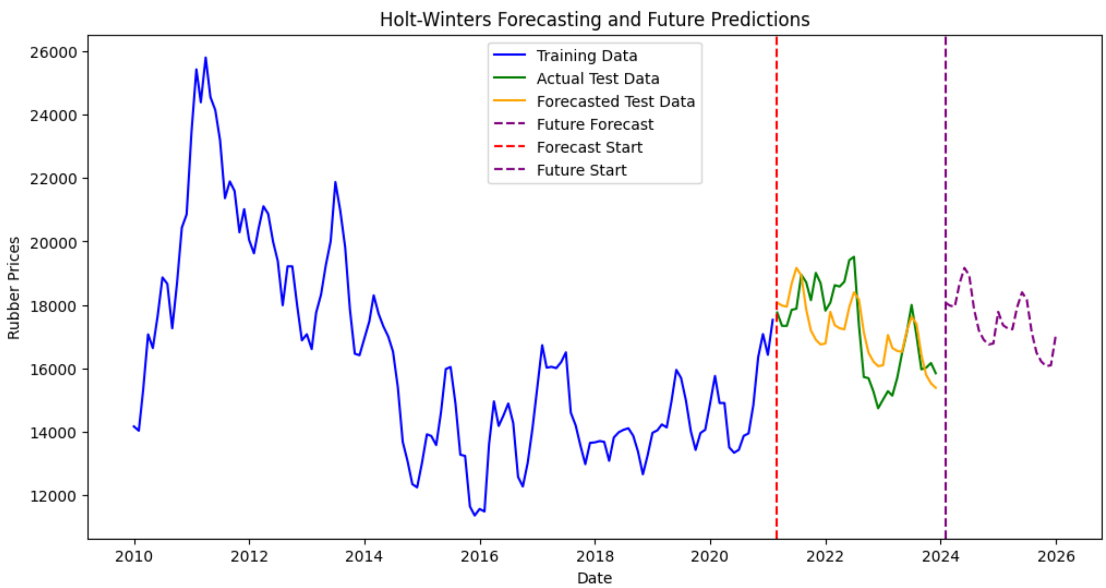

figure 5.8: Actual, Predicted and Forecasted values

The figure 5.8 is the plot of forecasted values using Holt Winter's model.

5.2 MODELLING OF RUBBER PRICE OF RSS3 USING HOLT-WINTER's MODEL

Holt's Winters Forecasting Procedure is used to forecast the rubber prices of RSS3. The parameter estimates for the model obtained are listed in the table 5.5

Table 5.5: Holt-Winter's model parameters

Parameter	Parameter Estimates
Alpha (Level)	0.845
Beta (Trend)	0.069
Gamma (Season)	0.002

Then the equation for the forecast value is given as:

$$L_t = 0.845(y_t - I_{t-s}) + (1 - 0.845)(L_{t-1} + T_{t-1})$$

$$T_t = 0.069(L_t - L_{t-1}) + (1 - 0.069)T_{t-1}$$

$$I_t = 0.002(y_t - L_t) + (1 - 0.002)I_{t-s}$$

$$\hat{y}_{t+h} = L_t + hT_t + I_{t+h-s}$$

Equation 5.2: Holt- Winter's Formula

Diagnostic Checking:

Diagnostic checking is a vital step in statistical modelling that ensures the model's accuracy, reliability, and effectiveness. It assesses the model's precision and helps refine it to achieve better prediction accuracy.

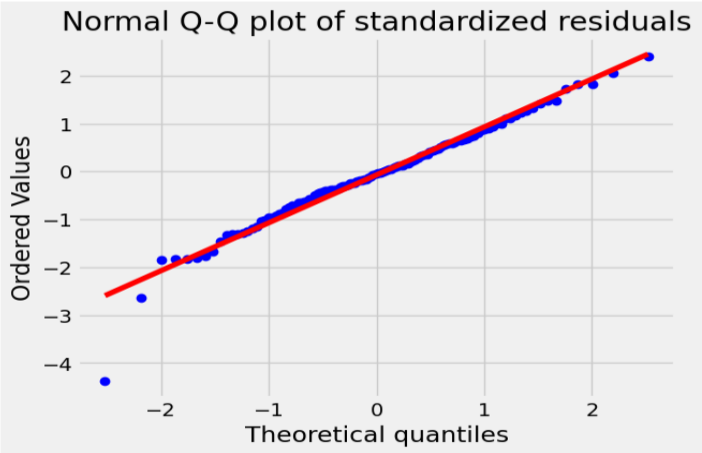

figure 5.9: Normal Q-Q plot of residuals

The figure 5.9 depicts the Q-Q plot, it is clear that the most of the residual values lie on the straight line, which indicates that residuals are approximately normally distributed.

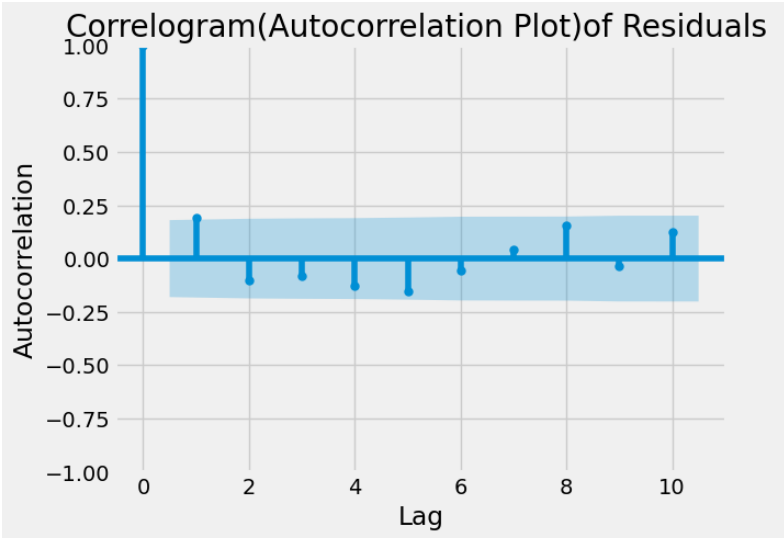

figure 5.10: Autocorrelation plot of Residuals- RSS3

The correlogram in figure 5.10 reveals that the autocorrelation at all lags is insignificant and decays to zero. This indicates that there is no substantial autocorrelation present in the data.

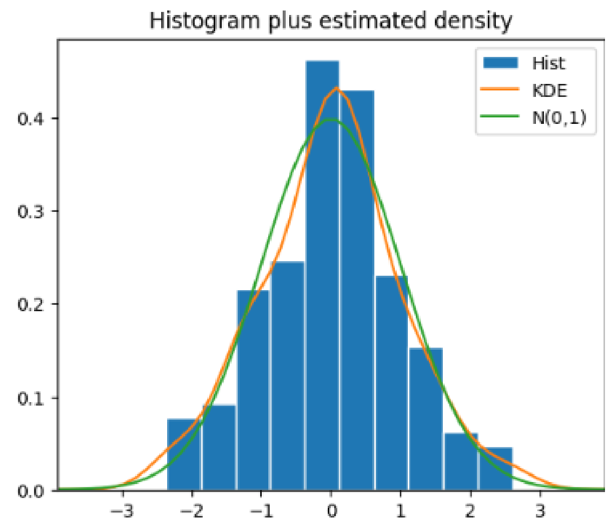

figure 5.11: Histogram of residuals

The figure 5.11 shows that the histogram matches the normal distribution curve, the residuals are approximately normal.

The diagnostic checking confirms that the fitted Holt-Winters model is statistically sound and adequate. Therefore, the model is suitable for forecasting monthly rubber prices.

In-sample Forecasting :

The fitted model is used to do In-sample forecasting. In-sample Forecasting done for the year 2023 is given in the below table 5.6

Table 5.6: In sample forecast values

Months	Actual Value	Predicted Value	Months	Actual Value	Predicted Values
January	14308	15727.5764	July	15616	16729.7945
February	14558	15888.3723	August	15004	16316.6359
March	14700	15891.1116	September	15131	15714.0856
April	15289	16102.2745	October	15413	15363.3154
May	16000	15845.4650	November	15638	15155.7924
June	15732	16294.2394	December	15376	15178.0221

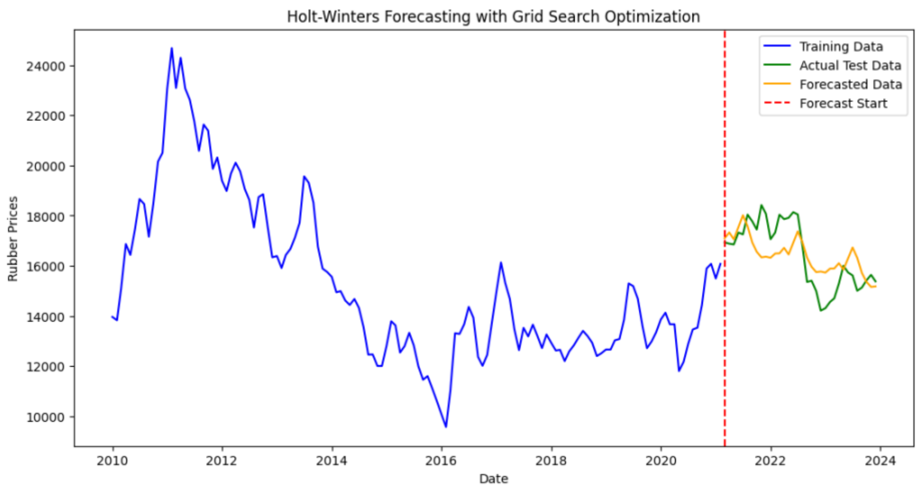

figure 5.12: Actual vs Predicted values plot

In figure 5.12 the blue line are the training data, green line is the actual testing data and orange line is the predicted testing data.

Forecasting of Rubber Prices using Holt-Winter's model:

Rubber Prices from January 2024 to December 2025 is forecasted using the model.

Table 5.7: Holt-Winter's forecast for RSS3

Months	Forecasted Values (Rs)	LCL	UCL	Months	Forecasted Values (Rs)	LCL	UCL
Jan 2024	17097.2327	12650.5432	22345.6789	Jan 2025	16494.1722	12482.1567	22143.9821
Feb 2024	17328.3000	10987.6543	20678.9012	Feb 2025	16715.2872	10785.6234	20548.7567

Mar 2024	17055.7767	11345.6789	19876.5432	Mar 2025	16450.6208	10172.3956	18632.4673
Apr 2024	17542.8032	12987.6543	21456.7890	Apr 2025	16918.5213	12867.9456	21369.1025
May 2024	18015.8392	13765.4321	23219.8765	May 2025	17372.8168	13505.9865	22121.8377
Jun 2024	17574.9508	12789.0123	22789.6543	Jun 2025	16945.7934	12511.7598	22546.1748
Jul 2024	16929.8393	11098.7654	20654.3210	Jul 2025	16321.9624	10935.3418	20477.3406
Aug 2024	16555.7750	10876.5432	19876.5432	Aug 2025	15959.5452	10632.5158	19625.4618
Sep 2024	16335.9614	10234.5678	19345.6789	Sep 2025	15745.8769	10063.6725	19296.5452
Oct 2024	16363.7693	9845.6789	19845.6789	Oct 2025	15770.8957	10586.5732	19639.9875
Nov 2024	16320.6795	11567.8901	18765.4321	Nov 2025	15727.5764	11376.4121	18560.9423
Dec 2024	16489.4281	11890.1234	19990.5678	Dec 2025	15888.3723	12134.9519	19754.8293

The Table 5.7 is the forecasted values and its LCL and UCL of rubber prices of RSS3 for January 2024 to December 2025.

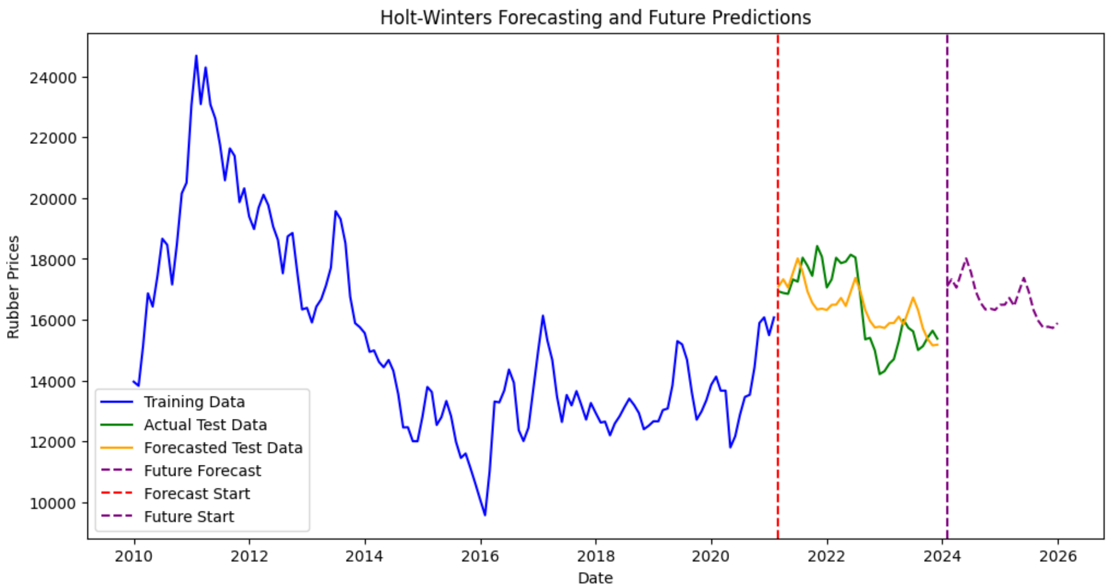

figure 5.13 : Actual, Predicted and Forecasted values

The figure 5.13 is the plot of forecasted values using Holt Winter's model.

5.3 MODELING AND FORECASTING OF RUBBER PRICES OF RSS1 USING SARIMA MODEL

The figure 5.1 presents the time series plot of monthly rubber price data, revealing noticeable seasonality in the data. However, visual inspection alone is insufficient to determine whether the changes in the mean are statistically significant. To assess this further, the ACF and PACF plots are examined, as shown in fig respectively.

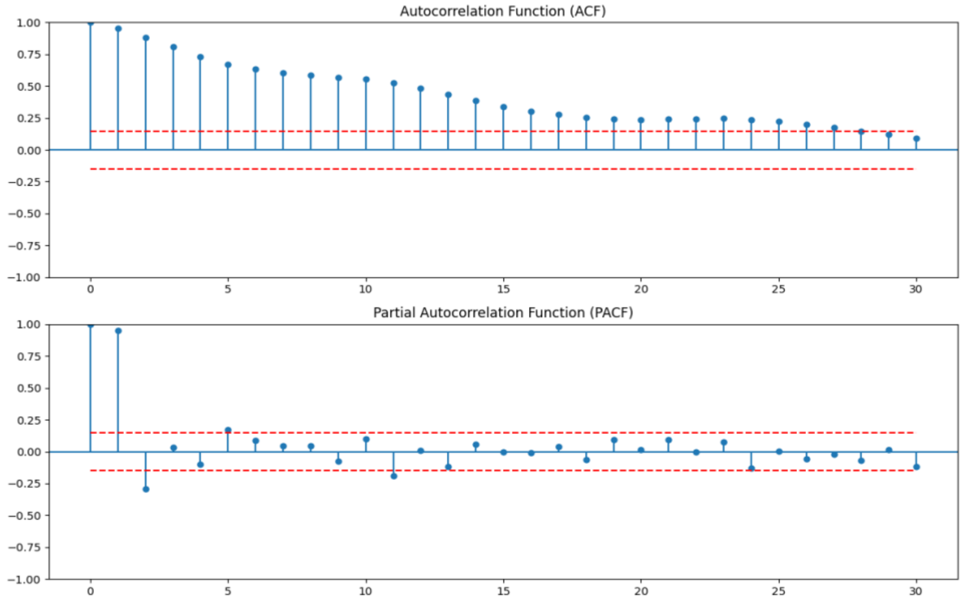

figure 5.14 : ACF and PACF of RSS1

The figure 5.14 indicates that the autocorrelation gradually decreases over time, but remains significant up to a considerable number of lags. This indicates the presence of **strong autocorrelation** and potential non-stationarity in the data. Now the seasonal index for various seasons can be obtained by ratio to moving average method.

Table 5.8: Seasonal Indices

Month	Seasonal Indices
January	0.967588
February	0.994633
March	1.013349
April	1.032182
May	1.031628
June	1.053073

July	1.064737
August	1.015458
September	0.970568
October	0.957002
November	0.946357
December	0.953433

Now set that, H_0 = *Data is non−stationary*

H_1 = *Data is stationary*

On performing the ADF test we get,

Dickey -Fuller : -2.6930401656524188

p- value : 0.07525038709459231

Since the p- value exceeds 0.05, we fail to reject the null hypothesis, indicating non-stationarity in the data. To achieve stationarity, we apply seasonal differencing by subtracting each value from its counterpart 12 periods prior (S=12), yielding the transformed data: $x_t = x_t - x_{t-12}$. The resulting time series plot of the seasonally differenced data of RSS1 is illustrated in the figure5.15.

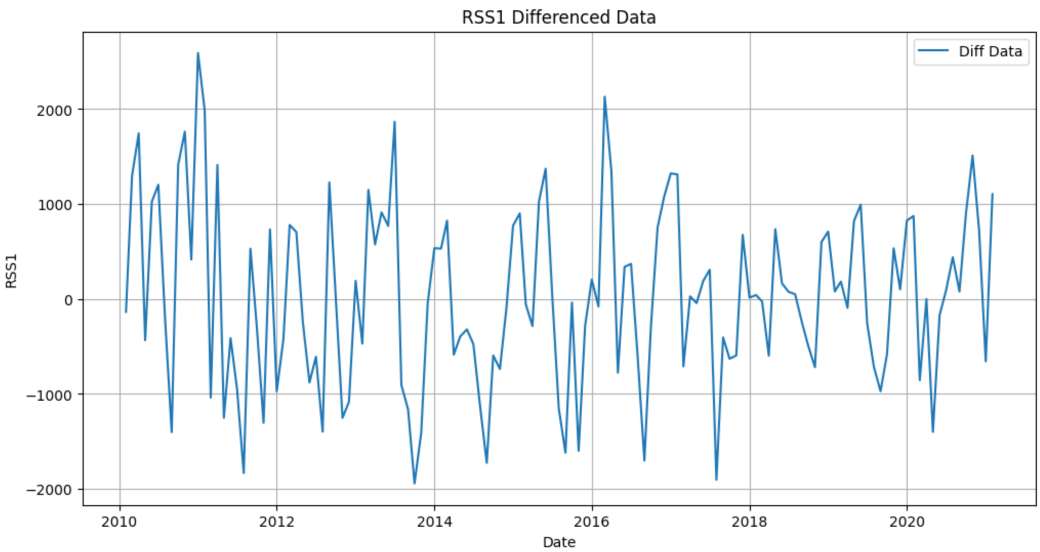

figure 5.15 : Differenced data

From the seasonal differenced figure 5.15 it is evident that the seasonal behaviour is removed from the series. Now again plot the seasonal differenced ACF and PACF.

ACF plot of Seasonally differenced data :

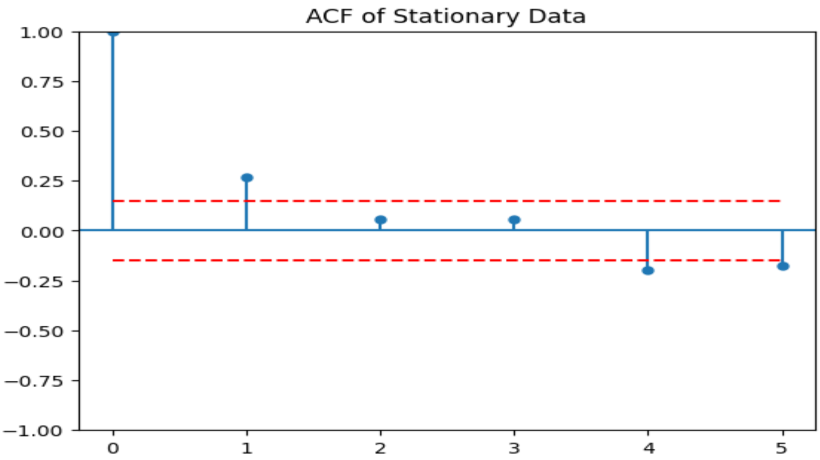

figure 5.16: ACF of Stationary data

PACF plot of Seasonally differenced data :

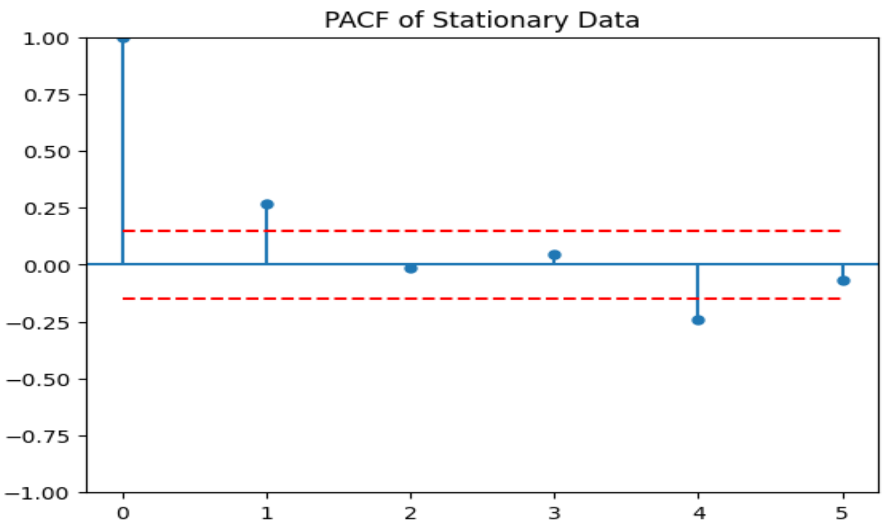

figure 5.17: PACF of Stationary data

Augmented Dickey-Fuller (ADF) Test :

To confirm that the differenced data is stationary, ADF test is done.

Dicky- Fuller : -4.283533020614704

p- value : 0.0004743009851349935

Since p-value is less than 0.05, it is clear that the differenced data is stationary. No more seasonal differencing is needed. Now D=1 and d=0 . Now we plot ACF and PACF of seasonally differenced data at seasonal lags to find P and Q (AR and MA order).

ACF of seasonally differenced data at seasonal lags :

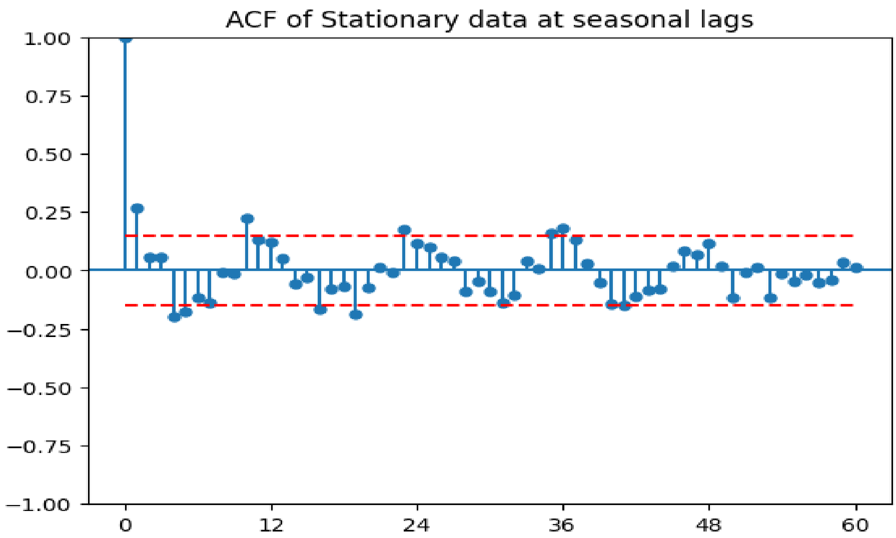

figure 5.18: ACF of Stationary data at Seasonal lags

PACF of seasonally differenced data at seasonal lags :

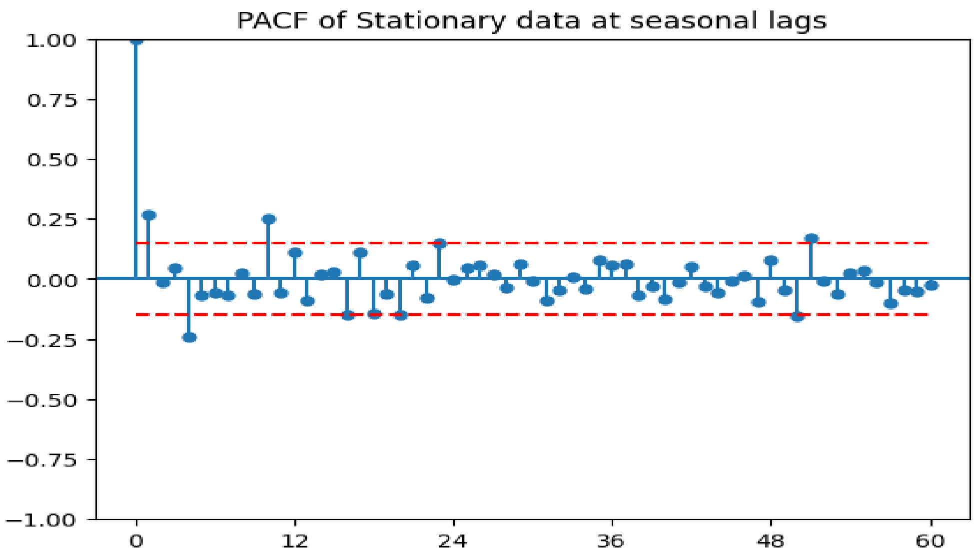

figure 5.19: PACF of Stationary data at Seasonal lags

From the above figures we get that, non-seasonal AR and MA order is maximum p=2, maximum q=3 and d=0. Similarly we get, seasonal AR and MA order is maximum P=3, maximum Q=4 and D=1.

Thus the possible time series models and their corresponding AIC statistics for the monthly rubber price data of RSS1 is given in Table 5.9.

Table 5.9: ARIMA models and their corresponding AIC values

NO.	ARIMA(p,d,q) x (P,D,Q)	AIC
1	ARIMA(0,0,0)(0,1,0)[12]	2069.365
2	ARIMA(1,0,0)(1,1,0)[12]	2029.511
3	ARIMA(0,0,0)(0,1,0)[12]	2067.645
4	ARIMA(1,0,0)(0,1,0)[12]	2063.344
5	ARIMA(1,0,0)(2,1,0)[12]	2008.914
6	ARIMA(1,0,0)(3,1,0)[12]	1999.322
7	ARIMA(1,0,0)(3,1,1)[12]	1993.827
8	ARIMA(1,0,0)(2,1,1)[12]	1992.300
9	ARIMA(1,0,0)(2,1,2)[12]	1996.907
10	ARIMA(0,0,0)(2,1,1)[12]	1995.808
11	ARIMA(2,0,0)(2,1,1)[12]	1993.736
12	ARIMA(0,0,1)(2,1,1)[12]	1991.523
13	ARIMA(0,0,1)(2,1,0)[12]	2008.394
14	ARIMA(0,0,1)(3,1,1)[12]	1993.166
15	ARIMA(0,0,1)(1,1,0)[12]	2028.331
16	ARIMA(0,0,1)(3,1,0)[12]	1999.217
17	<u>ARIMA(0,0,1)(2,1,1)[12]</u>	<u>1989.535</u>
18	ARIMA(0,0,1)(2,1,0)[12]	2006.482
19	ARIMA(0,0,1)(3,1,1)[12]	1991.170
20	ARIMA(0,0,1)(1,1,0)[12]	2026.489

Thus by evaluating the possible time series models we get that the model ARIMA(0 , 0 , 1) x (2 , 1 , 1)[12] with Akaike Information Criteria (AIC) value as 1989.535 as the more appropriate model.

The parameter estimates of model are

Table 5.10

Parameter	Coefficient	Standard Error	z	P> |z|
ar.L1	0.9673	0.029	32.897	0.00
ar.S.L12	-0.1665	0.135	-1.230	0.219
ar.S.L24	-0.0876	0.065	-1.342	0.179
ma.S.L12	-0.7914	0.170	-4.657	0.000
sigma2	6.065e+05	9.62e+05	6.306	0.000

The equation for the forecasted value is given as:

$$\boldsymbol{(1-0.9673\,B)(1-0.1665\,B^{12}-0.0876\,B^{24})(1-B^{12})Z_t=(1-0.\ 7914\,B^{12})\varepsilon_t}$$

Equation 5.3:SARIMA fitted equation

Diagnostic Checking :

Diagnostic checking is a vital step in statistical modelling that ensures the model's accuracy, reliability, and effectiveness. It assesses the model's precision and helps refine it to achieve better prediction accuracy.

figure 5.20 presents a Quantile-Quantile (Q-Q) plot, which compares the distribution of residuals to a normal distribution. The majority of residual values closely align with the straight line, suggesting that the residuals are approximately normally distributed.

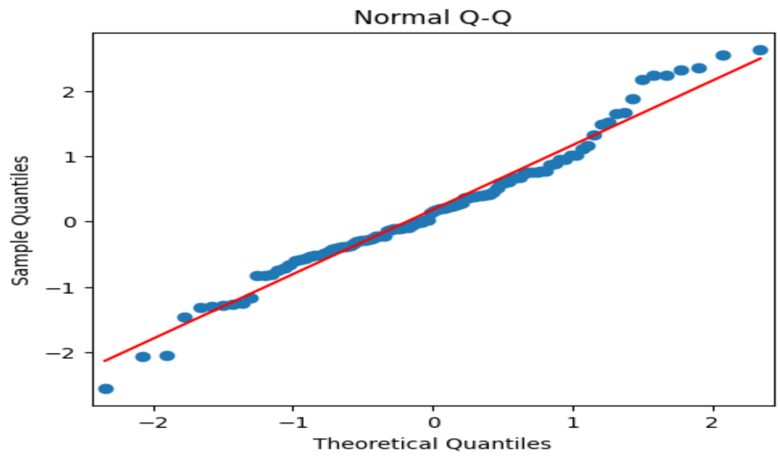

figure 5.20: Q-Q plot of Residuals

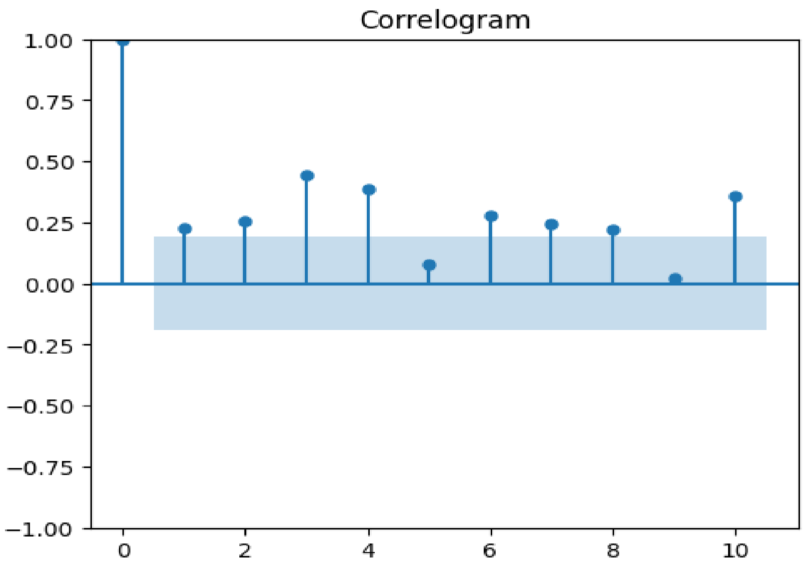

figure 5.21: Correlogram of Residuals

In the figure 5.21 of correlogram most spikes lie within the confidence interval (blue shaded region), suggesting no significant autocorrelation in residuals.

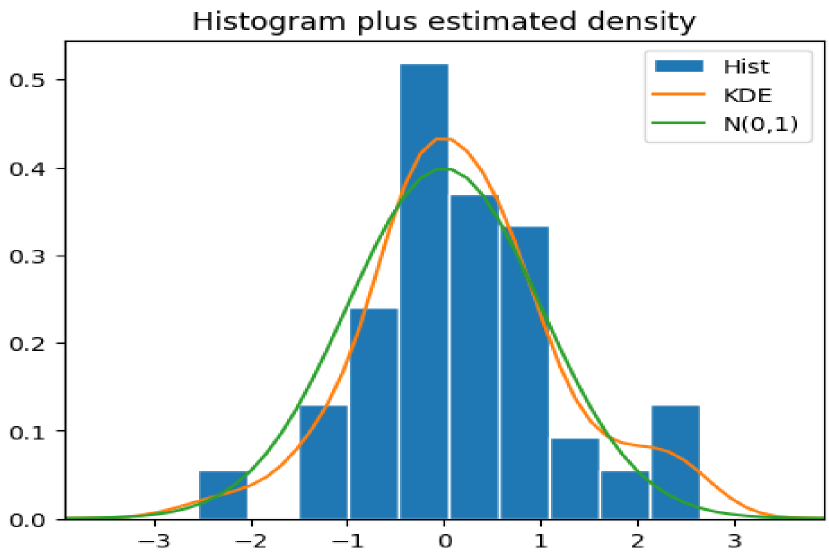

figure 5.22: Histogram of Residuals

In the figure 5.22, the histogram aligns reasonably well with the normal curve N(0,1), suggesting approximate normality of residuals.

Thus the diagnostic checking reveals that the fitted ARIMA (0,0,1) x (2,1,1)[12] model is statistically adequate. Also, the model satisfies stationary and invertibility requirements. So the model can be used to forecast the monthly Rubber price data of RSS1.

In-sample forecast :

The fitted time series model is now utilized for in-sample forecasting, focusing on the last year of the dataset, specifically January 2023 to December 2023.

Table 5.11: In sample forecast

Months	Actual Value	Predicted Value	Months	Actual Value	Predicted Value
January	15008	16016.9671	July	17996	16628.8798
February	15275	16552.3276	August	17040	16013.3611
March	15133	16513.7576	September	15962	15377.4795
April	15695	16525.4214	October	16013	15157.0373
May	16456	16367.9676	November	16162	15194.0329
June	17140	16608.1752	December	15840	15574.2999

Actual versus Predicted :

The actual versus the predicted values are plotted in fig5.23 and light blue line is the training data, orange line is the predicted values and dark blue is the test data.

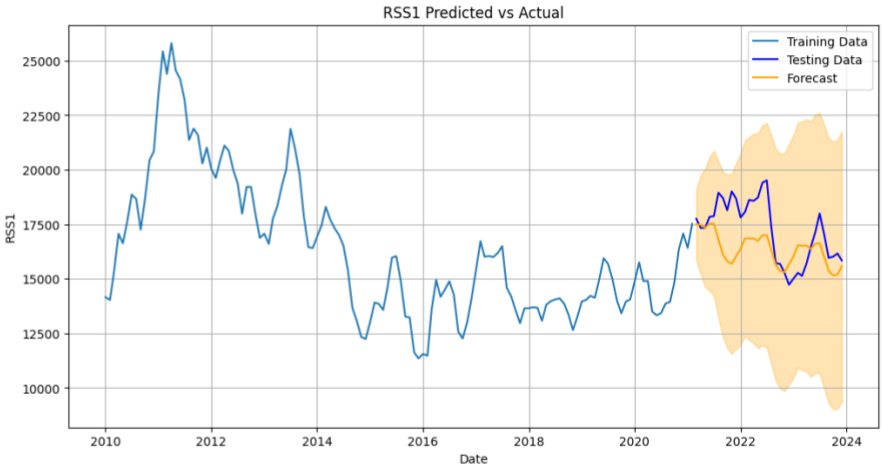

figure 5.23: Actual vs Predicted Values plot

Forecasting of Rubber prices using SARIMA model :

The prices of rubber of grade RSS1 from January 2024 to December 2025 is forecasted using the model fitted.

Table 5.12: SARIMA model forecast for RSS1

Months	Forecasted Values	LCL	UCL	Months	Forecasted Values	LCL	UCL
Jan 2024	17498.6479	12450.5432	22145.6789	Jan 2025	16849.7148	12345.1567	21943.9821
Feb 2024	17482.0050	10787.6543	20478.9012	Feb 2025	16850.0810	10685.6234	20348.7567
Mar 2024	17332.1911	11145.6789	19676.5432	Mar 2025	16745.9273	10072.3956	18432.4673
Apr 2024	17532.1075	12787.6543	21256.7890	Apr 2025	16997.0723	12767.9456	21169.1025
May 2024	17531.9102	13565.432	23019.8765	May 2025	17003.6264	13405.9865	21921.8377
Jun 2024	16815.4446	12589.0123	22589.6543	Jun 2025	16324.2084	12411.7598	22346.1748
Jul 2024	16117.2025	10998.7654	20454.3210	Jul 2025	15642.0920	10835.3418	20277.3406

Aug 2024	15807.6399	10676.5432	19676.5432	Aug 2025	15355.2587	10532.5158	19425.4618
Sep 2024	15686.0857	10034.5678	19145.6789	Sep 2025	15308.7335	10063.6725	19196.5452
Oct 2024	16031.6022	9625.6789	19645.6789	Oct 2025	15662.9013	10486.5732	19439.9875
Nov 2024	16365.2574	11367.8901	18565.4321	Nov 2025	16016.9671	11276.4121	18360.9423
Dec 2024	16858.5863	11690.1234	19790.5678	Dec 2025	16552.3276	12034.9519	19554.8293

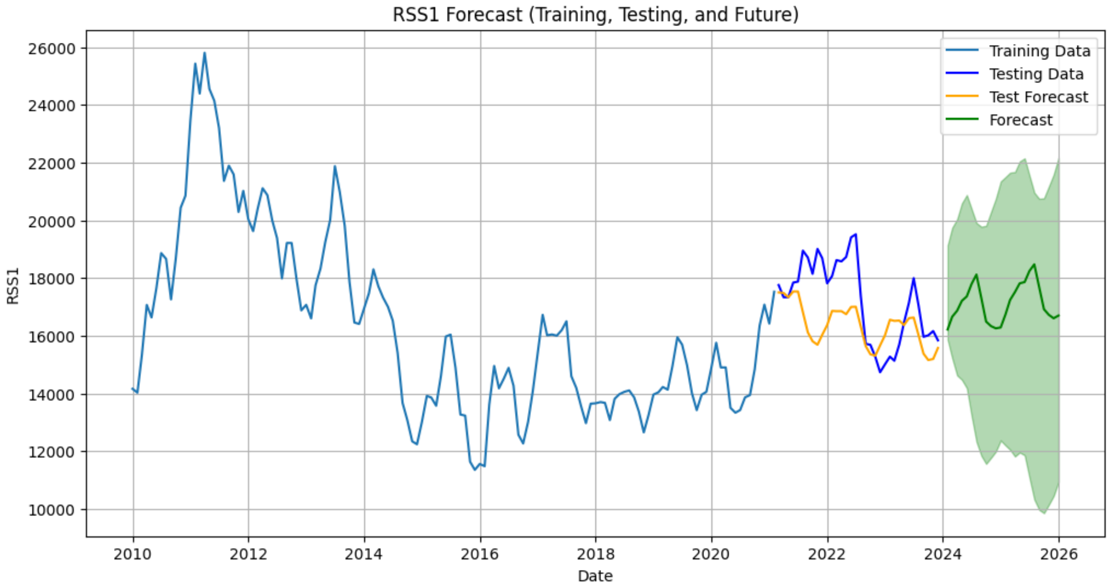

figure 5.24: Plot of forecasted values of RSS1

5.4 MODELING AND FORECASTING OF RUBBER PRICES OF RSS3 USING SARIMA MODEL

The figure 5.1 presents the time series plot of monthly rubber price data, revealing noticeable seasonality in the data. However, visual inspection alone is insufficient to determine whether the changes in the mean are statistically significant. To assess this further, the ACF and PACF plots are examined, as shown in fig respectively.

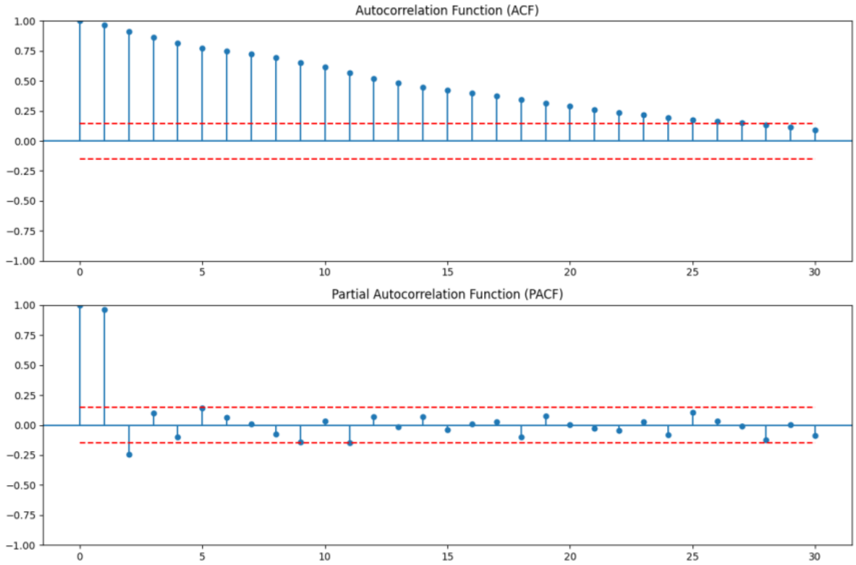

figure 5.25: ACF and PACF of RSS3

The fig 5.25 indicates that the autocorrelation gradually decreases over time, but remains significant up to a considerable number of lags. This indicates the presence of **strong autocorrelation** and potential non-stationarity in the data. Now the seasonal index for various seasons can be obtained by ratio to moving average method.

Table 5.13: Seasonal Indices

Month	Seasonal Indices
January	0.977924
February	0.990382
March	1.004616
April	1.021938
May	1.009572
June	1.024631
July	1.038888
August	1.012122
September	0.987418
October	0.980680
November	0.975853
December	0.975977

Now set that, $H_0 = Data\ is\ non{-}stationary$

$H_1 = Data\ is\ stationary$

On performing the ADF test we get,

Dickey -Fuller : -2.235191778203874

p- value : 0.19372162813031552

Since the p- value exceeds 0.05, we fail to reject the null hypothesis, indicating non-stationarity in the data. To achieve stationarity, we apply seasonal differencing by subtracting each value from its counterpart 12 periods prior (S=12), yielding the transformed data: $x_t = x_t - x_{t-12}$. The resulting time series plot of the seasonally differenced data of RSS3 is illustrated in the figure5.26.

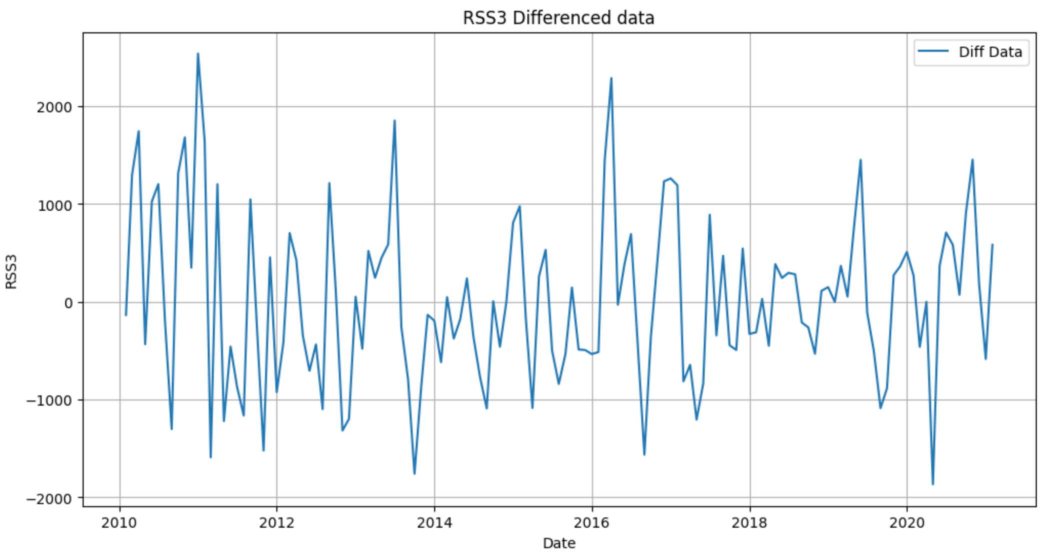

figure 5.26: Differenced data of RSS3

From the seasonal differenced figure 5.26 it is evident that the seasonal behaviour is removed from the series. Now again plot the seasonal differenced ACF and PACF.

ACF plot of Seasonally differenced data :

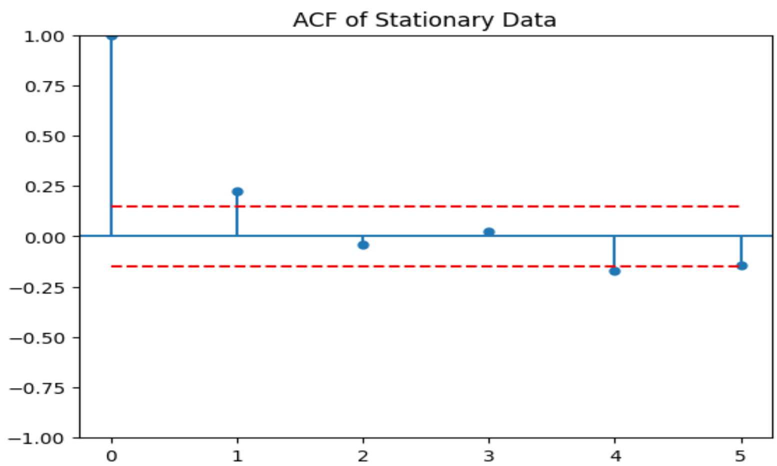

figure 5.27: ACF of Stationary data

PACF plot of Seasonally differenced data :

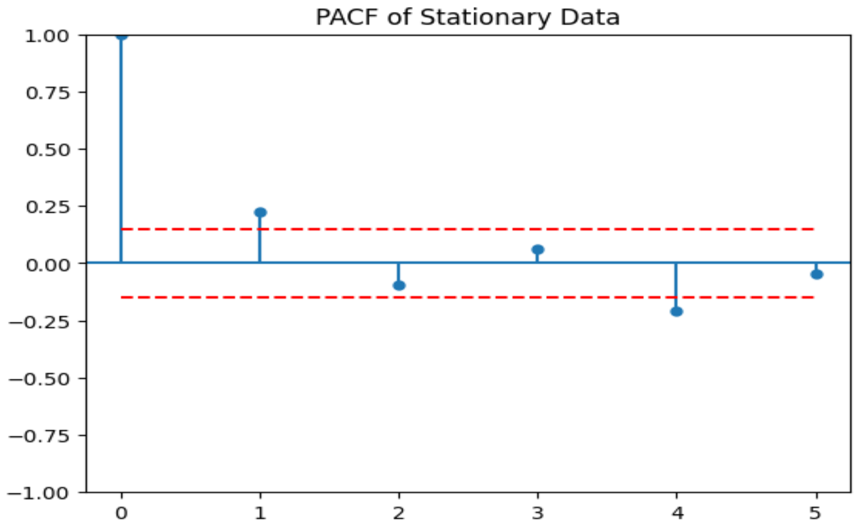

figure 5.28: PACF of Stationary data

Augmented Dickey-Fuller (ADF) Test :

To confirm that the differenced data is stationary, ADF test is done.

Dicky- Fuller : -7.0914751951317605

p- value : 4.398830065116639e-10

Since p-value is less than 0.05, it is clear that the differenced data is stationary. No more seasonal differencing is needed. Now D=1 and d=0 . Now we plot ACF and PACF of seasonally differenced data at seasonal lags to find P and Q (AR and MA order).

ACF of seasonally differenced data at seasonal lags :

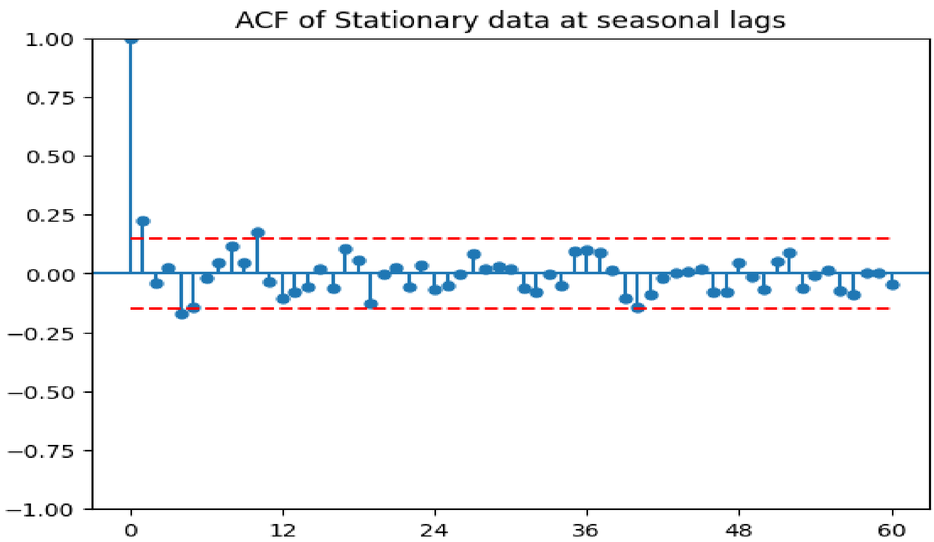

figure 5.29: ACF of Stationary data at seasonal lags

PACF of seasonally differenced data at seasonal lags :

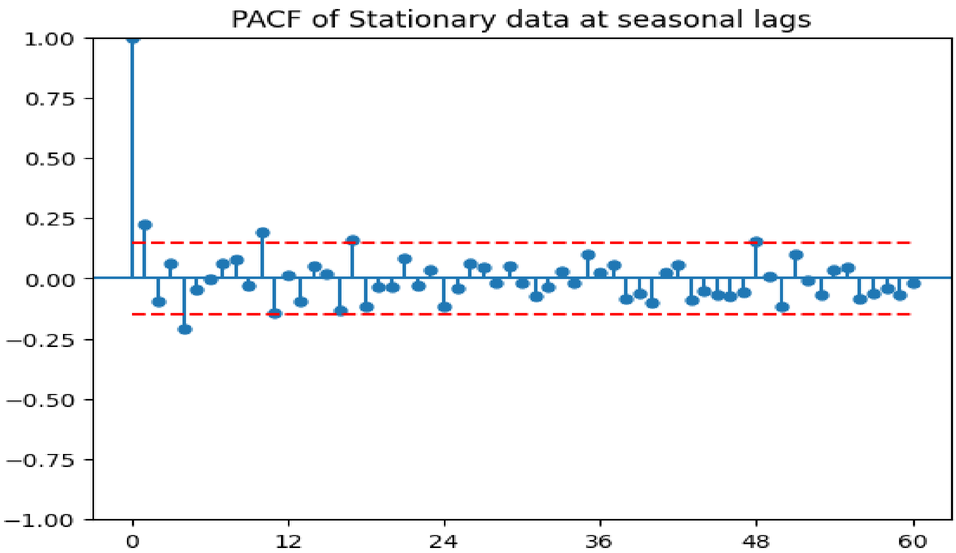

figure 5.30: PACF of Stationary data at Seasonal lags

From the above figures we get that, non-seasonal AR and MA order is maximum p = 2, maximum q = 3 and d = 0. Similarly we get, seasonal AR and MA order is maximum P = 4, maximum Q = 4 and D = 1.

Thus the possible time series models and their corresponding AIC statistics for the monthly rubber price data of RSS3.

Table 5.14: ARIMA models and their corresponding AIC values

NO.	ARIMA(p,d,q) x (P,D,Q)	AIC
1	ARIMA(0,0,0)(0,1,0)[12]	2064.691
2	ARIMA(1,0,0)(1,1,0)[12]	2031.964
3	ARIMA(0,0,0)(0,1,0)[12]	2063.013
4	ARIMA(1,0,0)(0,1,0)[12]	2062.335
5	ARIMA(1,0,0)(2,1,0)[12]	2017.469
6	<u>ARIMA(1,0,0)(3,1,0)[12]</u>	<u>1865.821</u>
7	ARIMA(0,0,0)(3,1,0)[12]	1997.351
8	ARIMA(2,0,0)(3,1,0)[12]	1989.252
9	ARIMA(1,0,1)(3,1,0)[12]	2003.017
10	ARIMA(0,0,1)(3,1,0)[12]	2079.293
11	ARIMA(2,0,1)(3,1,0)[12]	2189.359
12	ARIMA(1,0,0)(3,1,0)[12]	1989.356
13	ARIMA(1,0,0)(2,1,0)[12]	2133.654
14	ARIMA(0,0,0)(3,1,0)[12]	1975.482
15	ARIMA(2,0,0)(3,1,0)[12]	2045.687
16	ARIMA(1,0,1)(3,1,0)[12]	1905.367
17	ARIMA(0,0,1)(3,1,0)[12]	1952.147

18	ARIMA(2,0,1)(3,1,0)[12]	2099.349
19	ARIMA(1,0,0)(3,1,1)[12]	1962.517
20	ARIMA(1,0,0)(2,1,1)[12]	1956.241

Thus by evaluating the possible time series models we get that the model ARIMA(1, 0 , 0) x (3, 1 , 0)[12] with Akaike Information Criteria (AIC) value as 1865.821 as the more appropriate model.

The parameter estimates of model are

Table 5.15: ARIMA(1, 0 , 0) x (3, 1 , 0)[12] model parameters

Parameter	Coefficient	Standard Error	z	P> |z|
ma.L1	0.8489	0.073	11.694	0.000
ar.S.L12	-0.1248	0.113	-1.105	0.269
ar.S.L24	0.0828	0.121	0.684	0.494
ar.S.L36	-0.0064	0.080	-0.080	0.936
sigma2	1.841e+06	2.79e+05	6.602	0.000

The equation for the forecasted value is given as:

$$\mathbf{(1 - 0.1248\, B^{12} + 0.0828\, B^{24} - 0.0064\, B^{36})(1 - B^{12}) Z_t = (1 + 0.8489\, B)\varepsilon_t}$$

Equation 5.4: SARIMA fitted equation

Diagnostic Checking:

Diagnostic checking is a vital step in statistical modelling that ensures the model's accuracy, reliability, and effectiveness. It assesses the model's precision and helps refine it to achieve better prediction accuracy.

Figure 5.31 presents a Quantile-Quantile (Q-Q) plot, which compares the distribution of residuals to a normal distribution. The majority of residual values closely align with the straight line, suggesting that the residuals are approximately normally distributed.

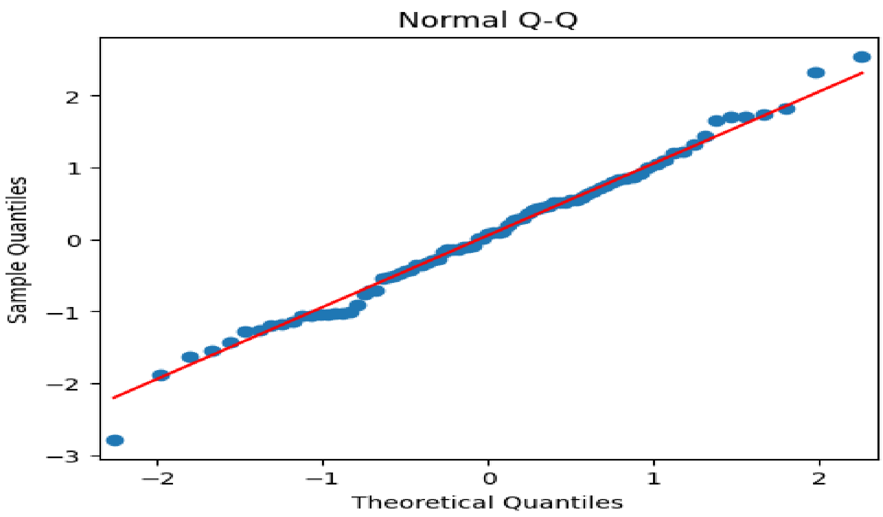

figure 5.31: Q-Q plot of Residuals

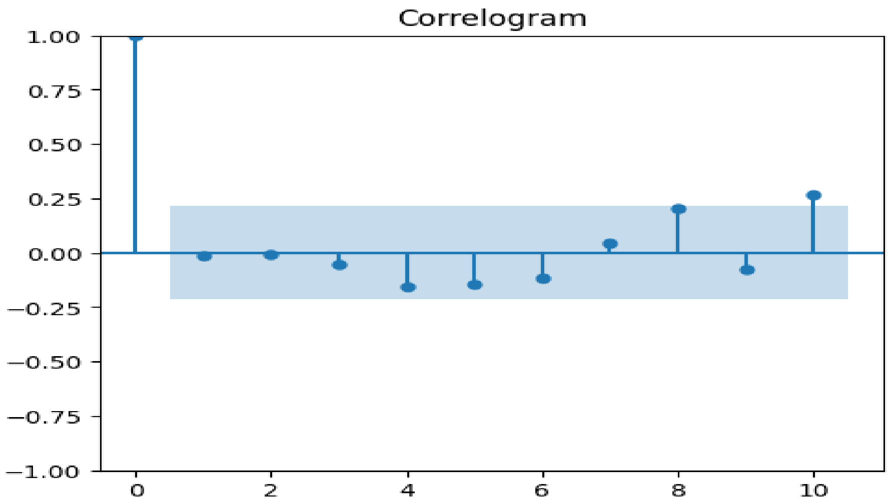

figure 5.32: Correlogram of Residuals

In the figure 5.32 of correlogram most spikes lie within the confidence interval (blue shaded region), suggesting no significant autocorrelation in residuals.

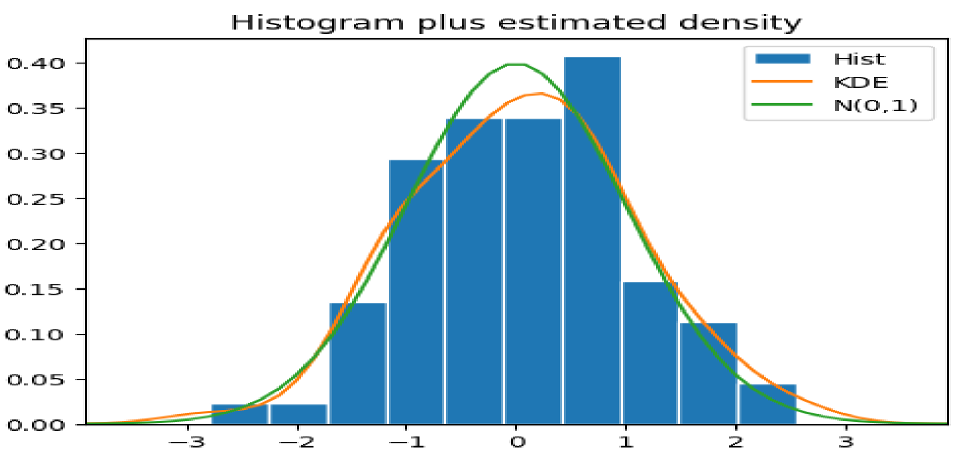

figure 5.33: Histogram of Residuals

In the figure 5.33 , the histogram aligns reasonably well with the normal curve N(0,1), suggesting approximate normality of residuals.

Thus the diagnostic checking reveals that the fitted ARIMA (0,0,1) x (3,1,0)[12] model is statistically adequate. Also, the model satisfies stationary and invertibility requirements. So the model can be used to forecast the monthly Rubber price data of RSS3.

In-sample forecast :

The fitted time series model is now utilized for in-sample forecasting, focusing on the last year of the dataset, specifically January 2023 to December 2023.

Table 5.16: In sample forecasts

Months	Actual Value	Predicted Value	Months	Actual Value	Predicted Value
January	14308	14943.4592	July	15616	15353.5088
February	14558	15093.5834	August	15004	15087.3847
March	14700	15077.7452	September	15131	14705.3299
April	15289	15136.8067	October	15413	14528.9594
May	16000	14747.8681	November	15638	14536.4954
June	15732	15054.1140	December	15376	14745.9669

Actual versus Predicted :

The actual versus the predicted values are plotted in figure 5.34 and light blue line is the training data, orange line is the predicted values and dark blue is the test data.

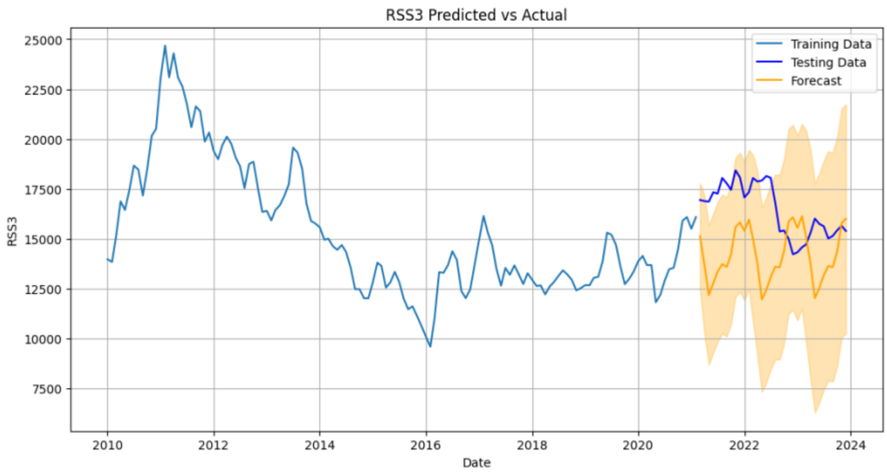

figure 5.34: Actual vs Predicted plot

Forecasting of Rubber prices using SARIMA model :

The prices of rubber of grade RSS3 from January 2024 to December 2025 is forecasted using the model fitted.

Table 5.17: SARIMA model forecasts for RSS3

Months	Forecasted Values	LCL	UCL	Months	Forecasted Values	LCL	UCL
Jan 2024	16092.8896	10548.2158	19268.0124	Jan 2025	15354.2853	10482.3698	19206.7432
Feb 2024	16146.5279	10984.2157	19035.3256	Feb 2025	15412.7351	10929.3698	19006.7432
Mar 2024	15924.6757	9147.6542	19243.6520	Mar 2025	15156.1739	9034.9876	19204.1234
Apr 2024	16093.8212	11035.3571	19987.3651	Apr 2025	15461.9496	11004.2157	19958.9012
May 2024	16334.5388	10654.2057	19872.3297	May 2025	15717.3358	10604.2057	19843.8765
Jun 2024	15907.1356	9958.2147	18832.6945	Jun 2025	15370.8530	9834.9876	18795.8765
Jul 2024	15486.8835	9831.0465	18624.3927	Jul 2025	14938.6067	9731.0465	18595.8765
Aug 2024	15148.0216	9364.8520	18643.9256	Aug 2025	14657.7940	9264.8520	18604.9012
Sep 2024	14829.7601	9047.3653	18035.9765	Sep 2025	14538.3998	8947.3653	18006.8765
Oct 2024	15016.0250	10584.3692	18976.2548	Oct 2025	14747.0195	10534.3692	18946.2548
Nov 2024	15245.4184	10684.3529	19163.6842	Nov 2025	14943.4592	10634.3529	19134.6842
Dec 2024	15334.1248	10023.3473	18354.0548	Dec 2025	15093.5834	10003.3473	18325.0548

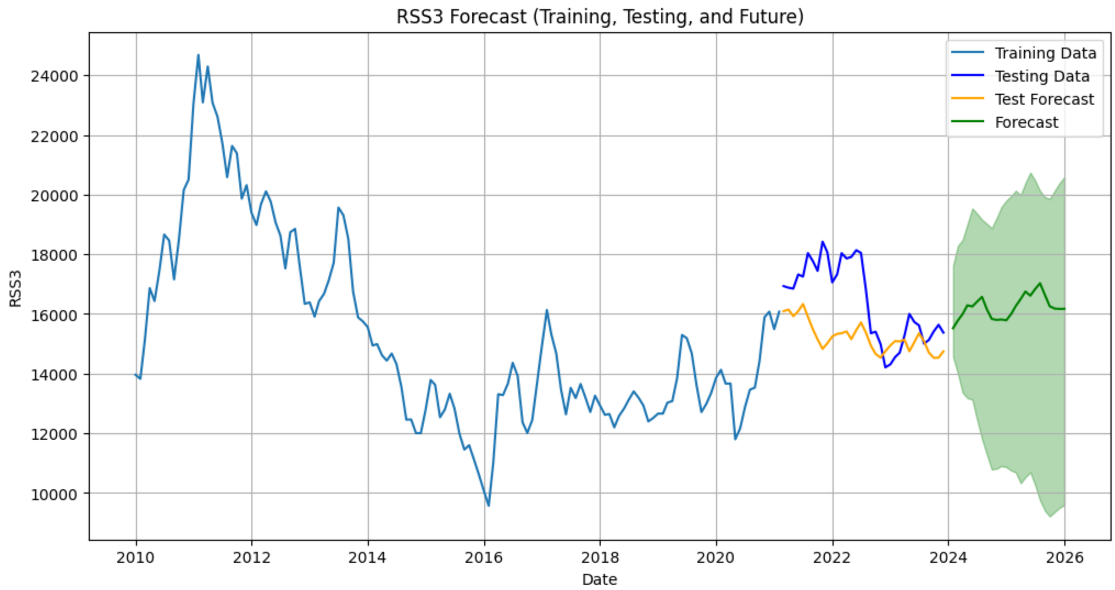

figure 5.35: Plot of forecasted values of RSS3

5.5 COMPARISON BETWEEN SARIMA AND HOLT-WINTER's MODEL

To determine the best model, performance metrices like Mean Absolute Error (MAE) and Root Mean Square Error is provided for the models

Table 5.18: RMSE and MAE values

		SARIMA Model	Holt-Winter's Model
RSS1	MAE	1028.25483167512	902.132895670896
	RMSE	1534.368953102475	1052.6545712051168
RSS3	MAE	1069.6584392175	841.373498511357
	RMSE	1684.5443951266758	984.7016247779933

Thus for both grades of rubber RSS1 and RSS3, Holt-Winter's have lower MAE and RMSE values, so Holt-Winter's model is a better performer than SARIMA model.

CONCLUSION

Two time series models were employed to forecast monthly rubber price values of two grades of rubber RSS1 and RSS3 from January 2024 to December 2025 utilising the historical data of two grades from January 2010 to December 2023. The time series models used are SARIMA and Holt-Winter's exponential smoothing technique. It was concluded that the Holt-Winter's model is the best model for both grades of rubber. In case of RSS1, the Holt-Winter's model has a low RMSE of 1052.655 and MAE of 902.133 when compared to that of SARIMA which has RMSE OF 1534.369 and MAE of 1028.254. In case of RSS3, the Holt-Winter's model has a low RMSE of 984.702 and MAE of 841.373 when compared to that of SARIMA which has RMSE OF 1684.544 and MAE of 1069.658. Both models provide effective forecasting solutions. However, if computational efficiency is a top priority, Holt-Winter's Exponential Smoothing is suitable for simpler datasets. For datasets with clear statistical properties and stationarity, SARIMA is a robust choice, but requires careful parameter tuning, which can be time-consuming.

In conclusion, this study sheds light on the historical trends and patterns in rubber prices, revealing valuable insights for stakeholders. While this study informs about the past, it also paves the way for future exploration in the realm of commodity price forecasting, empowering stakeholders to make informed decisions in an ever-changing market landscape.

REFERENCES

1. Anuja, A. R., Kar, A., Jha, G. K., & Kumar, R. (2013). Price dynamics and market integration of natural rubber under major trade regimes of India and abroad. *The Indian Journal of Agricultural Sciences*, *83*(5).
2. Chawla, D., & Jha, V. S. (2009). Forecasting production of natural rubber in India. *Paradigm*, *13*(1), 39-55.
3. George, J. G., & Chandrashekar, H. M. (2014). Growth and trends in production and marketing of natural rubber in Kerala, India. *International Journal of Current Research and Academic Review*, *2*(8), 53-61.
4. Kohjiya, S. (2015). *Natural rubber*. Smithers Rapra.
5. Mathew, S., & Murugesan, R. (2020). Indian natural rubber price forecast-an autoregressive integrated moving average (ARIMA) approach. *The Indian Journal of Agricultural Sciences*, *90*(2), 19-22.
6. Nithin, K. N., & Mahajanashetti, S. B. (2017). Trend and Growth Rate Analysis of Area, Production and Productivity of Natural Rubber in India. *Research Journal of Agricultural Sciences*, *8*(3), 731-734.
7. Sukiyono, K., Yuliarso, M. Z. S. P., Utama, E., Yuliarti, R., Novanda, R., & Priyono, B. S. (2019). Possible method For monthly naturalRubber price forecasting. *Journal of Advanced Research in Dynamical and Control Systems–Jardcs*, *11*, 387-395.
8. Vasagan, V. T., & Chakraborty, A. (2021). Demand and Supply Trend of Rubber in India: An Elucidative Analysis. *Indexed in*, *12*(2), 54-67.
9. Venkatachalam, P., Geetha, N., Sangeetha, P., & Thulaseedharan, A. (2013). Natural rubber producing plants: An overview. *African Journal of Biotechnology*, *12*(12).