Project Report

On

COMPARATIVE ANALYSIS OF ERNAKULAM RICE MARKET TRENDS USING TIME SERIES AND REGRESSION

Submitted

in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE

in

APPLIED STATISTICS AND DATA ANALYTICS

by

SUMA. A. PAI

(Reg No. SM22AS022)

(2022-2024)

Under the Supervision of

ANAKHA KURIAKOSE

ST. TERESA'S COLLEGE (AUTONOMOUS) ESTD 1925

DEPARTMENT OF MATHEMATICS AND STATISTICS (SF)

ST. TERESA'S COLLEGE (AUTONOMOUS)

ERNAKULAM, KOCHI – 682011

APRIL 2024

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, **COMPARATIVE ANALYSIS OF ERNAKULAM RICE MARKET TRENDS USING TIME SERIES AND REGRESSION** is a bonafide record of the work done by Ms. **SUMA.A.PAI** under my guidance as partial fulfilment 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: 29-04-24

Place: Ernakulam

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

Anakha Kuriakose
Assistant Professor,
Department of Mathematics and Statistics (SF)
St. Teresa's College (Autonomous)
Ernakulam.

Nisha Oommen
Assistant Professor& HOD
Department of Mathematics and Statistics (SF)
St. Teresa's College (Autonomous)
Ernakulam.

External Examiners

1. ....
CHINU JOSEPH
29/4/2024.

2. ....
LAKSHMI SURESH
29/04/2024

DECLARATION

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

Ernakulam

SUMA.A.PAI

Date: 29-04-24

SM22AS022

ACKNOWLEDGEMENTS

I must mention several individuals who encouraged me to carry out this work. Their Continuous invaluable knowledgeable guidance throughout this study helped me to Complete the work up to this stage.

I am very grateful to my project guide Anakha Kuriakose for the immense help during The period of work.

In addition, the very energetic and competitive atmosphere of the Department had much to do with this work. I acknowledge with thanks to the faculty, teaching, and non- teaching staff of the department and Colleagues.

I am also very thankful to HOD Mrs. Nisha Oommen for their valuable suggestions, critical examination of work during the progress.

Ernakulam SUMA.A.PAI

Date: SM22AS022

ABSTRACT

Rice price holds significant economic, social and political implications. Rice constitute 7.46% of production food grains in Indian agriculture. In this project, the focus was on forecasting rice prices of Ernakulam market of Ernakulam district of Kerala using statistical models like ARIMA (Autoregressive Integrated Moving Average) and linear regression. The study included collecting historical daily data of rice prices of Ernakulam market over a period of time. This data was analysed to identify patterns, trends and non seasonal variations. The ARIMA model was used to capture the complex time series nature of data, considering factors like auto correlation and moving average. Exponential smoothing model was used to smoothing data and for forecasting data points. Linear regression was used to understand the relationship between the variables. By combining ARIMA, Exponential Smoothing, and linear regression techniques, accurate forecasts were evolved for future rice prices. Finally comparison of all three models is made and best model was chosen.

ST.TERESA'S COLLEGE (AUTONOMOUS) ERNAKULAM

Certificate of Plagiarism Check for Dissertation

Author Name	SUMA. A. PAI
Course of Study	M.Sc. Applied Statistics& Data Analytics
Name of Guide	Ms. ANAKHA KURIAKOSE
Department	Post Graduate mathematics & Statistics
Acceptable Maximum Limit	20%
Submitted By	library@teresas.ac.in
Paper Title	COMPARATIVE ANALYSIS OF ERNAKULAM RICE MARKET TRENDS USING TIME SERIES AND REGRESSION
Similarity	0% AI 9%
Paper ID	1685619
Submission Date	2024-04-23 13:02:17

Signature of Student

Signature of Guide

Checked By
College Librarian

* This report has been generated by DrillBit Anti-Plagiarism Software

CONTENTS

CHAPTER 1

INTRODUCTION

Agricultural commodity prices give both producers and consumers a signal about how productions made and how output consumed changes in relative prices of various agricultural commodities. The affect of allocation especially crop area in particular, among agricultural products, generally by producers. Among other things, price plays a strategic role in influencing cultivation of crops. In fact, rice price analysis is very important for both producers as well as political decision makers when formulating development plans. Market plays an important role in facilitating the exchange of goods and services. Market also provides signals of true costs. The price of commodity also helps to spatially balance the value of crops after considering transfer costs between markets. Agricultural markets of most of the developing countries are not well integrated due to poor development. Recent advances in time series analysis, particularly relevant to market co-integration studies.

Rice is a staple food in Kerala, playing a vital role in the diet of millions of people across the district. The price of rice holds significant economic, social, and political implications. The correlation between rice prices and various factors in India is a topic of great interest and complexity.Rice constitute 7.46% of production of food grains in Indian agriculture. It is also a staple food in south and east part of India. The study of price of rice will give us an extensive outlook towards its market activities and also provide insights on inflation and other economic policies. This project will try to predict future rice price values for in Ernakulam market of Ernakulam District of Kerala. Under the study the semi dwarf variety of rice 'Jaya' which was derived from IR – 8 is considered. Jaya Rice is cultivated mainly in Andhra Pradesh and Thelungana. Jaya Rice is popular due to its ability to stay edible several hours after it cooked. Due to this characteristic, Jaya Rice is most popular in Southern parts of Kerala. From this study we can predict the prices for the years 2024 and 2025

1.1 About the Data

For the purpose of study, the data was collected from the official website (https://agmarket.gov.in). The data collected from the website consists of daily data of Ernakulam market for Ernakulam district in Kerala which contains Minimum and Maximum price of Jaya rice in quintal.

1.2 Objectives of the Study

- To model & forecast the rice prices of Ernakulam market of Kerala using ARIMA models
- To model & forecast the future values of rice price for Ernakulam District using Exponential smoothing
- To model & forecast the future values of rice price for Ernakulam District using linear regression
- To compare the forecast by ARIMA , Exponential smoothing & linear regression and to find best model

CHAPTER 2

REVIEW OF LITERATURE

Alexander et al. (1994) proposes improvements in using regression analysis to measure spatial market integration. They suggest an error-correction mechanism as an alternative to Ravallion's procedure, which is valid only under certain conditions. The new method allows for testing exogeneity and identifying the direction and strength of causality in price formation between markets. The authors demonstrate the approach using data on rice prices in various parts of the Indonesian market. The findings support the notion that supply sources have a greater influence on price fluctuations than demand sources.

Silvapulle and Surya (1994) propose the use of Johansen's multiple cointegration technique as a test for spatial market integration, which they argue surpasses previous methods' limitations. The application of this technique to the Philippines' rice markets reveals overall strong long-term integration with Manila as the dominant market, while also highlighting significant short-term inter-regional relationships.

Ismet et al. (1998) analyzes the spatial price relationships in Indonesian rice markets using multivariate cointegration tests on weekly price data from 1982 to 1993. They found that post-self-sufficiency, after 1984, had a smaller degree of market integration compared to the pre-self-sufficiency period. This shift in policy allowed the government to decrease intervention without significantly impacting market integration, indicating a responsive private sector. Government intervention through rice procurement positively influenced market integration in the post-self-sufficiency period, suggesting potential cost reductions by lowering procurement prices. Additionally, higher regional per capita income was related to increased market integration, indicating the possibility of reducing intervention during economic growth.

Ghosh (2000) conducted a study on rice markets in India using co-integration and ML methods to explore intra-state and inter-state spatial integration. The findings have crucial policy implications: If food markets are identified to be spatially integrated, the government can reduce its efforts to influence prices in those markets. Redundant programs like procurement or open market sales of food grains in all markets can be avoided, as scarcity in one market will be transmitted to others. Market integration ensures regional food security by balancing food-deficit, food-surplus, and non-food cash crop regions. Hence, unnecessary restrictions on inter-regional food grain movement should be eliminated. Additionally, market integration provides insights into incentive transmission across marketing chains, making it vital for successful price policy and market liberalization programs in India. Dorosh and Shahabuddin (2002) focus on the Government of Bangladesh's efforts to ensure food security for all households. To achieve this objective, the government employs various strategies, including market interventions for price stabilization, targeted food distribution to the poor, and emergency relief after natural disasters. The study examines the variability of domestic and international rice prices and analyzes the balance between government intervention and private sector involvement in rice markets. The research reveals that the stability of rice prices during the 1990s was largely attributed to private sector imports, which helped stabilize markets during periods of low production. On the other hand, domestic rice procurement had limited impact on raising producer prices at harvest time, involved only a small fraction of farmers, and incurred high costs due to excessive procurement prices set above market rates after successful harvests.

Abadan and Shabri (2014) proposes a hybrid methodology, combining empirical mode decomposition (EMD) and autoregressive integrated moving average (ARIMA), to forecast rice prices in Malaysian Ringgit per metric ton. The integration of EMD-ARIMA improves forecast accuracy compared to traditional ARIMA, as evident from lower root mean squared error (RMSE) and mean absolute error (MAE). This approach helps mitigate risks for rice producers, suppliers, consumers, and other stakeholders in the rice marketing and production industry

Ohyver and Pudjihastuti (2018) aims to develop a time series model using ARIMA to forecast rice prices. This model can assist the government in monitoring and controlling rice prices for the benefit of producers and consumers. The ARIMA (1, 1, 2) model proves suitable for medium quality rice data from January 2015 due to its good accuracy. Price stability is crucial for promoting healthy domestic trade.

Sarkar et al. (2022) addressed the issue of farmers facing profit losses due to price fluctuations caused by climatic changes and other factors. They emphasized the importance of price forecasting to reduce risks and manage fluctuations in the agricultural supply chain. The study focused on identifying effective data models for accurate and general price forecasts. The dataset used was from three districts in Bangladesh's Khulna Division. They applied Linear Regression and Neural Network methods to predict crop prices. The results showed that Linear Regression outperformed the Neural Network, but there is room for further improvement. The study highlights the significance of predictive analysis in mitigating the impact of uncertain climatic conditions and global warming on agricultural productivity. Sasi & Subramanian (2022) address the challenges of managing uncertainties in demand, customer behavior, and market trends during the pandemic. Focusing on rice sales demand in a Fair Price Shop

(FPS) in Kerala, India, the study utilizes time series techniques (ARIMA and DES) to model and forecast future demand. The forecast models significantly improve demand and inventory forecasting accuracy, outperforming empirical models, with ARIMA showing better performance for future forecasts. The research contributes valuable insights to enhance the efficiency of the Public Distribution System (PDS).

Reddy et al. (2022), it was found that price variances of horticultural commodities adversely impact the nation's GDP, causing significant financial and labor-related challenges for ranchers. To address this issue, the researchers proposed using time-series data analytics and machine learning models like SARIMA, Holt-Winter's Periodic technique, and Long Short-Term Memory (LSTM) Neural Networks to forecast rice prices in Andhra Pradesh, India. The results showed that the LSTM neural network method outperformed other approaches, making it the most suitable and accurate method for predicting rice prices from 2001 to 2020. This forecasting model can aid the farming production network in making informed decisions to manage and mitigate the risk associated with price fluctuations caused by climate change and global warming

CHAPTER 3

MATERIALS AND METHODS

3.1 TIME SERIES ANALYSIS

Time series consists of data that are collected, recorded and observed over regular intervals of time, time intervals mentioned here will be months, years or quarter. Time series analysis is used in various fields like research study, economics, business etc. It basically studies the data points collected over time and captures underlying patterns and trends.

The time series consists of four major components

- **Secular Trend(Tt)**
  Trend refers to long period changes. It shows definite and basic tendency of statistical data with the passage of time. It is smooth, regular and long term movement. It refers to general tendency of a statistical data to rise or fall or remain the same. For example in a series concerning population or national income an upward tendency can be noticed while in data on birth or death or illiteracy a downward tendency
- **Seasonal fluctuations(St)**
  Seasonal fluctuations are those fluctuations which occur with some degree of regularity within a specific period of one year or shorter. Climatic conditions, social customs, religious functions etc are the factors responsible for Seasonal fluctuations. For example the number of road accidents was very less during covid-19 due to lockdown and very high before covid-19 situation
- **Cyclic fluctuations(Ct)**
  Cyclic Fluctuations are periodic movements. These fluctuations occur at intervals (or periods) of more than a year. Business cycles and cycles work on Business and Economic series. These cyclic movements pass through different stages like prosperity, recession, depression and recovery. During these different stages, the time series show changes. These changes are called cyclic changes. Series related to prices production, demand etc undergo such cyclic changes.

- **Irregular fluctuations(It)**

  Irregular fluctuations are those caused by unusual, unexpected and accidental events. Effects of Earthquake, strike, flood etc lead to Irregular fluctuations. These events cause sudden transition of affairs from one sate to another. Irregular fluctuations are random in their nature. Their occurrence cannot be predicted unlike the other component of time series.

3.1.1 ARIMA Model

The abbreviation of ARIMA stands for Autoregressive Integrated Moving Average. The Autoregressive (AR) refers to autoregressive component in which it uses the past values of the variable being forecasted to predict future values. The Integrated (I) refers to integrated component in which observations undergo differencing ; subtracting each observation from its previous observation to make the time series stationary(stationarity is an assumption in time series which ensures that the statistical properties of the series do not change over time).The Moving Average (MA) component is used to model the error term of the time series, for that it finds the difference between the observed value and the predicted value by the autoregressive component.

Assumptions in ARIMA

- **Data should be stationary**- Stationarity is an assumption in time series which ensures that the statistical properties of the series do not change over time. A series with cyclic behavior and white noise can also be a stationary time series.
- **Data should univariate -** Auto regression (AR component) is the regression by the past values since ARIMA model works well with single variate data.

Steps to be followed in ARIMA

Step1: Exploratory Analysis- Exploratory Analysis involves three steps; Data Description, Data Visualization, Stationarity Check. In data description we should have a complete knowledge about our data domain, including its size, variables.

In data visualization, as the name suggest we present the data using line graph, bar graph, histograms, pie charts etc this is to explore more data features.

Stationarity check is the procedure of checking the stationarity of data. This is done using different tests such as Augmented Dickey-Fuller (ADF) Test, Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test, Phillips- Peron (PP) Test, Dickey-Fuller GLS (DF-GLS) Test. Other than these tests it can be also done by visualizing data points or visualizing Autocorrelation Function

(ACF) and Partial Autocorrelation Function (PACF) Plots. If the data is not stationary, we usually go for differencing; it is the process by which we compute the differences between consecutive observations.

Seasonality check is another procedure to check seasonality of data; this is done by plotting Autocorrelation Function (ACF) plot, Partial Autocorrelation Function (PACF) Plot, seasonal decomposition plot, seasonal subseries plot, seasonal box plot. To continue with ARIMA modeling it is necessary that the data should not contain seasonality i.e.; the data selected should not contain those fluctuations which occur with some degree of regularity within a specific period of one year or shorter.

Autocorrelation Function (ACF): The Autocorrelation Function (ACF) is a statistical tool used to measure the correlation between a time series and its lagged values, i.e. the ACF at lag k means the measures of the correlation between the time series and lagged values at lag k. In ARIMA model the ACF is particularly useful for determining the "AR" component. We first visualize the autocorrelation function by plotting the correlation coefficients against different lags. We try to identify the significant lags where the correlation coefficient is significantly different from zero. Then AR order is determined by the number of lags at which the autocorrelation is significant. AR order is denoted by ‘p’.

Partial Autocorrelation Function (PACF): The Partial Autocorrelation Function is another statistical tool used to measure the correlation between a specific lag and the current value, controlling the intermediate lags. We first visualize the partial autocorrelation function by plotting the partial correlation coefficients against different lags. We try to identify the significant lags where the correlation coefficient is significantly different from zero. Then AR order, usually denoted as p is determined by the number of lags at which the partial autocorrelation is significant.

Step2: Fitting the model- After checking the stationarity next step is to select the best (p, d, q) models from the above plots i.e.; Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) Plots to fit the best ARIMA model. Best (p, d, q) can be also determined from Information criteria, such as Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Grid Search and Hyper parameter Tuning etc. if the best model parameters are finalized divide the dataset completely into training and testing set. Typically the large set of points should be taken for training. Fit the model to the above chosen ARIMA model parameters for training data.

Step3: Diagnostic measures – The residuals are analyzed to ensure the errors and white noise in data set. For that we plot residuals in a graph, the graph obtained should be scattered. Validation of model performance is very crucial step after residual analysis. It helps us to calculate the accurate MSE and RMSE values. If correct MSE and RMSE values are obtained forecast the future values according to the requirement

3.1.2 SMOOTHING

In terms of Data analytics and Statistics smoothing refers to technique in which we can reduce the variability in data. By adopting smoothing techniques we can create a smooth representation of data. In this method, we usually calculate the average of nearby data points or use some mathematical model to the data trend. Smoothing methods are generally applied to those data sets which we are unable or difficult to find trend and patterns which may affect our prediction. These methods of smoothing are used in signal processing, time series analysis, machine learning etc

Exponential Smoothing Model

It is a most commonly used smoothing method for smoothing data and for forecasting data points as well in the time series analysis. It is effective in that data set which has a trend or seasonality. It is very understandable that today most of the data set in real world problems exhibit trend or seasonality, since this smoothing technique can be applicable to most of the datasets. In this method weights are assigned to all the observations in dataset. Usually the initial observations are given lesser weights than recent observations. This is a most popular method for forecasting since it is very simple and efficient.

Steps in Exponential smoothing model

Step1: Firstly we need an initial value to estimate the level of Time Series Analysis. This initial value is probably take as the 1st observation in a dataset or it can be determined by averaging the initial data points, taking mode of data points or using some statistical measure.

Step2: For each and every observation in dataset the smoothing value is calculated using the equation

$$S_t = \alpha \cdot Y_t + (1-\alpha) \cdot S_{t-1}$$

Where,

S_t = the smoothed value at time t

Y_t = actual observation at time t

S_{t-1} = smoothed ale of previous time

α = (alpha) smoothing parameter, known as smoothing constant

In the above equation the value of alpha plays a crucial role. It should be chosen according to the characteristics of the data and the desired level of smoothing. When we assign smaller value of alpha it results in more smoothing and vice versa

Step3: If the smoothed value for the current observation is computed successfully, the level estimate for the next iteration is updated and this value is considered as the smoothed value for next observation

Step4: Continue the above process for each observation, updating the level estimate and computing the smoothed value until reaching the last observation of the dataset.

Step5: Once all the above steps are completed we can forecast the future values as our requirements by utilizing the last smoothed value as the current level estimate and applying the above mentioned exponential smoothing formula.

Step6: After forecasting all the future values we can compute the mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE)

3.2 REGRESSION

Regression method in statistics is used to study and examine the relationship between one or more independent variables and a dependent variable. It is one of the commonly used methods to analyze and model the relationship among variables. The relationship between the variables studies or examined by fitting a mathematical model to observed data.

In regression analysis the dependent variable is termed as response variable, while the independent variable is termed as predictor variables or regressors. By fitting the previously mentioned mathematical we can determine how changes in independent variable affect the dependent variable.

3.2.1 Linear Regression Model

The Linear Regression is the commonly used regression analysis methods, where the relationship between the variables is assumed to be linear. By linear we mean that the dependent variable is proportional to changes in independent variables. But in some cases the variables will not be linear then also we can go with regression analysis techniques such as polynomial regression, exponential regression, and logistic regression.

<u>Steps in linear regression model</u>

Step1: Define the solving problem and the variable which you need to forecast or predict i.e.; define the dependent variable (target), and the independent variable predictor in the data set

Step2: All the historical data about the target variable and predictor variables should be collected. Make sure the data is accurate, relevant and covers the long time period to efficiently understand pattern and trends.

Step3: Data should be cleaned enough. It should not contain any errors, outliers or any missing values. These may entirely affect our forecasting results. If the data points are on different scales then transform the data variables to standardize form so that the variables are in the same domain or same range of values

Step4: After data preprocessing divide the data into training and testing set, the training set is used to train the regression model and testing set is used to validate the performance of the regression model chosen

Step5: Choose the best linear regression model i.e.; simple or multiple linear regressions based on the number of predictor variable. Fit the chosen model to the training dataset. The linear regression equation is,

$$Y = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + ... + \beta_n * X_n + \varepsilon$$

Where,

Y = Target variable

$x_{1,} x_{2,} \ldots\ldots\ldots\ldots, x_n$ = Predictor variable

β_0 = Intercept (constant term)

$\beta_1, \beta_2 \ldots\ldots.., \beta_n$ = Coefficients (weights) of the predictors.

ε = Error term

Step6: Using testing set evaluate the model performance, calculate the Mean Absolute Error (MAE), Mean Squared Error (MSE) and Root Mean Squared Error (RMSE). If the model is accurate go for forecasting the future values as requirements. Provide the input values for predictor variables for better prediction of future values.

3.3 Tools Used for Analysis

The present study make use of Microsoft excel and Python.

Microsoft Excel (2016) is widely used software which can be used for various statistical, mathematical, commercial computations. Data preprocessing, EDA forecasting etc is performed using this software.

Python is also widely used open source programming software which can be used for analysis of data in statistics. It is widely used among statistics and data minors for developing statistical software and analysis. Its libraries implement a wide variety of statistical and graphical techniques including linear and non linear modeling, time series analysis etc.

3.4 Tools Used for comparison

The present study uses Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) for comparison of models.

Mean Squared Error (MSE) is the average value of squared difference between actual and predicted values.

Root Mean Squared Error (RMSE) is the square root of Mean Squared Error (MSE).

Both Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) can be used to compare the difference model performance on certain data.

CHAPTER 4

ANALYSIS

4.1 DATA DESCRIPTION

For the purpose of the study, the data was collected from the official website (https://agmarket.gov.in). The data collected from the website consists of daily data of Ernakulam market for Ernakulam district in Kerala which contains Minimum and Maximum price of Jaya rice in quintal. For the convenience of the analysis the data was converted into monthly using Excel. The data converted consists of monthly data of the years 2003 to 2023

4.2 INTRODUCTION ABOUT THE CHAPTER

This analysis chapter of this study involves three important techniques used for forecasting and understanding patterns in data. As mentioned in previous chapter we use three methodologies; ARIMA Modelling, Exponential Smoothing & Linear Regression

ARIMA is the method that takes into account the non seasonal patterns in time series data. In this model it uses the historical trends and patterns and makes future predictions. In ARIMA modeling we assume that the data is stationary and univariate. Exponential Smoothing is effective in the data set which has a trend or seasonality. In this method weights are assigned to all the observations in dataset. Usually the initial observations are given lesser weights than recent observations. This is a most popular method for forecasting since it is very simple and efficient. Linear regression is used to study the relationship between two variables. Variables are assumed to be linear. By linear we mean that the dependent variable is proportional to changes in independent variables, where variables are assumed to be linear. By linear we mean that the dependent variable is proportional to changes in independent variables.

By using the above techniques our aim is to make useful insights and make predictions accurately

4.3 ARIMA MODEL

4.3.1 Minimum Price of Rice (1st variable)

4.3.1.1 Time series plot

The initial step in time series is to draw a time series plot. The time series plot of minimum price is given in Fig 4.3.1

Fig 4.3.1

4.3.1.2 Decomposition of time

The second step is to perform seasonal decomposition to capture the trend, seasonal and random components of time series. Plot the seasonal plot. Fig 4.3.2 depicts the seasonal plot

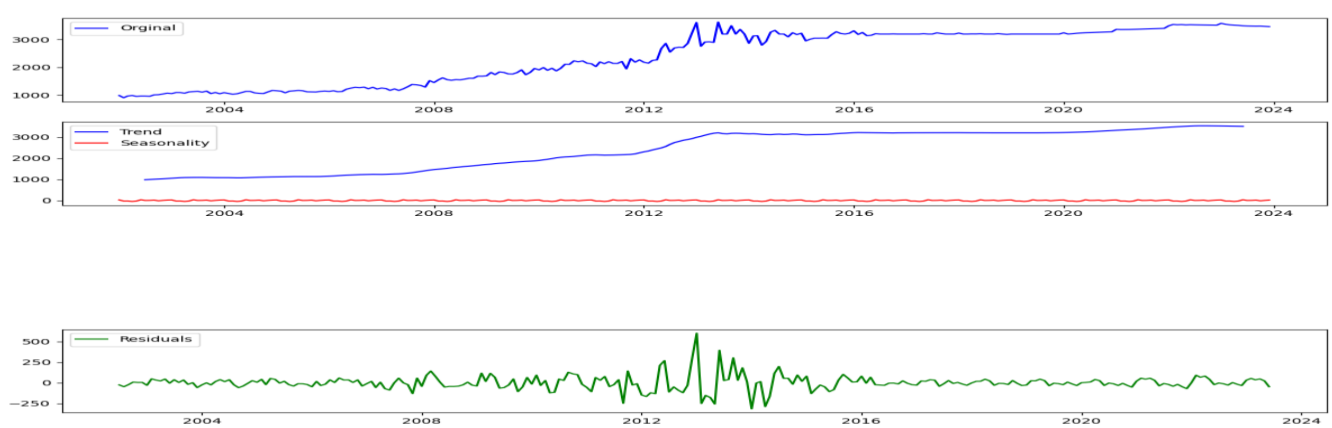

Fig 4.3.2

From the fig.4.3.2, it can conclude that the data has no seasonality

4.3.1.3 Stationarity check using Augmented Dickey- Fuller Test

To test the time series data for stationarity using ADF test, follows a hypothesis testing approach.

The null hypothesis H_0 is given by,

H_0: The data is non stationary.

The alternative hypothesis H_1 is given by,

H_1: The data is stationary.

The outcome achieved, ADF test statistic = -1.0788979085, Lag order = 6, p-value = 0.7234134

The ADF test gives the p-value 0.7234134, which is greater than 0.05, so we accept the null hypothesis i.e.; the data is non stationary, hence we perform n order differencing until we get time series stationary

We perform differencing with n = 1 Now we again check stationarity using ADF test.

Here we test the hypothesis,

H_0: The data is non-stationary.

Against

H_1: The data is stationary.

ADF test statistic = -3.424393960152931, Lag order = 6, p-value= 0.0101548655

The ADF test gives the p-value 0.0101548655, which is smaller than 0.05, so we fail to accept H_0 and Hence, we can conclude that data is stationary; Fig 4.3.3 shows the differenced minimum price

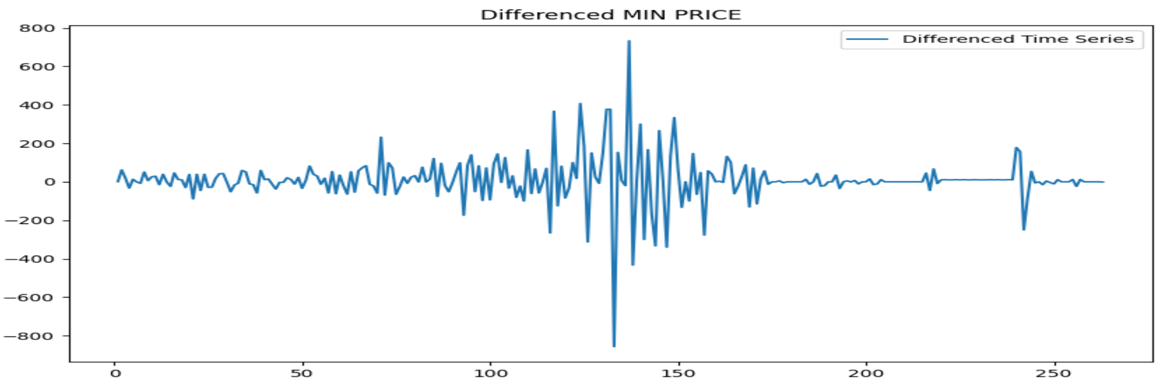

Fig 4.3.3

4.3.1.4 Autocorrelation and Partial Autocorrelation Function

Next step in Time Series Analysis is to plot and examine Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF). ACF &PACF Plot is given in Fig 4.3.4

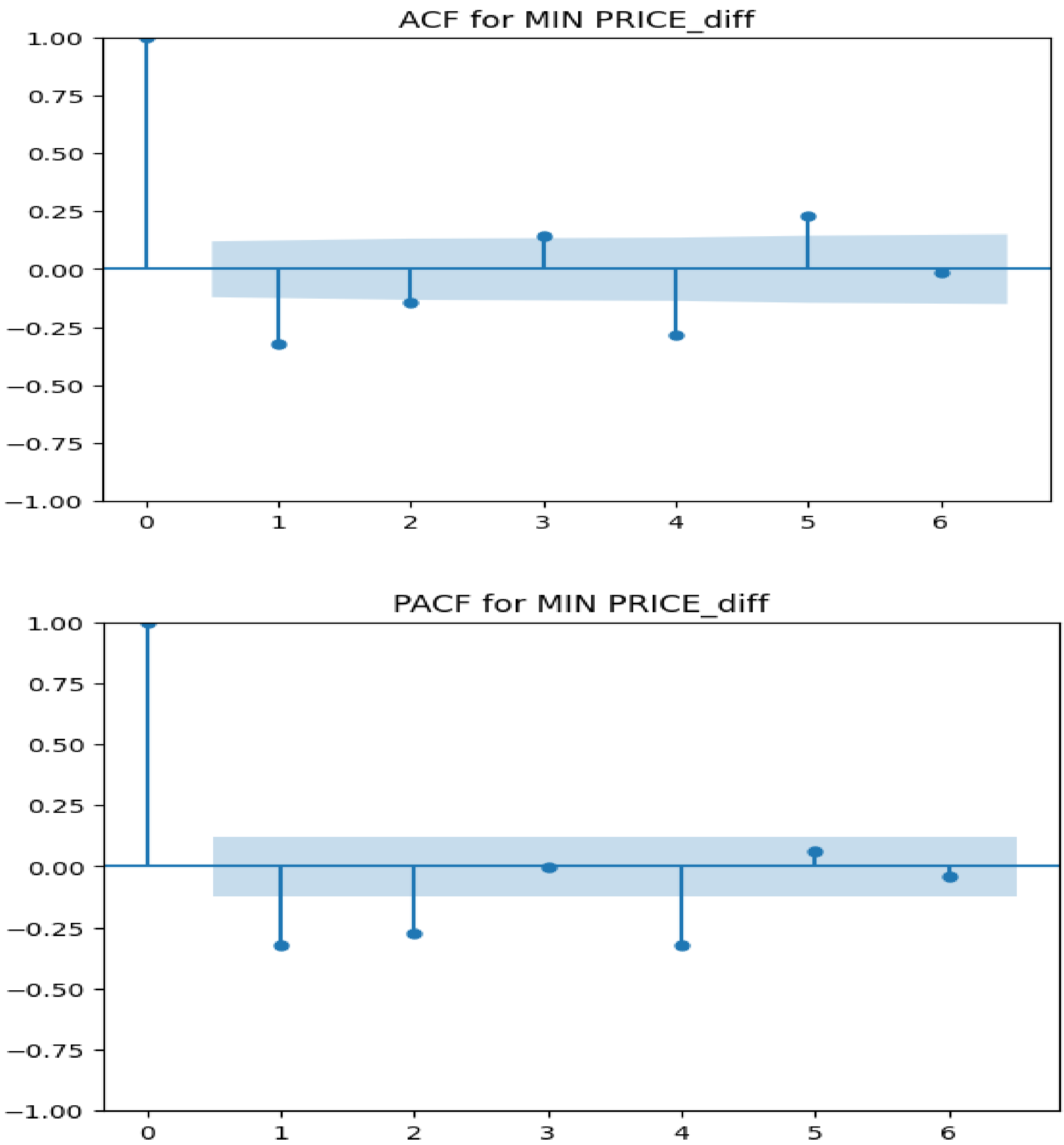

Fig 4.3.4

4.3.1.5 ARIMA Model for Minimum Price

In this step we choose the best model for forecasting the values. It is done by choosing one model from all possible models according to Akaike Information Criterion (AIC). The model with lowest AIC value is chosen as the best model. Below Table 4.3.1 shows the possible models with their AIC values

SL NO.	MODEL ARIMA (p, d, q) x (P, D, Q)	AIC
1	ARIMA(2, 1, 2)x(0, 0, 0)	3202.902
2	ARIMA(0, 1, 0)x(0, 0, 0)	3262.357
3	ARIMA(1, 1, 0)x(0, 0, 0)	3233.851
4	ARIMA(0, 1, 1)x(0, 0, 0)	3214.068
5	ARIMA(0, 1, 0)x(0, 0, 0)	3262.024
6	ARIMA(1, 1, 2)x(0, 0, 0)	3201.529
7	ARIMA(0, 1, 2)x(0, 0, 0)	3213.049
8	ARIMA(1, 1, 1)x(0, 0, 0)	3213.686
9	ARIMA(1, 1, 3)x(0, 0, 0)	3203.138
10	ARIMA(0, 1, 3)x(0, 0, 0)	3210.808
11	ARIMA(2, 1, 1)x(0, 0, 0)	3207.047
12	ARIMA(2, 1, 3)x(0, 0, 0)	3200.197
13	ARIMA(3, 1, 3)x(0, 0, 0)	3199.048
14	ARIMA(3, 1, 2)x(0, 0, 0)	3202.374
15	**ARIMA(4, 1, 3)x(0, 0, 0)**	**3193.412**
16	ARIMA(4, 1, 2)x(0, 0, 0)	3194.420
17	ARIMA(5, 1, 3)x(0, 0, 0)	3195.393
18	ARIMA(4, 1, 4)x(0, 0, 0)	3195.364
19	ARIMA(3, 1, 4)x(0, 0, 0)	3200.092
20	ARIMA(5, 1, 4)x(0, 0, 0)	3197.631

Table 4.3.1

Here the best model is ARIMA (4, 1, 3) x (0, 0, 0) with AIC value 3193.412

Coefficients:

	arL1	arL2	arL3	arL4	ma.L1	ma.L2	ma.L3
	-0.0604	-0.1765	-0.4264	-0.4388	-0.3421	-0.0239	0.5245
std err	0.092	0.075	0.061	0.059	0.087	0.107	0.075

4.3.1.6 Diagnostic checking

Diagnostics checking is performed for confirming the validity, effectiveness and reliability of statistical models. The main objective of it is to choose the right and best model

Diagnostic plot is given in fig.4.3.5

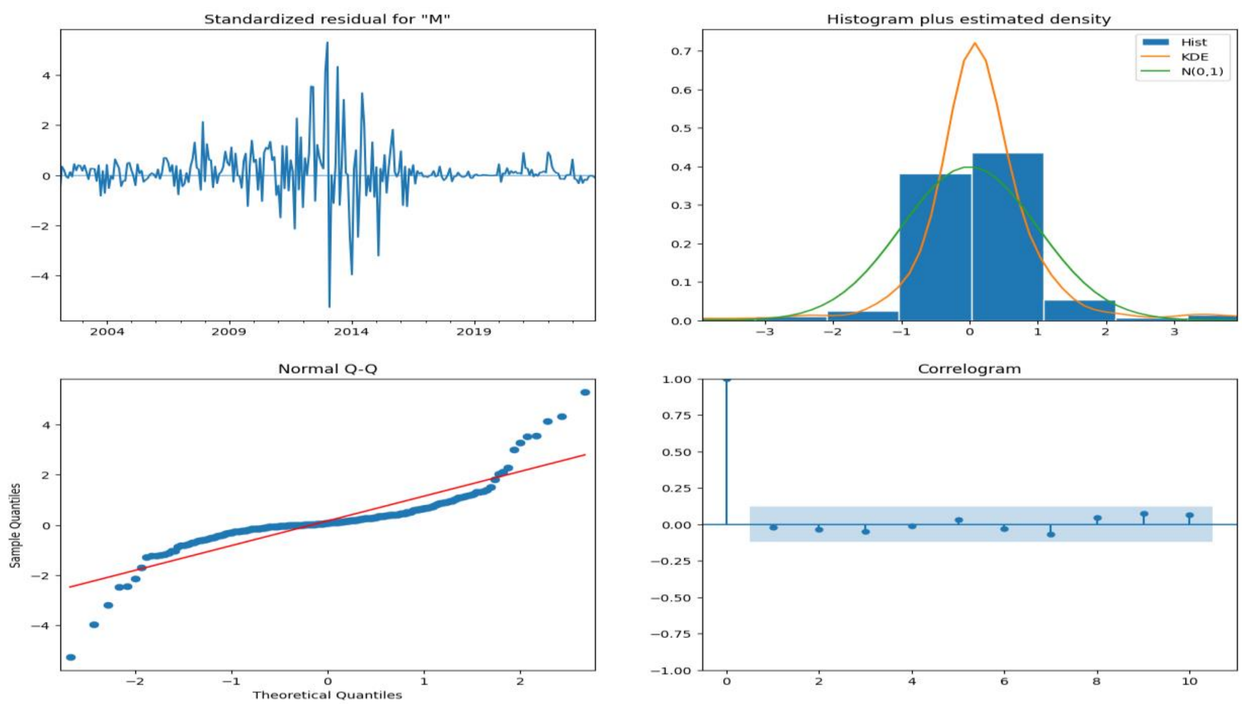

Fig 4.3.5

From Q-Q plot, it is clear that most of the residuals are on the same line and standard residual are normally fitted

4.3.1.7 Forecasting the Sample

Forecasting the Sample means to forecast the actual data points or the training data points. Here we can evaluate model performance on training dataset. The Table 4.3.2 is the actual and in sample forecasted values and Fig. 4.3.6 is the plot of actual and predicted values.

MONTH	ACTUAL MIN PRICE (QUINTAL)	PREDICTED MIN PRICE (QUINTAL)
2021-01-01	3369	3276.954084
2021-02-01	3363	3333.827073
2021-03-01	3363	3335.065045
2021-04-01	3365	3363.643597
2021-05-01	3368	3344.603993
2021-06-01	3373	3377.888001
2021-07-01	3379	3373.605384
2021-08-01	3384	3387.007048
2021-09-01	3389	3377.394611
2021-10-01	3394	3382.129017
2021-11-01	3399	3382.216848
2021-12-01	3405	3394.200716
2022-01-01	3497	3402.306437
2022-02-01	3547	3464.942442
2022-03-01	3536	3501.223094
2022-04-01	3542	3524.121926
2022-05-01	3531	3520.808627
2022-06-01	3538	3529.543497
2022-07-01	3534	3548.845809
2022-08-01	3532	3546.003871
2022-09-01	3528	3544.044664
2022-10-01	3525	3524.489583
2022-11-01	3523	3520.550200
2022-12-01	3520	3516.431973
2023-01-01	3588	3522.631623
2023-02-01	3553	3567.124216
2023-03-01	3534	3550.893801

2023-04-01	3521	3554.113281
2023-05-01	3513	3514.564753
2023-06-01	3499	3530.760132
2023-07-01	3491	3507.328468
2023-08-01	3488	3508.006261
2023-09-01	3484	3488.424371
2023-10-01	3486	3486.886766
2023-11-01	3477	3480.552279
2023-12-01	3469	3478.852475

Table 4.3.2

Plot of actual values vs predicted values of min price is given in Fig 4**.3.**6

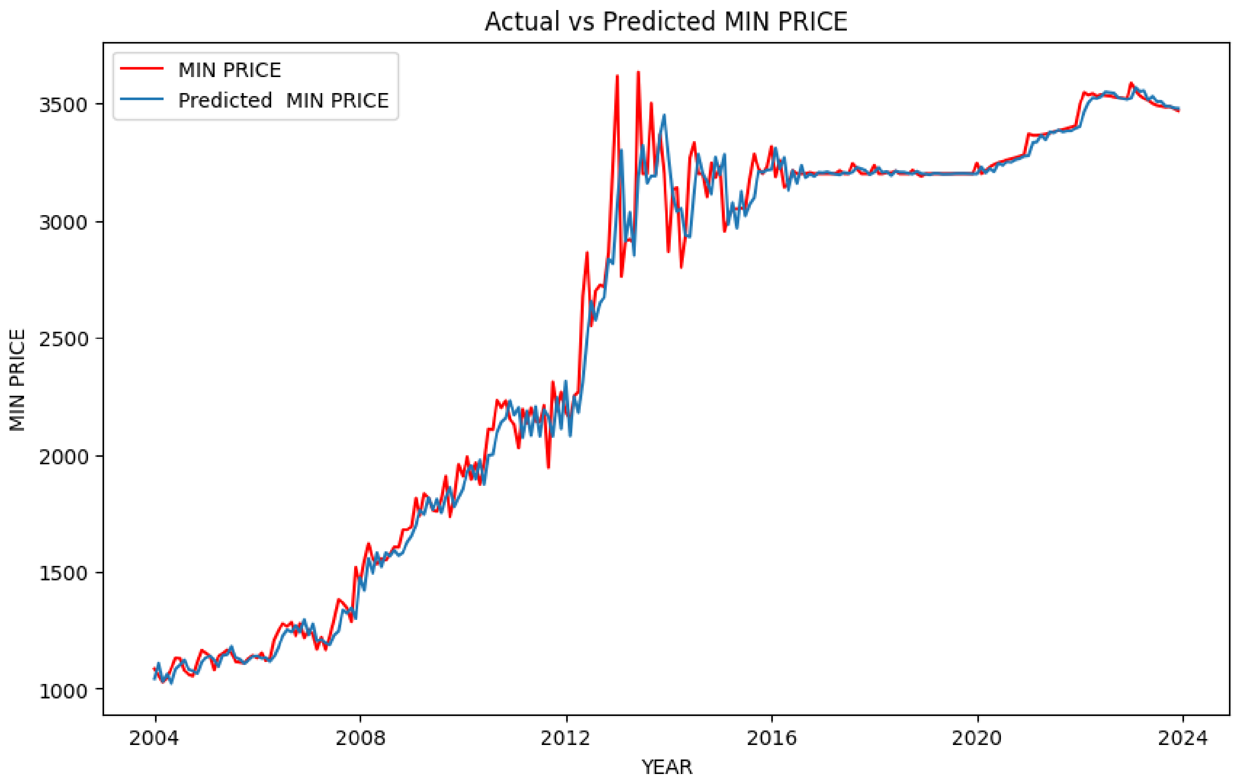

Fig 4**.3.**6

4.3.1.8 Forecasting the Future Values

The forecasted Minimum price of rice for the months of 2024 and 2025 is given in Table 4.3.3

MONTH	FORECASTED MIN PRICE (QUINTAL)	LCL	UCL
2024-01-01	3474.633533	3273.644805	3675.622260
2024-02-01	3477.106056	3240.661310	3713.550802
2024-03-01	3478.007776	3224.854745	3731.160808
2024-04-01	3478.466776	3195.049677	3761.883876
2024-05-01	3474.714866	3180.479394	3768.950339
2024-06-01	3473.299642	3154.317254	3792.282031
2024-07-01	3473.421270	3129.683860	3817.158680
2024-08-01	3475.166262	3109.111715	3841.220808
2024-09-01	3477.349393	3090.161461	3864.537324
2024-10-01	3477.513294	3076.312245	3878.714343
2024-11-01	3476.276510	3061.619022	3890.933998
2024-12-01	3474.548711	3045.713967	3903.383454
2025-01-01	3473.812547	3028.871755	3918.753340
2025-02-01	3474.637214	3011.990186	3937.284242
2025-03-01	3476.055543	2997.371556	3954.739531
2025-04-01	3476.938055	2984.543211	3969.332898
2025-05-01	3476.598229	2972.357383	3980.839074
2025-06-01	3475.454031	2959.677153	3991.230910
2025-07-01	3474.543035	2946.152931	4002.933139
2025-08-01	3474.552119	2932.519053	4016.585186
2025-09-01	3475.379394	2919.676180	4031.082608
2025-10-01	3476.255306	2908.039825	4044.470788
2025-11-01	3476.465996	2897.131580	4055.800413
2025-12-01	3475.923647	2886.148943	4065.698350

Table 4.3.3

The Fig 4.3.7 shows the graph of forecasted min price values of the months in 2024 and 2025

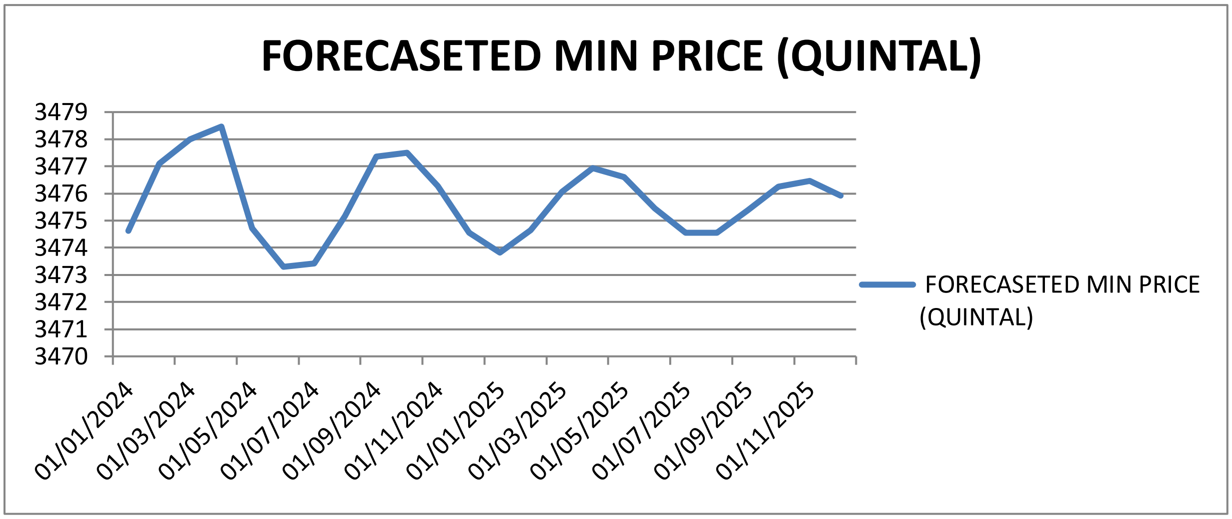

Fig 4.3.7

4.3.2 Maximum Price of Rice (2nd variable)

4.3.2.1 Time series plot

The initial step in time series is to draw a time series plot. The time series plot of minimum price is given in Fig 4.3.8

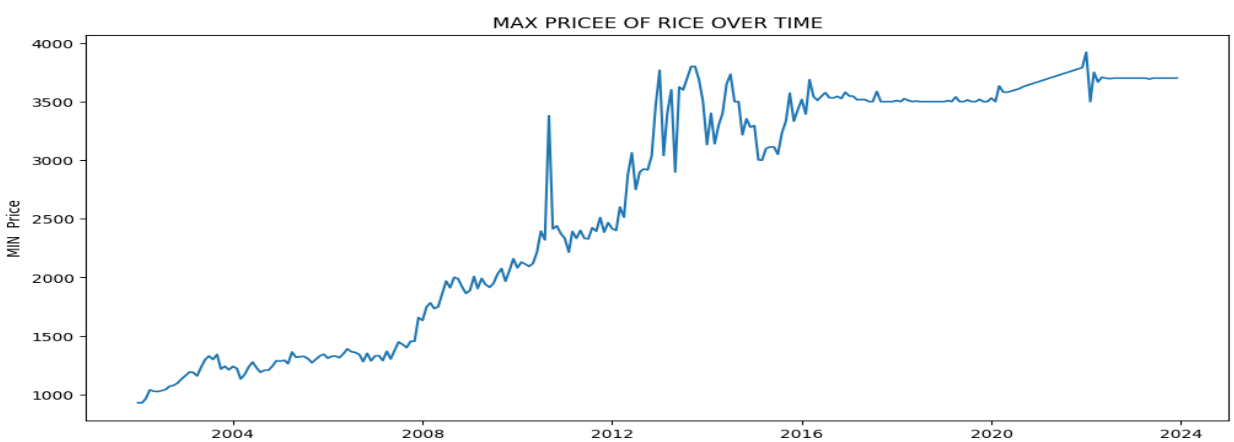

Fig 4.3.8

4.1.2.2 Decomposition of time

The second step is to perform seasonal decomposition to capture the trend, seasonal and random components of time series. Plot the seasonal plot. Fig 4.3.9 depicts the seasonal plot

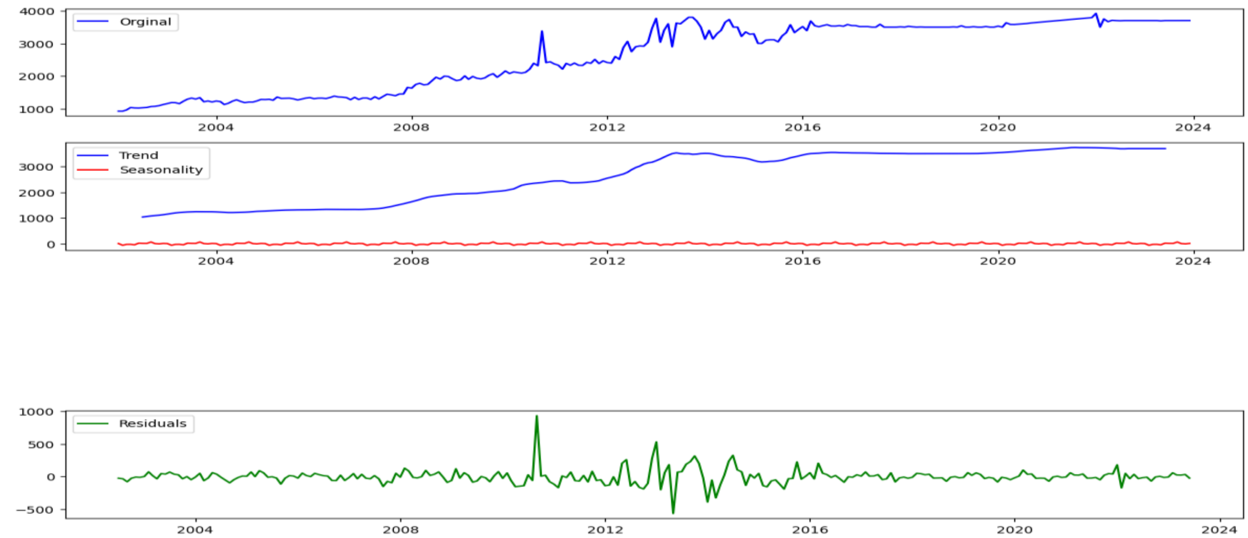

Fig 4.3.9

From the fig.4.3.9, it can conclude that data has no seasonality

4.1.2.3 Stationarity check using Augmented Dickey- Fuller Test

To test the time series data for stationarity using ADF test, follows a hypothesis testing approach.

The null hypothesis H_0 is given by,

H_0: The data is non stationary.

The alternative hypothesis H_1 is given by,

H_1: The data is stationary.

The outcome achieved, ADF test statistic = -1.3283661419, Lag order = 6, p-value = 0.6160969433

The ADF test gives the p-value 0.616096943, which is greater than 0.05, so we accept the null hypothesis i.e.; the data is non stationary, hence we perform n order differencing until we get time series stationary

We perform differencing with n = 1, Now we again check stationarity using ADF test.

Here we test the hypothesis

H_0: The data is non-stationary.

Against

H_1: The data is stationary.

ADF test statistic = -9.47953852930256, Lag order = 6, p-value = 3.9064332103672186e-16

The ADF test gives the p-value 3.9064332103672186e-16, which is smaller than 0.05, so we fail to accept H_0 and hence; we can conclude that data is stationary.

Fig 4.3.10 shows the differenced minimum price

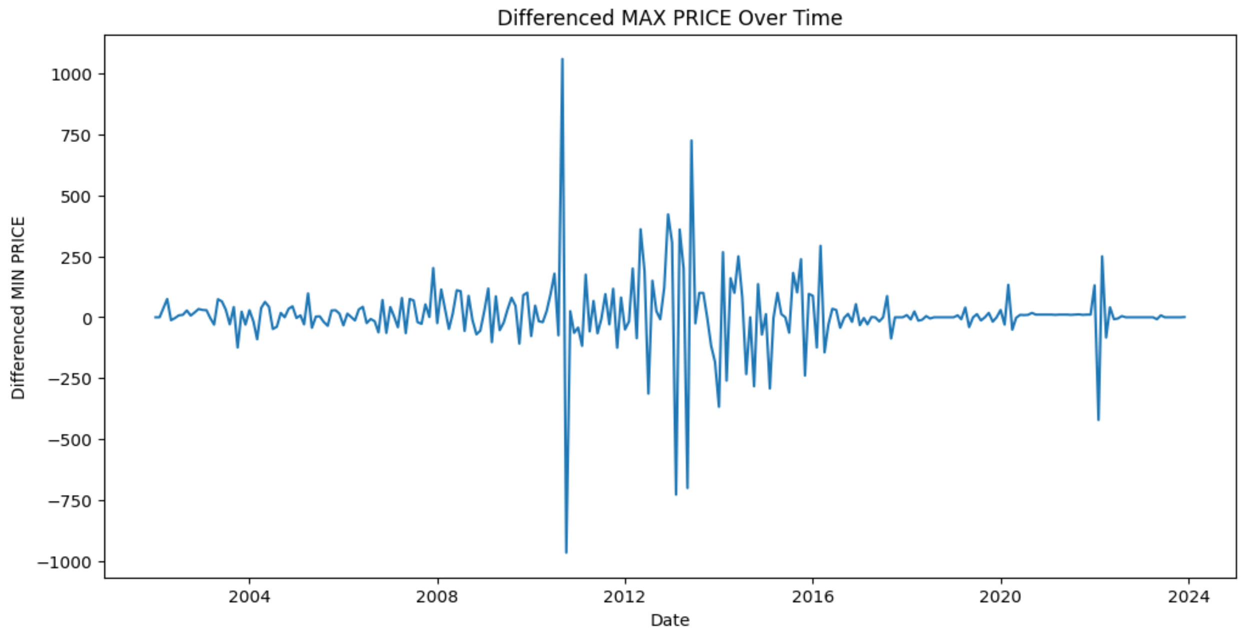

Fig 4.3.10

4.1.2.4 Autocorrelation and Partial Autocorrelation Function

Next step in Time Series Analysis is to plot and examine Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF). ACF &PACF Plot is given in Fig 4.3.11

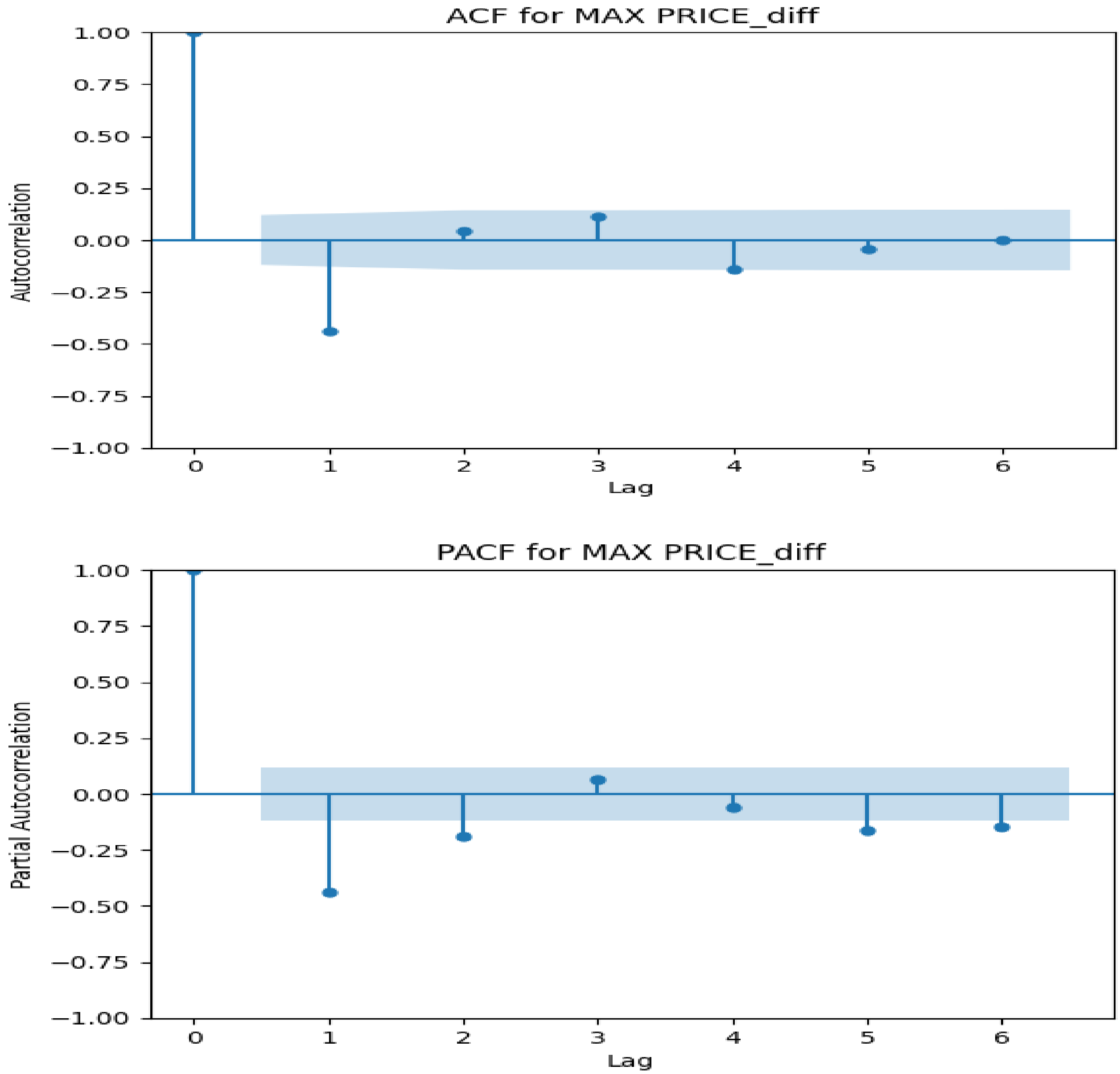

Fig 4.3.11

4.1.2.5 ARIMA Model for Maximum Price

In this step we choose the best model for forecasting the values. It is done by choosing one model from all possible models according to Akaike Information Criterion (AIC). The model with lowest AIC value is chosen as the best model. Below Table 4.3.4 shows the possible models with their AIC values

SL NO.	MODEL ARIMA (p, d, q) x (P, D, Q)	AIC
1	ARIMA(2, 1, 2) x (0, 0, 0)	3335.424
2	ARIMA(0, 1, 0) x (0, 0, 0)	3396.193
3	ARIMA(1, 1, 0) x (0, 0, 0)	3341.419
4	**ARIMA(0, 1, 1) x (0, 0, 0)**	**3335.052**
5	ARIMA(0, 1, 0) x (0, 0, 0)	3395.438
6	ARIMA(1, 1, 1) x (0, 0, 0)	3335.833
7	ARIMA(0, 1, 2) x (0, 0, 0)	3335.997
8	ARIMA(1, 1, 2) x (0, 0, 0)	3337.820

Table 4.3.4

Here the best model is ARIMA (0, 1, 1) x (0, 0, 0) with AIC value 3335.052

Coefficients:

	ma.L1
	-0.4938
Std Err	0.026

4.1.2.6 Diagnostic checking

Diagnostics checking is performed for confirming the validity, effectiveness and reliability of statistical models. The main objective of it is to choose the right and best model

Diagnostic plot is given in Fig 4.3.12

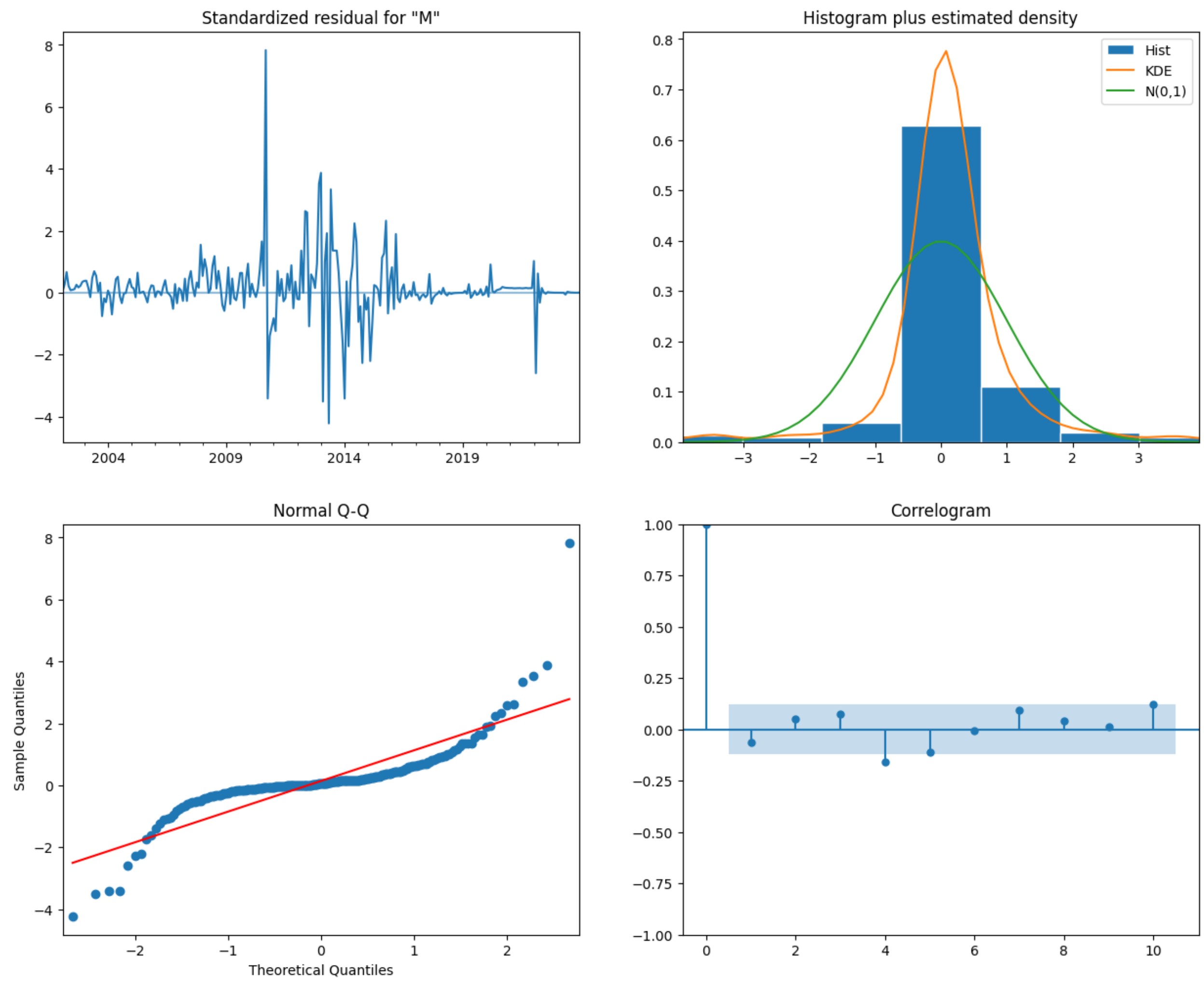

Fig 4.3.12

From Q-Q plot, it is clear that most of the residuals are on the same line and standard residuals are normally fitted

4.1.2.7 Forecasting the Sample

Forecasting the Sample means to forecast the actual data points or the training data points. Here we can evaluate model performance on training dataset. The Table 4.3.5 is the actual and in sample forecasted values and Fig 4.3.13 is the plot of actual and predicted values.

MONTH	ACTUAL MAX PRICE (QUINTAL)	PREDICTED MAX PRICE (QUINTAL)
2021-01-01	3671	3650.275033
2021-02-01	3682	3661.401527
2021-03-01	3692	3672.460112
2021-04-01	3703	3682.950380
2021-05-01	3714	3693.714305
2021-06-01	3725	3704.604970
2021-07-01	3735	3715.554333
2021-08-01	3746	3725.994017
2021-09-01	3758	3736.734515
2021-10-01	3768	3748.151194
2021-11-01	3779	3758.807309
2021-12-01	3790	3769.648044
2022-01-01	3921	3780.574282
2022-02-01	3500	3855.963834
2022-03-01	3750	3664.859566
2022-04-01	3750	3710.568424
2022-05-01	3667	3687.178093
2022-06-01	3708	3698.356631
2022-07-01	3700	3699.238897
2022-08-01	3695	3696.963185
2022-09-01	3700	3698.593542
2022-10-01	3700	3699.348619
2022-11-01	3700	3699.698322
2022-12-01	3700	3699.860282
2023-01-01	3700	3699.935292
2023-02-01	3700	3699.970031
2023-03-01	3700	3699.986120
2023-04-01	3700	3699.993572

2023-05-01	3692	3699.997023
2023-06-01	3700	3695.703707
2023-07-01	3700	3698.010233
2023-08-01	3700	3699.078468
2023-09-01	3700	3699.573206
2023-10-01	3700	3699.802336
2023-11-01	3700	3699.908455
2023-12-01	3701	3699.957602

Table 4.3.5

Plot of actual values vs predicted values of max price is given in Fig 4.3.13

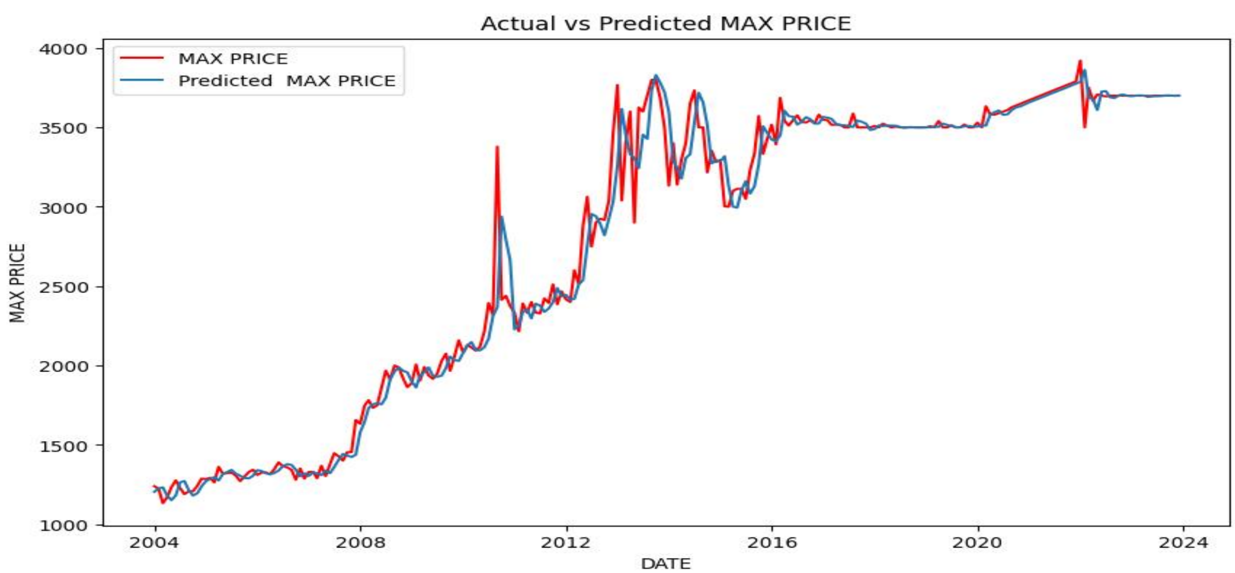

Fig 4.3.13

4.1.2.8 Forecasting the Future Values

The forecasted Maximum price of rice for the months of years 2024 and 2025 is given in Table 4.3.6

MONTH	FORECASTED MAX PRICE (QUINTAL)	LCL	UCL
2024-01-01	3698.564738	3431.724467	3969.305214
2024-02-01	3698.789666	3395.678038	4005.351643
2024-03-01	3697.463728	3363.464862	4037.564819
2024-04-01	3697.479776	3334.072622	4066.957059
2024-05-01	3698.534876	3306.868920	4094.160761
2024-06-01	3700.299642	3281.427352	4119.602328
2024-07-01	3699.621270	3257.444266	4143.585415
2024-08-01	3698.166262	3234.694332	4166.335348
2024-09-01	3697.349393	3213.004884	4188.024797
2024-10-01	3698.634593	3192.240142	4208.789539
2024-11-01	3696.206510	3172.291043	4228.738638
2024-12-01	3697.345544	3153.068412	4247.961269
2025-01-01	3698.374465	3134.498230	4266.531451
2025-02-01	3696.973611	3116.518252	4284.511429
2025-03-01	3699.695661	3099.075544	4301.954136
2025-04-01	3699.934187	3082.124641	4318.905040
2025-05-01	3699.996922	3065.626149	4335.403531
2025-06-01	3699.569473	3049.545669	4351.484011
2025-07-01	3699.074188	3033.852952	4367.176729
2025-08-01	3700.345544	3018.521230	4382.508451
2025-09-01	3704.533607	3003.526680	4397.503001
2025-10-01	3710.655801	2988.847989	4412.181692
2025-11-01	3711.465996	2974.465999	4426.563681
2025-12-01	3715.923647	2960.363416	4440.666265

Table 4.3.6

The Fig 4.3.14 shows the graph of forecasted max price values of the months in 2024 and 2025

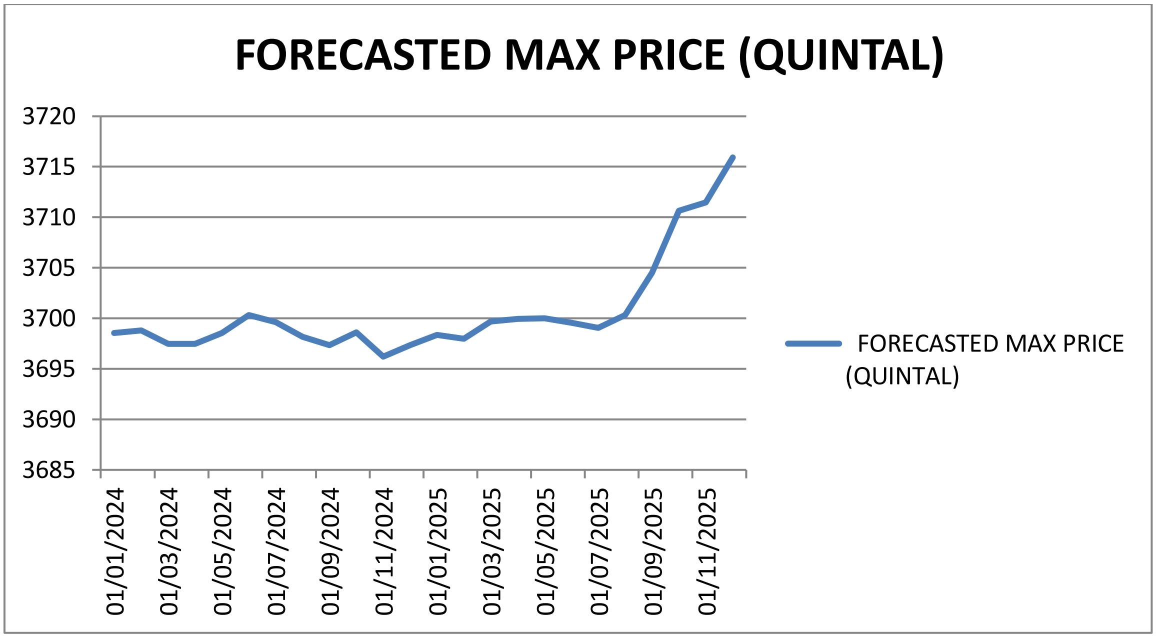

Fig 4.3.14

4.4 EXPONENTIAL SMOOTHING MODEL

4.4.1 Minimum Price of Rice (1st variable)

4.4.1.1 Choosing smoothing parameter

Smoothing parameter is the only constant value chosen by a statistician according to the characteristics of the data. When we choose smaller value of alpha it results in more smoothing and vice versa. In this study the value of alpha (smoothing parameter) is given as 0.1 to increase the smoothing effect i.e. to remove noise and fluctuations to understand and evaluate the patterns and trends in data.

4.4.1.2 Forecasting future values

Using python and the equation of single exponential smoothing we forecasted the minimum price of rice (quintal) the months of the years 2024 & 2025 the Table 4.4.1 below shows forecasted min price values

MONTH	FORECASTED MIN PRICE (QUINTAL)
2024-01-01	3514.66100
2024-02-01	3469.98132
2024-03-01	3463.42561
2024-04-01	3483.80306
2024-05-01	3509.16048
2024-06-01	3576.88548
2024-07-01	3546.23855
2024-08-01	3553.23855
2024-09-01	3569.20710
2024-10-01	3545.72255
2024-11-01	3573.16700
2024-12-01	3597.94627
2025-01-01	3611.13123
2025-02-01	3574.45156
2025-03-01	3595.89584
2025-04-01	3588.27329
2025-05-01	3613.63071
2025-06-01	3681.35571

2025-07-01	3640.70878
2025-08-01	3658.35357
2025-09-01	3673.67734
2025-10-01	3650.19279
2025-11-01	3677.63724
2025-12-01	3702.41651

Table 4.4.1

The Existing vs Forecasted graph is formulated from python is depicted in Fig 4.4.1

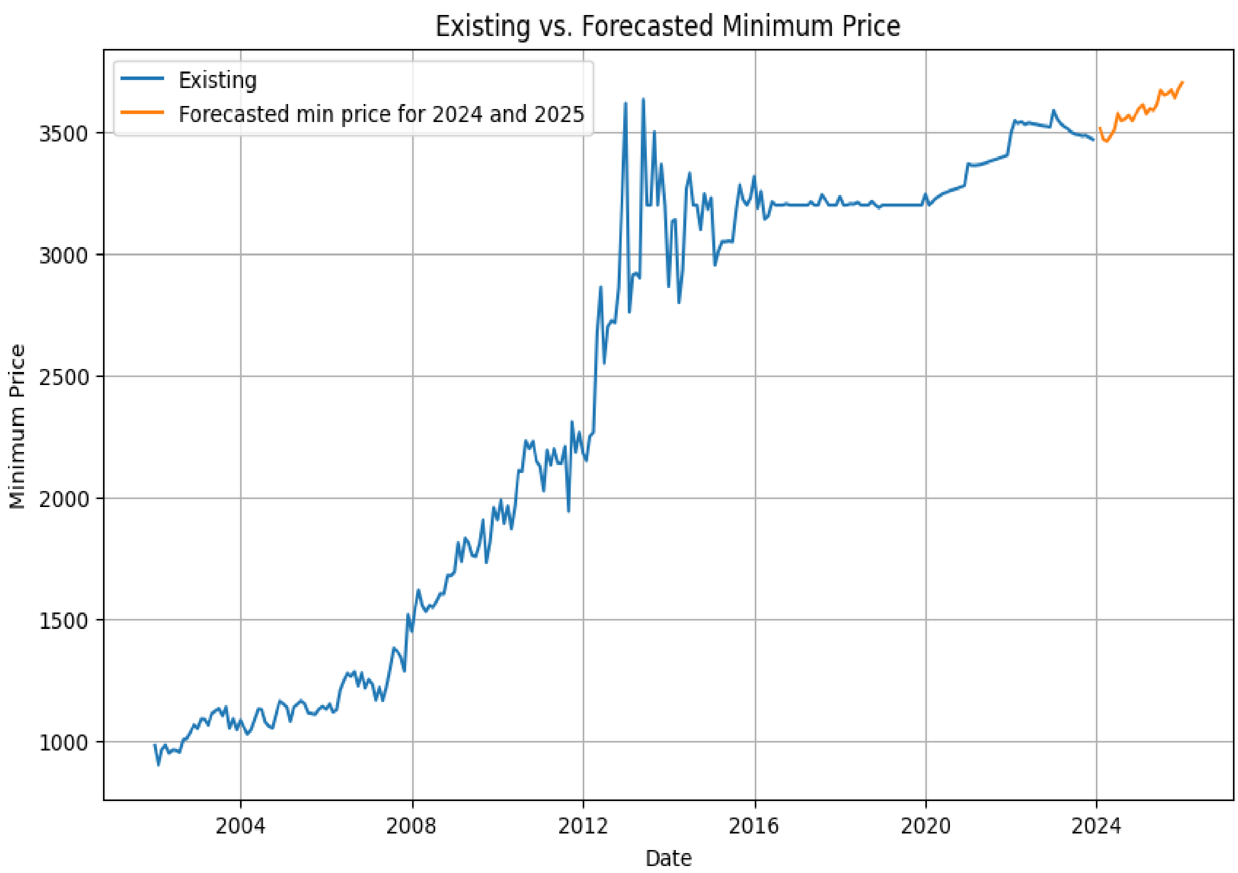

Fig 4.4.1

The below Fig 4.4.2 is formulated in excel. It depicts only the forecasted region of Fig 4.4.1

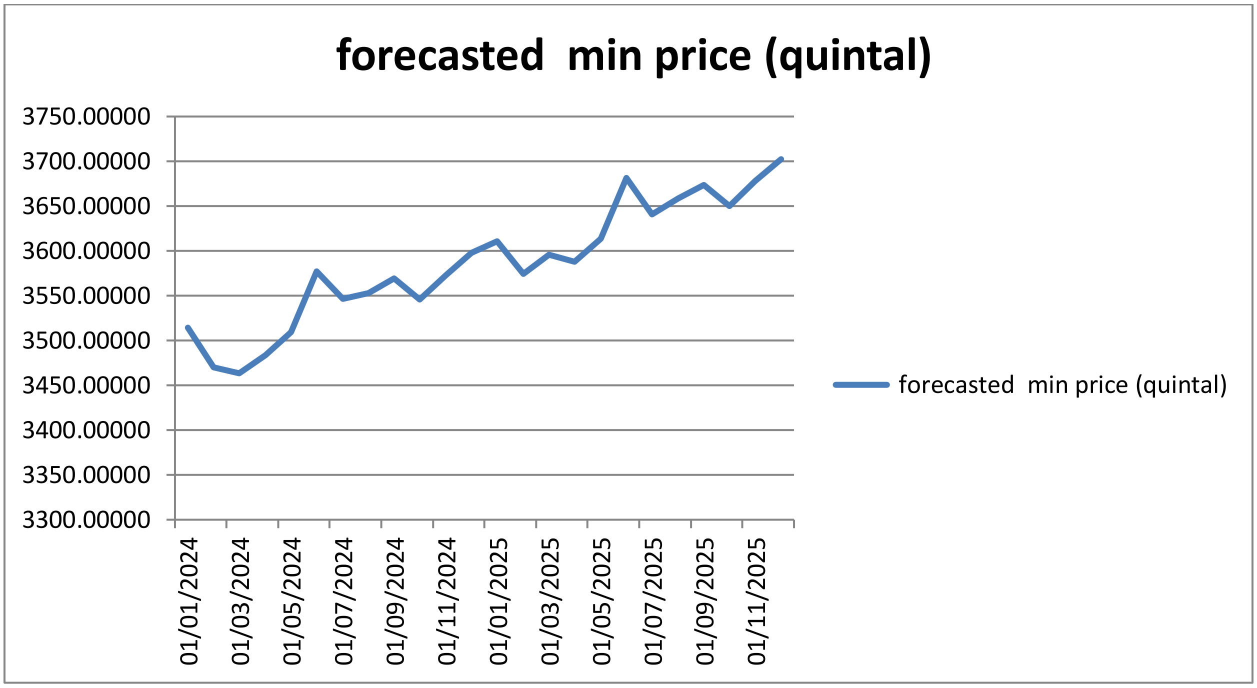

Fig 4.4.2

4.4.2 Maximum Price of Rice (2st variable)

4.4.2.1 Choosing smoothing parameter

Smoothing parameter is the only constant value chosen by a statistician according to the characteristics of the data. When we choose smaller value of alpha it results in more smoothing and vice versa. In this study the value of alpha (smoothing parameter) is given as 0.1 to increase the smoothing effect i.e. to remove noise and fluctuations to understand and evaluate the patterns and trends in data.

4.4.2.2 Forecasting future values

Using python and the equation of single exponential smoothing we forecasted the maximum price (quintal) the months of the years 2024 & 2025 .The Table 4.4.2 below shows forecasted max price values

MONTH	FORECASTED MAX PRICE (QUINTAL)
2024-01-01	3,703.88
2024-02-01	3,683.61
2024-03-01	3,680.33
2024-04-01	3,671.63
2024-05-01	3,679.16
2024-06-01	3,737.89
2024-07-01	3,727.22
2024-08-01	3,721.31
2024-09-01	3,724.33
2024-10-01	3,713.15
2024-11-01	3,707.40
2024-12-01	3,732.32
2025-01-01	3,735.44
2025-02-01	3,724.62
2025-03-01	3,712.89
2025-04-01	3,721.06
2025-05-01	3,703.98
2025-06-01	3,762.71
2025-07-01	3,752.05
2025-08-01	3,746.14
2025-09-01	3,795.15
2025-10-01	3,737.98
2025-11-01	3,732.23
2025-12-01	3,757.14

Table 4.4.2

The Existing vs Forecasted graph is formulated from python is depicted in Fig 4.4.3

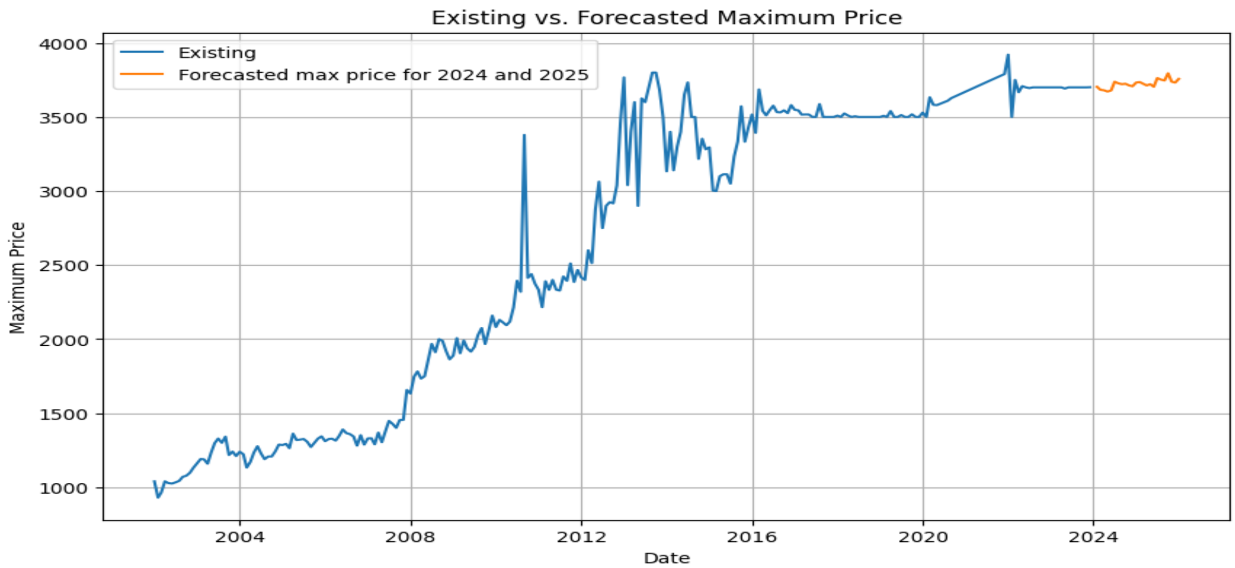

Fig 4.4.3

The below Fig 4.4.4 is formulated in excel. It depicts only the forecasted region of Fig 4.4.3

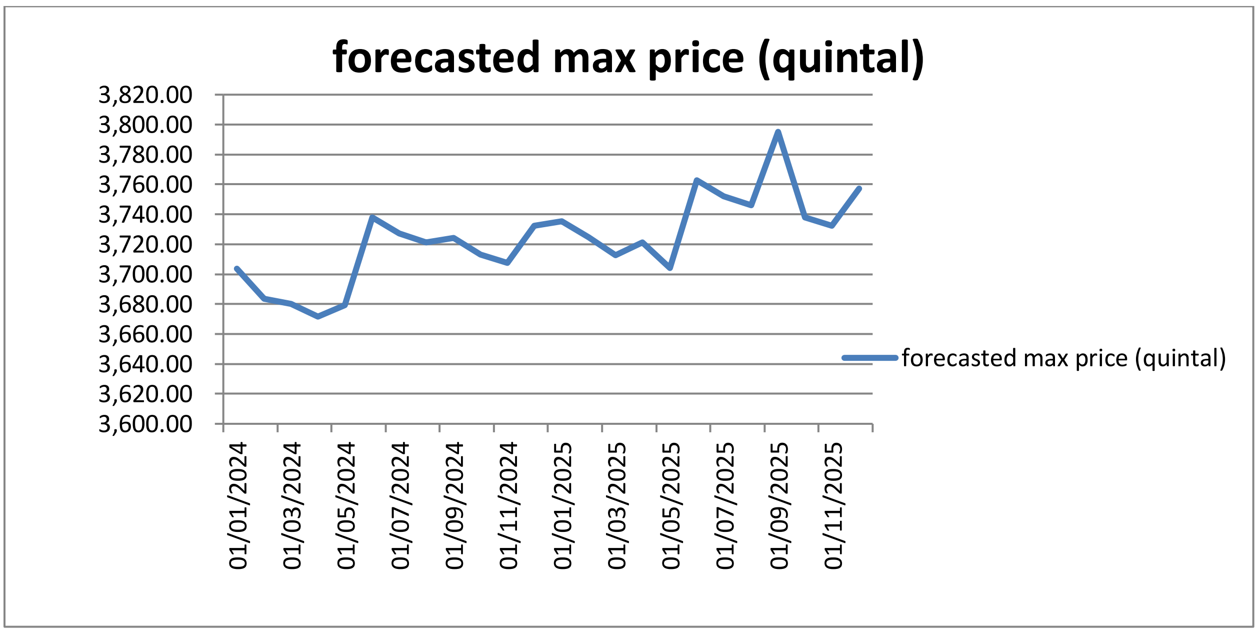

Fig 4.4.4

4.5 LINEAR REGRESSION MODEL

4.5.1 Minimum Price of Rice (1st variable)

4.5.1 .1 Forecasting the Sample

Forecasting the Sample means to forecast the actual data points or the training data points. Here we can evaluate model performance on training dataset. The Table 4.5.1 is the actual and in sample forecasted min price values and Fig 4.5.1 is the plot of actual and predicted values formulated in python

MONTH	ACTUAL MIN PRICE (QUINTAL)	PREDICTED MIN PRICE (QUINTAL)
2021-01-01	3369	3553.49375509
2021-02-01	3363	3565.23994072
2021-03-01	3363	3576.98612636
2021-04-01	3365	3588.73231199
2021-05-01	3368	3600.47849762
2021-06-01	3373	3600.47849762
2021-08-01	3384	3623.97086889
2021-09-01	3389	3635.71705452
2021-10-01	3394	3647.46324015
2021-11-01	3399	3659.20942579
2021-12-01	3405	3670.95561142
2022-01-01	3497	3683.07012734
2022-02-01	3547	3694.81631297
2022-03-01	3536	3706.5624986
2022-04-01	3542	3718.30868423
2022-05-01	3531	3730.05486987
2022-06-01	3538	3741.8010555
2022-07-01	3534	3753.54724113
2022-08-01	3532	3765.29342676
2022-09-01	3528	3777.0396124
2022-10-01	3525	3788.78579803

2022-11-01	3523	3800.53198366
2022-12-01	3520	3812.27816929
2023-01-01	3588	3824.39268521
2023-02-01	3553	3836.13887085
2023-03-01	3534	3847.88505648
2023-04-01	3521	3859.63124211
2023-05-01	3513	3871.37742774
2023-06-01	3499	3883.12361338
2023-07-01	3491	3894.86979901
2023-08-01	3488	3906.61598464
2023-09-01	3484	3918.36217027
2023-10-01	3486	3930.10835591
2023-11-01	3477	3941.85454154
2023-12-01	3469	3953.60072717

Table 4.5.1

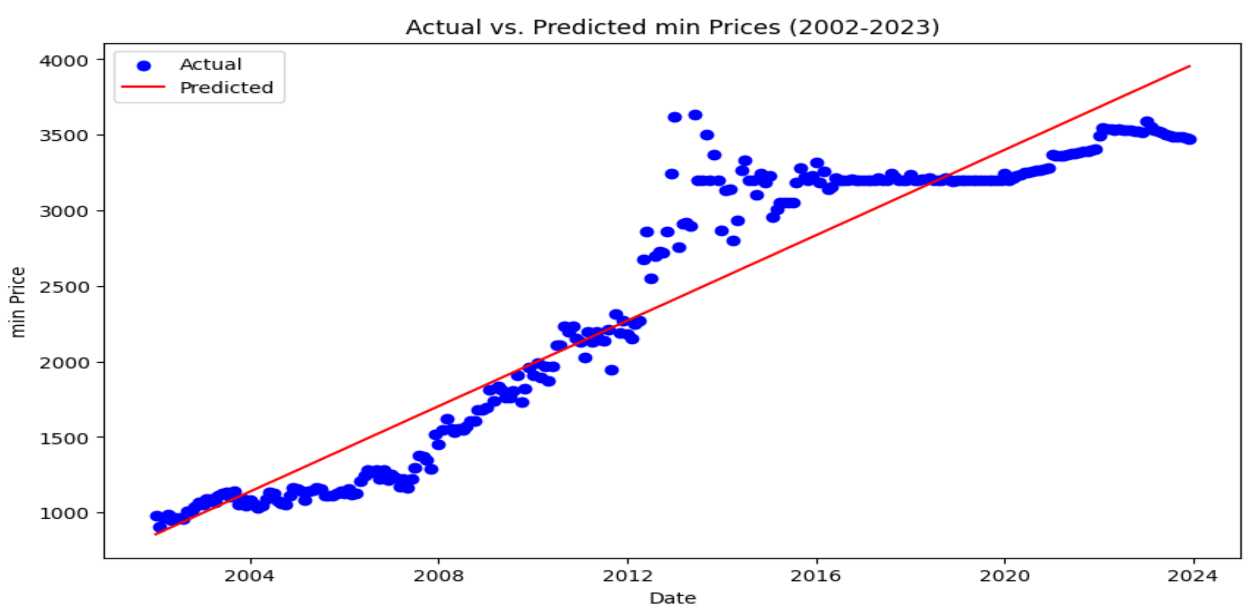

Fig 4.5.1

4.5.1.2 Forecasting the Future Values

Using python, we forecasted the minimum price of the months of the years 2024 & 2025 the Table 4.5.2 below shows forecasted min price values

MONTH	FORECASTED MIN PRICE (QUINTAL)
2024-01-01	3965.7275
2024-02-01	3977.4632
2024-03-01	3989.1989
2024-04-01	4000.9346
2024-05-01	4012.6703
2024-06-01	4024.4060
2024-07-01	4036.1417
2024-08-01	4047.8774
2024-09-01	4059.6130
2024-10-01	4071.3487
2024-11-01	4083.0844
2024-12-01	4094.8201
2025-01-01	4107.0441
2025-02-01	4118.7798
2025-03-01	4130.5155
2025-04-01	4142.2512
2025-05-01	4153.9869
2025-06-01	4165.7226
2025-07-01	4177.4583
2025-08-01	4189.1940
2025-09-01	4200.9297
2025-10-01	4212.6654
2025-11-01	4224.4011
2025-12-01	4236.1368

Table 4.5.2

The below Fig 4.5.2 is formulated in excel. It depicts the forecasted min price values for 2024 and 2025

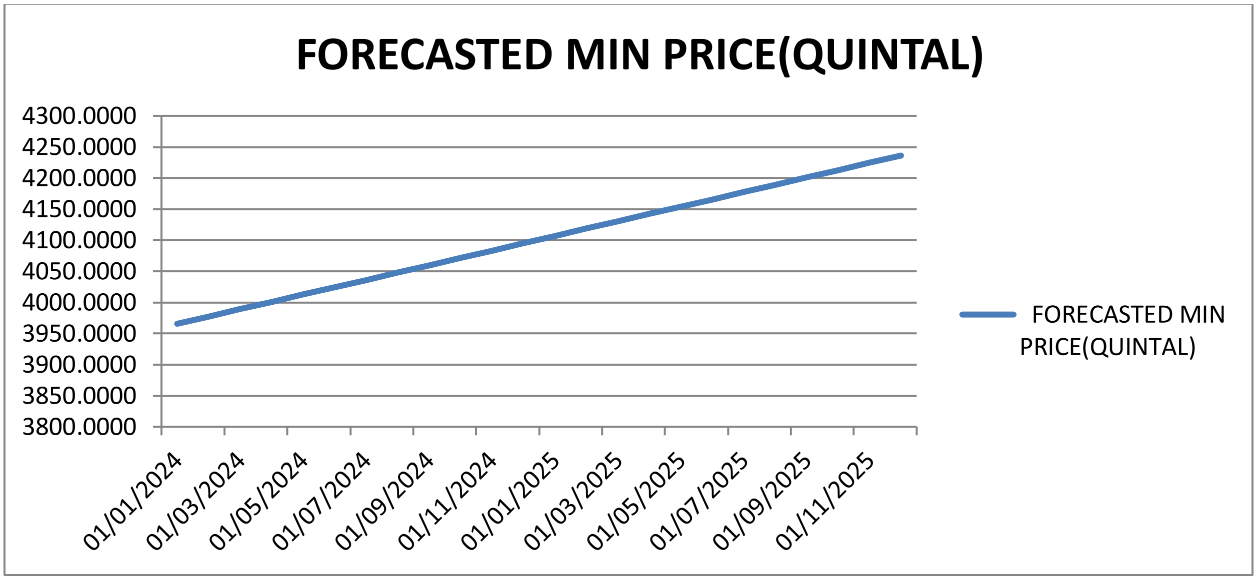

Fig 4.5.2

4.5.2 Maximum Price of Rice (1st variable)

4.5.2 .1 Forecasting the Sample

Forecasting the Sample means to forecast the actual data points or the training data points. Here we can evaluate model performance on training dataset. The Table 4.5.3 is the actual and in sample forecasted max price values and Fig 4.5.3 is the plot of actual and predicted values formulated in python

MONTH	ACTUAL MAX PRICE (QUINTAL)	PREDICTED MAX PRICE (QUINTAL)
2021-01-01	3671	3819.80881256
2021-02-01	3682	3834.74603443
2021-03-01	3692	3849.6832563
2021-04-01	3703	3864.62047817
2021-05-01	3714	3879.55770004
2021-06-01	3725	3894.49492191
2021-07-01	3735	3909.43214378
2021-08-01	3746	3924.36936565

2021-09-01	3758	3939.30658752
2021-10-01	3768	3954.24380939
2021-11-01	3779	3969.18103125
2021-12-01	3790	3984.11825312
2022-01-01	3921	3969.46093942
2022-02-01	3500	3984.39816129
2022-03-01	3750	3999.33538316
2022-04-01	3750	4014.27260503
2022-05-01	3667	4029.2098269
2022-06-01	3708	4044.14704877
2022-07-01	3700	4059.08427064
2022-08-01	3695	4074.02149251
2022-09-01	3700	4088.95871438
2022-10-01	3700	4103.89593624
2022-11-01	3700	4118.83315811
2022-12-01	3700	4133.77037998
2023-01-01	3700	4119.11306628
2023-02-01	3700	4134.05028815
2023-03-01	3700	4148.98751002
2023-04-01	3700	4163.92473189
2023-05-01	3692	4178.86195376
2023-06-01	3700	4193.79917563
2023-07-01	3700	4208.7363975
2023-08-01	3700	4223.67361936
2023-09-01	3700	4238.61084123
2023-10-01	3700	4253.5480631
2023-11-01	3700	4268.48528497
2023-12-01	3701	4283.42250684

Table 4.5.3

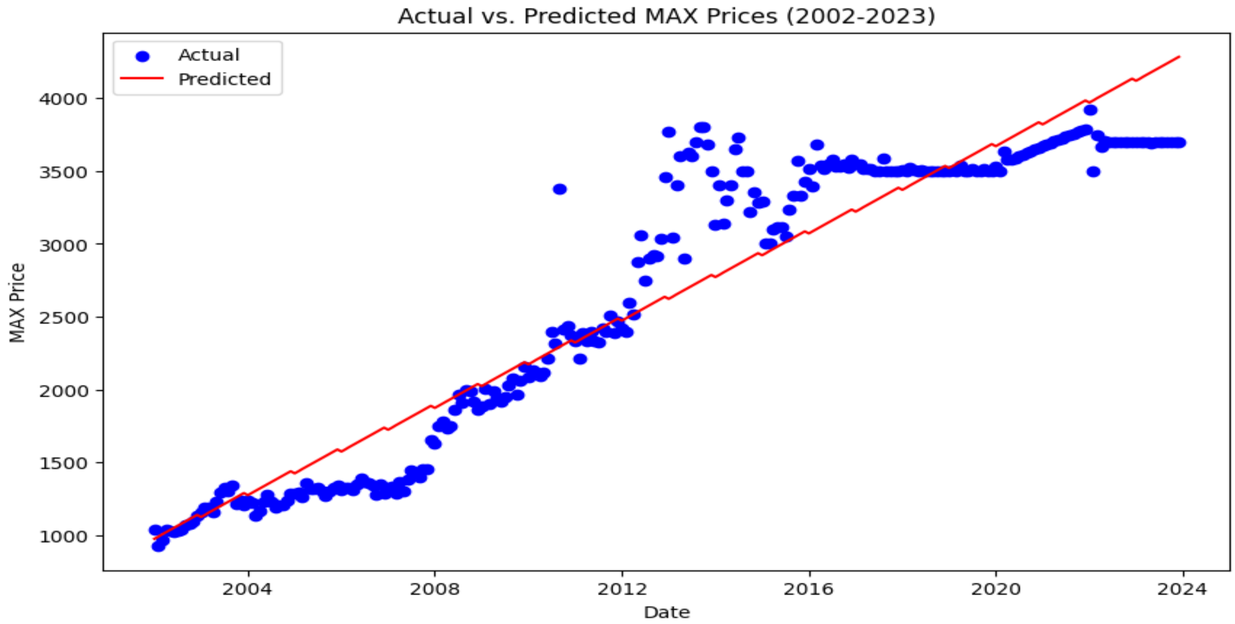

Fig 4.5.3

4.5.2.2 Forecasting the Future Values

Using python, we forecasted the maximum price (quintal) the months of the years 2024 & 2025 the Table 4.5.4 below shows forecasted max price values

MONTH	FORECASTED MAX PRICE (QUINTAL)
2024-01-01	4268.765193
2024-02-01	4283.702415
2024-03-01	4298.639637
2024-04-01	4313.576859
2024-05-01	4328.514081
2024-06-01	4343.451302
2024-07-01	4358.388524
2024-08-01	4373.325746
2024-09-01	4388.262968
2024-10-01	4403.20019
2024-11-01	4418.137412
2024-12-01	4433.074634
2025-01-01	4418.41732

2025-02-01	4433.354542
2025-03-01	4448.291764
2025-04-01	4463.228986
2025-05-01	4478.166207
2025-06-01	4493.103429
2025-07-01	4508.040651
2025-08-01	4522.977873
2025-09-01	4537.915095
2025-10-01	4552.852317
2025-11-01	4567.789539
2025-12-01	4582.726761

Table 4.5.4

The below Fig 4.5.4 is formulated in excel. It depicts the forecasted max price values for 2024 and 2025

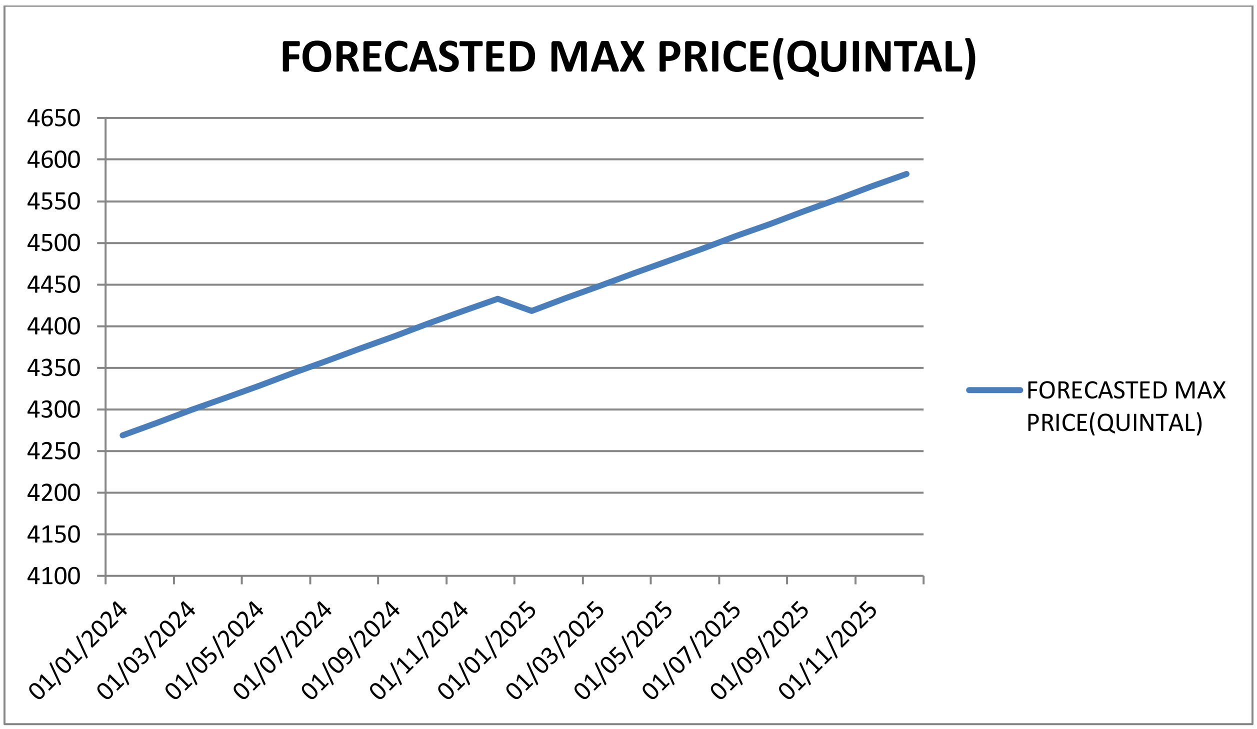

Fig 4.5.4

4.6 Comparison of RMSE and MSE values

In order to determine best model from the above three i.e. ARIMA model, Exponential smoothing model and Linear Regression model comparison of the RMSE and MSE values is done for each model. Mean Squared Error (MSE) is the average value of squared difference between actual and predicted values. Root Mean Squared Error (RMSE) is the square root of Mean Squared Error (MSE). Comparison of MSE and RMSE values of ARIMA model, Exponential smoothing model and Linear Regression model is given in the following Table 4.6.1

	ARIMA model		**Exponential smoothing model**		**Linear Regression model**	
	Min price (quintal)	Max price (quintal)	Min price (quintal)	Max price (quintal)	Min price (quintal)	Max price (quintal)
MSE	11389.86	20354.72	10510.78	16244.60	88394.59	109952.10
RMSE	106.72	142.67	102.52	127.45	297.31	331.59

Table 4.6.1

From above Table 4.6.1 it is clear that the Exponential smoothing model has the smaller MSE and RMSE values compared to ARIMA model and Linear Regression model. Based on the provided MSE and RMSE values, it is to conclude that the Exponential smoothing model would perform better than the others since it can be considered as the best model for the study in terms of RMSE and MSE

CHAPTER 5

CONCLUSION

In this study it was aimed to forecast minimum prices and maximum prices in the future months of the years 2024 and 2025 of Ernakulam market, situated in the Ernakulam district of Kerala. For forecasting, three different models were chosen ARIMA (Auto Regressive Integrated Moving Average), Exponential Smoothing and Linear Regression. The first three Objectives of this study were to forecast the minimum price and maximum price of rice in quintal for the months of the year's 2024 and 2025. By carefully observing the forecasted values obtained from the above mentioned three models we can conclude that the rice prices for the year's 2024 and 2025 are increasingly fluctuating.

The final Objective of this study was to compare the forecast by ARIMA, Exponential smoothing & linear regression and to find most effective model which can accurately predict the prices of Rice in the Ernakulam Market.

After precise Analysis procedures and the comparison of Mean Squared Error (MSE) and Root Mean Squared Error (RMSE), it was conclusively determined that the Exponential smoothing model would perform better than the others. Exponential smoothing is a time series forecasting model showed superior performance by capturing the trends and patterns in the rice prices of Ernakulam market. This model's ability to adapt with underlying or changing patterns and trends contributed to its accuracy and effectiveness in accurate predictions. While the other models i.e. The ARIMA model and Linear Regression models are tools which can be also used in forecasting, their performance may vary depending on the nature of the data and the underlying patterns.

According to the insights gained from the Exponential smoothing model we can clearly conclude that in the future years 2024 and 2025 the rice prices in Ernakulam market are going to increase exponentially. Accurate Rice predictions enable the market to make decisions regarding the production cycles, pricing strategies, procurement and inventory management etc to maximize profits.

In conclusion this project of study concludes by the effectiveness of Exponential smoothing model in predicting minimum and maximum rice prices in the Ernakulam market.

REFERENCES

1. Abadan, S., & Shabri, A. (2014). Hybrid empirical mode decomposition-ARIMA for forecasting price of rice. Applied Mathematical Sciences, 8(63), 3133-3143.
2. Alexander, C., & Wyeth, J. (1994). Cointegration and market integration: An application to the Indonesian rice market. The Journal of Development Studies, 30(2), 303-334.
3. Dorosh, P. A., & Shahabuddin, Q. (2002). Rice price stabilization in Bangladesh: an analysis of policy options (No. 595-2016-39987).
4. Ghosh, M. (2000). Cointegration tests and spatial integration of rice markets in India. Indian Journal of Agricultural Economics, 55(4), 616-626.
5. Ismet, M., Barkley, A. P., & Llewellyn, R. V. (1998). Government intervention and market integration in Indonesian rice markets. Agricultural Economics, 19(3), 283-295
6. Ohyver, M., & Pudjihastuti, H. (2018). ARIMA model for forecasting the price of medium quality rice to anticipate price fluctuations. Proscenia Computer Science, 135, 707-711.
7. Reddy, P. C. S., Suryanarayana, G., & Yadala, S. (2022, November). Data analytics in farming: rice price prediction in Andhra Pradesh. In 2022 5th International Conference on Multimedia, Signal Processing and Communication Technologies (IMPACT) (pp. 1-5). IEEE.
8. Sarkar, J. P., Raihan, M., Biswas, A., Husain, K. A., Sarder, K., Majumder, N…... & Sana, K. (2022). Paddy price prediction in the south-western region of Bangladesh. In Intelligent Computing & Optimization: Proceedings of the 4th International Conference on Intelligent Computing and Optimization 2021 (ICO2021) 3 (pp. 258-267). Springer International Publishing
9. Sasi, A., & Subramanian, T. (2022). Comparative analysis of ARIMA and double exponential smoothing for forecasting rice sales in fair price shop. Journal of Statistics and Management Systems, 25(7), 1601-1619
10. Silvapulle, P., & Jayasuriya, S. (1994). Testing for Philippines rice market integration: A multiple cointegration approach. Journal of Agricultural Economics, 45(3), 369-380

.