Project Report

On

ANALYZING AND FORECASTING INDIAN CONSUMER PRICE INDEX (CPI) OF MEAT AND FISH - RURAL AND URBAN

Submitted

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

in

APPLIED STATISTICS AND DATA ANALYTICS

by

NAMITHA ELIZABETH JOHN

(Register No. SM22AS011)

(2023-2024)

Under the Supervision of

ANU MARY JOHN

DEPARTMENT OF MATHEMATICS AND STATISTICS

ST. TERESA'S COLLEGE (AUTONOMOUS) ERNAKULAM, KOCHI- 682011

MAY 2024

ST. TERESA'S COLLEGE (AUTONOMOUS), ERNAKULAM

CERTIFICATE

This is to certify that the dissertation entitled, **ANALYZING AND FORECASTING INDIAN CONSUMER PRICE INDEX (CPI) OF MEAT AND FISH - RURAL AND URBAN** is a Bonafide record of the work done by **NAMITHA ELIZABETH JOHN** 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: 29.04.2024

Place: Ernakulam

ANU MARY JOHN

Assistant Professor,

Department of Mathematics and Statistics,

St. Teresa's College (Autonomous),

Ernakulam.

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

Ms. NISHA OOMMEN

Assistant Professor & HOD

Department of Mathematics and Statistics,

St. Teresa's College (Autonomous),

Ernakulam.

External Examiners

1: CHINU JOSEPH

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 **ANU MARY JOHN**, Assistant Professor, Department of Mathematics, St. Teresa's College (Autonomous), Ernakulam has not been included in any other project submitted previously for the award of any degree.

Ernakulam.

NAMITHA ELIZABETH JOHN

Date: 29.04.2024.

SM22AS011

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. Anu Mary John 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: 29.04.2024

NAMITHA ELIZABETH JOHN

SM22AS011

ABSTRACT

In this study, the historical data of CPI for the commodity Meat and Fish was taken as data from Jan 2013 to May 2023 for rural and urban India. This data was analyzed using Seasonal ARIMA and Holt Winters Exponential Smoothing method. Using the model created, the CPI for the coming months till 2025-05-01 was forecasted for urban and rural India. Then the two models were compared using Root Mean Square Error (RMSE) and Mean Square Error (MSE). It was clear that both the RMSE and MSE value less for SARIMA model for urban and rural India. Hence, SARIMA is the best model for forecasting rural and urban India.

ST.TERESA'S COLLEGE (AUTONOMOUS) ERNAKULAM

Certificate of Plagiarism Check for Dissertation

Author Name	NAMITHA ELIZABETH JOHN
Course of Study	M.Sc. Applied Statistics & Data Analytics
Name of Guide	Ms. ANU MARY JOHN
Department	Post Graduate Mathematics & Statistics
Acceptable Maximum Limit	20%
Submitted By	library@teresas.ac.in
Paper Title	ANALYZING AND FORECASTING INDIAN CONSUMER PRICE INDEX (CPI) OF MEAT AND FISH - RURAL AND URBAN
Similarity	5% AI 8%
Paper ID	1664502
Submission Date	2024-04-19 12:39:21

Signature of Student

Signature of Guide

Checked By
College Librarian

TABLE OF CONTENTS

Project Report

On

ANALYZING AND FORECASTING INDIAN CONSUMER PRICE INDEX (CPI) OF MEAT AND FISH - RURAL AND URBAN

Submitted

in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE

in

APPLIED STATISTICS AND DATA ANALYTICS

by

NAMITHA ELIZABETH JOHN

(Register No. SM22AS011)

(2023-2024)

Under the Supervision of

ANU MARY JOHN

DEPARTMENT OF MATHEMATICS AND STATISTICS

ST. TERESA'S COLLEGE (AUTONOMOUS) ERNAKULAM, KOCHI- 682011

MAY 2024

ST. TERESA'S COLLEGE (AUTONOMOUS), ERNAKULAM

CERTIFICATE

This is to certify that the dissertation entitled, **ANALYZING AND FORECASTING INDIAN CONSUMER PRICE INDEX (CPI) OF MEAT AND FISH - RURAL AND URBAN** is a Bonafide record of the work done by **NAMITHA ELIZABETH JOHN** 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:

Place: Ernakulam

ANU MARY JOHN

Assistant Professor,

Department of Mathematics and Statistics,

St. Teresa's College (Autonomous),

Ernakulam.

Ms. NISHA OOMMEN

Assistant Professor & HOD,

Department of Mathematics and Statistics,

St. Teresa's College (Autonomous),

Ernakulam.

External Examiners

1: .... 2: ....

DECLARATION

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

Ernakulam. **NAMITHA ELIZABETH JOHN**

Date: **SM22AS011**

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. Anu Mary John 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. **NAMITHA ELIZABETH JOHN**

Date: **SM22AS011**

TABLE OF CONTENTS

ABSTRACT

In this study, the historical data of CPI for the commodity Meat and Fish was taken as data from Jan 2013 to May 2023 for rural and urban India. This data was analyzed using Seasonal ARIMA and Holt Winters Exponential Smoothing method. Using the model created, the CPI for the coming months till 2025-05-01 was forecasted for urban and rural India. Then the two models were compared using Root Mean Square Error (RMSE) and Mean Square Error (MSE). It was clear that both the RMSE and MSE value less for SARIMA model for urban and rural India. Hence, SARIMA is the best model for forecasting rural and urban India.

CHAPTER 1
INTRODUCTION

The Consumer Price Index (CPI) is a statistical composite indicator that measures the average change in the prices over time that consumers pay for a basket of goods and services (Fernando, 2021). It's basically used to measure the inflation rate and overall economic stability. It is an inflation indicator, carefully assessing the rate at which the pricing of goods or services shifts. This thus helps policy makers to identify consumer power and make better judgements on financial policies and other decisions affecting the economic stability. Furthermore, CPI helps to maintain living standards by adjusting the cost of living. Moreover, it lays the groundwork for budgeting and financial planning for consumers and businesses by providing foresights about future price changes. Additionally, CPI serves as a universal standard for international comparisons, facilitating assessments of inflation rates and cost-of-living differences across different countries.

In this project, we are specifically analyzing and forecasting the CPI of the commodity "Meat and Fish". As we know, it is a staple food item and an essential component of diets in India. By looking into the CPI of this commodity, we get the crucial information about the affordability and accessibility of this commodity for different sectors of the Indian population- Rural and Urban India. Also, analysing CPI for meat and fish enables policymakers to measure inflationary pressures within the food commodities. Price fluctuations in these commodities are influenced by a lot of factors, including supply and demand dynamics, production costs, and market conditions. Looking at these dynamics provides policymakers to understand their impact on overall inflation rates and consumer spending patterns, thereby providing strategic decisions regarding economic policies and interventions. Furthermore, CPI analysis for meat and fish unfolds broader socio-economic trends and differences within society. The affordability and availability of these commodities often reflect underlying socio-economic challenges, such as food insecurity and income inequality. By identifying these challenges, policymakers are able to design targeted interventions and social safety circle to take note of vulnerabilities and promote equitable access to nutritious food options. This ensures taking care of inclusive growth and enhancing the well-being of populations of India.

1.1. About the data

For this research, the data was collected from https://data.gov.in/resource/all-india-consumer-price-index-ruralurban-upto-may-2023 for the years 2013-01-01 to 2023-01-01 on the CPI index of the commodity- Meat and Fish of Rural and Urban India.

1.2. Objectives:

The project mainy focuses on attaining the following objectives:

- To model and forecast the CPI of Meat and Fish in both rural and urban areas using Seasonal ARIMA.
- To model and forecast the CPI of Meat and Fish in both rural and urban areas using Holt Winters exponential smoothening.
- To compare Sarima and Holt Winters algorithm.

To conclude, analyzing and forecasting the Consumer Price Index (CPI) for "Meat and Fish" represents a vital oversight in economic research, especially considering India's socio-economic context. By investigating the affordability and availability of these fundamental food commodity in both rural and urban areas, this study seeks to shed light on essential aspects of consumer influence, inflation patterns, and wider socio-economic discrepancies.

CHAPTER 2

LITERATURE REVIEW

Lidiema (2017) in his paper considered two models of forecasting- the Box-Jenkins procedure employing the SARIMA and the Holt-Winters triple exponential smoothing. Published Consumer Price Index Data from Kenya National Bureau of Statistics (KNBS) for the period November 2011 to October 2016 was used. In this paper the researcher equated the forecasted values of both the models and chose the best model based on the least mean Absolute square error (MASE), mean absolute error (MAE) and mean absolute percentage error (MAPE). The three-step model building for Box-Jenkins was first employed, followed by the Holts-Winters triple exponential smoothing. The study found that the SARIMA Model was a better model than the Holt-winters triple exponential smoothing as per the obtained results using MASE, MAE and MAPE.

Gjika *et. al* (2018) in their work analysed the CPI index as the official index to measure inflation in Albania, Harmonized Indices of Consumer Prices (HICPs) as the bases for comparative measurement of inflation in European countries and other financial indicators that may affect CPI. This study attempted to model CPI based on combination of multiple regression model with time series forecasting models. In the first approach, time series models were used directly on the CPI time series index to obtain the forecast. In the second approach, the time series models (SARIMA, ETS) were used to model and simulate forecast for each subcomponent with significant correlation to CPI. Then it is used the multiple regression model to obtain CPI forecast. The projection of this indicator is important for understanding the country's economic and social development. This study would help researchers in the field of time series modelling, economic analysis and investments.

Boniface & Martin (2019) states that the knowledge of economic and financial indicators is the basis of making right decisions and sound judgment with respect to investment and allocation scare of resources. Such important indicators include the consumer price index, which measures the change in the prices paid by households for goods and services consumed. A trigger in the consumer price in Ghana causes inflation which affects the purchasing power of its citizens. Knowledge of the trend of the CPI is crucial in economic planning. The study therefore sought to construct the appropriate time series model for the CPI and then use the model to predict the next nine months CPI. The study further sought to determine the type of trend model that characterizes the CPI. The Box-Jenkins methodology was adopted. The results of analysis showed SARIMA (2, 1, 1) x (1, 0, 0)12 as most fitted time series model and was used to predict the consumer price index for the next nine months. The S-model was also found to be the appropriate trend model for the CPI. The SARIMA (2, 1, 1) x (1, 0, 0)12 model is recommended for forecasting consumer price index in Ghana as per the study.

Efrilia (2021) on his research paper looks into the Consumer Price Index (CPI) as an essential economic index that shows the level of prices for goods and services consumed by the public in a certain period in a specific region. Forecasting the ICP is needed to find out the pattern of economic movement in the area. The purpose of this study was to determine the forecasting rate for CPI from July 2021 to June 2022 by comparing two forecasting methods, i.e., ARIMA and Exponential Smoothing Holt-Winters. The data used in this study is Tegal City CPI data for January 2014 - June 2021. The year 2018 was set as the base year equals 100 with a time series of 90 observations. The back casting technique was implemented to the CPI figures of January 2014 – December 2019 (Base Year 2012=100) to adjust the new Base Year following 2018 on Classification of Individual Consumption According to Purpose (COICOP). The results from the two methods showed that the Exponential Holt-Winters method has a minor Mean Absolute Percentage Error (MAPE) value, which is 0.281 compared to the MAPE of ARIMA value of 0.311. Hence, the Exponential Holt-Winters Additive method is chosen as the best CPI forecasting model for Tegal City.

Wanjuki *et. al* (2021) believed that attaining price stability is one of the objectives of monetary policy in any economy to protect both consumers' and producers' interest. Unpredictable food and beverages prices make it difficult for consumers to plan for their expenditure in case of unexpected inflation. However, low prices may hurt producers as they may not be able to protect their profit margins. It is therefore imperative to develop a precise and accurate model to forecast Kenya's commodity prices. Therefore, the current sought to model the commodities price of food and beverage in Kenya using a Seasonal Autoregressive Integrated Moving Average (SARIMA). SARIMA model takes into account the seasonal periodic fluctuations in a series that usually recur with about the same time interval. Secondary data on monthly food price index was obtained from the KNBS website. The data covered the period from January 1991 to June 2017 with a total of 318 monthly observations. Data analysis was carried out using the R-statistical software. Using the Maximum Likelihood Estimation method, the SARIMA (0,1,2) x (0,1,1)12 model had better forecasts accuracy than other competing orders based on the Bayesian Information Criterion (BIC=1638.42) criterion with MAE of 2.25 in its forecasting ability. The two-year predictions of food and beverages price index showed an oscillatory behaviour with an increasing trend. The forecasts can help consumers adjust expenditure in preparation for periods of inflation. The research concluded that policymakers should make priorities to ensure stability of future commodity prices.

Konarasinghe (2022) in his research considered the CPI which attempts to quantify the overall price levels of goods and services in an economy and to measure the purchasing power of a respective country's currency unit. The CPI is used in the calculation of many key economic indicators that require a country's unit of currency measures, including estimates of income, earnings, productivity, output, and poverty. Thailand's CPI showed an upward trend. This study is designed to forecasts the CPI of Thailand. Monthly CPI data of Thailand for the period of May 2012 to October 2021 were obtained from the International Monitory Fund (IMF) database. Auto-Regressive Distributed Lag Model (ARDLM), Double Exponential Smoothing (DES), and Autoregressive Integrated Moving Average (ARIMA) models were tested to forecast the CPI of Thailand. Auto Correlation Function (ACF), Anderson Darling test and Ljung-Box Q (LBQ) test were applied to test the model assumptions. The relative and absolute measurements of errors were applied to assess the forecasting ability of the model. Results of the study revealed that the ARDLM with lags 1 and 2 was the most suitable model to forecast the CPI of Thailand. The future values of the CPI could be forecasted by the value of the last month and the month before. It was also strongly recommended to design more studies on modelling the CPI for other countries furthermore.

Zhang (2022) in his research paper considered the CPI as a vital indicator to measure the level of inflation in China. The CPI showed the impact of commodity price changes on the daily lives of the residents. He observed that the CPI data had obvious seasonal time series characteristics. By extracting a total of 59 months from January 2017 to November 2021, a SARIMA model was established for analysis and prediction. The results showed that SARIMA (0, 1, 0) x (0, 1, 1) 12 had the highest degree of fitting and could reflect the future trend of his country's CPI better. Based on that, he used this model to predict the trend of his country's CPI in 2022. He hence found that China's CPI will remain stable at about 102% in 2022. This will provide a certain reference for the decision-making of the government, businesses and other market entities.

Majhi et. al (2023) in their research looks at the prediction of household food price index as a significant challenge for the food industry. This is so especially in developing countries like India, where the majority of the population depends on agriculture for their livelihoods. In their project, they aimed to develop a food price index prediction system for household food items like cereals, millets, and pulses using three popular time-series forecasting models- SARIMA, ETS, and FB Prophet. They used historical price index data to build the forecasting models. The performance of each method was assessed using MAE and RMSE. The results showed that all three methods can effectively predict the demand for food items with high accuracy. However, FB Prophet had better performance than the other two methods when it comes to forecasting accuracy and computation time. The study highlighted the effectiveness of time series forecasting techniques such as SARIMA, ETS, and FB Prophet in predicting the demand for household food items. This can aid in reducing food wastage and improving food supply chain management. Additionally, this study provided valuable insights into the application of time series forecasting methods in the food industry.

Qureshi *et. al* (2023) in his research looks at economic theory where a steady consumer price index (CPI) and its associated low inflation rate (IR) are mostly preferred. CPI is considered a major metric in measuring the IR of a country. These indices are those of price changes and have major significance in monetary policy decisions. In this study, different conventional and machine learning methodologies had been applied to model and forecast the CPI of Pakistan. Pakistan's yearly CPI data from 1960 to 2021 were modelled using seasonal autoregressive moving average (SARIMA), neural network autoregressive (NNAR), and multilayer perceptron (MLP) models. Several forms of the models were compared by employing the root mean square error (RMSE), mean square error (MSE), and mean absolute percentage error (MAPE) as the key performance indicators (KPIs). It is observed that the 20-hidden-layered MLP model appeared as the best-performing model for CPI forecasting based on the key performance indicators. Forecasted values of Pakistan's CPI from 2022 to 2031 showed a notable increase in value which is might be unpleasant to consumers and economic management. Also, the research concludes that if the increasing CPI trend observed if not addressed it will trigger a rising purchasing power, thereby causing higher commodity prices. It was also recommended that the government put vibrant policies in place to address this alarming situation.

Njenga (2024) in his paper applied Holt-Winters exponential smoothing approach to model and forecast monthly CPI in Kenya and South Africa. Monthly data from January 2000 to December 2023 was obtained from Central Bank of Kenya and South Africa department of statistics. Time series decomposition showed that the trend component is the most dominant component in both countries. Kenya Holt-Winters estimated model had parameters 0.6756, 0.0077 and 1 for level smoothing, trend smoothing and seasonal smoothing respectively. On the other hand, South Africa estimated model had parameters 0.8917, 0.1057 and 1 for level smoothing, trend smoothing and seasonal smoothing respectively. The estimated models were efficient and effective as on average the fitted values were less than one percent off the observed values. The initial values for level smoothing, trend smoothing and seasonal smoothing were approximately equal in both countries. The estimated models were then used to predict CPI for the next twelve months. Over the forecast period, South Africa will experience a lower index as compared to Kenya. In both countries, it is expected that monthly CPI will rise according to the researcher.

CHAPTER 3

MATERIALS AND METHODS

3.1. Data Collection:

Data on Consumer Price Index (CPI) of Meat and Fish on urban and rural India is collected for this research from data.gov.in. The data collected is: "D:\(csv)originaldata.csv" from the website:https://data.gov.in/resource/all-india-consumer-price-index-ruralurban-upto-may-2023

3.2. Data Description:

The data collected is the monthly CPI of the commodity Meat and Fish on rural and urban regions of India from January 2013 to May 2023. The attributes include the sector, month, year and CPI of the commodity - Meat and Fish. There are 376 data value.

The first 10 observations of the rural dataset are given by:

date	Meat and fish
01-01-2013	106.3
01-02-2013	108.7
01-03-2013	108.8
01-04-2013	109.5
01-05-2013	109.8
01-06-2013	112.1
01-07-2013	114.9
01-08-2013	115.4
01-09-2013	115.7
01-10-2013	115.4

The first 10 observations of the urban dataset are given by:

date	Meat and fish
01-01-2013	109.1
01-02-2013	112.9
01-03-2013	111.4
01-04-2013	113.4
01-05-2013	114.2
01-06-2013	120.1
01-07-2013	119.2
01-08-2013	120.4
01-09-2013	119.1
01-10-2013	118.1

3.3. Tools for Analysis and Forecasting:

3.3.1. SARIMA (Seasonal Autoregressive Integrated Moving Average) model

A seasonal autoregressive integrated moving average, or SARIMA, is a statistical analysis model that uses time series data to either better understand the data set or to predict future trends.

A seasonal autoregressive integrated moving average mode is a form of regression analysis that gauges the strength of one dependent variable relative to other changing variables. A SARIMA model can be understood by outlining each of its components as follows:

- Seasonal (S): refers to a model with seasonal component.
- Autoregression (AR): refers to a model that shows a changing variable that regresses on its own lagged, or prior, values.
- Integrated (I): represents the differencing of raw observations to allow the time series to become stationary (i.e., data values are replaced by the difference between the data values and the previous values).
- Moving average (MA): incorporates the dependency between an observation and a residual error from a moving average model applied to lagged observations.

3.3.2. Holt-Winters exponential smoothing technique

Holt-Winters is a model of time series behaviour. Forecasting always requires a model, and Holt-Winters is a way to model three aspects of the time series: a typical value (average), a slope (trend) over time, and a cyclical repeating pattern (seasonality).

The Holt-Winters method uses exponential smoothing to encode lots of values from the past and use them to predict "typical" values for the present and future. Exponential smoothing refers to the use of an exponentially weighted moving average (EWMA) to "smooth" a time series.

The process of fitting the model and forecasting is done using Python.

3.4. Tools for Comparison:

3.4.1. Akaiki Information Criterion(AIC)

The Akaike information criterion (AIC) is a mathematical method for evaluating how well a model fits the data it was generated from. In statistics, AIC is used to compare different possible models and determine which one is the best fit for the data. AIC is calculated from:

- the number of independent variables used to build the model.
- the maximum likelihood estimates of the model (how well the model reproduces the data).

The best-fit model according to AIC is the one that explains the greatest amount of variation using the fewest possible independent variables.

3.4.2. Root Mean Square Error (RMSE)

The Root Mean Squared Error (RMSE) is one of the two main performance indicators for a regression model. It measures the average difference between values predicted by a model and the actual values. It provides an estimation of how well the model is able to predict the target value (accuracy).

The formula for RMSE is:

$$RSME = \sqrt{\frac{\sum(y_i - \hat{y}_i)^2}{N - P}}$$

Where,

y_i is the actual value for the i^{th} observation.

$\hat{y}_i$ 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.

3.4.3. Mean Square Error (MSE)

Mean squared error (MSE) measures the amount of error in statistical models. It assesses the average squared difference between the observed and predicted values.

When a model has no error, the MSE equals zero. As model error increases, its value increases. The mean squared error is also known as the mean squared deviation (MSD). The formula for MSE is:

$$MSE = \frac{\sum(y_i - \hat{y}_i)^2}{n}$$

Where,

y_i is the i^{th} observed value.

$\hat{y}_i$ is the corresponding predicted value.

n is the number of observations.

CHAPTER 4

TIME SERIES

Time series analysis is a particular way of analyzing a sequence of data points collected over an interim of time. In time series analysis, examiners record data points at reliable intervals over a set period of time rather than just recording the data points irregularly or randomly. However, this sort of analysis is not just the act of collecting data over time. What sets time series data apart from other data is that the analysis can appear how factors change over time. In other words, time is a significant variable since it appears how the information alters over the course of the information focuses as well as the last comes about. It gives an extra source of data and a set arrange of conditions between the information.

Time series analysis typically requires a huge number of data points to guarantee consistency and reliability. A broad data set guarantees you have a representative sample size and that analysis can cut through noisy data. It moreover guarantees that any trends or patterns found are not exceptions and can account for regular change. Moreover, time series data can be utilized for forecasting—predicting future data based on historical data.

4.1. SARIMA:

SARIMA (Seasonal Auto-Regressive Integrated Moving Average) is an extension of the ARIMA (Autoregressive Integrated Moving Average) model that incorporates seasonality in addition to the non-seasonal components. ARIMA models are widely used for time series analysis and forecasting, while SARIMA models are specifically designed to handle data with seasonal patterns.

It is represented as,

SARIMA (p, d, q) x (P, D, Q) m

where,

p: Trend autoregression order.

d: Trend difference order.

q: Trend moving average order.

P: Seasonal autoregressive order.

D: Seasonal difference order.

Q: Seasonal moving average order.

m: The number of time steps for a single seasonal period.

We use uppercase notation for the seasonal parts of the model, and lowercase notation for the non-seasonal parts of the model.

The seasonal component of SARIMA models adds the following three components:

- Seasonal Autoregressive (P): This component captures the relationship between the current value of the series and its past values, specifically at seasonal lags.
- Seasonal Integrated (D): Similar to the non-seasonal differencing, this component accounts for the differencing required to remove seasonality from the series.
- Seasonal Moving Average (Q): This component models the dependency between the current value and the residual errors of the previous predictions at seasonal lags.

The previous equation represents p, d and q as follows: P represents the order of the seasonal AR model, D represents the number of seasonal variations, Q is the order of the seasonal MA, and s is length of the season (periodicity). In addition, the ωt and B represent the white noise value at period t and the backward shift operator, respectively. Taking into account the relationship between the data, the SARIMA (p, d, q) x (P, D, Q) s model is effectively applied to various time series due to its comparatively small order. Based on the dataset, the period value of the time series s (seasonality) is determined.

4.2. Holt- Winters Exponential Smoothing:

Holt winters' exponential smoothing method is an expansion to Holt's method that finally permits for the capturing of a seasonal component. Since Winter's exponential smoothing is built on top of both single and double exponential smoothing, Winter's method is thus also known as triple exponential smoothing.

Winter's method assumes that the time series has a level, trend and seasonal component. A estimate with Winter's exponential smoothing can be expressed as:

$$F_{t+k}=L_t+kT_t+S_{t+k-M}$$

where,

L_t is the level estimate for time t,

k is the number of estimates into the future,

T_t is the trend assess at time t,

St is the seasonal estimate at time t

M is the number of seasons.

The forecast condition is the extenuation of both the SES and HES methods, finally expanded with the incorporation of the seasonal, S, component.

Just like with Holt's method, the forecasting equation has numerous varieties for each of the types of time series - Additive and Multiplicative

Additive Seasonality

$$\boxed{F_{t+k} = L_t + (k * T_t) + S_{t+k-M}}$$

Multiplicative Seasonality

$$\boxed{F_{t+k} = [L_t + (k * T_t)] * S_{t+k-M}}$$

The additive method is preferred when the seasonal variations are roughly constant through the series, while the multiplicative method is preferred when the seasonal variations are changing proportional to the level of the series. With the additive method, the seasonal component is expressed in absolute terms in the scale of the observed series, and in the level equation the series is seasonally adjusted by subtracting the seasonal component. Within each year, the seasonal component will add up to approximately zero. With the multiplicative method, the seasonal component is expressed in relative terms (percentages), and the series is seasonally adjusted by dividing through by the seasonal component. Within each year, the seasonal component will sum up to approximately m.

4.2.1. Holt-Winters' additive method-

The component for additive method is:

$$\begin{aligned}\hat{y}_{t+h|t} &= \ell_t + hb_t + s_{t+h-m(k+1)} \\ \ell_t &= \alpha(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\ b_t &= \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1} \\ s_t &= \gamma(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m},\end{aligned}$$

where k is the integer part of $(h-1)/m$, which ensures that the estimates of the seasonal indices used for forecasting come from the final year of the sample. The level equation shows a weighted average between the seasonally adjusted observation $(y_t - s_{t-m})$ and the non-seasonal forecast $(\ell_{t-1} + b_{t-1})$ for time t. The trend equation is identical to Holt's linear method. The seasonal equation shows a weighted average between the current seasonal index, $(y_t - \ell_{t-1} - b_{t-1})$, and the seasonal index of the same season last year (i.e., m time periods ago).

The equation for the seasonal component is often expressed as:

$$s_t = \gamma^*(y_t - \ell_t) + (1-\gamma^*)s_{t-m}.$$

If we substitute ℓ_t from the smoothing equation for the level of the component form above, we get

$$s_t = \gamma^*(1-\alpha)(y_t - \ell_{t-1} - b_{t-1}) + [1 - \gamma^*(1-\alpha)]s_{t-m},$$

which is identical to the smoothing equation for the seasonal component we specify here, with $\gamma = \gamma^*(1-\alpha)$. The usual parameter restriction is $0 \le \gamma^* \le 1$, which translates to $0 \le \gamma \le 1-\alpha$.

4.2.2. Holt-Winters' multiplicative method-

The component form for the multiplicative method is:

$$
\begin{aligned}
\hat{y}_{t+h|t} &= (\ell_t + hb_t)s_{t+h-m(k+1)} \\
\ell_t &= \alpha \frac{y_t}{s_{t-m}} + (1-\alpha)(\ell_{t-1} + b_{t-1}) \\
b_t &= \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1} \\
s_t &= \gamma \frac{y_t}{(\ell_{t-1} + b_{t-1})} + (1-\gamma)s_{t-m}
\end{aligned}
$$

CHAPTER 5

RESULTS AND DISCUSSIONS

In this section, the CPI data is analyzed and forecasted till 2025-05-01 using two methods- Seasonal ARIMA and Holts Winter Exponential Smoothing Method and these two methods are compared to find the best method.

Sarima models is used to analyze and forecast seasonal time series data using historical data. Holts winter exponential smoothing on the other hand is used in time series data having both trend and seasonality in them. We then used Root Mean Square Error (RMSE) and Mean Square Error (MSE). The model with the least RMSE and MSE value is the best model for forecasting the CPI.

5.1. Modelling and Forecasting CPI using SARIMA

5.1.1. Rural India

The basic statistics was done on the dataset and we got the following initial details as given in fig 5.1.

	Meat and fish
count	125.000000
mean	155.415000
std	32.450513
min	106.300000
25%	130.600000
50%	144.400000
75%	187.500000
max	217.200000

Fig. 5.1.

Line chart was plotted with date on x axis and CPI on y axis. Then we got the following graph fig 5.2.

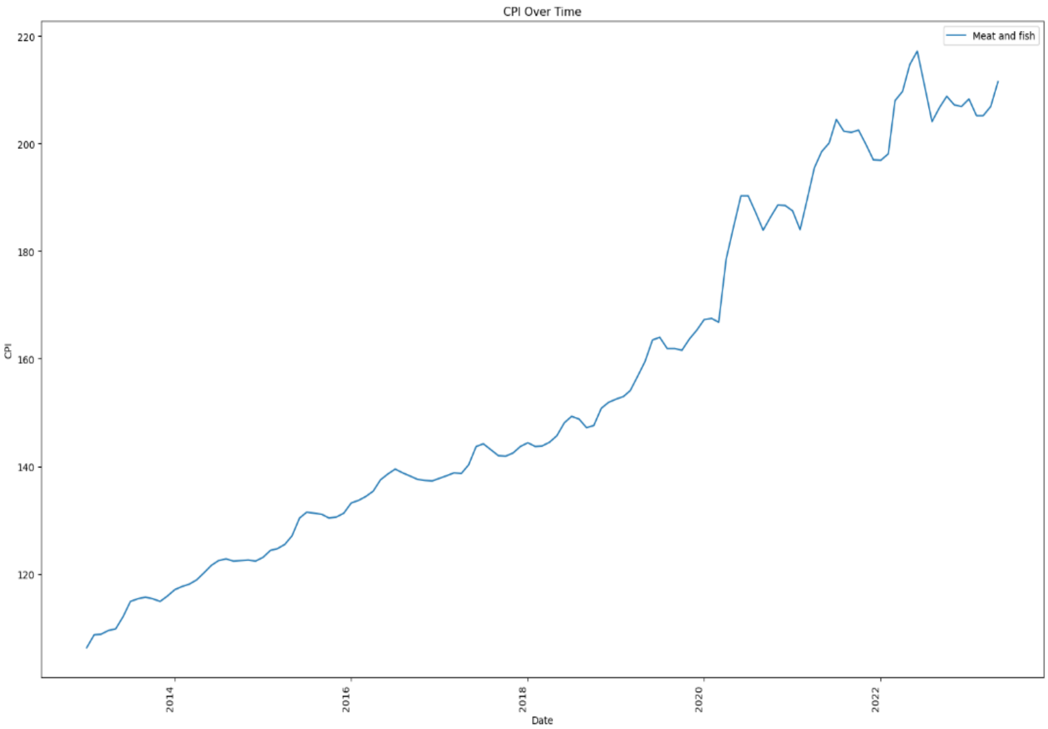

Fig. 5.2.

ADF test for the commodity ‘Meat and Fish’ was done and we got the following result:

ADF Statistic for Commodity: -0.3530455846156664

p-value for Commodity: 0.9176482493840616

Critical Values for Commodity: {'1%': -3.490683082754047, '5%': -2.8879516565798817, '10%': -2.5808574442009578}

Fail to reject the null hypothesis that the time series is non-stationary.

Therefore, the time series is not stationary.

Seasonal decomposition was performed and plotted as in fig. 5.3.

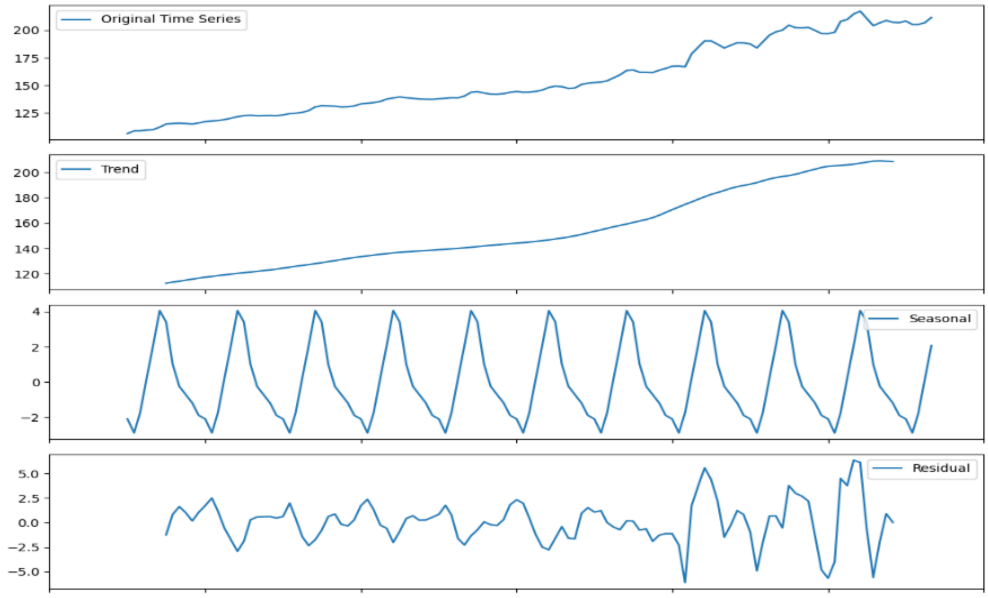

Fig.5.3.

We can observe an upward trend with seasonality in the data.

Using grid search, all models and their AIC values are found out. Few of them are given as in Table 5.1.

Table 5.1.

No	Model	AIC
1	**ARIMA(0,1,1)x(0,1,1)$_{12}$**	**449.646**
2	ARIMA(0,1,1)x(1,1,1)$_{12}$	451.646
3	ARIMA(1,1,1)x(1,1,1)$_{12}$	453.213
4	ARIMA(1,1,0)x(0,1,1)$_{12}$	454.339
5	ARIMA(0,1,0)x(0,1,1)$_{12}$	455.036
6	ARIMA(1,1,0)x(1,1,1)$_{12}$	456.339
7	ARIMA(0,1,0)x(1,1,1)$_{12}$	457.036
8	ARIMA(1,0,1)x(1,1,1)$_{12}$	457.135
9	ARIMA(1,0,0)x(1,1,1)$_{12}$	462.501

From this, it is clear that ARIMA(0, 1, 1)x(0, 1, 1)$_{12}$ is the best model.

Next, the model is trained and its performance is examined through summary table given by Table 5.2.

Table 5.2.

	coef	std err	z	P>|z|	[0.025	0.975]
ma.L1	0.2166	0.084	2.586	0.010	0.052	0.381
ma.S.L12	-0.6881	0.069	-9.994	0.000	-0.823	-0.553
sigma2	5.2569	0.377	13.946	0.000	4.518	5.996

The diagnostic plot for the model is observed and it is found that we can proceed with the model. The diagnostic plot is given as in fig.5.4.

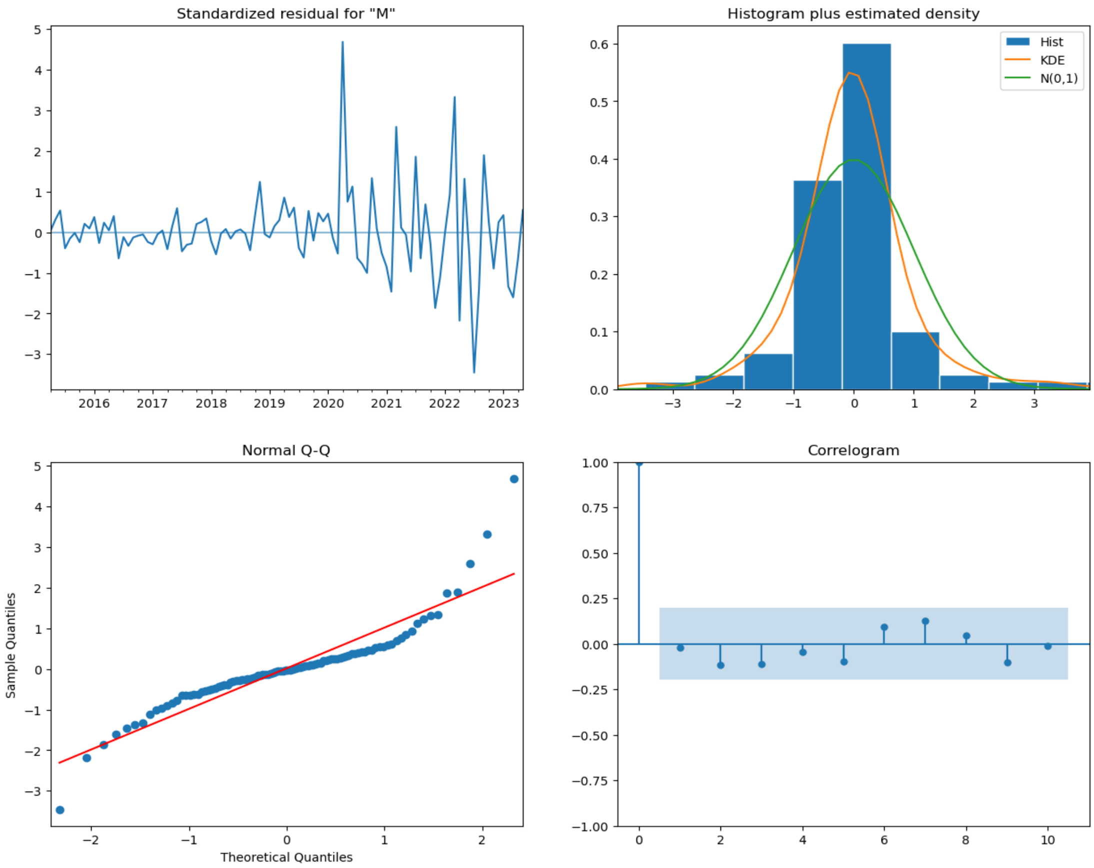

Fig.5.4.

The actual vs predicted CPI was found out using the best model obtained from above. It is given in the Table 5.3.

Table 5.3.

DATE	ACTUAL CPI	PREDICTED CPI
2018-01-01	144.4	144.903168
2018-02-01	143.7	144.963094
2018-03-01	143.8	143.880310
2018-04-01	144.5	144.320206
2018-05-01	145.7	146.058307
…	…	…
2023-01-01	208.3	207.336842
2023-02-01	205.2	208.257719
2023-03-01	205.2	208.875452
2023-04-01	206.9	208.397057
2023-05-01	211.5	210.244949

The actual vs predicted graph is plotted in fig.5.5.

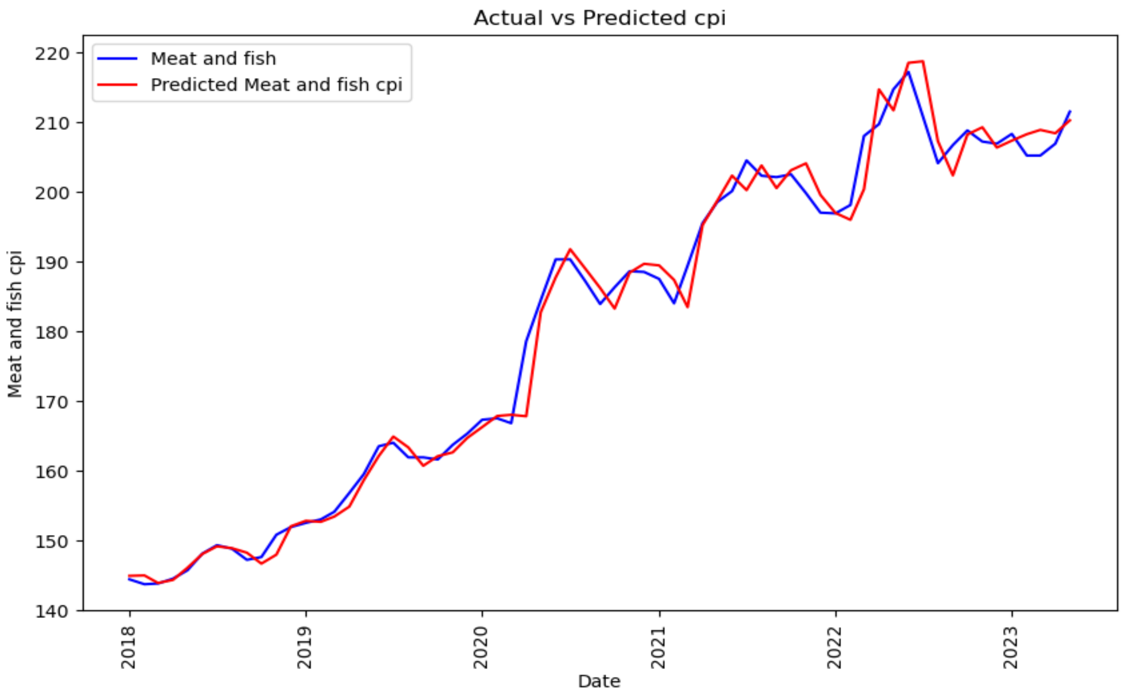

Fig.5.5.

Using this model, the monthly CPI values for the next 2 years from 2023-06-01 to2025-05-01 was also obtained along with its LCL and UCL as Table 5.4. as given.

Table 5.4.

DATE	FORECASTED VALUE	LCL	UCL
2023-06-01	214.71	210.22	219.20
2023-07-01	213.95	206.87	221.03
2023-08-01	210.60	201.66	219.55
2023-09-01	210.69	200.20	221.17
2023-10-01	211.73	199.91	223.56
2023-11-01	211.46	198.43	224.48
2023-12-01	211.08	196.95	225.20
2024-01-01	211.73	196.58	226.88
2024-02-01	210.59	194.48	226.70
2024-03-01	213.57	196.57	230.58
2024-04-01	216.85	198.99	234.72
2024-05-01	220.81	202.13	239.50
2024-06-01	223.84	203.93	243.75
2024-07-01	223.08	201.92	244.24
2024-08-01	219.73	197.39	242.08
2024-09-01	219.81	196.35	243.28
2024-10-01	220.86	196.32	245.40
2024-11-01	220.59	195.02	246.15
2024-12-01	220.21	193.65	246.76
2025-01-01	220.86	193.35	248.37
2025-02-01	219.72	191.29	248.15
2025-03-01	222.70	193.39	252.02
2025-04-01	225.98	195.80	256.16
2025-05-01	229.94	198.92	260.96

It is then plotted as following with date as x axis and CPI as y axis as fig 5.6.

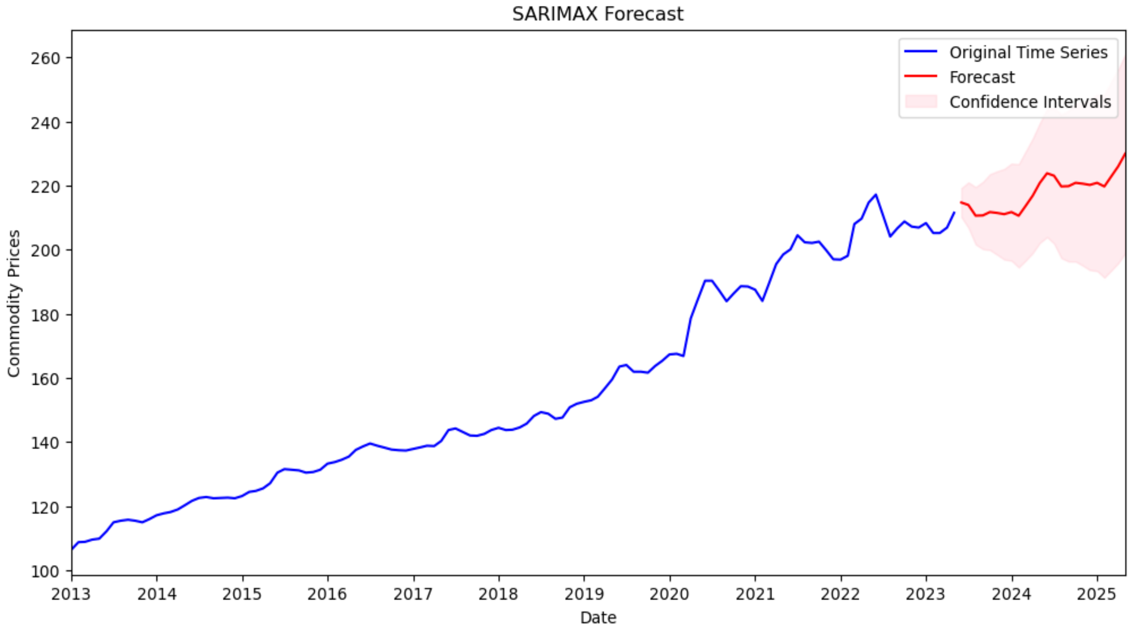

Fig 5.6.

The forecasted CPI of Rural India for Meat and Fish shows an upward trend in the coming years as per the forecast. Hence, we predicted the CPI for Meat and Fish for 2 years of Rural India using Seasonal ARIMA.

5.1.1. Urban India

The basic statistics was done on the dataset and we got the following initial details as given in fig 5.7.

	Meat and fish
count	125.000000
mean	158.235200
std	34.127563
min	109.100000
25%	129.800000
50%	144.300000
75%	195.500000
max	223.400000

Fig.5.7.

Line chart of the points was plotted with date on x axis and CPI on y axis as in fig.5.8.

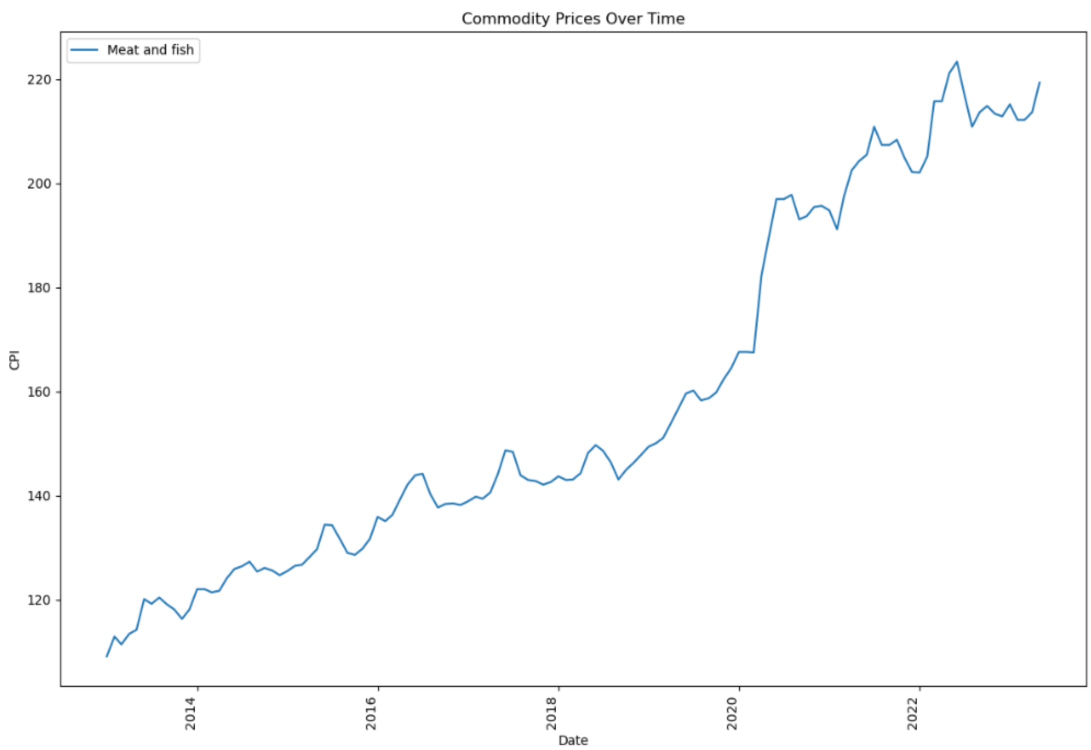

Fig.5.8.

ADF test for the commodity 'Meat and Fish' was done and we got the following result:

ADF Statistic for Commodity: -0.2988149442513617

p-value for Commodity: 0.9256720039547217

Critical Values for Commodity: {'1%': -3.490683082754047, '5%': -2.8879516565798817, '10%': -2.5808574442009578}

Fail to reject the null hypothesis that the time series is non-stationary.

Therefore, the time series is non-stationary.

Seasonal decomposition was performed and plotted as given in fig.5.9.

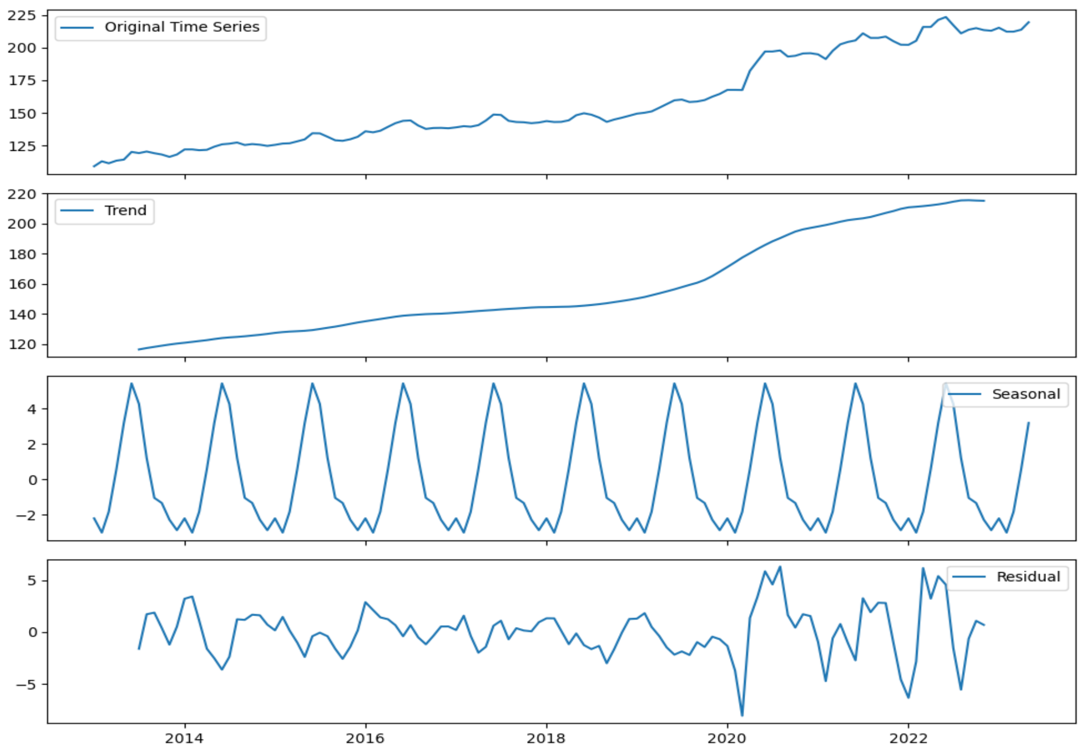

Fig.5.9.

Using grid search, all models and their AIC values are found out. Few of them are given as table 5.5.

Table 5.5.

No	Model	AIC
1	**ARIMA(0,1,1)x(0,1,1)$_{12}$**	**485.708**
2	ARIMA(1,1,1)x(0,1,1)$_{12}$	487.325
3	ARIMA(0,1,1)x(1,1,1)$_{12}$	487.683
4	ARIMA(1,1,1)x(1,1,1)$_{12}$	489.315
5	ARIMA(0,1,0)x(0,1,1)$_{12}$	489.793
6	ARIMA(1,1,0)x(0,1,1)$_{12}$	490.306
7	ARIMA(0,1,0)x(1,1,1)$_{12}$	491.743
8	ARIMA(1,1,0)x(1,1,1)$_{12}$	492.286
9	ARIMA(1,0,1)x(1,1,1)$_{12}$	493.810

From this, it is clear that ARIMA(0, 1, 1)x(0, 1, 1)$_{12}$ is the best model.

The model is trained and its performance is examined through summary table given as in table 5.6.

Table 5.6.

	coef	std err	z	P>|z|	[0.025	0.975]
ma.L1	0.1197	0.091	1.313	0.189	0.059	0.298
ma.S.L12	-0.8293	0.094	-8.796	0.000	-1.014	-0.645
sigma2	7.2708	0.561	12.958	0.000	6.171	8.371

The diagnostic plot for the model is observed and it is found that we can proceed with the model. The diagnostic plot is given as in fig 5.10.

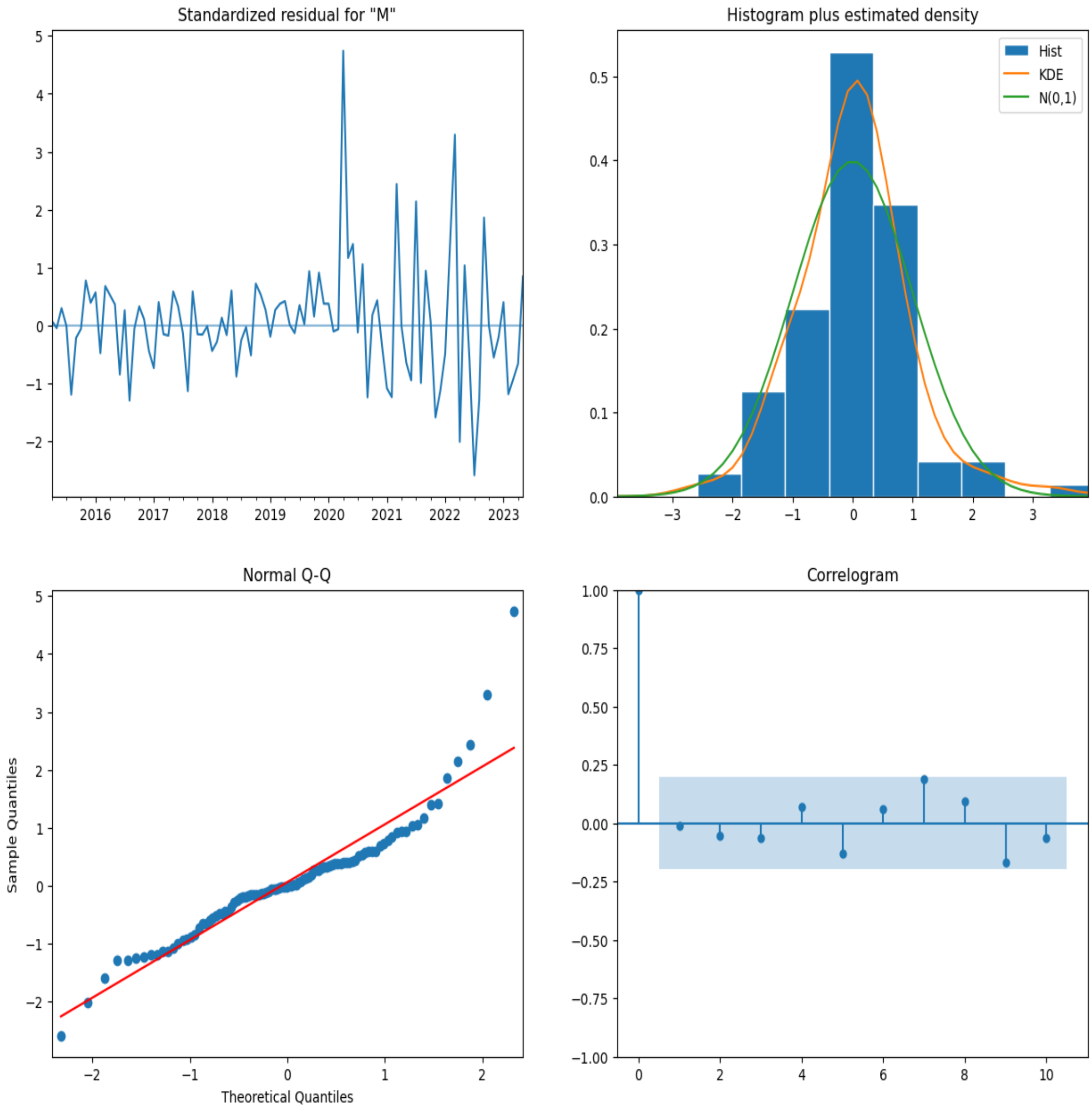

Fig 5.10.

The actual vs predicted CPI was found out using the best model obtained from above and noted in table 5.7.

Table 5.7.

DATE	ACTUAL CPI	PREDICTED CPI
2018-01-01	143.7	144.941312
2018-02-01	143.0	143.802213
2018-03-01	143.1	142.734517
2018-04-01	144.3	144.761063
2018-05-01	148.2	146.539707
…	…	…
2023-01-01	215.2	214.118051
2023-02-01	212.2	215.420880
2023-03-01	212.2	214.719416
2023-04-01	213.7	215.493110
2023-05-01	219.4	217.119192

It is plotted as in fig.5.11.

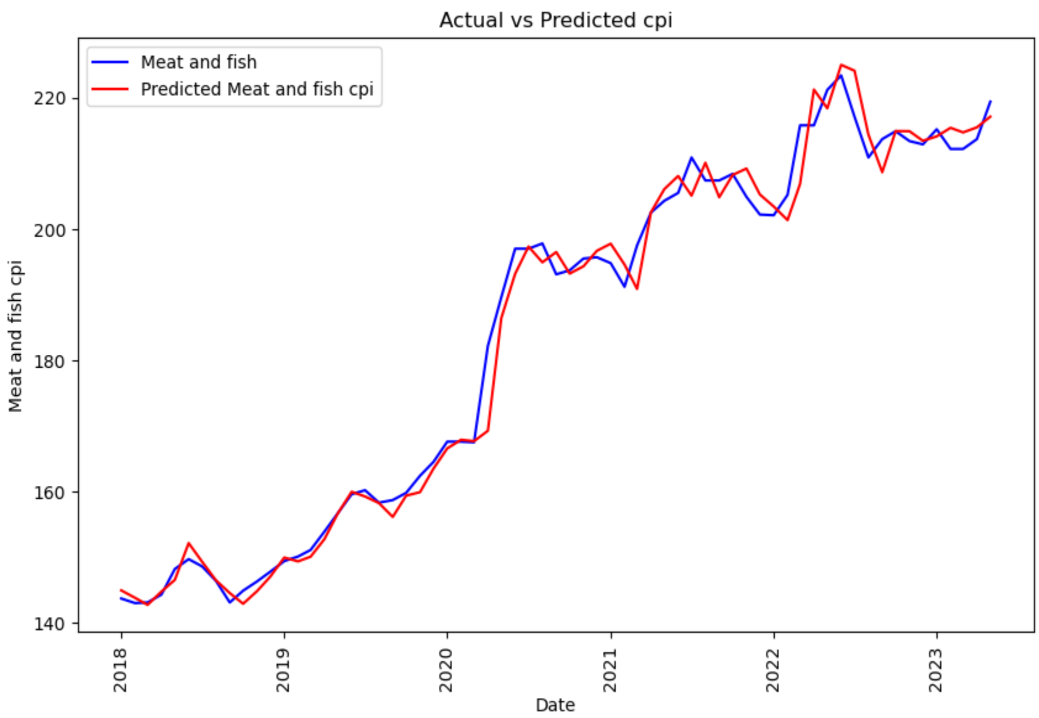

Fig.5.11.

Using this model, the monthly CPI values for the next 2 years from 2023-06-01 to 2025-05-01 was also obtained along with its LCL and UCL as in table 5.8.

Table 5.8.

DATE	FORECASTED VALUE	LCL	UCL
2023-06-01	222.92	217.61	228.22
2023-07-01	222.53	214.57	230.49
2023-08-01	219.88	209.95	229.82
2023-09-01	218.89	207.32	230.47
2023-10-01	219.61	206.60	232.62
2023-11-01	219.35	205.05	233.65
2023-12-01	219.44	203.95	234.93
2024-01-01	220.91	204.32	237.49
2024-02-01	220.44	202.83	238.06
2024-03-01	222.83	204.24	241.42
2024-04-01	226.06	206.54	245.57
2024-05-01	230.06	209.66	250.45
2024-06-01	233.35	211.83	254.87
2024-07-01	232.97	210.34	255.59
2024-08-01	230.32	206.64	254.00
2024-09-01	229.33	204.64	254.02
2024-10-01	230.05	204.39	255.71
2024-11-01	229.79	203.20	256.38
2024-12-01	229.88	202.38	257.37
2025-01-01	231.34	202.98	259.71
2025-02-01	230.88	201.67	260.09
2025-03-01	233.27	203.24	263.30
2025-04-01	236.49	205.66	267.32
2025-05-01	240.49	208.89	272.10

The above forecasted values are then plotted as following with date as x axis and CPI as y axis as in fig.5.12.

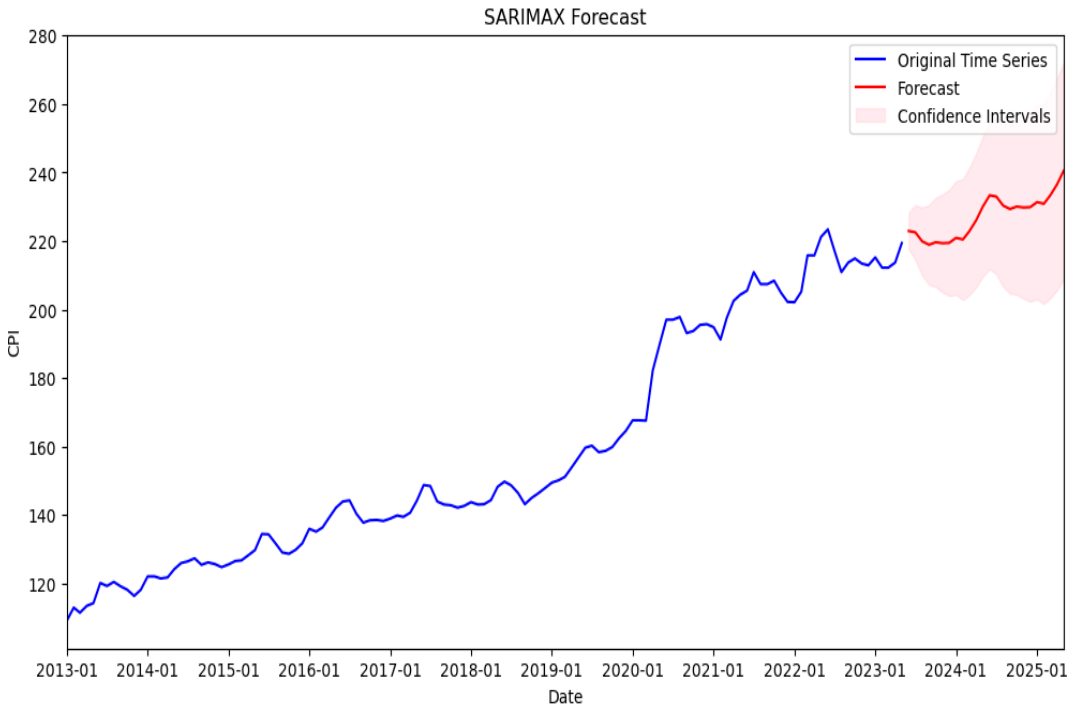

Fig.5.12

5.2. Modelling and Forecasting CPI using Holt-Winters

5.2.1. Rural India

After setting the training and testing dataset, an initial model was found out using Holt method. Using this model, a graph was plotted as in fig.5.13.

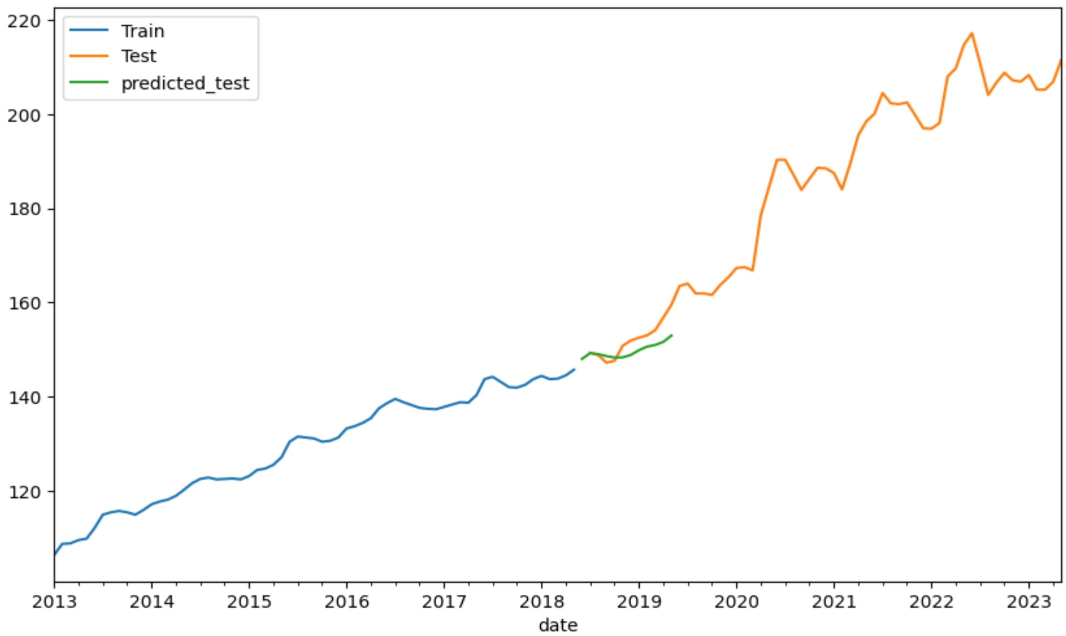

Fig.5.13.

After this, seasonal decomposition was done and was plotted as in fig.5.14.

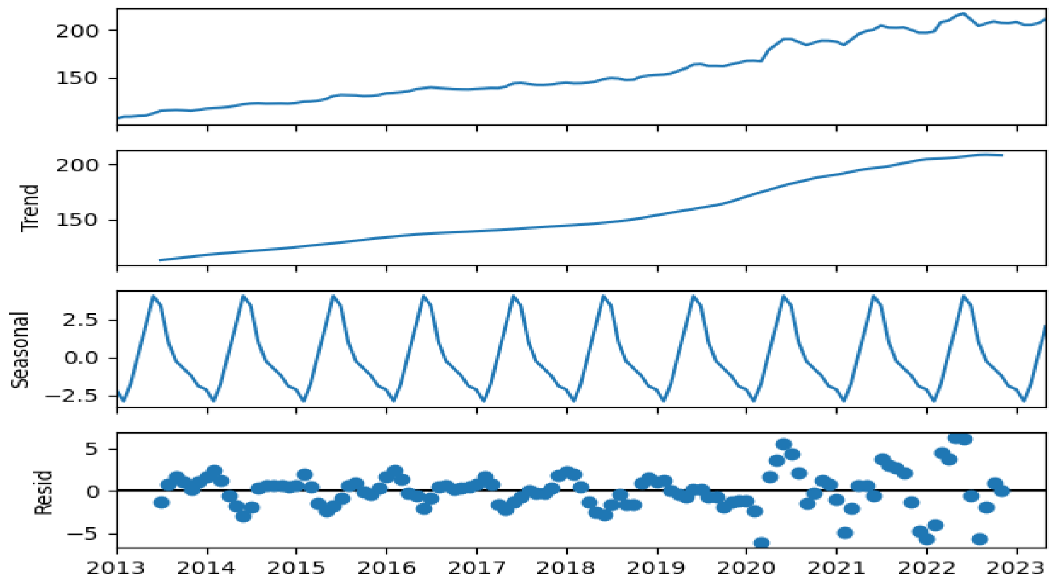

Fig.5.14.

The final model was formed and using this a diagnostic plot were plotted as given in fig.5.15.

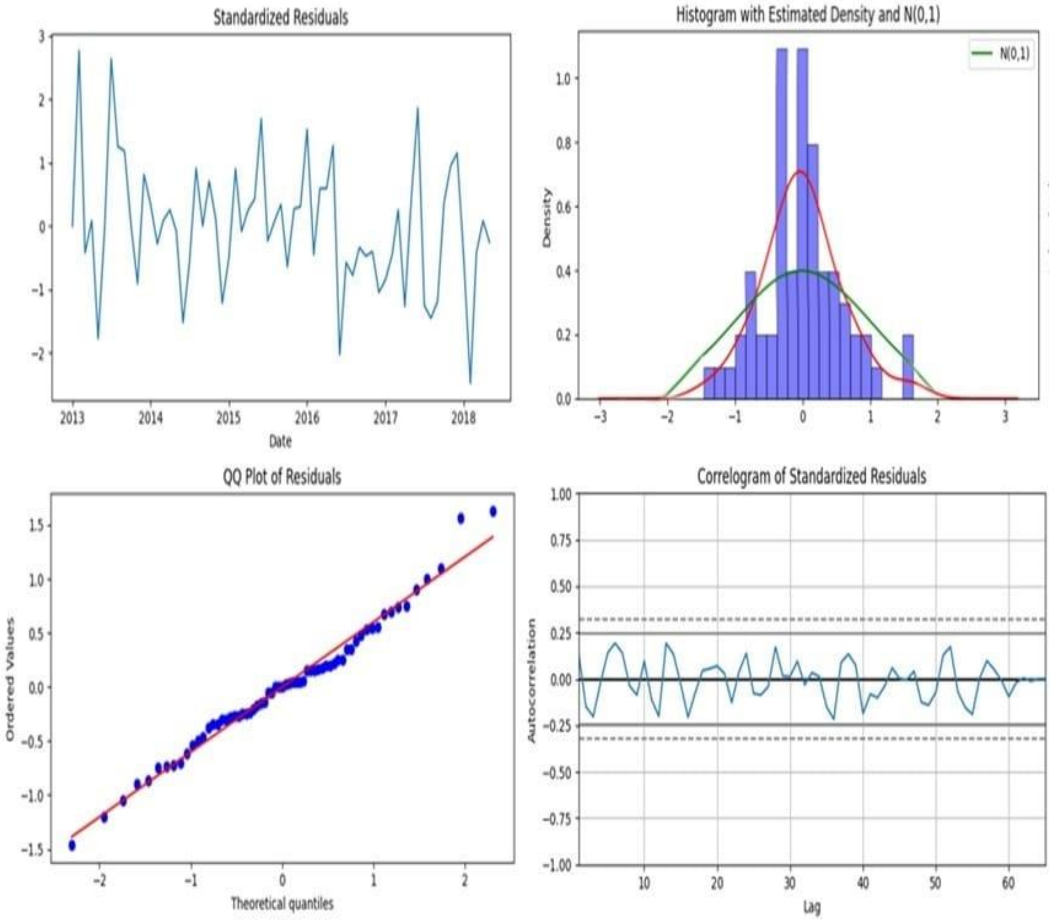

Fig.5.15.

The actual vs predicted CPI was found out using the best model as given in table 5.9.

Table 5.9.

DATE	ACTUAL CPI	PREDICTED CPI
2013-01-01	106.3	106.299661
2013-02-01	108.7	106.290916
2013-03-01	108.8	110.318220
2013-04-01	109.5	111.331840
2013-05-01	112.1	112.161358
…	…	…
2023-01-01	208.3	207.689950
2023-02-01	205.2	208.290916
2023-03-01	205.2	206.818220
2023-04-01	206.9	207.731840
2023-05-01	211.5	209.561358

These points were plotted in the graph as shown in fig 5.16.

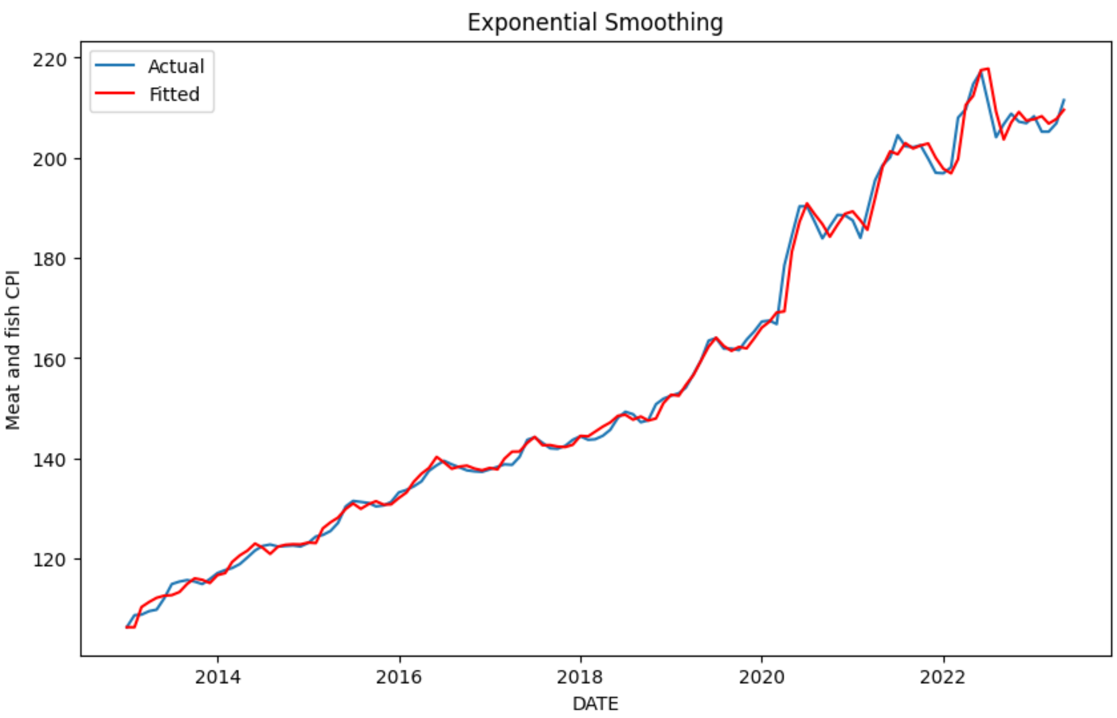

Fig.5.16.

Using this model, the monthly CPI values for the next 2 years was also obtained along with its LCL and UCL as in table 5.10.

Table 5.10.

DATE	FORECASTED VALUE	LCL	UCL
2023-06-01	214.28	205.05	223.50
2023-07-01	214.87	205.64	224.09
2023-08-01	213.29	204.06	222.51
2023-09-01	212.84	203.61	222.06
2023-10-01	213.18	203.95	222.40
2023-11-01	213.53	204.30	222.75
2023-12-01	213.74	204.51	222.96
2024-01-01	214.53	205.30	223.75
2024-02-01	214.52	205.29	223.74
2024-03-01	216.14	206.91	225.36
2024-04-01	218.67	209.44	227.89
2024-05-01	221.33	212.10	230.55
2024-06-01	224.12	214.89	233.34
2024-07-01	224.71	215.48	233.93
2024-08-01	223.13	213.90	232.35
2024-09-01	222.68	213.45	231.90
2024-10-01	223.02	213.79	232.24
2024-11-01	223.37	214.14	232.59
2024-12-01	223.58	214.35	232.80
2025-01-01	224.37	215.14	233.59
2025-02-01	224.36	215.13	233.58
2025-03-01	225.98	216.75	235.20
2025-04-01	228.51	219.28	237.73
2025-05-01	231.17	221.94	240.39

It is then plotted as following with date as x axis and CPI as y axis as in fig.5.17.

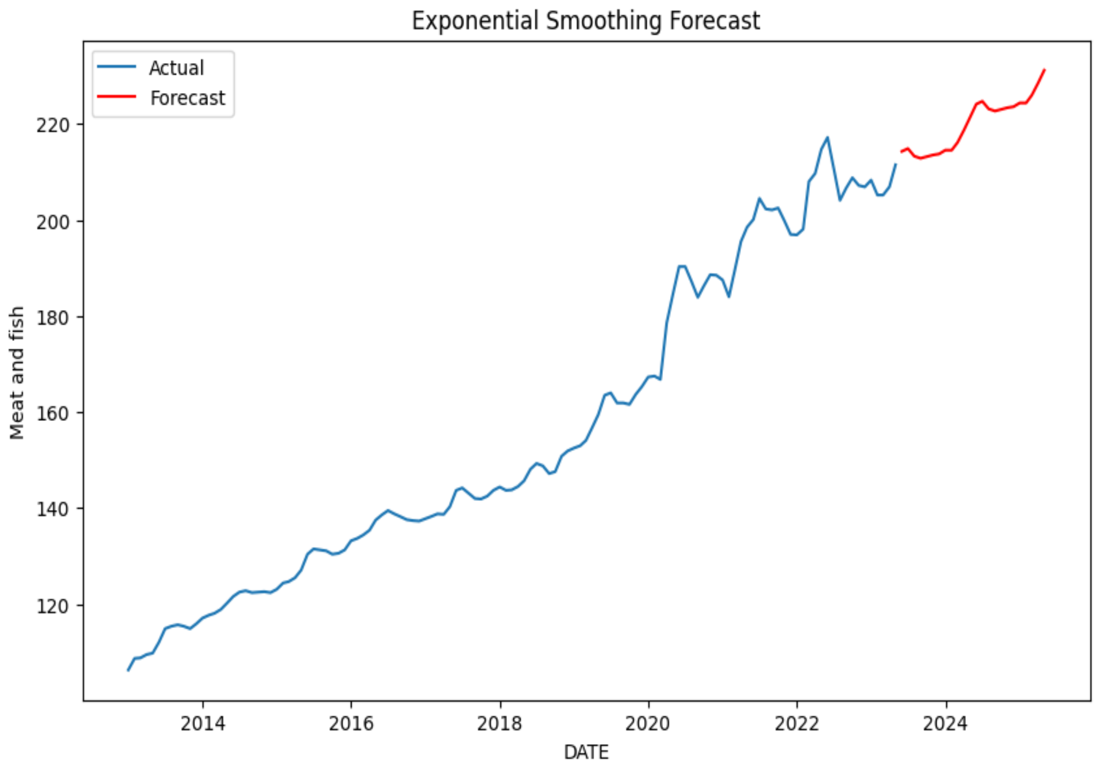

Fig.5.17.

5.2.2. Urban India

After setting the training and testing dataset, an initial model was found out using Holt method. Using this model, a graph was plotted as in fig.5.18.

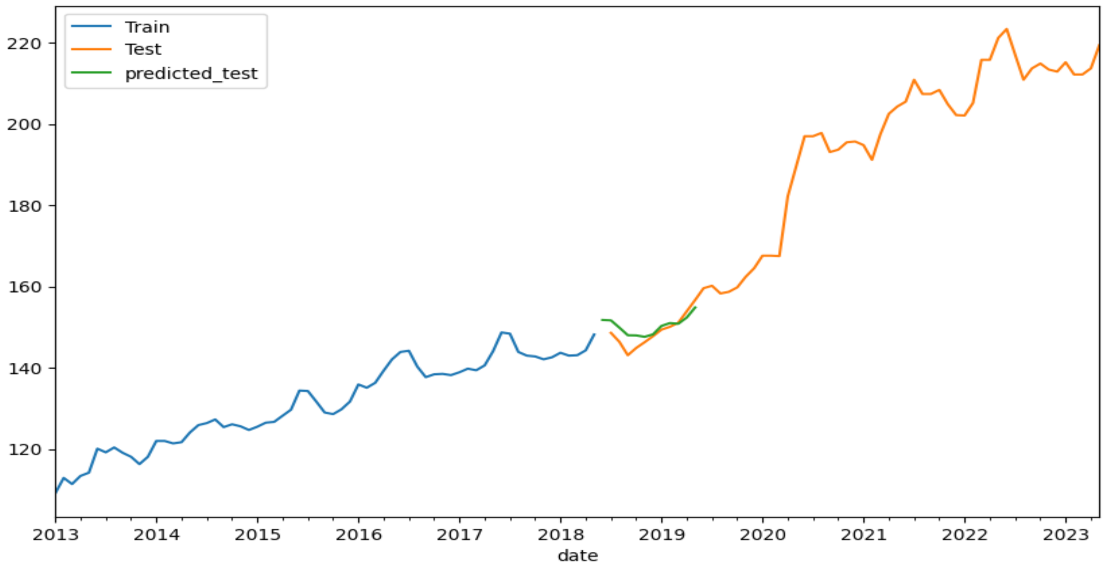

Fig.5.18.

After this, seasonal decomposition was done and was plotted as in fig.5.19.

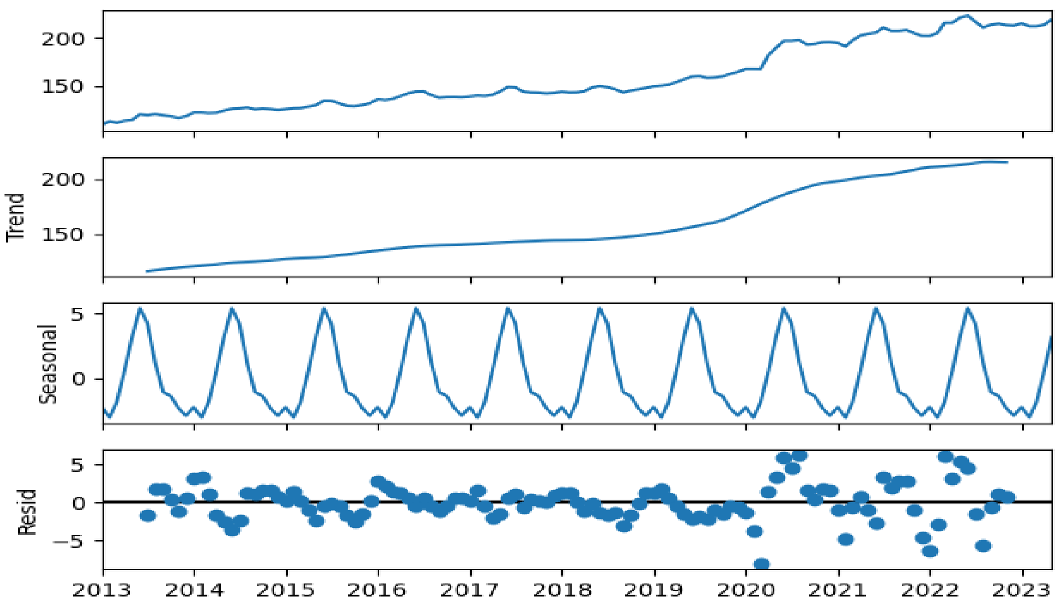

Fig.5.19.

The final model was formed and using this a diagnostic plot were plotted as given in fig.5.20.

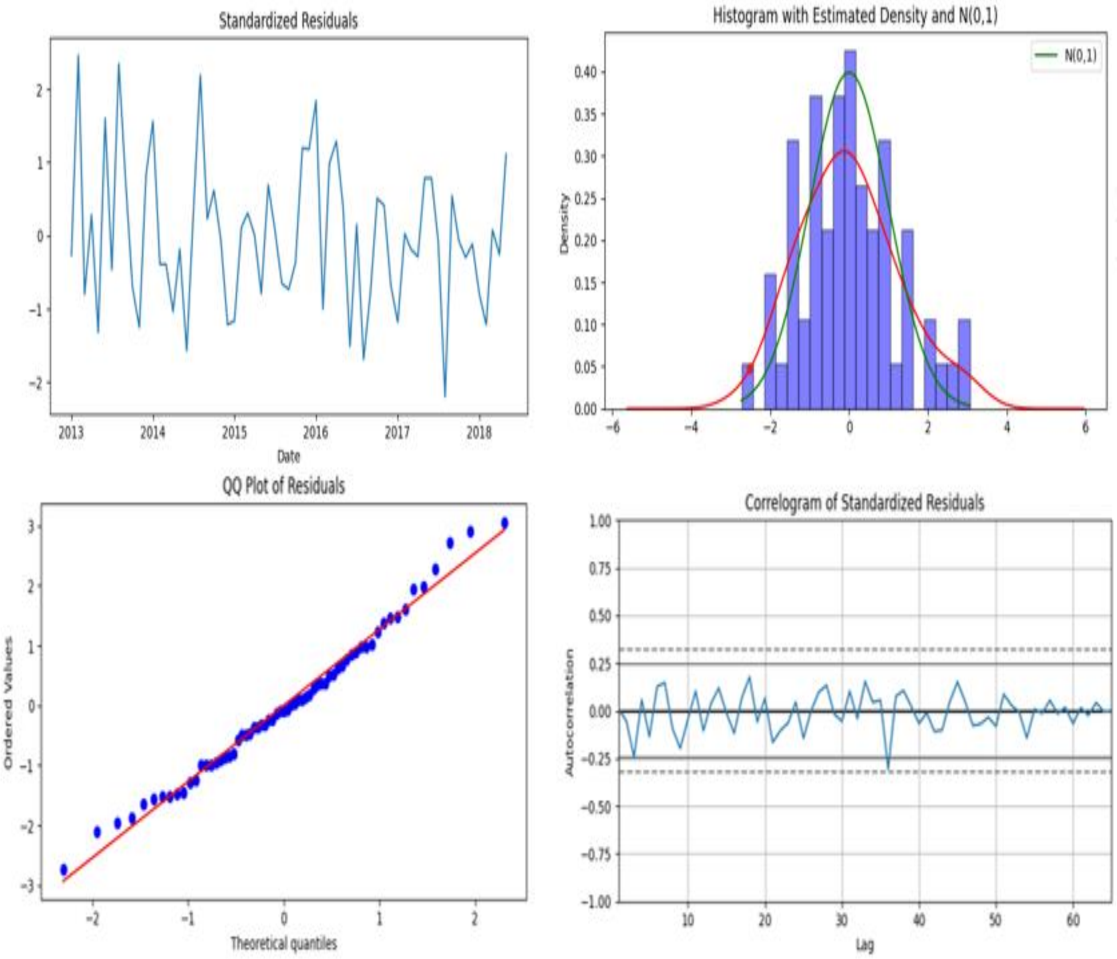

Fig.5.20.

The actual vs predicted CPI was found out using the best model obtained from above as in table 5.11.

Table 5.11.

DATE	ACTUAL CPI	PREDICTED CPI
2013-01-01	109.1	109.100016
2013-02-01	112.9	109.227266
2013-03-01	111.4	114.427217
2013-04-01	113.4	114.418210
2013-05-01	114.2	116.854527
…	…	…
2023-01-01	215.2	214.570019
2023-02-01	212.2	215.327266
2023-03-01	212.2	213.727217
2023-04-01	213.7	215.218210
2023-05-01	219.4	217.154527

The above points were plotted and the graph was obtained as in fig 5.21.

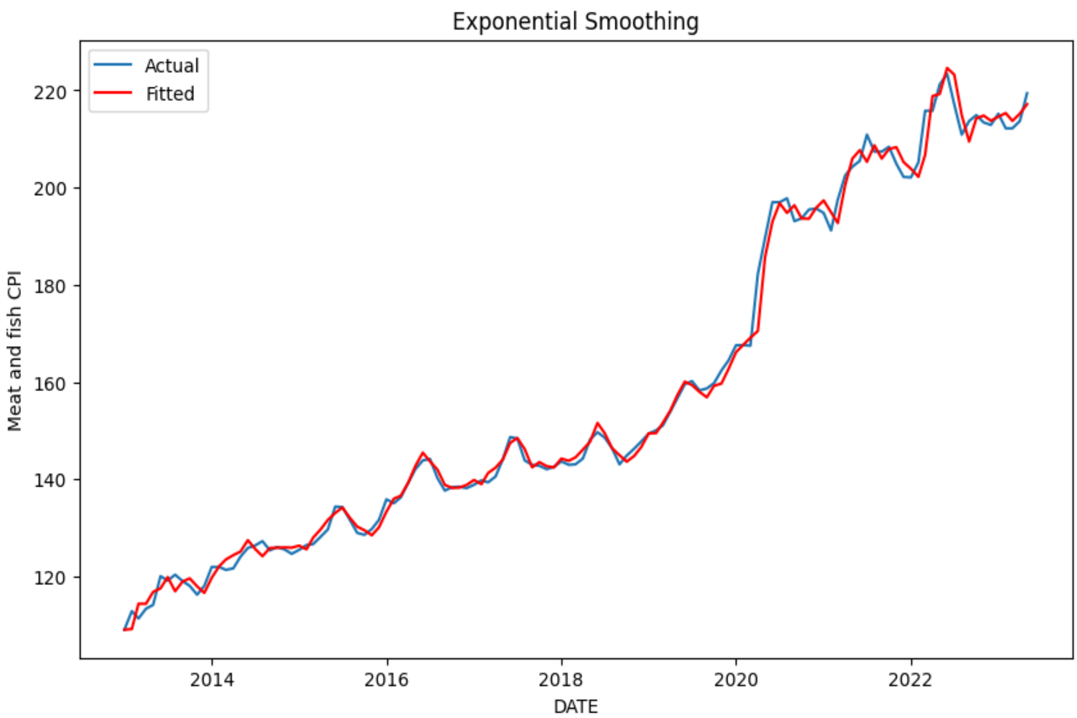

Fig.5.21.

Using this model, the monthly CPI values for the next 2 years was also obtained along with its LCL and UCL as given in Table 5.12.

Table 5.12.

DATE	FORECASTED VALUE	LCL	UCL
2023-06-01	222.80	215.90	229.70
2023-07-01	222.61	215.71	229.51
2023-08-01	220.42	213.52	227.32
2023-09-01	219.00	212.10	225.90
2023-10-01	219.55	212.65	226.45
2023-11-01	219.46	212.56	226.36
2023-12-01	219.82	212.92	226.72
2024-01-01	221.49	214.59	228.39
2024-02-01	221.62	214.72	228.51
2024-03-01	223.14	216.25	230.04
2024-04-01	226.16	219.27	233.06
2024-05-01	229.62	222.72	236.51
2024-06-01	233.02	226.12	239.91
2024-07-01	232.83	225.93	239.72
2024-08-01	230.64	223.74	237.53
2024-09-01	229.22	222.32	236.11
2024-10-01	229.77	222.87	236.66
2024-11-01	229.68	222.78	236.57
2024-12-01	230.04	223.14	236.93
2025-01-01	231.71	224.81	238.60
2025-02-01	231.83	224.94	238.73
2025-03-01	233.36	226.47	240.26
2025-04-01	236.38	229.48	243.28
2025-05-01	239.83	232.94	246.73

It was then plotted as following with date as x axis and CPI as y axis as given below in fig.5.22

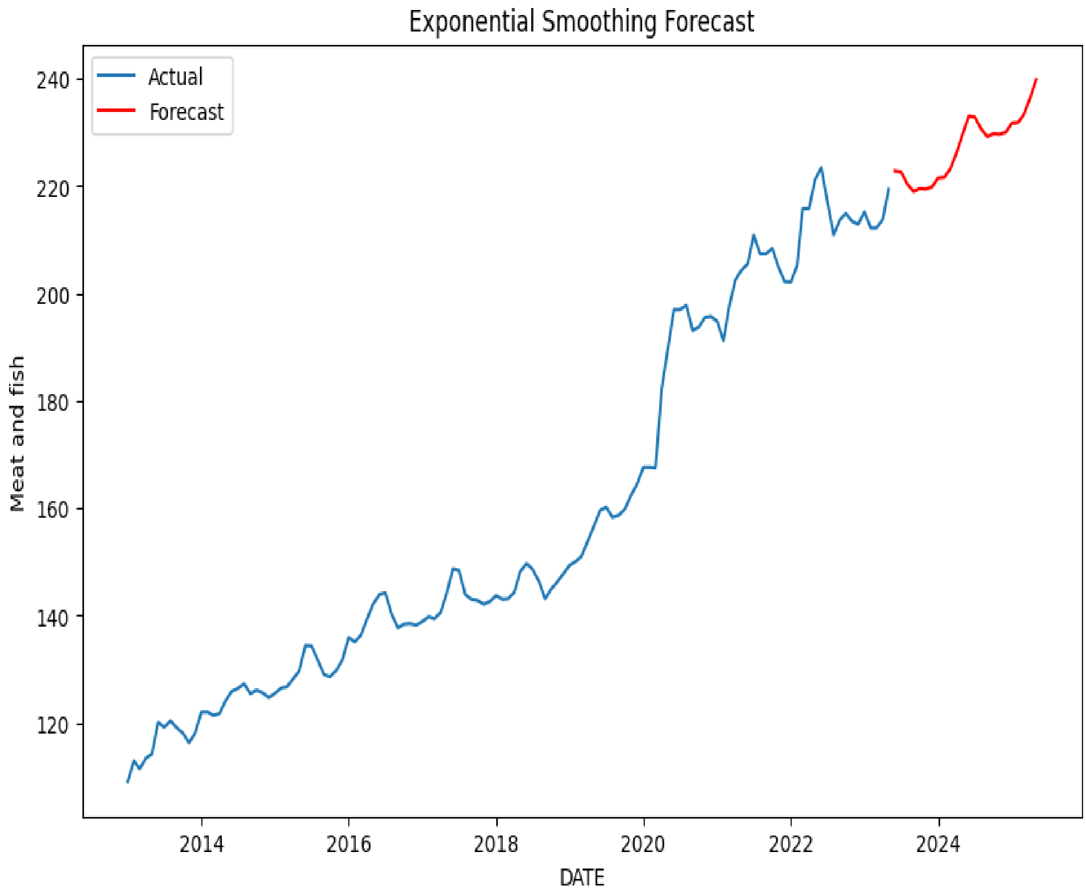

Fig.5.22.

5.3. Comparing SARIMA and Holt winters model

The RMSE and MSE values of both models are compared to find the best method. It was observed as follows:

	RURAL-SARIMA	RURAL-HOLT WINTERS	URBAN-SARIMA	URBAN-HOLT WINTERS
RMSE	2.7749	4.7300	3.1778	3.5184
MSE	7.70058	15.3733	10.0989	12.3793

From the above table it is clear that the SARIMA model is better for both Rural and Urban India for forecasting CPI as both for rural and urban, the RMSE and MSE value is less for SARIMA than Holt Winters.

CHAPTER 6

CONCLUSION

In this project, the dataset includes monthly CPI of the commodity Meat and Fish on rural and urban regions of India from January 2013 to May 2023. The attributes include the sector, month, year and CPI of the commodity - Meat and Fish. There were 376 data value. SARIMA and Holt-Winters Exponential Smoothing methods were used to model and predict the CPI for "Meat and Fish" for the two upcoming years. Both these were used in modelling and forecasting future values. Using performance indicators like Root Mean Square Error (RMSE) and Mean Square Error (MSE), it can be concluded that for both the urban and rural values, SARIMA model was the best as it provided better accuracy.

Also, from this project one can understand the future CPI values for the coming years that will help policymakers to understand the overall consumer spending pattern for the commodity- "Meat and Fish" and hence introduce better economic policies and interventions for the country.

To conclude, to model and forecast CPI values, the best model would be SARIMA as it provided with more accurate results.

CHAPTER 7

REFERENCES

1. Boniface, A., & Martin, A. (2019). Time Series Modeling and Forecasting of Consumer Price Index in Ghana. Journal of Advances in Mathematics and Computer Science, 32(2), 1-11.
2. Efrilia, I. (2021). Comparison of ARIMA and Exponential Smoothing Holt-Winters Methods for Forecasting CPI in The Tegal City, Central Java. Jurnal Ekonomi Pembangunan, 19(02), 97-106.
3. Gjika, E., Puka, L., & Zaçaj, O. (2018, May). Forecasting consumer price index (CPI) using time series models and multi regression models (Albania case study). In *10th International Scientific Conference "Business and Management 2018* (pp. 466-472).
4. Hassan, A., Alam, M. M., & Faeique, A. (2024). Forecasting Monthly Inflation in Bangladesh: A Seasonal Autoregressive Moving Average (SARIMA) Approach. *Journal of Economics and Financial Analysis*, *7*(2), 25-43.
5. Konarasinghe, K. M. U. B. (2022). Modeling Consumer Price Index of Thailand. Konarasinghe, KMUB (2022). Modeling Consumer Price Index of Thailand. Journal of New Frontiers in Education and Social Sciences, 2(1), 1-15.
6. Lidiema, C. (2017). Modelling and forecasting inflation rate in Kenya using SARIMA and Holt-Winters triple exponential smoothing. American Journal of Theoretical and Applied Statistics, 6(3), 161-169.
7. Majhi, S. K., Bano, R., Srichandan, S. K., Acharya, B., Al-Rasheed, A., Alqahtani, M. S., ... & Soufiene, B. O. (2023). Food price index prediction using time series models: A study of Cereals, Millets and Pulses.
8. Njenga, J. K. (2024). Analysis and Forecasting of Consumer Price Index (CPI) in Kenya and South Africa using Holt Winter Model. Asian Journal of Economics, Business and Accounting, 24(4), 322-331.
9. Qureshi, M., Khan, A., Daniyal, M., Tawiah, K., & Mehmood, Z. (2023). A comparative analysis of traditional SARIMA and machine learning models for CPI data modelling in Pakistan. Applied Computational Intelligence and Soft Computing, 2023.

10. Wanjuki, T. M., Wagala, A., & Muriithi, D. K. (2021). Forecasting Commodity Price Index of Food and Beverages in Kenya Using Seasonal Autoregressive Integrated Moving Average (SARIMA) Models. European Journal of Mathematics and Statistics, 2(6), 50-63.
11. Zhang, X. (2022, December). Forecast and Analysis of China's CPI Based on SARIMA Model. In 2022 International Conference on Bigdata Blockchain and Economy Management (ICBBEM 2022) (pp. 1354-1361). Atlantis Press.