Project Report

On

TIME SERIES ANALYSIS AND PREDICTION OF GOLD RATE

Submitted

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

in

APPLIED STATISTICS AND DATAANALYTICS

by

NIYA PHILIP

(Reg No. SM22AS012)

(2022-2024)

Under the Supervision of

ANAKHA KURIAKOSE

DEPARTMENT OF MATHEMATICS AND STATISTICS (SF)

ST. TERESA’S COLLEGE (AUTONOMOUS) ERNAKULAM,

KOCHI – 682011

APRIL 2024

ST. TERESA'S COLLEGE (AUTONOMOUS), ERNAKULAM

CERTIFICATE

This is to certify that the dissertation entitled, **TIME SERIES ANALYSIS AND PREDICTION OF GOLD RATE** is a Bonafide record of the work done by **NIYA PHILIP** 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

Ms. **ANAKHA KURIAKOSE**

Assistant Professor

Department of Mathematics and Statistics(SF)

St. Teresa's College (Autonomous)

Ernakulam.

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

External Examiner

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 **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 **NIYA PHILIP**

Date: **SM22AS012**

ACKNOWLEDGEMENT

I would like to express my sincere gratitude and appreciation to the following individuals who have contributed to the completion of this research work.

Firstly, I would like to thank my supervisor, Ms. ANAKHA KURAIKOSE, for her valuable guidance, support, and encouragement throughout the research process. Her insightful comments and suggestions have been instrumental in shaping this research work.

I want to express my heartfelt gratitude to Ms. Nisha Oommen, HOD, Department of Mathematics and Statistics (SF) , who guided me throughout this journey.

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

Finally, I want to express my gratitude to my family and friends for their constant support and encouragement throughout this journey. Their love and encouragement have kept me motivated and focused on completing this research work. Thank you all for your invaluable contributions to this research work.

Ernakulam **NIYAPHILIP**

Date: **SM22AS012**

ABSTRACT

This paper fully based on time series model for forecasting the gold rate in India and USA . Forecasting crop yield are crucial since they help to establish expectation for the future yield of the crop. In this paper only one variable is present that is price . Analysis of gold rate is taken in account. The variable has taken the monthly value so it is monthly data. By using seasonal differencing make the data has stationary. The data show seasonal behaviour. Seasonal ARIMA and Holt-winters models are fitted for the time series data and the best time series models are selected using the RMSE and MSE. Forecast of the data sets are made using selected seasonal ARIMA and Holt-winters exponential smoothing model forecast value of agriculture crop yield are predicted till the year 2023 and reported in the study.

Keywords: Time series, seasonal ARIMA, Holts-winter exponential smoothing, gold rate and forecasting

CONTENTS

.

CHAPTER -1

INTRODUCTION

Statistical analysis can offer insights into the variables that affect gold rate. And also aid in forecasting future price movements by observing past patterns and behaviors. Applying various statistical methods to observing historical data and patterns in gold prices is statistical analysis of the gold rate. Examine the dataset of historical gold prices over a defined time period , it ranging from daily or monthly or yearly intervals , is the first step in analysis. The other variables such as market variables , economic indicators , or geographical events that may affect in gold rate .

Gold , a precious metal with a long history worth and attraction . The gold rate on a particular month referred to as the month gold rate . The monthly gold rate is a crucial component of a signal that attracts traders and international financial markets investors , and people all over the world . Gold rate used to protect assets during uncertain economic times and indicate as a symbol of richness and prosperity . The price is usually expressed in terms of currency per unit weight, such as US Dollars per ounce .

The monthly gold rate is examine and studied by financial institutions, analysts, and the general public. It can have a significant impact on consumer behavior, investment choices, and even the mood of the financial markets as a whole . The gold rate is susceptible to changes brought on by a wide range of variables, geopolitical developments, including but not limited to economic indicators, dynamics of supply and demand, inflation, currency movements, monetary policies, and market sentiment . A common means of commerce, store of value, and hedge against inflation, gold has maintained its significance across cultures and civilizations.

1.1 OBJECTIVES

- To model and forecast monthly gold rate of India and United States using SARIMA model.
- To model and forecast monthly gold rate of India and United States using Holt-winters exponential smoothing technique.
- To compare the forecast by SARIMA and Holt-winters exponential smoothing of India and United States.

CHAPTER 2
REVIEW OF LITERATURE

This chapter presents a review of literature related to STATISTICAL ANALYSIS OF GOLD RATE data which are based on published information. A literature review is the writing process of summarizing synthesizing and/or critiquing the literature found as a result of a literature search . It may be used as background or context for a primary research project .

Mills (2004) propose the statical of daily gold price data . This paper investigates the statistical behaviour of daily gold price data from 1971 to 2002. The observations are characterised by short run persistence and scaling with a break point of 15 days, i.e., three working weeks. Daily returns are highly leptokurtic, with multi-period returns only recovering Gaussianity after 235 days (approximately eleven working months). Volatility also scales, with long-run correlations being particularly important .

Baur (2011) studied the explanatory mininig for gold: Contrasting evidence from simple and multiple regressions . Gold has traditionally served as a store of value and an inflation hedge, but it has also gained importance as a hedge against uncertainty and a safe haven. However, this study reveals that many properties associated with gold hold true only in a simple regression framework, while they significantly change in a multiple regression framework. Analyzing monthly data from 1979 to 2011, the study demonstrates that gold mainly acts as a hedge against a weaker US dollar and higher commodity prices, rather than consumer price inflation. Furthermore, the empirical results suggest that gold has recently evolved as a safe haven asset.

Simakavo (2011) studied the analysis of the relationship between oil and gold prices . This article focuses on the relationship between oil and gold prices. The aim of this article is to analyze and determine the character of the co-movement between price levels. This article also presents the basic characteristic and determinants of current price trends. This work uses methods of analysis and synthesis of theoretical knowledge from literature, published articles and other publications. There is also included a quantitative analysis of the variables, such as Granger causality test, Johansen cointegration test and Vector Error Correction model. Thispaper reveals the existence of a long-term relationship between analyzed variables

Gangopadhyay at el . (2016) studied the forecasting the price of gold :An error correction approach . Gold prices in the Indian market may be influenced by a multitude of factors such as the value of gold in investment decisions, as an inflation hedge, and in consumption motives. Develop a model to explain and forecast gold prices in India, using a vector error correction model. Dentify investment decision and inflation hedge as prime movers of the data. Also present out-of-sample forecasts of our model and the related properties.

Dadhich (2017) presented an analysis of volatility of macro economic variables on gold price . Gold price volatility is influenced by Sensex, a market-weighted stock index of reputable companies on the Bombay Stock Exchange. It is a complex task to accurately predict gold's movement due to various interconnected factors and the added challenge of macroeconomic variables' volatility. Sensex and the dollar price also impact gold pricesin India. A researcher conducted a study on the relationship among these three variables using statistical tests like Correlation, Augmented Dickey Fuller (ADF), Unit root tests, Cointegration test, and Granger Causality test in a specific time period.

Ul Sami and Junejo (2017) studied predicting future gold rates using machine learning approach . This study explores the historical and present significance of gold as a trade medium and reserve asset. Currently, central banks hold gold to ensure foreign debt repayment and control inflation, indicating a country's financial strength. Multinational companies and individuals also invest in gold reserves. Asian traditions involve gifting gold and using it in marriages. The performance of leading economies strongly influences gold rates. Using machine learning and 22 market variables, we accurately predict daily gold rates, offering valuable insights for investors and central banks in deciding when to invest in this commodity.

Parimi (2018) investigate the factor influencing the gold prices: An empirical investigation in the India contex . The most precious metals is Gold for long and its value has been used as the standard for many currencies also known as gold standard. The consumption of gold has increased drastically with strong economic growthand promising movements in gold prices in 1990s during liberalization of gold import policy. The gold pricesin India are continuously increasing due to domestic demand based on security, liquidity and diversified portfolio. There would be various factors that influence the prices of the gold. This study aims at understandingand analyzing the various factors which influence the gold prices in the Indian context .

Qian at el . (2019) propose the analysis of factors affecting global gold price . Gold's price has garnered increasing attention as a favored investment product among global investors. This study assesses the factors influencing global gold prices, using a reverse process of response surface methodology (RSM). Six factors, including the dollar index, federal funds rate, CPI, exchange rate, oil price, and S&P500, were analyzed. The findings reveal that all factors, except CPI, have a negative impact on gold prices. Additionally, the significance level of CPI and oil price's impact on the response variable is not significant at 5%, leading to complexity in their interaction effect.

Vanitha and Saravanakumar (2019) investigated the usage of gold and the investment analysis based on gold rate in India .Gold is a favored commodity for customers looking to invest their money, offering better interest compared to banks. In the Indian context, it is commonly purchased for children's marriages. Gold's appeal liesin its easy conversion to cash from banks and gold merchants. Its value is influenced by factors such as fixed deposits, provident funds, international crude oil prices, stock markets, and mutual funds. Conducting a comparative analysis of gold against other investment options empowers customers to make informed decisions for their hard-earned money, aiming for promising future returns.

Vidya and Hari (2020) studied the gold price prediction and modelling using deep learning . The demandfor gold is perpetual, making it one of the most attractive investment options. To make informed decisions, predicting the gold rate trend becomes crucial. This paper focuses on utilizing the LSTM Network for gold rate prediction. Gold prices exhibit a nonlinear nature, making accurate predictions essential for sound financial and investment strategies. The fluctuation in gold rates can be represented by an exponential curve. Convolutional Neural Networks, particularly RNNs, excel at addressing nonlinearity in data and are well-suitedfor time series predictions. The study employs a dataset from the World Gold Council and demonstrates that the proposed architecture is among the top financial forecasting methods.

Aruna et al . (2021) analysis prediction of potential gold prices using machine learning approach . Gold has historically been a widely used form of payment for trading transactions globally. Certain states have accumulated and fortified their gold deposits, earning reputations as wealthy and progressive nations. Gold prices are heavily influenced by the success of the world's leading economies, in addition to market demand and availability of goods. By employing machine learning techniques and analyzing 21 market factors, researchers have accurately predicted potential daily gold prices. These forecast models can prove valuable forinvestors and central banks in determining when to invest in this asset.

CHAPTER – 3

MATERIALS AND METHODOLOGY

3.1 DATA DISCRIPTION

This chapter discuss a comparative study of time series modelling and forecasting of monthly gold rate in India and United states (USD) using SARIMA model and HOLT- WINTERS EXPONENTIAL SMOOTHING method . The data comprises 368 observations starting from January 1991 to AUGUST 2021. Data used in this study is collected from the kaggle .

3.2 METHODOLOGY

The initial and crucial step in this analysis is to study the data in detail . The first step of the analysis remove the unwanted rows and columns using excel .Then the main purpose of the study is to forecast the future gold rate using time series models. Finally, the model was forecasted using SARIMA and HOLT-WINTERS methods , and the MSE and RMSE values of both were calculated .

3.3 TOOLS FOR ANALYSIS & FORECASTING

1. Seasonal ARIMA (Autoregressive Integrated Moving Average)
2. HOLT-WINTERS

3.4 TOOLS FOR COMPARISON

1. Mean Squared Error (MSE)
2. Root Mean Squared Error (RMSE)
3. Akaike Information Criterion (AIC)

3.5 PYTHON SOFTWARE

Python is a popular programming language used to create websites and software , automate tasks and analysethe data . In this study , python is used to forecast the gold rate using SARIMA and HOLT- WINTERS methods . SARIMA is reliable for catching seasonal patterns and trends in time series data .

CHAPTER – 4

TIME SERIES

A time series is a set of observations measured sequentially through time . These measurements may be made continuously through time or be taken at a discrete set of time points . A time series is said to be continuous when observations are made continuously through time and discrete when observations are taken only at specific times , usually equally spaced . Arrangement of data in chronological order . i.e., in accordance with occurrence of time is known as "Time Series". In time which may be either year, month, week, day, and hour or even- minute or seconds. Examples of time series are share prices on successive days , company profits in successive years ect .

Time series are very frequently plotted via line chats . Time series are used in statistics , econometrics, mathematical finance , weather forecasting , astronomy , signal processing intelligent transport and trajectory forecasting ,earthquake , electroencephalography, astronomy , control engineering , astronomy ,, and largely in any domain of applied science and engineering which involves temporal measurement.

4.1 TIME SERIES ANALYSIS

A time series is defined as $\{Xt, t \in T\}$ of observations for a random variable X that is indexed to time, typically represented by t, where T is a time index set. The organization of statistical data in a time series is based on the passage of time. The various underlying structures that can be seen within a time series are referred to as time series components. These elements support comprehension and modeling.

Making future predictions is typically the aim of time series analysis, where time is frequently the independent variable. When observations are made consistently and on schedule, a time series is considered continuous. When observations in a time series are made exclusively at certain, often equally spaced times, the data is considered discrete.

4.1.1 Trend: It shows the series' overall growth or fall over a long period of time. The data's long-term movement or orientation. A trend may be nonexistent, nonlinear, or linear.

4.1.2 Seasonality: Periodic, recurring patterns in data that show up at regular intervals. For example, peaking in the winter, sales of winter apparel usually follow a seasonal trend. Holidays, weather, and the time of year are some examples of factors that typically affect seasonality.

4.1.3 Cyclical variation :. Cycles are longer-term patterns that recur irregularly and are typically impacted by political, economic variables or social. Lacks a set period and is comparable to seasonality In contrast to seasonality, cycles are more difficult to forecast.

4.1.4 Irregular/Residual Component: Forecasting is the primary goal of time series analysis The random or erratic fluctuations that cannot be explained by seasonality, trend or cyclic components are represented by the irregular/residual component. We project the series' future values when we forecast.These irregular variations may be the consequence of measurement errors, unanticipated events, or other uncontrollable variables.

4.2 STATIONARY TIME SERIES

This suggests that the fundamental mechanisms producing the data don't change over time, which facilitates modeling and value prediction. Over the course of its duration, a stationary time series displays autocorrelation structure, constant mean, variance, and among other statistical traits. The application of methods like autoregressive integrated moving average (ARIMA) models is frequently made possible by stationary time series, which presuppose a stable environment.

4.3 NON- STATIONARY TIME SERIES

Modeling and prediction can be difficult with non-stationary data because developing patterns may not be well captured by conventional techniques. Non-stationary time series show how their means, trends, and irregular fluctuations change over time, among other statistical characteristics. Seasonality, intrinsic instability oroutside influences in the data-generating process are some of the possible causes of these variationsTo eliminate seasonality or trends, differencing entails deducting each data point from its correlated counterpart. Once changed, the data can frequently be efficiently modeled with stationary approaches. . Differentiating the data into a stationary form is a typical way to handle non-stationarity.

4.4 AUTOCORRELATION FUNTION (ACF)

Time series is said to be stationary if the joint probability distribution of any two observations say Yt and Yt+kis same for any two time periods t and t+k . The nature of the time series and the useful information can be obtained by plotting the scatter digram of the data that are separated by the same interval k . The interval k is called the lag .

The Autocorrelation Function (ACF) is a graphical tool used to assess the correlation between a time series and its lagged versions. In other words, it quantifies the similarity between the values of a time series at different time points. By plotting the ACF, one can identify whether there is a significant correlation between the current observation and observations at various lags. A sharp drop in autocorrelation after a certain lag suggests the presence of a seasonality component with that lag. An ACF that gradually declines indicates a potential AR component in time series. The ACF is pivotal for identifying the order of moving average (MA) term in a model like ARIMA.

4.5 PARTIAL AUTOCORRELATION FUNCTION (PACF)

The partial autocorrelation function between Yt and Yt-k is the autocorrelation function between Yt and Yt-k after adjusting for Yt-1, Yt-2,……, Yt-k+1. Hence for an AR(p) model the partial autocorrelation function between Yt and Yt-k for k>p should be equal to zero .

The Partial Autocorrelation Function (PACF) extends the concept of autocorrelation by capturing the direct correlation between two observations while accounting for the indirect correlations introduced by the intermediate lags. It essentially measures the correlation between the current observation and observations at specific lags, after removing the effects of the intervening lags. By plotting the PACF, analysts can identify the lag values where the direct correlation is significant. A sharp cut off in the PACF after a certain lag indicates a potential autoregressive (AR) component in the time series. Just like the ACF, the PACF assists in determining the appropriate parameters for models like ARIMA.

4.6 ARIMA (AUTOREGRESSIVE INTEGRATED MOVING AVERAGE)

ARIMA, which stands for Auto-Regressive Integrated Moving Average, is a powerful time series forecasting model used to capture the patterns, trends, and dependencies present in time series data. ARIMA combines three main components: autoregression (AR), differencing (I for integrated), and moving average (MA). It is designed to work with stationary and weakly stationary data, which means that the mean, variance, and autocorrelation structure remain relatively constant over time.

4.6.1 AUTOREGRESSIVE (AR) PROCESSES

In time series analysis, also referred to as the Autoregressive (AR) process , it is regarded as a particular kind of linear regression model. A stochastic model is used to describe the behavior of a variable over time. The present value of the variable is modeled as a linear combination of the previous values plus a random error term. An AR(p) process of order p can be expressed mathematically as

$$Zt= \phi1\ Zt\text{-}1 + \phi2\ Zt\text{-}2 +\ldots..+ p1\ Zt\text{-}p + \varepsilon t$$

Where $\{\varepsilon t\}$ – purely random process having mean of 0 and variance $\sigma 2$

4.6.2 DIFFERENCING (I) COMPONENT

The differencing component is used to make the data stationary by subtracting each observation from its lagged observation. It accounts for trends and seasonality that might be present in the data. The "d" in ARIMA denotes the orderof differencing required to achieve stationarity.

4.6.3 MOVING AVERAGE (MA) PROCESSES

AR processes model the present value of a variable as a linear combination of its previous values whereas MA processes model the present value as a linear combination of previous error terms. In order to explain the behaviour of a variable with time, a different type of stochastic model is used which is the Moving Average (MA) process. Mathematically, it can be represented as:

$X_t = \mu + \varepsilon_t + \theta_1 \times \varepsilon_{t-1} + \theta_2 \times \varepsilon_{t-2} + \ldots + \theta_q \times \varepsilon_{t-q}$

Where:

X_t - value of the time series at time t.

μ - constant term.

ε_t - white noise error term at time t representing the random fluctuations or shocks in the data.

$\theta_1, \theta_2 \ldots$..are the moving average parameters

4.7 SARIMA (SEASONAL AUTOREGRESSIVE INTEGRATED MOVING AVERAGE)

SARIMA is an enhanced variant of ARIMA which comprises of seasonal components as well to account for the periodic patterns in the data. ARIMA (p, d, q) × (P, D, Q) represents a seasonal ARIMA model. If the data is monthly or shows seasonality, SARIMA is typically employed. Seasonal moving average (SMA) and seasonal autoregressive (SAR) components make up this system.

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

4.7.1 Seasonal Auto-Regressive (SAR) Component: This component models the relationship between the current observation and past observations at the same lag in previous seasons. It captures the effect of seasonality on the data. The order of the seasonal AR component is denoted as "P."

4.7.2 Seasonal Differencing (S) Component: Similar to the non-seasonal differencing, the seasonal differencing component is used to remove the seasonality from the data. The order of seasonal differencing is denoted as "D."

4.7.3 Seasonal Moving Average (SMA) Component`: This component models the relationship between the current observation and past errors or residuals at the same lag in previous seasons. It accounts for past shocks that persist across seasons. The order of the seasonal MA component is denoted as "Q." SARIMA is denoted as SARIMA (p, d, q) (P, D, Q, s), where "s" represents the season's length

4.8 AUGMENTED DICKEY-FULLER (ADF) TEST

This test is frequently employed to examine the stationarity of a sequence. This test falls within the "Unit Root Test" category. It is among the most effective ways to determine whether a time series of data is stationar. It examines the subsequent scenarios:

H0: Non-stationarity exists in the time series.
H1: There is stationarity in the time series.

A p-value of less than a certain value (e.g., 0.05) indicates that the time series is stationary and the null hypothesis is rejected.

4.9 KWIATKOWSKI-PHILLIPS-SCHMIDT-SHIN (KPSS) TEST

It is another method to check stationarity of the time series data. It tests the following cases:H0:

The time series is stationary.

H1: The time series is non-stationary.

If the p-value is greater than a particular value (eg.,0.05), then the null hypothesis is rejected an concludes that the time series is non-stationary.

4.10 AKAIKE INFORMATION CRITERION (AIC)

AIC compares the goodness of fit of different models while balancing for the complexity and model fit of the model. AIC means Akaike Information Criterion which is used for choosing of model among a set of other models. AIC is commonly used in different fields ,such as machine learning ,statistics etc .

4.11 FORECASTING

Time series forecasting can be done using a variety of methods and models, depending on the features of the data. Future value predicting is a major challenge in many domains, sales forecasting, stock control and including economics. Predicting future values from past data is the process of forecasting time series. We may select the best model and approach for predicting based on the particular goals of the analysis. The historical data can be used to draw important conclusions and gain new insights, allowing for the prediction of the future. The following are a few methods for forecasting time series data:
Simple Moving Average (SMA): SMA is unable to identify any trends or seasonality in the data. This is an easy-to-implement approach that can help reduce random fluctuations in the data. By averaging the historical observations over a predetermined period of time, it predicts the values of the future.

Exponential Smoothing methods: It is capable of capturing trends and seasonality in the data, depending on the smoothing parameter selected. When prior measurements are given exponentially decreasing weights, exponential smoothing is used to forecast future results.

Types of Exponential smoothing methods are:

i) Exponential Smoothing (SES)
ii) Double Exponential Smoothing (Holt's method)
iii) Triple Exponential Smoothing (Holt-Winters method)

HOLT-WINTERS FORECASTING

This method is suitable for time series data having trend and seasonal components. Three types of smoothing techniques are involved: level smoothing, trend smoothing, and seasonal smoothing.

i) Level smoothening: It signifies the smoothed level of the time series at t time. The level smoothing component is often denoted as L_t .The level smoothing equation is given by:

$$L_t = \alpha \times Y_t + (1-\alpha) \times (L_{t-1} + T_{t-1})$$

Where Y_t - the observed value of the time series value

T_{t-1} - the trend component from previous time period
L_{t-1} - smoothed level from the previous time period
α - level smoothing parameter and lies between 0 and 1

ii) Trend Smoothing: The smoothed trend of the time series at time t is represented by the trend smoothing. Long-term changes in the data are identified by trend smoothing. The trend smoothing component is denoted as T_t. The trend smoothing equation is given by:

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

Where L_t and L_{t-1} are the smoothed levels for the present and previous time periods respectively
T_{t-1} is the smoothed trend from the previous time period
β is the trend smoothing parameter ranging from 0 to 1

iii) Seasonal Smoothing: Seasonal variation is represented by this smoothing. Repetitively occurring cycles are identified. S_t represents the seasonal smoothing component. The equation for seasonal smoothing is given by,

$$S_t = \gamma \times (Y_t - L_t) + (1-\gamma) \times S_{t-m}$$

Where Y_t - the observed value of the time series value
L_t - smoothed levels for the present period
S_{t-m} - is the smoothed seasonal component from the same season in the previous year.

CHAPTER 5

DATA ANALYSIS AND FORECASTING

The initial step of time series analysis is to draw a time series plot . The first step of the analysis remove the unwanted rows and columns using excel . The row is from January 1979 to august 1991. Then it is reduce to analysis the from January 1991 to august 2021.

5.1 TIME SERIES MODELLING

This chapter presents the time series modelling of gold rate India and USA from 1991 to 2021. The process of model fitting was done by using Python programming language.

5.1.1 GOLD RATE PREDICTION IN INDIA USING SARIMA MODEL

Time series plot of monthly gold rate of India from 1991 to 2021 in given in fig 5.1.

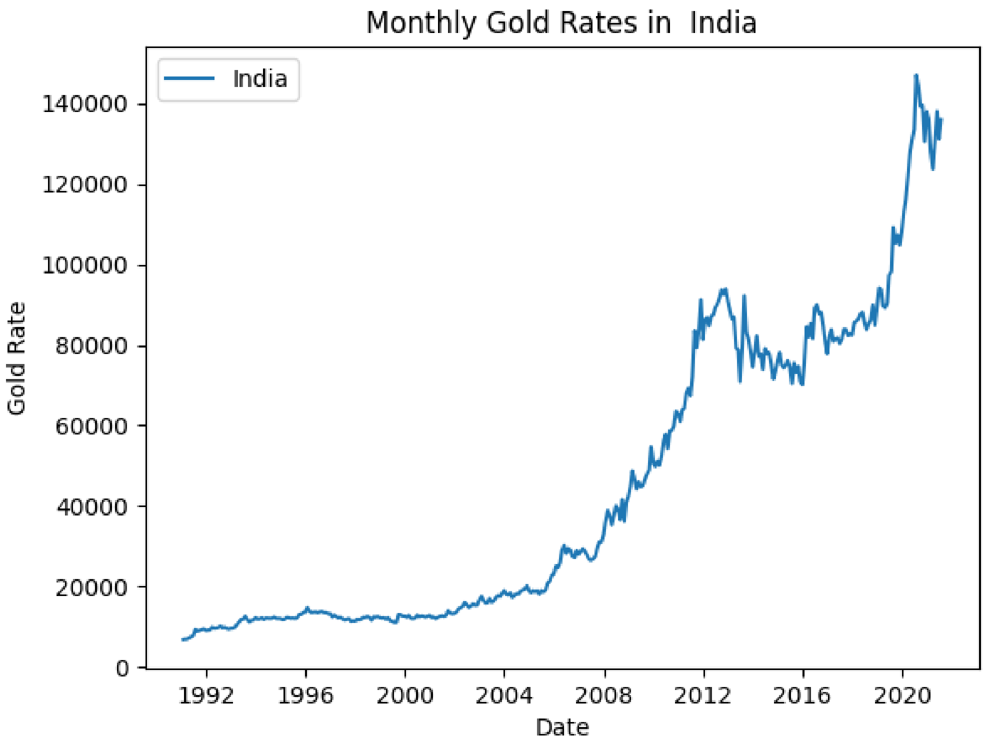

Fig 5.1

5.1.1.2 DECOMPOSITION OF TIME

Seasonal decomposition is performed for evaluation of trend, seasonality and random components. Fig 5.2 shows the seasonal decomposition

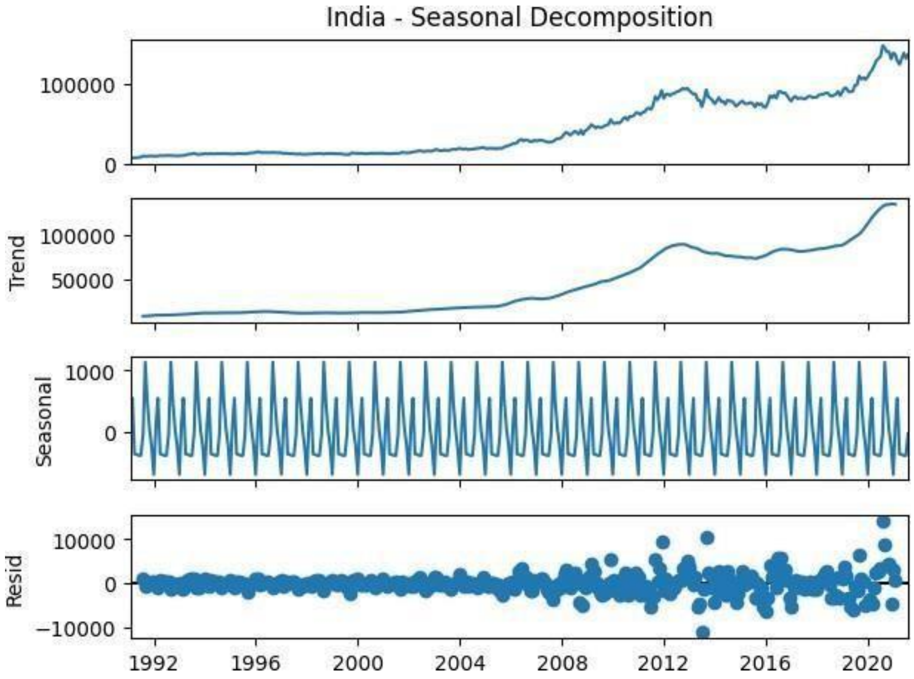

Fig 5.2

5.1.1.3 SEASONAL DIFFERENCING

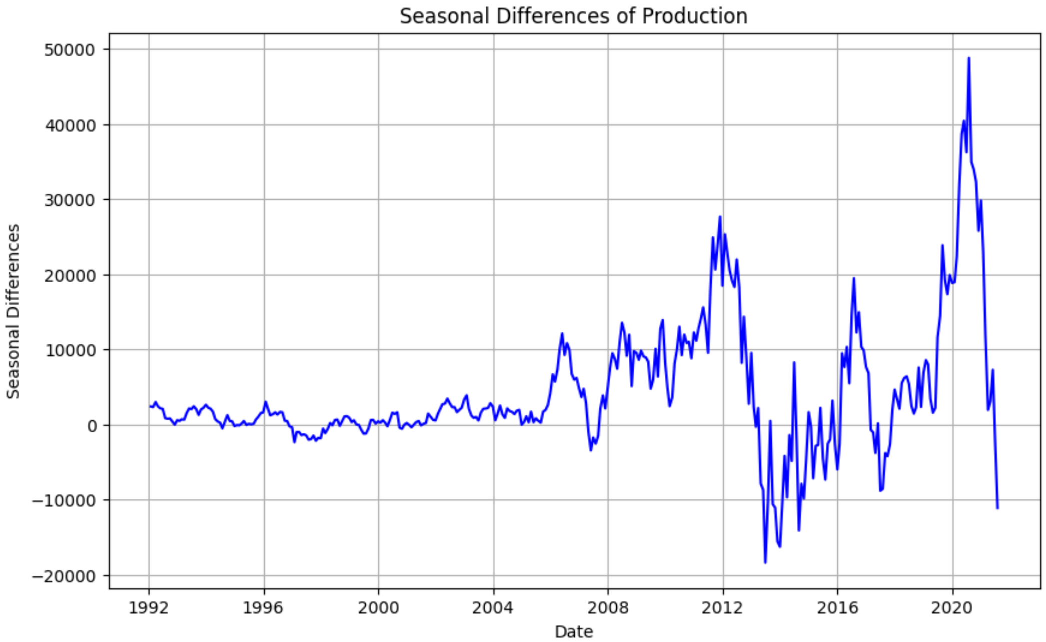

Fig 5.3

5.1.1.4 STATIONAY USING AUGMENT DICKEY-FULLER TEST

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

. H0: The data is non stationary.

H1: The data is stationary.

ADF TEST STATISTIC	-3.6540222691296487
P -VALUE	0.004808370119907616

Table 5.1

We have obtained from the test that the p value is which is less than 0.05.

Thus, we reject the null hypothesis "Time series is not stationary", which means the time series is stationary.

5.1.1.5 STATIONARITY USING KPSS TEST

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

H0: The data is non stationary.

H1: The data is stationary.

KPSS Statistic	0.02878843496432
P-value	0.1

Table 5.2

The KPSS test gives the p-value 0.1, so fail to accept H0 and hence can conclude that data is stationary.

5.1.1.6 SARIMA MODEL FOR GOLD RATE IN INDIA

The best model for forecasting. The model the lowest AIC value is the best model. Choose the best model from all possible model according to Akaike Information Criterion (AIC). Thus,the possible time series models along with their corresponding AIC statistic,

p- The number of lags in the observation.

q- order of the moving average.

d-Order of differencing.

Values obtained through AIC can be used for forecasting. This process helps to save time and improve the accuracy of the forecasting models

SL .no	MODEL ARIMA(p,d,q)*ARIMA(P,D,Q)	AIC VALUES
1	(0, 0, 0)x(0, 0, 0, 12)	9066.244716675634
2	(0, 0, 0)x(0, 0, 1, 12)	8553.589719318064
3	(0, 0, 0)x(0, 1, 0, 12)	7508.138242643556
4	(0, 0, 0)x(0, 1, 1, 12)	7235.941009694927
5	(0, 0, 0)x(1, 0, 0, 12)	7428.012818893425
6	(0, 0, 0)x(1, 0, 1, 12)	7407.994943462617
7	(0, 0, 0)x(1, 1, 0, 12)	7239.132478790047
8	(0, 0, 0)x(1, 1, 1, 12)	7180.47005276418
9	(0, 0, 1)x(0, 0, 0, 12)	8778.790484788875
10	(0, 0, 1)x(0, 0, 1, 12)	8431.003764332745
11	(0, 0, 1)x(0, 1, 0, 12)	7168.266416052278
12	(0, 0, 1)x(0, 1, 1, 12)	6933.179135294041
13	(0, 0, 1)x(1, 0, 0, 12)	8448.627524198062
14	(0, 0, 1)x(1, 0, 1, 12)	8403.046332070304
15	(0, 0, 1)x(1, 1, 0, 12)	6970.112131091196
16	(0, 0, 1)x(1, 1, 1, 12)	7003.651063042705
17	(0, 1, 0)x(0, 0, 0, 12)	6829.418811052446
18	(0, 1, 0)x(0, 0, 1, 12)	6615.497785263082
19	(0, 1, 0)x(0, 1, 0, 12)	6791.98688279629
20	(0, 1, 0)x(0, 1, 1, 12)	6414.737944511049
21	(0, 1, 0)x(1, 0, 0, 12)	6631.628468223853
22	(0, 1, 0)x(1, 0, 1, 12)	6606.499351373276
23	(0, 1, 0)x(1, 1, 0, 12)	6440.497866770607
24	(0, 1, 0)x(1, 1, 1, 12)	6402.809495877162
25	(0, 1, 1)x(0, 0, 0, 12)	6808.863847421237
26	(0, 1, 1)x(0, 0, 1, 12)	6594.278387338141
27	(0, 1, 1)x(0, 1, 0, 12)	6766.731171529791
28	(0, 1, 1)x(0, 1, 1, 12)	6412.189446028961
29	(0, 1, 1)x(1, 0, 0, 12)	6627.817306164785
30	(0, 1, 1)x(1, 0, 1, 12)	6584.1485293338155
31	(0, 1, 1)x(1, 1, 0, 12)	6440.88213877187
32	(0, 1, 1)x(1, 1, 1, 12)	6380.764629095162
33	(1, 0, 0)x(0, 0, 0, 12)	6842.563696574256
34	(1, 0, 0)x(0, 0, 1, 12)	6630.320355239939
35	(1, 0, 0)x(0, 1, 0, 12)	6800.1394544936575
36	(1, 0, 0)x(0, 1, 1, 12)	6452.227744221042
37	(1, 0, 0)x(1, 0, 0, 12)	6629.420084669355
38	(1, 0, 0)x(1, 0, 1, 12)	6623.880370357228
39	(1, 0, 0)x(1, 1, 0, 12)	6439.26126570626
40	(1, 0, 0)x(1, 1, 1, 12)	6451.156268289846
41	(1, 0, 1)x(0, 0, 0, 12)	6818.818926325064
42	(1, 0, 1)x(0, 0, 1, 12)	6606.494594704512
43	(1, 0, 1)x(0, 1, 0, 12)	6778.509931493008
44	(1, 0, 1)x(0, 1, 1, 12)	6431.078605162569
45	(1, 0, 1)x(1, 0, 0, 12)	6623.2398923737965
46	(1, 0, 1)x(1, 0, 1, 12)	6601.441309039716
47	(1, 0, 1)x(1, 1, 0, 12)	6462.36541896398
48	(1, 0, 1)x(1, 1, 1, 12)	6431.290030324246

49	(1, 1, 0)x(0, 0, 0, 12)	6827.0234083260475
50	(1, 1, 0)x(0, 0, 1, 12)	6612.708657110819
51	(1, 1, 0)x(0, 1, 0, 12)	6786.027607957005
52	(1, 1, 0)x(0, 1, 1, 12)	6429.963394171897
53	(1, 1, 0)x(1, 0, 0, 12)	6610.821854905681
54	(1, 1, 0)x(1, 0, 1, 12)	6602.359174778068
55	(1, 1, 0)x(1, 1, 0, 12)	6423.133983410769
56	(1, 1, 0)x(1, 1, 1, 12)	6397.440933520391
57	(1, 1, 1)x(0, 0, 0, 12)	6810.70046070786
58	(1, 1, 1)x(0, 0, 1, 12)	6596.05966732774
59	(1, 1, 1)x(0, 1, 0, 12)	6768.491593601231
60	**(1, 1, 1)x(0, 1, 1, 12)**	**6380.035587318118**
61	(1, 1, 1)x(1, 0, 0, 12)	6611.871425717545
62	(1, 1, 1)x(1, 0, 1, 12)	6585.949216496866
63	(1, 1, 1)x(1, 1, 0, 12)	6423.726552709396
64	(1, 1, 1)x(1, 1, 1, 12)	6380.973346930188

Table 5.3

According to Akaike Information Criterion (AIC), ARIMA (1,1,1) × (0,1,1,12) model is considered as the most appropriate one.

5.1.1.7 DIAGNOSTIC CHECKING

Diagnostic checking is the method of evaluating the capability of a fitted model by examining the residuals.The below figures show that the data is normally distributed. Standard residuals for ‘E’ show how much outliers are present in the dataset and the Q-Q plot shows the normality of the data. The main aim of diagnostic checking is to ensure that the assumptions of the model are met and the residuals exhibit certain expedient properties

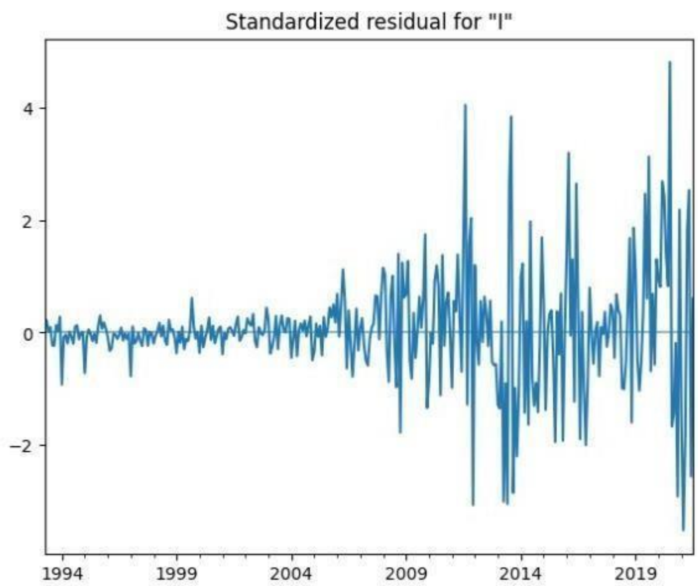

Fig 5.4

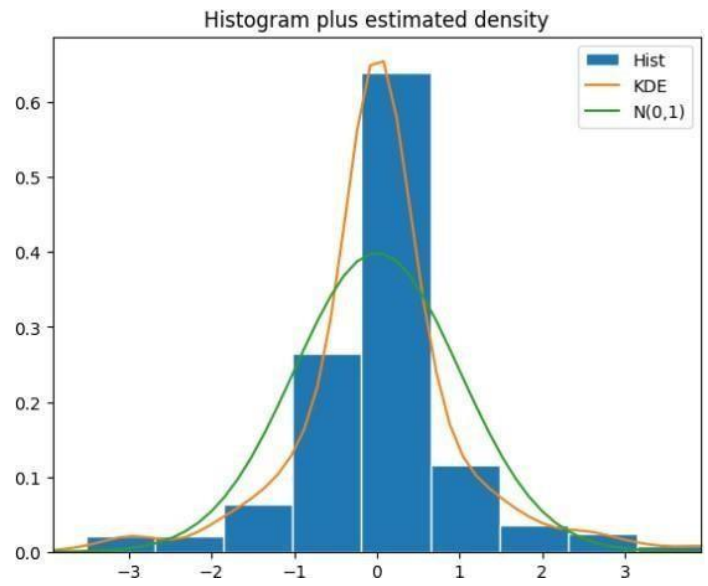
Histogram plus estimated density
Hist
KDE
N(0,1)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
−3
−2
−1
0
1
2
3

Fig 5.5

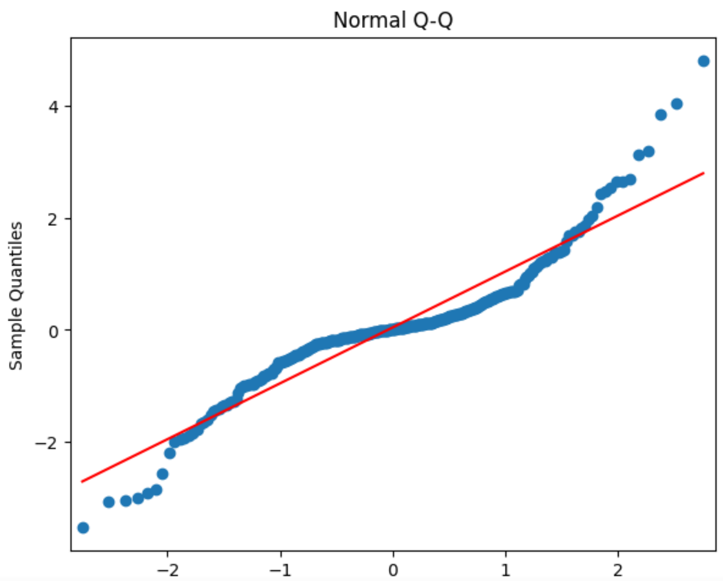
Normal Q-Q
Sample Quantiles
4
2
0
−2
−2
−1
0
1
2

Fig 5.6

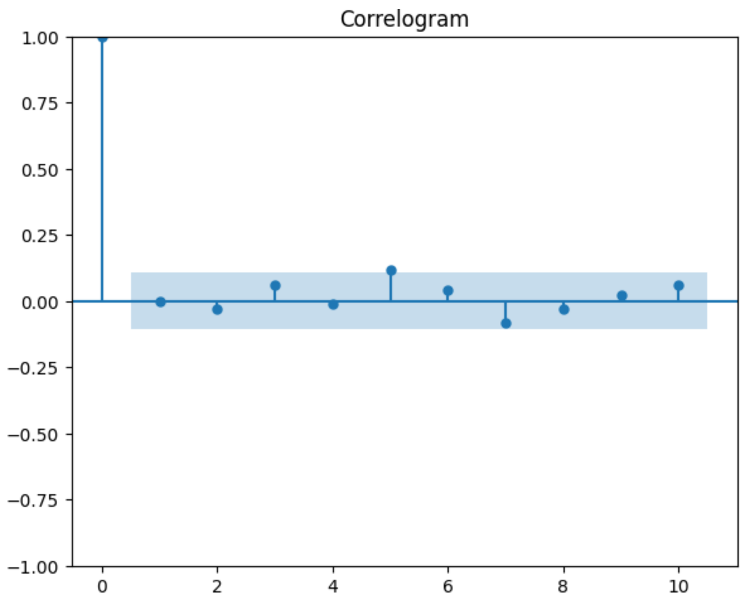

Fig 5.7

5.1.1.8 FORECASTING THE SAMLE

The In-sample forecast is obtained as follows:

DATE	ACTUAL VALUE	PREDICTED VALUE
1991-01-31	6810.6	0.0
1991-02-28	6917.0	7104.61979462692
1991-03-29	6942.2	6973.020524357834
1991-04-30	7226.6	6952.359727474049
1991-05-31	7527.3	7202.092010582692
1991-06-28	7785.6	7493.503863460676
1991-07-31	9348.0	7753.718839550263
1991-08-30	8953.7	9193.751527595734
1991-09-30	9120.5	8956.326403765393
1991-10-31	9241.4	9105.381366960199
1991-11-29	9466.7	9226.634264461156
1991-12-31	9080.6	9442.182767297061
…………………………	…………………………..	……………………………
2021-01-	135982.9	140289.17691607992
2021-02-26	128073.3	138019.206438799
2021-03-31	123639.0	128870.93949460509
2021-04-30	130934.3	125760.67345522299
2021-05-31	137979.1	130812.20907712042
2021-06-30	131054.9	138290.2254591468
2021-07-30	135863.2	134157.34677153244

Table 5.4

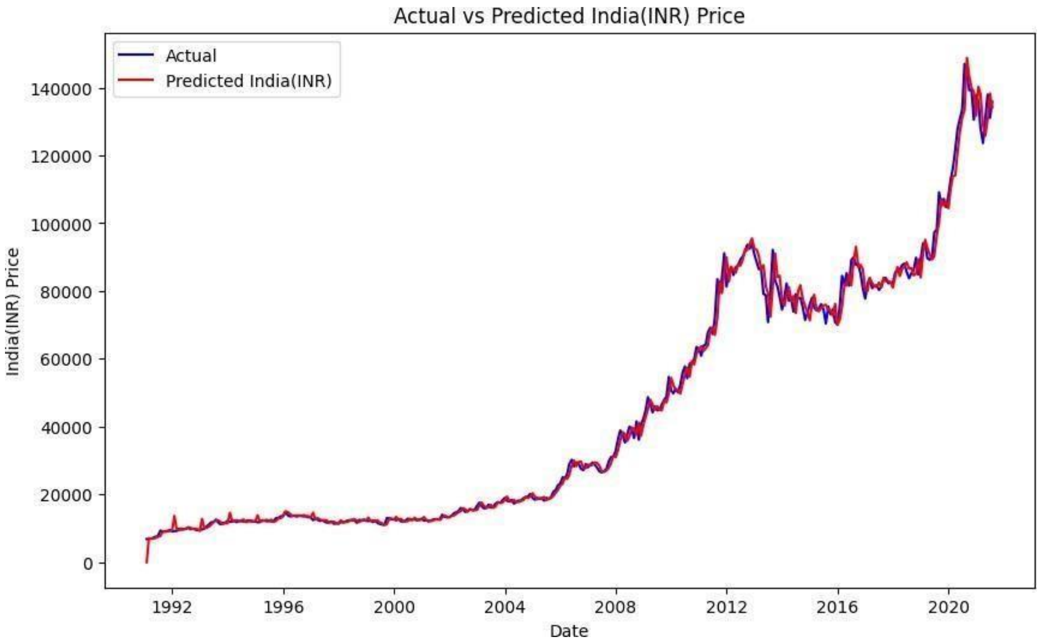

Fig 5.8

5.1.1.9 FORECATING THE FUTURE VALUES

The LCL and UCL values are given in the following table,

LCL	UCL
-3394.757202228515	3394.757
901.3962344205111	13307.84
750.5458792137242	13195.495
728.9169029987379	13175.80
978.6007353573641	13425.583
1270.0101642108657	13716.997
1530.2250190263449	13977.21
2970.257701004469	15417.245
2732.832576870578	15179.820
2881.8875400501965	15328.8
3003.1404375503953	15450.128
3218.6889403862615	15665.67
……………………	………..................
134743.508714855783	145834.845117302
132473.5323078506	143564.8769799092
123325.27677628657	134416.6022129236
120215.01115963816	131306.33575080783
125266.5468021002	136357.87135214065
132744.56318523493	143835.88773307865
128611.68449767404	139703.00904539082

Table 5.5

DATE	FUTURE PREDICTION
2021-08-31	138464.686045
2021-09-30	136741.077832
2021-10-29	136916.638915
2021-11-30	134875.829989
2021-12-31	136036.028926
2022-01-31	138353.551810
2022-02-28	138490.734012
2022-03-31	137304.991728
2022-04-29	139432.672816
2022-05-31	140875.178481
2022-06-30	141006.821567
2022-07-29	143640.511037
2022-08-31	146346.040952
2022-09-30	144635.907744
2022-10-31	1448113.214012
2022-11-30	142772.631109
2022-12-30	143932.859319
2023-01-31	146250.385994
2023-02-28	146387.568688
2023-03-31	145201.826467
2023-04-28	147329.507563
2023-05-31	148772.013229
2023-06-30	148903.656316
2023-07-31	151537.345786
2023-08-31	154242.875700
2023-09-29	152532.742492
2023-10-31	152710.048760
2023-11-30	150669.465857
2023-12-29	151829.694067
2024-01-31	154147.220742
2024-02-29	154284.403436
2024-03-29	153098.661215
2024-04-30	155226.342311
2024-05-30	156668.847978
2024-06-28	156800.491064
2024-07-31	159434.180534
2024-08-30	162139.710448
2024-09-30	160429.577241
2024-10-31	160606.883508
2024-11-29	158566.300605
2024-12-31	159726.528816
2025-01-31	162044.055491
2025-02-28	162181.238184
2025-03-31	160995.495963
2025-04-30	163123.177059
2025-05-30	164565.682726
2025-06-30	164697.325812
2025-07-31	167331.015282

Table 5.6

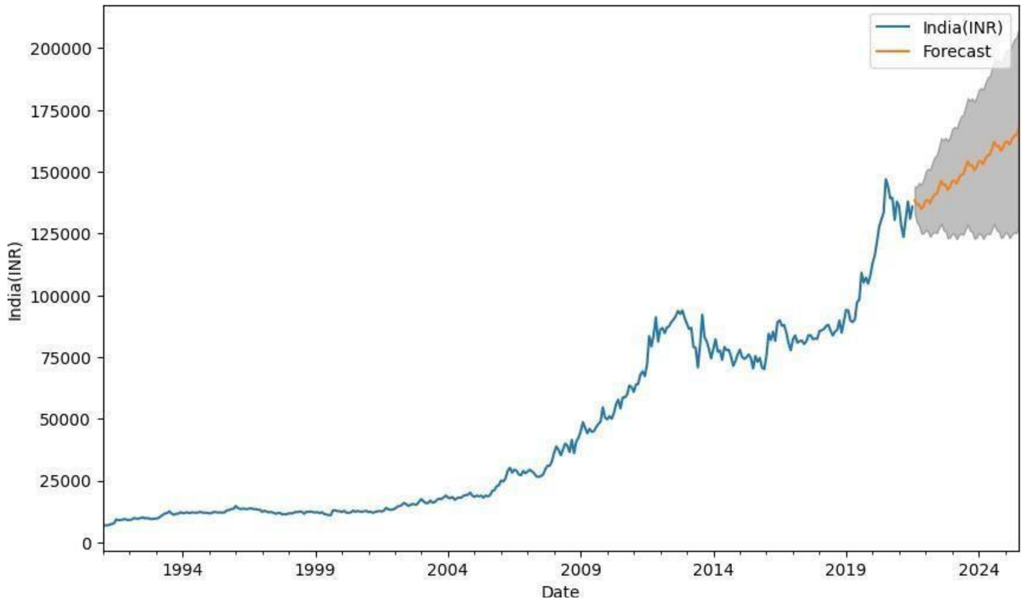

Fig 5.9

5.2 GOLD RATE PREDICTION USING HOLT-WINTERS

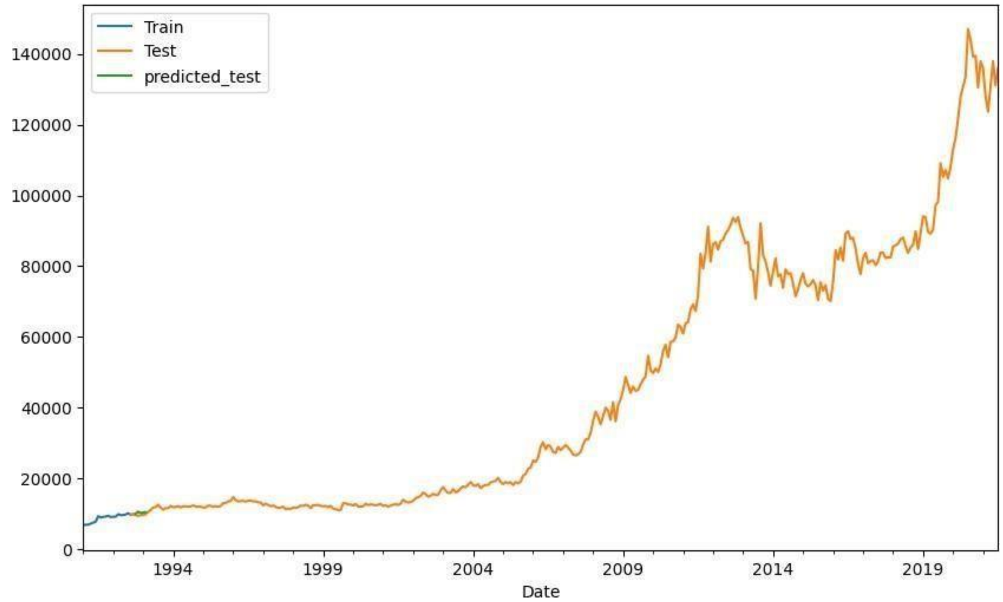

Fig 5.10

5.2.1 DECOMPOSITION OF TIME

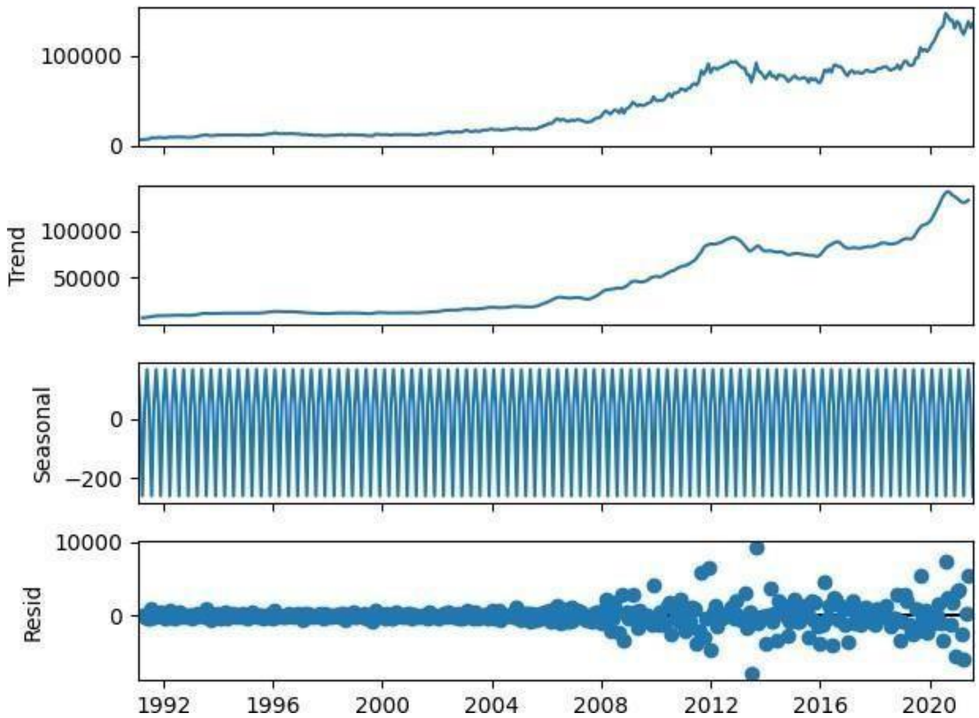

Fig 5.11

5.2.2 FORECASTING THE SAMLE

The In-sample forecast is obtained as follows:

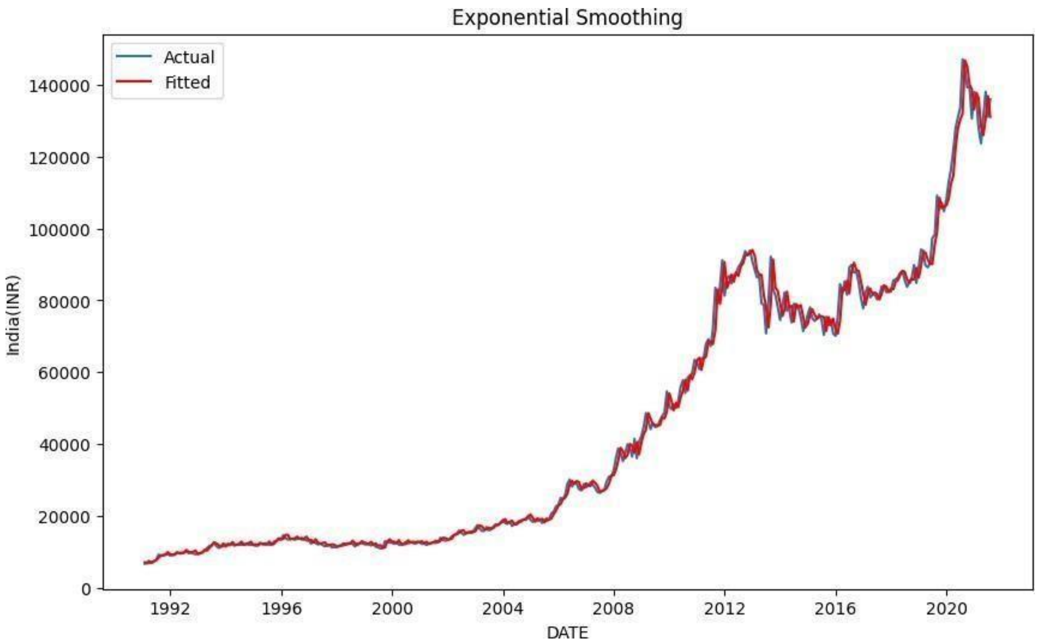

Fig 5.12

5.2.3 FORECATING THE FUTURE VALUES

DATE	PREDICTED VALUE
2021-08-31	137115.665044
2021-09-30	137477.306538
2021-10-29	137003.864860
2021-11-30	136221.581318
2021-12-31	138199.129082
2022-01-31	138561.486453
2022-02-28	138082.185123
2022-03-31	137291.638907
2022-04-29	139282.593120
2022-05-31	139645.666369
2022-06-30	139160.505386
2022-07-29	138361.696495
2022-08-31	140366.057157
2022-09-30	140729.846284
2022-10-29	140238.825649
2022-11-30	139431.754083
2022-12-31	141449.521195
2023-01-31	141814.026200
2023-02-28	141317.145911
2023-03-31	140501.811671
2023-04-29	142532.985233
2023-05-31	142898.206115
2023-06-30	142395.466174
2023-07-31	141571.869259
2023-08-31	143616.449270
2023-09-29	143982.386031
2023-10-31	143473.786437
2023-11-30	142641.926847
2023-12-29	144699.913308

2024-01-31	145066.565946
2024-02-29	144552.106700
2024-03-29	143711.984435
2024-04-30	145783.377346
2024-05-31	146150.745862
2024-06-28	145630.426963
2024-07-31	144782.042023
2024-08-30	146866.841383
2024-09-30	147234.925777
2024-10-31	146708.747226
2024-11-29	145852.099611
2024-12-31	145852.099611
2025-01-31	148319.105693
2025-02-28	147787.067489
2025-03-31	146922.157199
2025-04-30	149033.769459
2025-05-30	149403.285608
2025-06-30	148865.387752
2025-07-31	147992.214788

Table 5.7

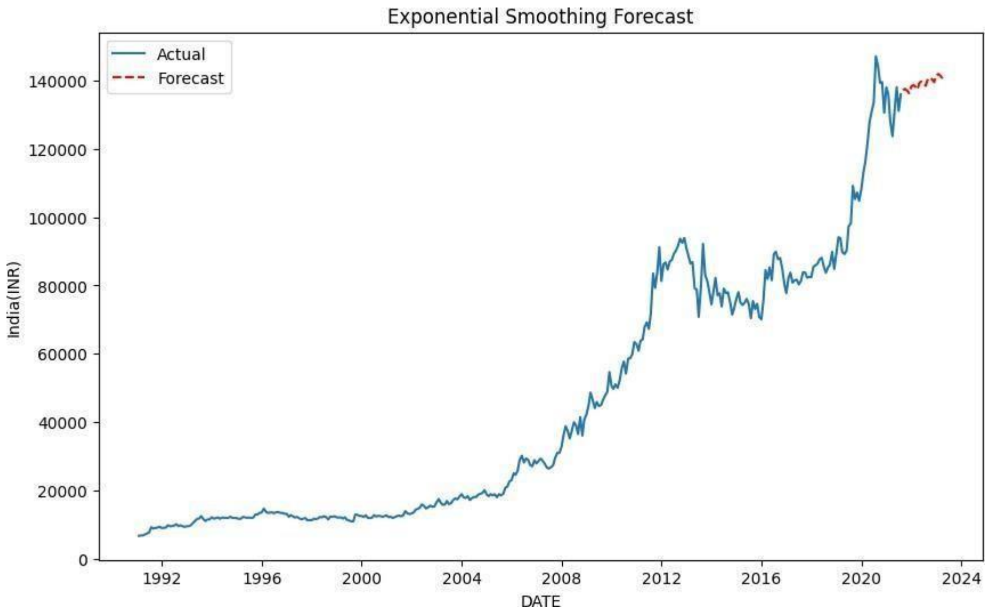

Fig 5.13

5.3 GOLD RATE PREDICTION IN UNITED STATES USING SARIMA MODEL

Time series plot of monthly gold rate of USA from 1991 to 2021 in given in fig 5.13.

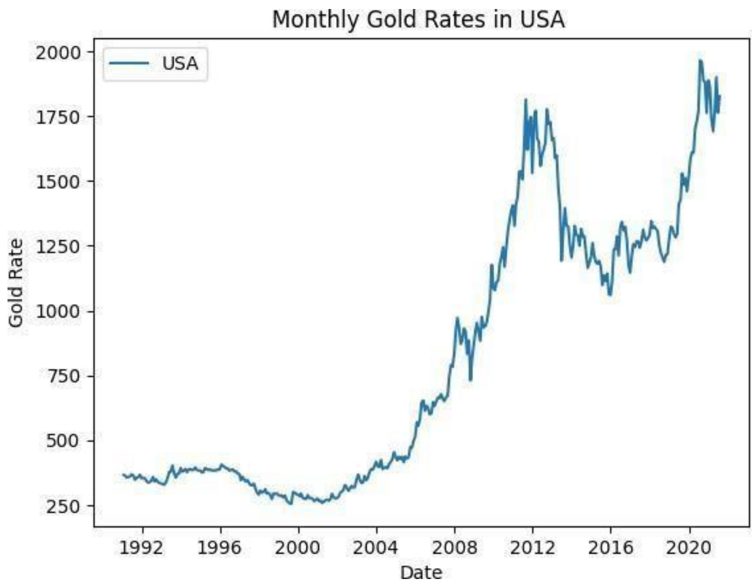

Fig 5.14

5.3.1 DECOMPOSITION OF TIME

Seasonal decomposition is performed for evaluation of trend, seasonality and random components. Fig 5.2 shows the seasonal decomposition

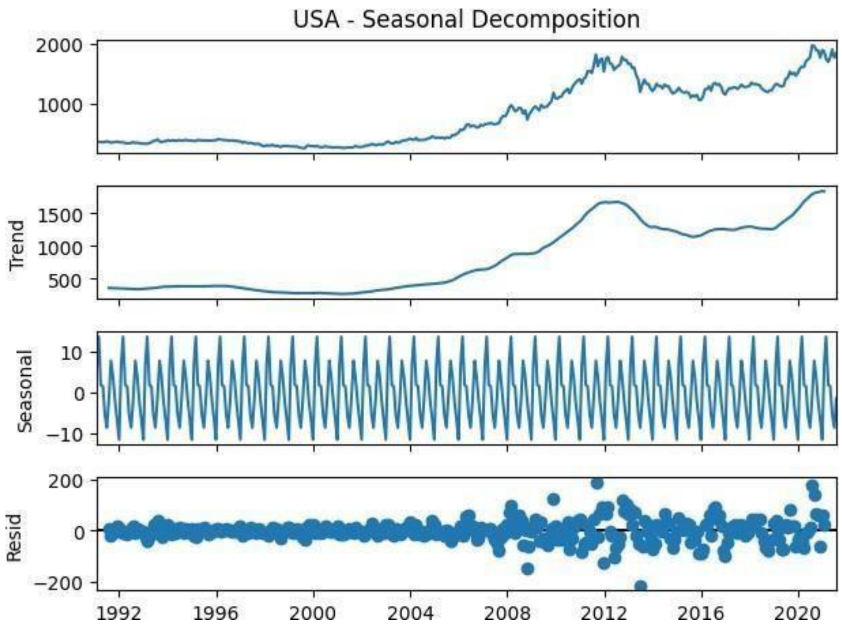

Fig 5.15

5.3.2 SEASONAL DIFFERENCING

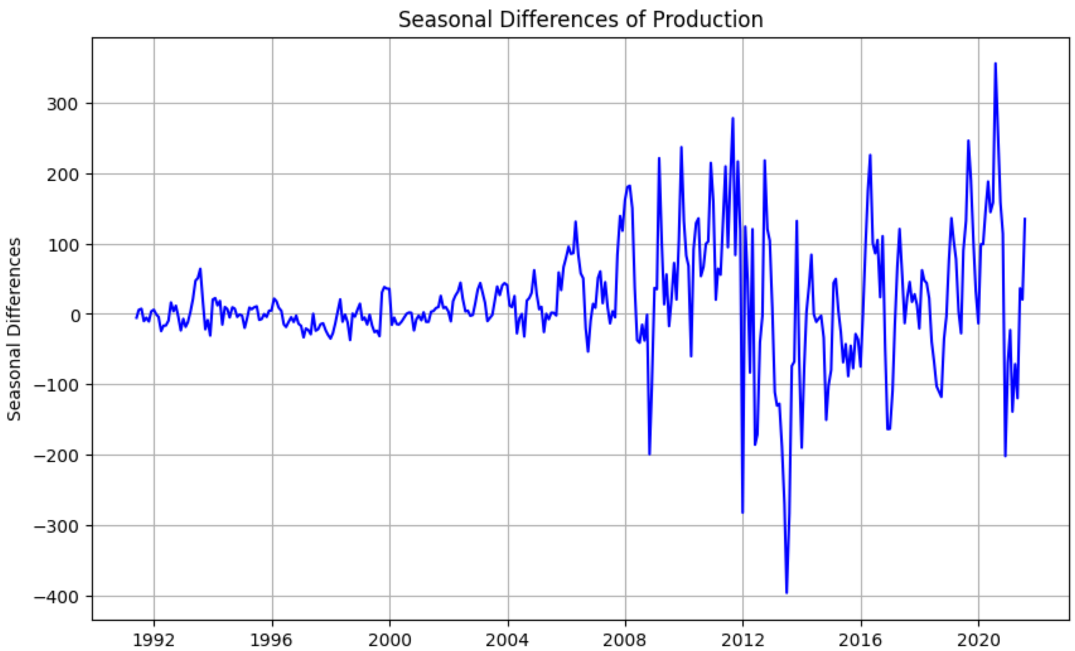

Table 5.16

5.3.3 STATIONARY USING AUGMENTED DICKEY FULLER (ADF)

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

H0: The data is non stationary.

H1: The data is stationary.

Dickey-Fuller	-2.668524728731647
P-value	0.00766450176066903

Table 5.8

We have obtained from the test that the p value is which is less than 0.05.

Thus, we reject the null hypothesis "Time series is not stationary", which means the time series is stationary.

5.3.4 STATIONARITY USING KPSS TEST

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

H0: The data is non stationary.

H1: The data is stationary.

KPSS Statistic	0.02675729714961
P-value	0.1

Table 5.9

The KPSS test gives the p-value 0.1, so fail to accept H0 and hence can conclude that data is stationary.

5.3.5 SARIMA MODEL FOR GOLD RATE IN USA

The best model for forecasting. The model the lowest AIC value is the best model. Choose the best model from all possible model according to Akaike Information Criterion (AIC). Thus,the possible time series models along with their corresponding AIC statistic,

p- The number of lags in the observation.

q- order of the moving average.

d-Order of differencing.

SL .no	MODEL ARIMA(p,d,q)*ARIMA(P,D,Q)	AIC VALUES
1	(0, 0, 0)x(0, 0, 0, 12)	6065.536301267894
2	(0, 0, 0)x(0, 0, 1, 12)	5478.095401542911
3	(0, 0, 0)x(0, 1, 0, 12)	4589.8852637762575
4	(0, 0, 0)x(0, 1, 1, 12)	4410.326732954182
5	(0, 0, 0)x(1, 0, 0, 12)	4572.258153187725
6	(0, 0, 0)x(1, 0, 1, 12)	4539.2719242939565
7	(0, 0, 0)x(1, 1, 0, 12)	4417.641438190378
8	(0, 0, 0)x(1, 1, 1, 12)	4407.666974815214
9	(0, 0, 1)x(0, 0, 0, 12)	5572.189873704369
10	(0, 0, 1)x(0, 0, 1, 12)	5033.550444443763
11	(0, 0, 1)x(0, 1, 0, 12)	4286.169025138087
12	(0, 0, 1)x(0, 1, 1, 12)	4148.610767874876
13	(0, 0, 1)x(1, 0, 0, 12)	4292.455388017976
14	(0, 0, 1)x(1, 0, 1, 12)	4271.127785605179
15	(0, 0, 1)x(1, 1, 0, 12)	4170.819850970678
16	(0, 0, 1)x(1, 1, 1, 12)	4149.951831110515
17	(0, 1, 0)x(0, 0, 0, 12)	3871.1627718938394
18	(0, 1, 0)x(0, 0, 1, 12)	3757.4109001962934
19	(0, 1, 0)x(0, 1, 0, 12)	3974.232471553639
20	(0, 1, 0)x(0, 1, 1, 12)	3659.5304365052816
21	(0, 1, 0)x(1, 0, 0, 12)	3767.02452445441
22	(0, 1, 0)x(1, 0, 1, 12)	3759.337537372388
23	(0, 1, 0)x(1, 1, 0, 12)	3738.6887644215267
24	(0, 1, 0)x(1, 1, 1, 12)	3661.2550468158565
25	(0, 1, 1)x(0, 0, 0, 12)	3856.119264894281
26	(0, 1, 1)x(0, 0, 1, 12)	3742.0263732779404
27	(0, 1, 1)x(0, 1, 0, 12)	3945.3230721010923
28	(0, 1, 1)x(0, 1, 1, 12)	3643.596564109097
29	(0, 1, 1)x(1, 0, 0, 12)	3761.166452424517
30	(0, 1, 1)x(1, 0, 1, 12)	3743.659244461699
31	(0, 1, 1)x(1, 1, 0, 12)	3735.1911214693046
32	(0, 1, 1)x(1, 1, 1, 12)	3645.5291143241184
33	(1, 0, 0)x(0, 0, 0, 12)	3881.004151784765
34	(1, 0, 0)x(0, 0, 1, 12)	3767.9652626049374
35	(1, 0, 0)x(0, 1, 0, 12)	3970.473768780047
36	(1, 0, 0)x(0, 1, 1, 12)	3671.30462940249
37	(1, 0, 0)x(1, 0, 0, 12)	3767.534607796274
38	(1, 0, 0)x(1, 0, 1, 12)	3769.3540926400606
39	(1, 0, 0)x(1, 1, 0, 12)	3734.456214237999
40	(1, 0, 0)x(1, 1, 1, 12)	3673.0393350478257
41	(1, 0, 1)x(0, 0, 0, 12)	3864.6713300348783
42	(1, 0, 1)x(0, 0, 1, 12)	3751.48468880269
43	(1, 0, 1)x(0, 1, 0, 12)	3948.531319617132
44	(1, 0, 1)x(0, 1, 1, 12)	3655.651154298128
45	(1, 0, 1)x(1, 0, 0, 12)	3760.5902542433137
46	(1, 0, 1)x(1, 0, 1, 12)	3752.786145566474
47	(1, 0, 1)x(1, 1, 0, 12)	3732.4362930212087
48	(1, 0, 1)x(1, 1, 1, 12)	3657.572255327884
49	(1, 1, 0)x(0, 0, 0, 12)	3866.599402133973
50	(1, 1, 0)x(0, 0, 1, 12)	3752.679237358937

51	(1, 1, 0)x(0, 1, 0, 12)	3959.1197500482876
52	(1, 1, 0)x(0, 1, 1, 12)	3654.432353340708
53	(1, 1, 0)x(1, 0, 0, 12)	3752.562633971972
54	(1, 1, 0)x(1, 0, 1, 12)	3754.3550460807855
55	(1, 1, 0)x(1, 1, 0, 12)	3725.5335554329304
56	(1, 1, 0)x(1, 1, 1, 12)	3656.33837600289
57	(1, 1, 1)x(0, 0, 0, 12)	3857.7919324193244
58	(1, 1, 1)x(0, 0, 1, 12)	3743.6936255173096
59	(1, 1, 1)x(0, 1, 0, 12)	3947.174630186558
60	(1, 1, 1)x(0, 1, 1, 12)	3645.399042188345
61	(1, 1, 1)x(1, 0, 0, 12)	3753.2081470654875
62	(1, 1, 1)x(1, 0, 1, 12)	3745.3278365046845
63	(1, 1, 1)x(1, 1, 0, 12)	3727.1738645793066
64	(1, 1, 1)x(1, 1, 1, 12)	3647.388847887716

Table 5.10

According to Akaike Information Criterion (AIC), ARIMA (0, 1, 1)x(0, 1, 1, 12)model is considered as the most appropriate one.

5.3.6 DIAGNOSTIC CHECKING

Diagnostic checking is the method of evaluating the capability of a fitted model by examining the residuals.The below figures show that the data is normally distributed. Standard residuals for 'E' show how much outliers are present in the dataset and the Q-Q plot shows the normality of the data. The main aim of diagnostic checking is to ensure that the assumptions of the model are met and the residuals exhibit certain expedient properties

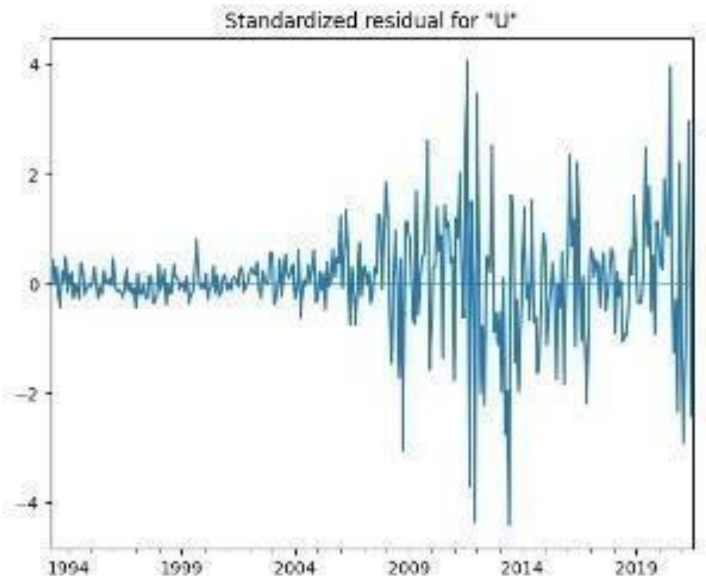

Fig 5.17

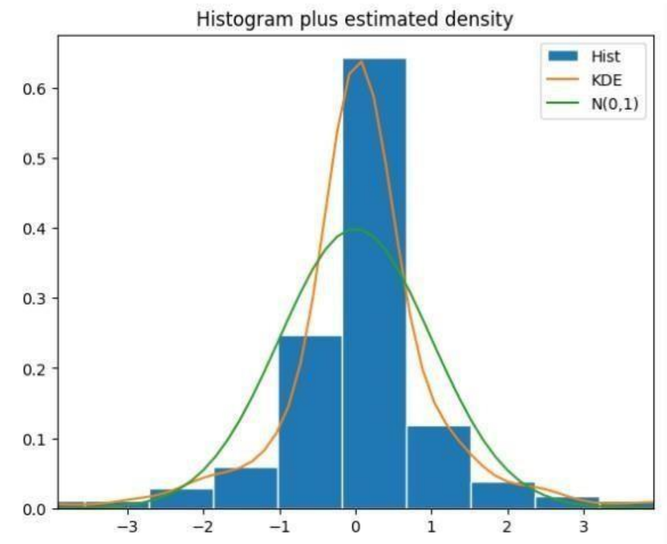

Fig 5.18

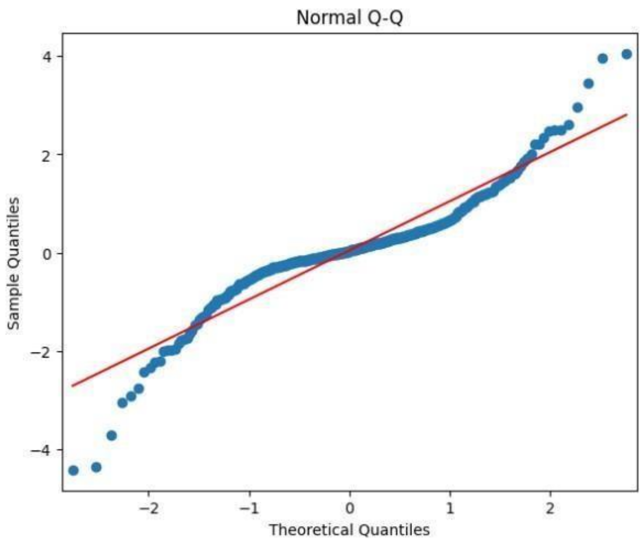

Fig 5.19

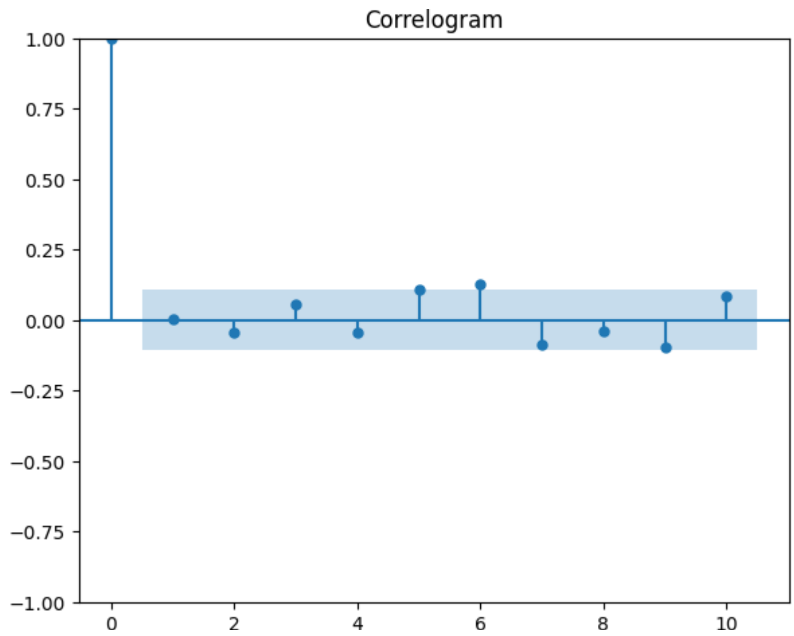

Fig-5.20

5.3.7 FORECASTING THE SAMLE

The In-sample forecast is obtained as follows:

DATE	ACTUAL VALUE	PREDICTED VALUE
1991-01-31	366.0	0.0
1991-02-28	362.7	366.0000000000006
1991-03-29	355.7	362.70054665125303
1991-04-30	357.8	355.7011596229861
1991-05-31	360.4	357.7996523323582
1991-06-28	368.4	360.399569258934
1991-07-31	362.9	368.3986747487234
1991-08-30	347.4	362.90091084167415
1991-09-30	354.9	347.4025676866984
1991-10-31	357.5	354.89875806915086
1991-11-29	366.3	357.49956911080153
1991-12-31	353.2	366.2985422307064
…………………………	…………………………...	…………………………….
2021-01-29	1863.8	1909.1623751351352565
2021-02-26	1742.9	1887.848050227502
2021-03-31	1691.1	1755.8627581197898
2021-04-30	1767.7	1710.3232202096408
2021-05-31	1900.0	1752.4925502565532
2021-06-30	1763.2	1883.76399437511689
2021-07-30	1825.8	1806.1559527813333

Table 5.11

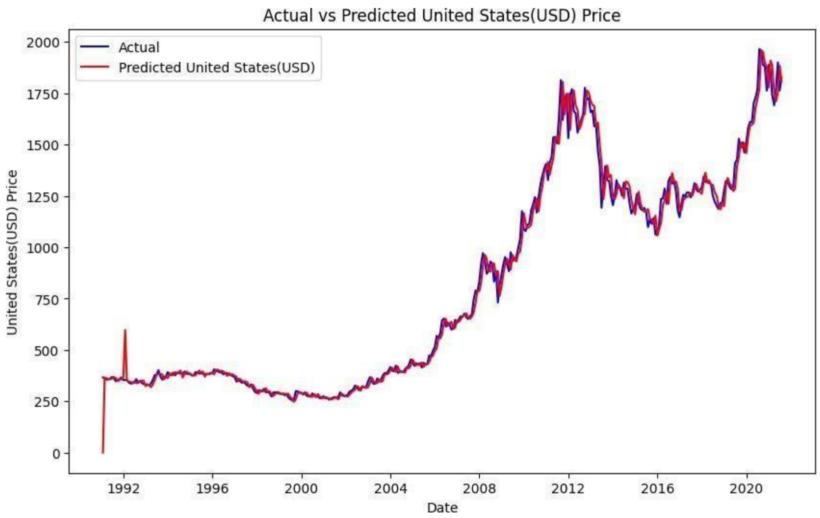

Fig 5.21

5.3.8 FORECATING THE FUTURE VALUES

The LCL and UCL values are given in the following table,

LCL	UCL
-3394.757202228515	3394.757202228515
-2407.237637637424328	3139.237637424328
-2410.5739575767802	3135.975050879286
-2417.5733446060585	3128.975663852031
-2415.4748518966867	3131.074156561403
-2412.874934970111	3133.6740734879786
-2404.8758294803215	3141.673178977768
-2410.373593387371	3136.175415070719
-2425.871965420607	3120.677071916029
-2418.375746159894	3128.173262298196
-2415.7749351182433	3130.7740733398464
-2406.9759619883383	3139.5730494597514
……………………	………...................
1811.268907146171	2007.0558431243421
1789.9547453759278	1985.7413550790761
1657.9752569218201	1853.7502593177594
1612.4358684999202	1808.2105719193614
1654.6052024058056	1850.3798981073007
1785.8765960000392	1981.6512915022986
1708.2686050327752	1904.0433005298914

Table 5.12

DATE	PREDICTED VALUE
2021-08-31	1845.1348022407017
2021-09-30	1832.1017459642492
2021-10-29	1831.0701494409707
2021-11-30	1815.7954943572192
2021-12-31	1823.7591136826175
2022-01-31	1857.27106338474375
2022-02-28	1861.544700642275
2022-03-31	1847.4644021931963
2022-04-29	1862.422329004475
2022-05-31	1868.9604764496814
2022-06-30	1863.356168223465
2022-07-29	1890.4753765419339
2022-08-31	1921.6804693220713
2022-09-30	1899.6474130456188
2022-10-31	1898.6158165223405
2022-11-30	1883.341161438589
2022-12-30	1891.3047807639869
2023-01-31	1924.8167309288076
2023-02-28	1929.090367723645
2023-03-31	1915.0100692745661
2023-04-28	1929.9679960818173
2023-05-31	1936.5061435310513
2023-06-30	1930.901835304835
2023-07-31	1958.0210436233037
2023-08-31	1980.2261364034414
2023-09-29	1967.1930801269887
2023-10-31	1966.1614836037104
2023-11-30	1950.886828519959
2023-12-29	1958.8504478453567
2024-01-31	1992.3623980101775
2024-02-29	1992.3623980101775
2024-03-29	1982.555736355936
2024-04-30	1997.5136631631872
2024-05-31	2004.0518106124211
2024-06-28	1998.4475023862049
2024-07-31	2025.5667107046736
2024-08-30	2047.7718034848112
2024-09-30	2034.7387472083585
2024-10-31	2033.7071506850803
2024-11-29	2018.4324956013288
2024-12-31	2026.3961149267266
2025-01-31	2059.908065091547
2025-02-28	2064.1817018863844
2025-03-31	2050.1014034373056
2025-04-30	2065.059330244557
2025-05-30	2071.5974776937906
2025-06-30	2065.9931694675743
2025-07-31	2093.112377786043

Table 5.13

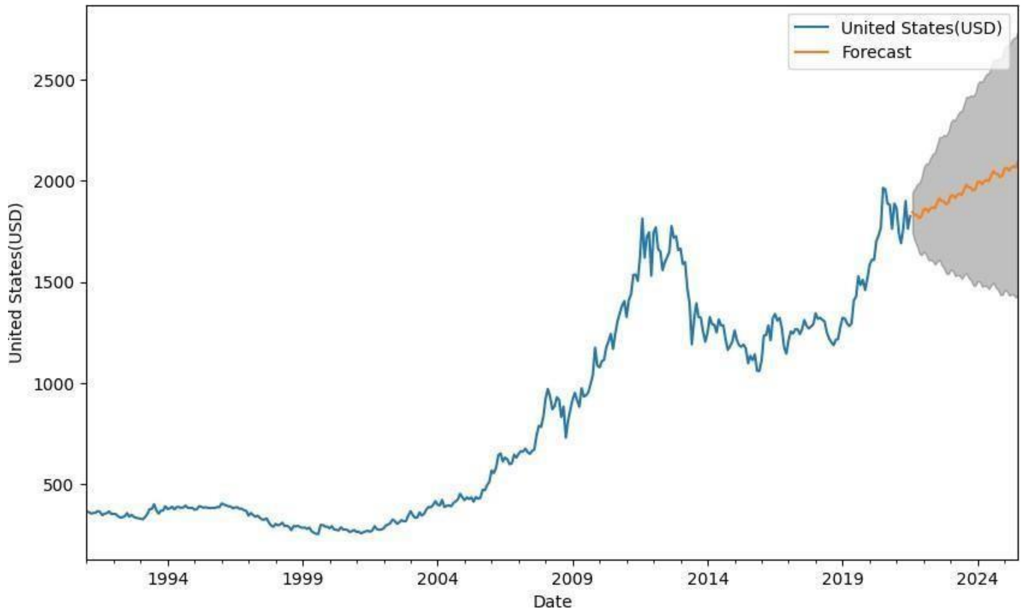

Fig 5.22

5.4 GOLD RATE PREDICTION USING HOLT-WINTERS

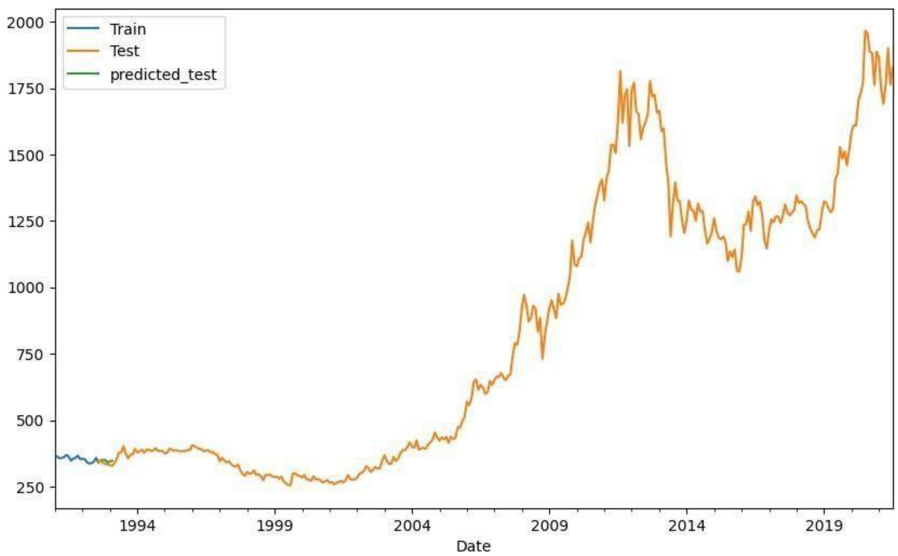

Fig 5.23

5.4.1 DECOMPOSITION OF TIME

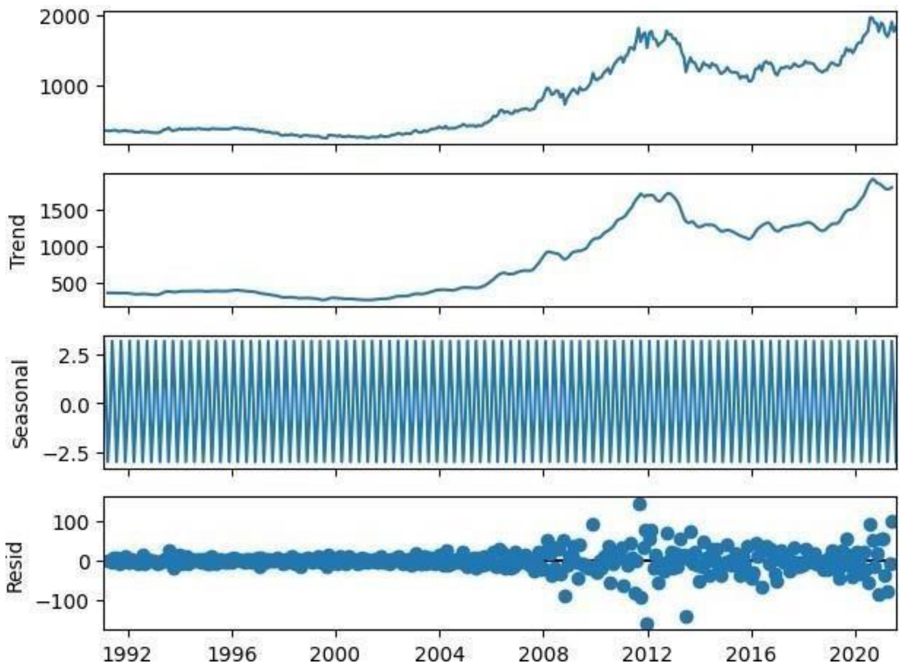

Fig 5.24

5.4.2. FORECASTING THE SAMLE

The In-sample forecast is obtained as follows:

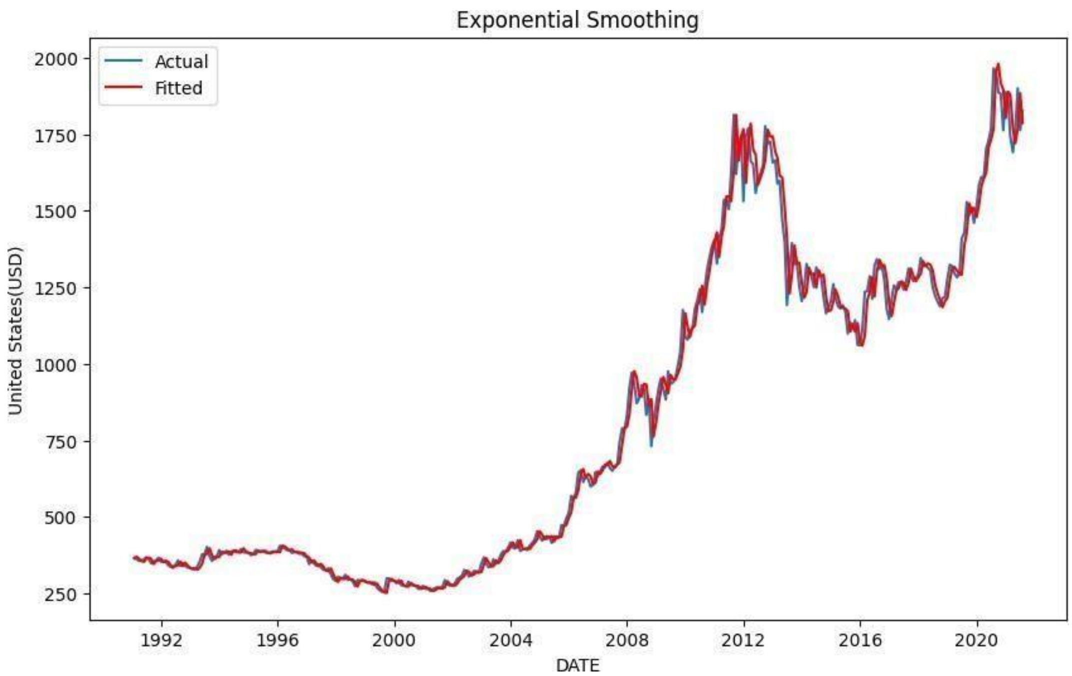

Fig 5.25

5.4.3 FORECATING THE FUTURE VALUES

DATE	PREDICTED VALUE
2021-08-31	1838.873600
2021-09-30	1851.475320
2021-10-29	1853.172146
2021-11-30	1857.326091
2021-12-31	1877.720937
2022-01-31	1890.383387
2022-02-28	1891.912344
2022-03-31	1895.951263
2022-04-29	1916.568274
2022-05-31	1929.291454
2022-06-30	1930.652542
2022-07-29	1934.576435
2022-08-31	1955.415612
2022-09-30	1968.199521
2022-10-31	1969.392739
2022-11-30	1973.201607
2022-12-30	1994.262949
2023-01-31	2007.107589
2023-02-28	2008.132937
2023-03-31	2011.826779
2023-04-28	2033.110287
2023-05-31	2046.015656
2023-06-30	2046.873135
2023-07-31	2050.451952
2023-08-31	2071.957624
2023-09-29	2084.923723
2023-10-31	2085.613332
2023-11-30	2089.077124
2023-12-29	2110.804961
2024-01-31	2123.831790
2024-02-29	2124.353530
2024-03-29	2127.702296
2024-04-30	2149.652299
2024-05-31	2162.739858
2024-06-28	2163.093728
2024-07-31	2166.327468
2024-08-30	2188.499636
2024-09-30	2201.647925
2024-10-31	2201.833926
2024-11-29	2204.952640
2024-12-31	2227.346973
2025-01-31	2240.555992
2025-02-28	2240.574123
2025-03-31	2243.577813
2025-04-30	2266.194311
2025-05-30	2279.464059
2025-06-30	2279.314321
2025-07-31	2282.202985

Table 5.14

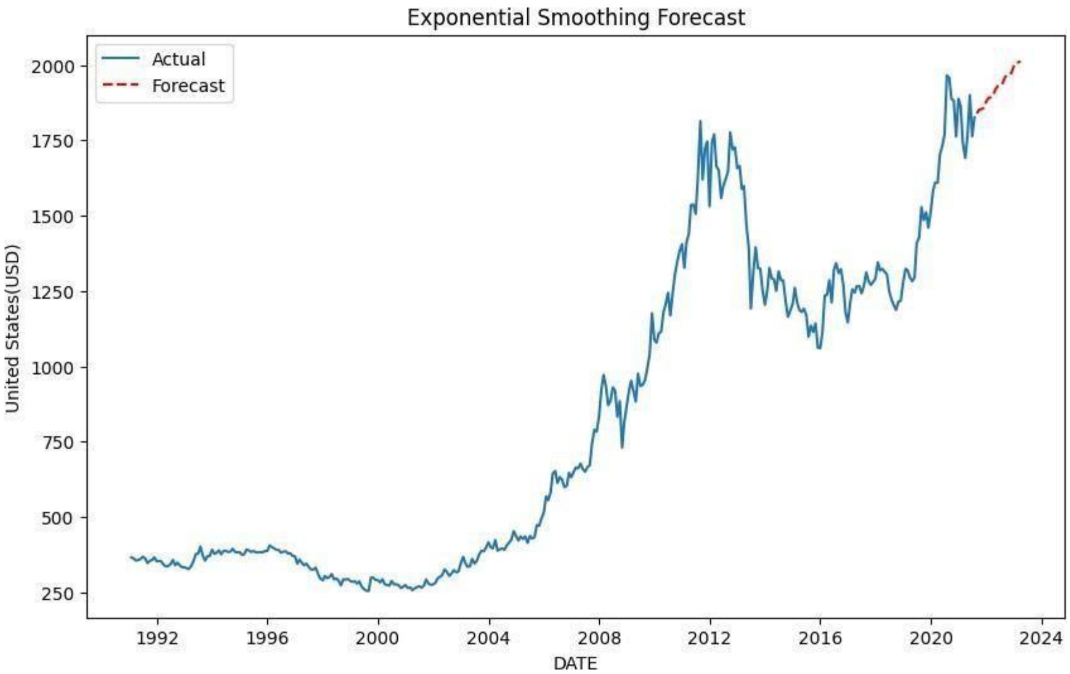

Fig 5.26

5.5 COMPARISON OF FORECAST OF SARIMA AND HOLT-WINTERS MODEL

The forecast of both SARIMA and Holt-Winters models are compared with the help of MSE and RMSE values. The model having least error will be considered as the best one.

IN INDIA

	SARIMA	**HOLT-WINTERS**
MSE	7626381.92	156392566.0141645
RMSE	2761.5904697240317	12505.701340355306

Table 5.15

IN UNITED STATES

	SARIMA	HOLT-WINTERS
MSE	2854.14	35666.28159577199
RMSE	53.42410531151533	188.8551868384133

Table 5.16

We obtain that the Mean Squared Error for SARIMA model is less than that of the Holt-Winters model in both India and USA. Thus, the SARIMA model is considered as the best model

CHAPTER 6

CONCLUSION

The analyses of two time series model resulted forecasting of future 4 years monthly gold rate of India and USA .The historical gold rate data of India and USA from January 1991 to august 2021 were analyzed and forecasted using SARIMA and Holt-Winter's Exponential Smoothening. SARIMA model was got as the best model with a low RMSE of India is 2761.5904697240317 and MSE of India is 7626381.92 . SARIMA best model with a low RMSE of USA is 53.42410531151533 and MSE of USA is 2854.14. Holt's winter model had a RMSE of India is 12505.701340355306 and MSE of India is 156392566.0141645 . Holt's winter model had a RMSE of USA is 188.8551868384133 and MSE of USA is 35666.28159577199 . Both models offer viable forecasting solutions, with ARIMA emphasizing precision and Holt-Winters providing a balanced trade-off between accuracy and simplicity

In accordance with the obtained results, it is evident that the error observed for the SARIMA model is less than that for Holt-Winters model. This means that SARIMA model can be used for the forecasting of gold rate in India and USA . The accuracy of the values forecasted can depend on the data quality and regular updating in the data is also necessary. From the data, it is also visible that the rate of gold rate increasing in future.

CHAPTER -7

REFERENCES

- Aruna, S., Umamaheswari, P., & Sujipriya, J. (2021). Prediction of potential gold prices using machine learning approach. *Annals of the Romanian Society for Cell Biology*, 1385-1396.

- Baur, D. G. (2011). Explanatory mining for gold: Contrasting evidence from simple and multiple regressions. *Resources Policy*, *36*(3), 265-275

- Dadhich, M. (2017). An analysis of volatility of macro economic variables on gold price. *Pacific Business Review International*, *9*(12), 21-25.

- Gangopadhyay, K., Jangir, A., & Sensarma, R. (2016). Forecasting the price of gold: An error correction approach. *IIMB management review*, *28*(1), 6-12

- Mills, T. C. (2004). Statistical analysis of daily gold price data. *Physica A: Statistical Mechanics and its Applications*, *338*(3-4), 559-566.

- Parimi, S. (2018). Factor influencing the gold prices: An empirical investigation in the Indian context. *Theoretical Economics Letters*, *8*(15), 3444-3456.

- Qian, Y., Ralescu, D. A., & Zhang, B. (2019). The analysis of factors affecting global gold price. *Resources Policy*, *64*, 101478

- Šimáková, J. (2011). Analysis of the relationship between oil and gold prices. *Journal of finance*, *51*(1), 651-662

- ul Sami, I., & Junejo, K. N. (2017). Predicting future gold rates using machine learning approach. *International Journal of Advanced Computer Science and Applications*, *8*(12).

- Vanitha, S., & Saravanakumar, K. (2019). The usage of gold and the investment analysis based on gold rate in India. *International Journal of Electrical and Computer Engineering*, *9*(5), 4296.

- Vidya, G. S., & Hari, V. S. (2020, December). Gold price prediction and modelling using deep learning techniques. In *2020 IEEE Recent Advances in Intelligent Computational Systems (RAICS)* (pp. 28-31). IEEE

- Website : (Monthly Gold Prices (1979-2021) (kaggle.com))