An Empirical Analysis of Inflation Rate in India

Dr. V. Sumalatha,
Assistant Professor, Department of Statistics, Keshav Memorial Institute of Commerce and Sciences, Narayanaguda, Hyderabad.

***Abstract*—This project aims to conduct a comprehensive time series analysis of inflation rate, focusing on major factors such as Food & Beverages, Housing, Clothing & Footwear, Fuel & light, Pan & Tabacco, Miscellaneous.**
This study aims to investigate the trends and patterns of inflation rates in India and to forecast the inflation rate in India using statistical methods. Inflation, a critical economic variable affects purchasing power and overall economic stability. Secondary data was obtained from Ministry of Statistics and Programme Implementation, Government of India. For the given data some of the statistical techniques were applied such as Time series Analysis to forecast the inflation rate Seasonal Autoregressive Integrated Moving Average (SARIMA) model was applied. Some external factors also influence the inflation such as economic conditions and some policies.
This project involves data collection, data pre-processing and the applying forecasting model to predict future inflation rates.
The findings will contribute to a deeper understanding of inflation factors in the context of a rapidly evolving Indian Economy.

***Index Terms*—Inflation, Consumer Price Index (CPI), Forecasting, Time series, SARIMA.**

I. INTRODUCTION

Inflation is a crucial economic indicator that affects purchasing power, economic stability, and policy decisions. In India, inflation is measured using the Consumer Price Index (CPI), which includes major components such as Food & Beverages, Housing, Clothing & Footwear, Fuel & Light, Pan & Tobacco, and Miscellaneous items. Inflation patterns are influenced by seasonal variations, economic conditions, and policy interventions.
This study analyzes trends and patterns in India's inflation rates using secondary data from the Ministry of Statistics and Programme Implementation, Government of India. A time series approach is employed, with the Seasonal Autoregressive Integrated Moving Average (SARIMA) model used to forecast monthly inflation rates for the next five years. The findings aim to enhance understanding of inflation dynamics in the context of India's evolving economy.

II. PROBLEM STATEMENT

This project will analyse India's inflation history to forecast inflation rates for the next five years, focusing on trend identification and key factor analysis.

III. OBJECTIVES

1. To analyse the trends and patterns of Inflation Rates in India.
2. Forecasting the Inflation rates for next 5 years (monthly wise).

IV. MATERIALS AND METHODS

4.1 Data collection
The Dataset is of Secondary datatype. The Dataset has been collected from the website Ministry of Statistics and Programme Implementation, Government of India. "All India Inflation Rate".

4.2 Description of data
The Dataset contains the Monthly Inflation Rates from 2014-2024. The original dataset includes 120 different inflation rates each having 7 columns including factors such as Food & Beverages, Housing, Clothing & Footwear, Fuel & light, Pan & Tabacco, Miscellaneous. The dataset is pre-processed selected all the factors from the original dataset
After preprocessing the dataset each year have the following number of rows and columns: each year had 12 rows and 8 columns.

The Dataset consists of 9 columns (Features) such as Date, Month, Food & Beverages, Housing, Clothing & Footwear, Fuel & light, Pan & Tabacco, Miscellaneous.

Date: The date on which the inflation rates was recorded.

Month: The inflation rates is calculated on first of every month.
Food & Beverages: It tells us about the value of Food & Beverages in every month.

Housing: It tells us about the value of Housing in every month.
Clothing & Footwear: It tells us about the value of Clothing & Footwear in every month.
Fuel & Light: It tells us about the value of Fuel & Light in every month.

Pan & Tabacco: It tells us about the value of Pan & Tabacco in every month.

Miscellaneous: It tells us about the value of Miscellaneous in every month.
Consumer Food Price Index: It tells us about the value of Consumer Price Index in every month.

4.3 Time Series Analysis
A time series is data collected on a variable over time, such as sales per month or temperature per day. Each value is linked to a specific time period. Time series data help us understand how a variable changes over time.
The main goals of time series analysis are:
To understand past patterns in the data
To predict future values
Usually, time series models describe values as a function of time along with some random error.

4.3.1 Components of Time Series
A time series is made up of four main components:

4.3.1.1 Trend
The trend shows the long-term direction of the data. It may increase, decrease, or remain constant over a long period. Trends reflect long-term influences such as economic growth or population changes.

4.3.1.2. Seasonal Variation
Seasonal variations are regular patterns that repeat within a year. These can occur monthly, quarterly, weekly, or daily and are often caused by weather, festivals, or routine business activities.

4.3.1.3. Cyclical Variation
Cyclical variations are long-term ups and downs that last for more than one year. They are commonly linked to business cycles like boom, recession, depression, and recovery. These cycles are not fixed and can vary in length.

4.3.1.4. Irregular Variation
Irregular variations are unexpected and random changes caused by events such as natural disasters, wars, strikes, or epidemics. These changes are unpredictable and do not repeat.

4.4 Forecasting
Forecasting is the process of making predictions based on past and present data. Later these can be compared against what happens. Prediction is similar but more general ter. Forecasting might refer to specific formal statistical methods employing time series, cross-sectional or longitudinal data, or alternatively to less formal judgmental methods or the process of prediction and resolution itself. Usage can vary between areas of application: for example, in hydrology the terms “forecast” and “forecasting” are sometimes reserved for estimates of values at certain specific future times, while the term “prediction” is used for more general estimates, such as the number of times floods will occur over a long period.

4.4.1 Auto-correlation function & partial auto-correlation function

4.4.1.1 Auto-correlation function (ACF):
The autocorrelation function measures the correlation between observations of a time series separated by various time lags. It helps identify temporal dependence, repeating patterns, and the presence of stochastic memory in the data. ACF is commonly visualized through correlograms, where correlation coefficients are plotted against lag values.

$$\rho_k = \frac{\mathrm{Cov}\left(e_t, e_{t+k}\right)}{\sqrt{\mathrm{Var}\left(e_t\right)\mathrm{Var}\left(e_{t+k}\right)}}$$

A correlated process, such that ARMA or ARIMA, has non-zero values at lags other than zero to indicate a correlation between different lagged observations.
ACF plots the correlation coefficient against the lag, and it's a visual representation of autocorrelation.
The correlation coefficient is measured either by Pearson's correlation coefficient or by Spearman's rank correlation coefficient. The correlation coefficient can range from -1 to +1.

4.4.1.2 Partial Auto-correlation function (PACF):
The partial autocorrelation function measures the direct correlation between a time series and its lagged values while controlling for the effects of intermediate lags. PACF is particularly useful for identifying the appropriate order of autoregressive models and is a key component of the Box–Jenkins methodology. The use of this function was introduced as part of the Box-Jenkins approach to time series modelling, whereby plotting the partial autocorrelative functions one could determine the appropriate lags p in an AR (p) model or in an extended ARIMA (p, d, q) model.
The PACF graph is constructed by plotting all the values of PACF obtained from regressions at different lags.

4.4.2 Autoregressive (AR) Component & Seasonal Autoregressive (SAR) Component

4.4.2.1. Autoregressive (AR) Component:
Autoregression is a statistical technique used in time-series analysis that assumes that the current value of a time series is a function of its past values. Autoregressive models use similar mathematical techniques to determine the probabilistic correlation between elements in a sequence.
The autoregressive non-seasonal component represented by (1−ϕ1B) (1−ϕ1 B) captures the relationship between the current observation and a certain number of lagged observations (previous values in the time series). The B term represents the backshift operator is commonly used in time series analysis. It represents the lag operator, which shifts the time series backward by a certain number of time period. The order of the autoregressive component, denoted by (p), determines the number of past values considered in the model.
Autoregressive models apply linear regression with lagged variables of its output taken from previous steps. Unlike linear regression, the autoregressive model doesn't use other independent variables except the previously predicted results. Consider the following formula.

$$p(x) = \prod_{i=1}^{n} p(x_i | x_1, x_2, \ldots x_{i-1}) = \prod_{i=1}^{n} p(x_i | x_{<i})$$

When expressed in the probabilistic term, an autoregressive model distributes independent variables over n-possible steps, assuming that earlier variables conditionally influence the outcome of the next one.
We can also express autoregressive modelling with the equation below.

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \epsilon_t$$

Here, y is the prediction outcome of multiple orders of previous results multiplied by their respective coefficients, ϕ. The coefficient represents weights or parameters influencing the predictor's importance to the new result. The formula also considers random noise that may affect the prediction, indicating that the model is not ideal and further improvement is possible.

4.4.2.2 Seasonal Autoregressive (SAR) Component
This component captures the relationship between the current value of the series and its past values, specifically at seasonal lags.
The seasonal autoregressive component is represented by (1−Φ1Bs) (1−Φ1 Bs) This component captures the relationship between the current observation and a certain number of lagged observations at seasonal intervals. The BsBs term represents the backshift operator applied to the seasonal lagged observations.
The SAR component has three main parameters. They are:
P (Seasonal Autoregressive Terms):
D (Degree of Seasonal Differencing):
Q (Seasonal Moving Average Terms):

4.4.3 SARIMA (Seasonal Autoregressive Integrated Moving Average)
SARIMA, which stands for Seasonal Autoregressive Integrated Moving Average, is a versatile and widely used time series forecasting model. It's an extension of the non-seasonal ARIMA model, designed to handle data with seasonal patterns. SARIMA captures both short-term and long-term dependencies within the

data, making it a robust tool for forecasting. It combines the concepts of autoregressive (AR), integrated (I), and moving average (MA) models with seasonal components.

SARIMA Model is represented as,

$$SARIMA\underbrace{(p,d,q)}_{non-seasonal}\underbrace{(P,D,Q)_m}_{seasonal}$$

were,

m = number of observations per year;

P = Number of seasonal AR terms;

D = Number of seasonal differences;

Q = Number of seasonal MA terms

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

4.4.3.1 The Components of SARIMA

To grasp SARIMA, let's break down its components:

4.4.3.1.1 Seasonal Component: The "S" in SARIMA represents seasonality, which refers to repeating patterns in the data. This could be daily, monthly, yearly, or any other regular interval. Identifying and modelling the seasonal component is a key strength of SARIMA.

4.4.3.1.2 Autoregressive (AR) Component: The "AR" in SARIMA signifies the autoregressive component, which models the relationship between the current data point and its past values. It captures the data's autocorrelation, meaning how correlated the data is with itself over time.

4.4.3.1.3 Integrated (I) Component: The "I" in SARIMA indicates differencing, which transforms non-stationary data into stationary data. Stationarity is crucial for time series modelling. The integrated component measures how many differences are required to achieve stationarity.

4.4.3.1.4 Moving Average (MA) Component: The "MA" in SARIMA represents the moving average component, which models the dependency between the current data point and past prediction errors. It helps capture short-term noise in the data.

4.5 Evaluate the Model

Let's evaluate the forecasted sales values by comparing them to the observed sales data using two common metrics for this evaluation: Mean Absolute Error (MAE) and Mean Squared Error (MSE).

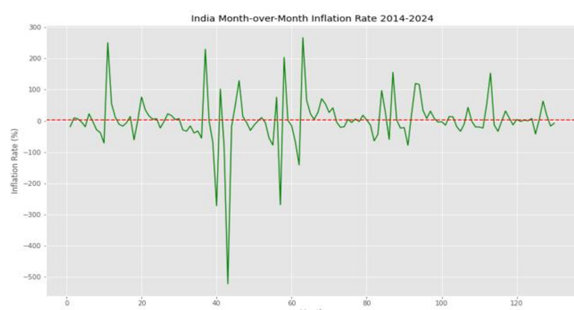
MAE (Mean Absolute Error) measures the average absolute difference between the observed and forecasted values. It provides a simple and easily interpretable measure of the model's accuracy.

MSE (Mean Squared Error) measures the average of the squared differences between the observed and forecasted values. MSE gives more weight to large errors and is sensitive to outliers.

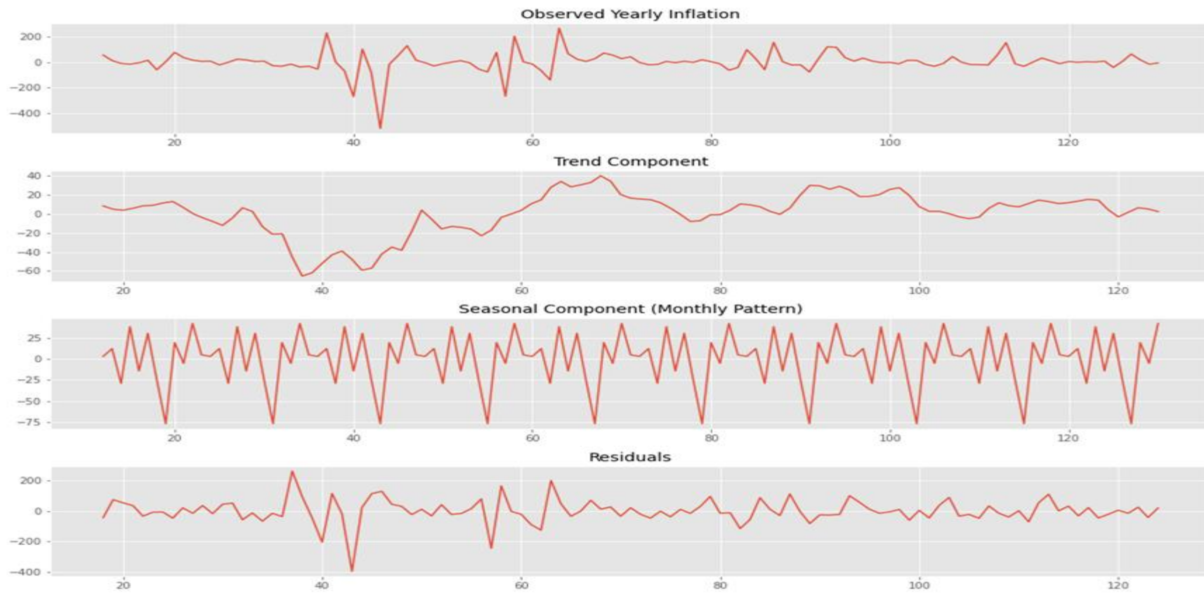
Lower values indicate better performance.

V. RESULTS AND DISCUSSIONS

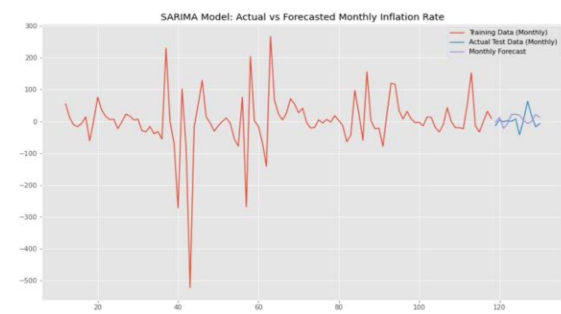
This project explored and forecasted the monthly inflation rates for next 5 years (2025-2029). By analysing historical data and employing time series forecasting technique Seasonal Autoregressive Integrated Moving Average (SARIMA).



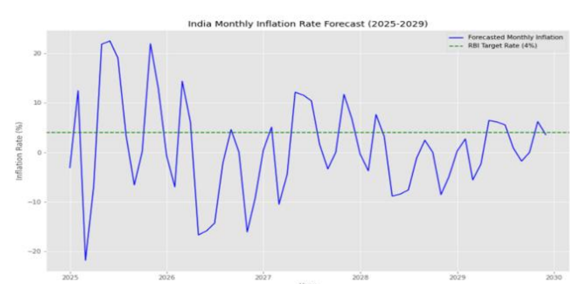
- X- Axis represents the time periods (in months) from 2014-2024.The numbers on the X-Axis represents the month index
- Y- Axis represents the month – over - month inflation rate, expressed as a percentage.
- Green line shows the actual month - over - month inflation rate over the specified time period.
- Red line represents the zero-inflation rate. Serves as a baseline to quickly identify periods of inflation and deflation.



- From the 1st plot X- Axis represents the time index and Y- Axis represents the observed yearly rates.
- This plot shows the raw, observed yearly inflation data. It exhibits fluctuations and volatility, suggesting the presence of trends, seasonality, and random noise.
- From 2nd plot X- Axis represents the time index and Y- Axis represents the Trend component.
- This plot shows the long-term movement or direction of the data. It smooths out the short-term fluctuations and highlights the overall upward or downward trajectory. We can see a period of decline followed by a rise and then a period of relative stability.
- From 3rd plot X- Axis represents the time index and Y- Axis represents the seasonal component.
- This plot shows the recurring patterns or fluctuations that occur at regular intervals. The consistent peaks and troughs suggest a strong seasonal pattern in the inflation data.
- From 4th plot X- Axis represents the time index and Y- Axis represent the residuals.
- This plot shows the remaining variations I the data after removing the trend and seasonal components. Ideally, the residuals it suggests that the decomposition has not fully captured the underlying structure of the data.



- X- Axis represents the time period in months.
- Y- Axis represents the monthly inflation rate, expressed as a percentage.
- Red line (Training data) exhibits significantly volatility and fluctuations, indicating a non-stationary time series.
- Blue line (Actual Test Data) continues the pattern of volatility and fluctuations seen in the training data.
- Grey line (Monthly Forecast) attempts to capture the general trend and fluctuations in the test data.



- X- Axis represents the years from 2025 to 2030, showing the timeframe of the forecast
- Y- Axis represents the monthly inflation rate, expressed as percentage.
- Blue line (Forecasted monthly inflation) shows significant volatility and fluctuations, exhibits both positive and negative inflation rates.
- Green dotted line (RBI Target rate) represents the RBI's desired inflation rate of 4%.
- In 2025 inflation rate was very low and then suddenly it went to peak stage in the same year.

VI. CONCLUSION

This project explored and forecasted the monthly inflation rates for next 5 years (2025-2029). By analysing historical data and employing time series forecasting technique Seasonal Autoregressive Integrated Moving Average (SARIMA)The forecasts show significant swings in monthly inflation, with both substantial positive & negative values. The analysis identified Food & Beverages as the most influential factor with a positive impact on inflation rate. This implies that changes in "food & beverages" prices have a substantial impact on the overall inflation rate. "Clothing & Footwear" and "Pan & Tabacco" also have influence but when compared with food & beverages it shows less impact on inflation rate.

REFERENCES

[1] Banerjee, A. (2021). Forecasting price levels in India – An ARIMA framework. Academy of Marketing Studies Journal, 25(1). https://www.abacademies.org/articles/forecasting-price-levels-in-india-an-arima-framework-9703.html Allied Business Academies

[2] Kothadia, K. D., & Nayak, D. N. (2020). Modeling inflation in India: A univariate approach. Arthshastra Indian Journal of Economics & Research. https://doi.org/10.17010/aijer/2020/v9i2-3/155602 indianjournalofeconomicsandresearch.com

[3] Laxmi Prasanna, P. (2021). Forecasting inflation rate (WPI & CPI) in India with time series models – A statistical recurrent neural network approach. International Advanced Research Journal in Science, Engineering and Technology (IARJSET). IARJSET

[4] Muthu, N. S., Senthamarai Kannan, K., Deneshkumar, V., & Thangasamy, P. (2021). SARIMA model for forecasting price index fluctuations. European Journal of Mathematics and Statistics, 2(6), 1–8. https://doi.org/10.24018/ejmath.2021.2.6.67 ej-math.org

[5] Singh, R., & Sen, S. (2022). Forecasting CPI-C of India using ARIMA, SARIMA and LSTM: A comparative approach. International Research Journal of Management Science and Technology, 13(3), 174–186. irjmst.com

[6] Sharma, J. L., Maheshwari, R. K., & Agarwal, S. (2024). Forecasting inflation in India using the SARIMA model. International Journal of Research in Economics and Social Sciences, 14(01), 1–15. euroasiapub.org

[7] Henderi, H., & Sofiana, S. (2025). Comparative study of traditional and modern models in time series forecasting for inflation prediction. International Journal for Applied Information Management, 5(3), 155–167. ijaim.net

[8] Box, G. E. P., & Jenkins, G. M. (1976). Time series analysis: Forecasting and control (Rev. ed.). Holden-Day.

[9] Ministry of Statistics and Programme Implementation. (2024). Consumer Price Index (CPI) data. Government of India. https://www.mospi.gov.in

[10] Reserve Bank of India. (2023). Handbook of statistics on Indian economy. https://www.rbi.org.in