Term Paper: Forecasting Inflation Rate of (2014-2024)

1. Introduction

Inflation is a critical economic indicator that impacts the purchasing power of consumers and the overall economic stability of a country. Accurate forecasting of inflation rates can assist policymakers and financial institutions in making informed decisions. This paper applies the Box-Jenkins (BJ) methodology to forecast the inflation rate of India using a univariate time series model based on data from January 2014 to January 2024.

2. Data Description and Preprocessing

The dataset consists of monthly inflation rates in India from January 2014 to January 2024, spanning 121 observations. The time series object is created using the ts() function in R with a frequency of 12, corresponding to monthly data.

2.1 Summary Statistics

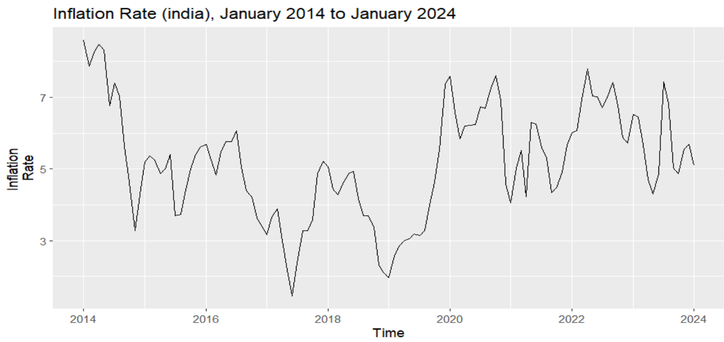
The summary statistics show that the inflation rate ranges from 1.46% to 8.6%, with a mean of approximately 5.16%. These statistics provide an initial understanding of the central tendency and dispersion of the data.

Statistic	Min	1st Qu.	Median	Mean	3rd Qu.	Max
Value	1.46	4.17	5.1	5.162	6.23	8.6

3. Exploratory Data Analysis

3.1 Time Series Plot

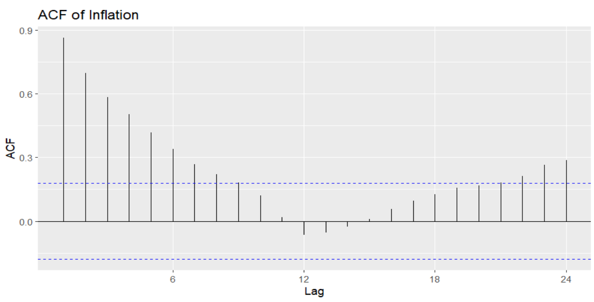
The time series plot reveals fluctuations in the inflation rate over the ten-year period, with some apparent seasonality and a slight downward trend over the years. The inflation rate peaked at 8.6% in early 2014 and declined to its lowest point at 1.46% later in the series.

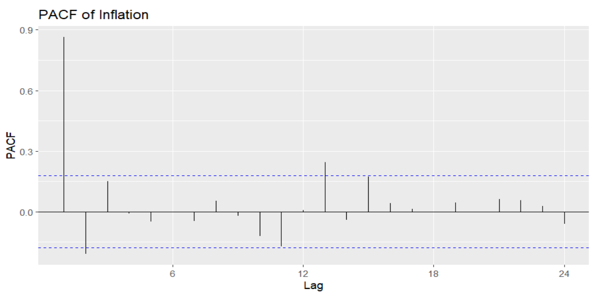


data shows a repeating pattern at regular intervals, it indicates seasonality.

3.2 Seasonality Testing …

3.2.1 ACF and PACF of data



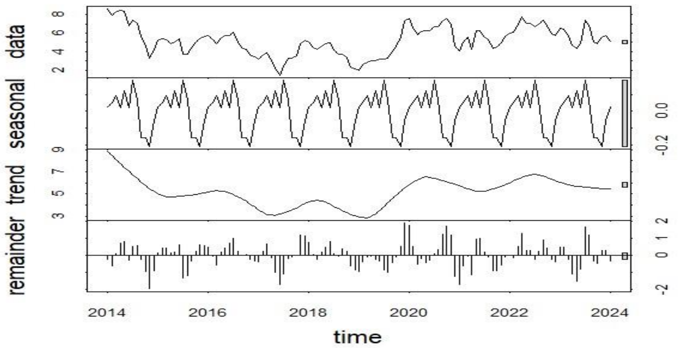


- Acf [**Seasonality**: The presence of spikes at lags 12 and 24 suggests a **yearly seasonality** in the inflation rate. This means that the inflation rate exhibits a repeating pattern every 12 months, which is typical for economic indicators that are influenced by seasonal factors (e.g., changes in consumer demand, agricultural cycles).
   Non-seasonal Autocorrelation: The gradual decay of autocorrelation at lower lags indicates that the inflation rate might also have a strong autoregressive component, meaning that past values have a substantial influence on future values, aside from the seasonal effect.]

- Pacf • significant lag at 12 suggests that the inflation rate 12 months ago has a direct influence on the current inflation rate, which is a strong indication of annual seasonality.

3.2.2 Decomposition testing

- decomposition in time series analysis is a technique used to break down a time series into its constituent components, typically into trend, seasonality, and residual (or noise) components. This helps in understanding and analyzing the underlying structure of the time series data

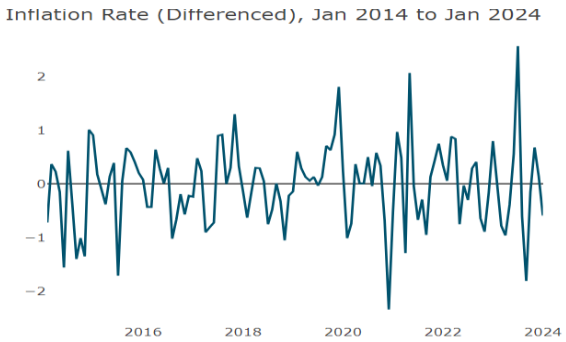


Box-Ljung test

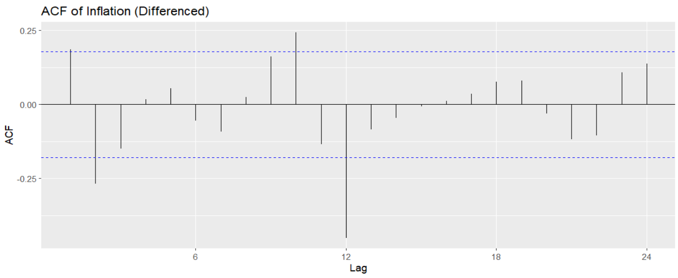
```
data:  inf
X-squared = 288.86, df = 12, p-value < 2.2e-16
```

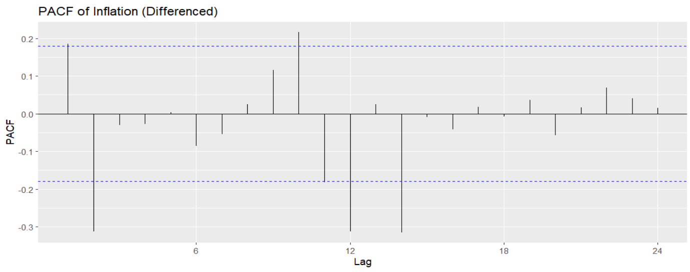
the data does not appear to be independently distributed, and some autocorrelation structure likely exists.
the Box-Ljung test tells you about autocorrelation, the specific pattern of that autocorrelation (detected through ACF or PACF) is what helps confirm seasonality.

First order difference of label data



3.2.3 Differenced ACF and PACF of data





The ACF at lag 1 is positive and significant, which suggests that the differenced series is positively correlated with its immediately preceding value. This is common in many time series that exhibit a random walk behavior before differencing.

- The significant lag 1 autocorrelation, combined with the PACF plot you provided earlier, suggests that an AR(1) term could be appropriate in an ARIMA model.
- The significant spike at lag 12 indicates that the model might need to account for seasonality or a cyclical component, which could suggest including a seasonal ARIMA component or exploring models like SARIMA (Seasonal ARIMA).

PACF : lag 12 has a significant negative PACF. This might indicate a seasonal component or another cyclical behavior in the data.

To check the stationary of DATA

4.1 Augmented Dickey-Fuller (ADF) Test

The ADF test is conducted with different lag orders to determine whether the series is stationary. The results indicate mixed outcomes. At lag order 1, the p-value is 0.02266, suggesting the series is stationary. However, at higher lag orders, the p-values exceed 0.05, indicating non-stationarity.

Data	Dickey-Fuller	Lag Order	p-value	Alternative Hypothesis
inf	-3.5414	4	0.04152	stationary
inf	-3.337	2	0.06838	stationary
inf	-3.7729	1	0.02266	stationary
dinf	-5.2314	4	0.01	stationary

4.2 Phillips-Perron (PP) Test

The PP test shows a p-value of 0.1372 for the original series, suggesting non-stationarity. However, after differencing the series, the p-value drops below 0.01, indicating stationarity.

Test Type	Data	Test Statistic	Truncation Lag Parameter	p-value	Alternative Hypothesis
Phillips-Perron Unit Root Test	inf	-16.923	4	0.1372	stationary
Phillips-Perron Unit Root Test	dinf	-77.615	4	0.01	stationary

4.3 KPSS Test

The KPSS test for level stationarity gives a p-value of 0.08622 for the original series and 0.1 for the differenced series. These results suggest that the original series is close to being non-stationary, but the differenced series is stationary.

Test Type	Data	Test Statistic	Truncation Lag Parameter	p-value	Alternative Hypothesis
KPSS Test for Level Stationarity	inf	0.37897	4	0.08622	stationary
KPSS Test for Level Stationarity	dinf	0.11739	4	0.1	stationary

5. Model Identification and Estimation

5.1 In-Sample Forecasting and Validation OR Model Performance

#Partition the data into test data and training data

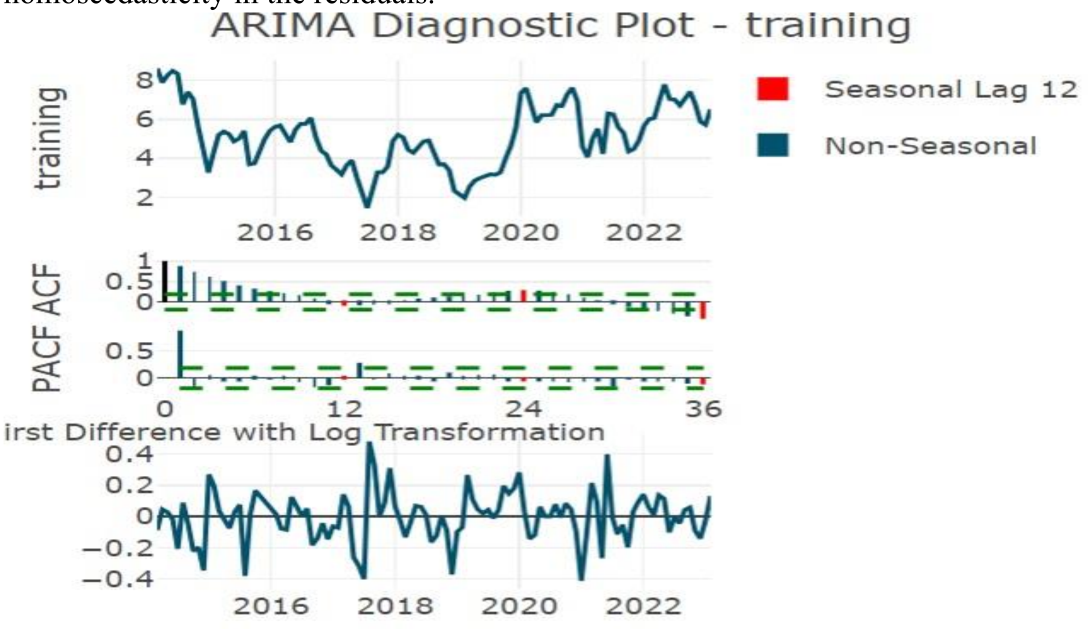
The data was split into training (109 observations) and testing sets (12 observations) to validate the models

#Using an ARIMA Diagnostic Plot on the Training Dataset

Given the results of the stationarity tests, the series is differenced to achieve stationarity, and the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots of the differenced series are examined to identify the appropriate model.

Model Diagnostics

Each model was evaluated using residual diagnostics to check for autocorrelation, normality, and homoscedasticity in the residuals.



6. ARIMA Model Identification

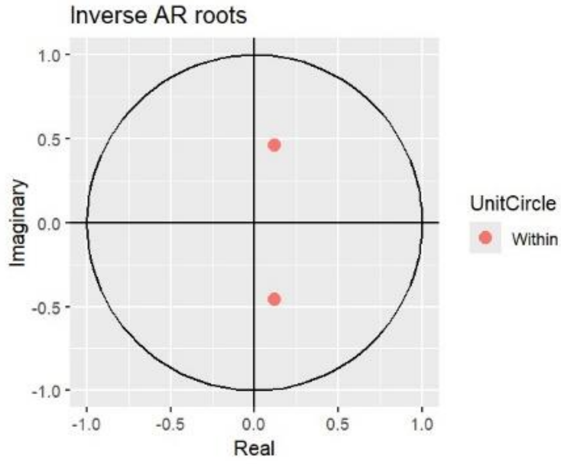
Two models were considered:

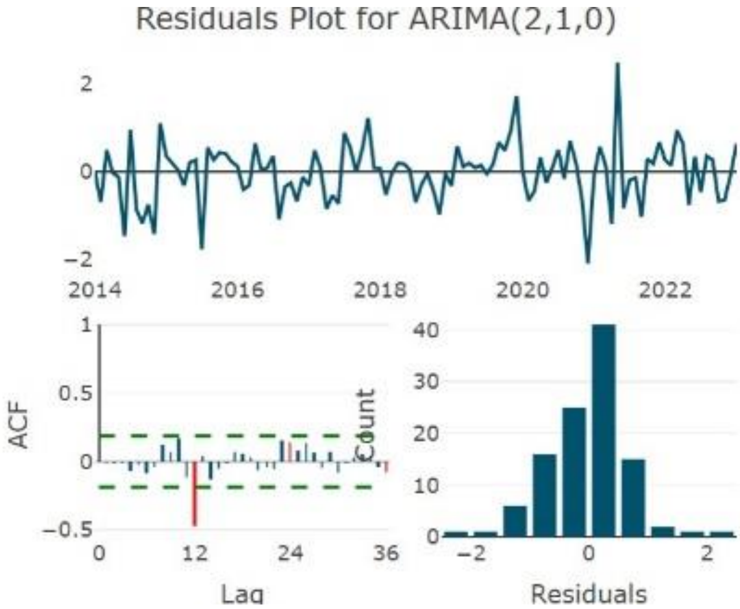
1. **ARIMA(2,1,0)**: This model was selected based on ACF and PACF plots showing significant lags at the second autoregressive term (AR2) and no moving average.

```
 arima(x = training, order = c(2, 1, 0))

Coefficients:
         ar1      ar2
      0.2424  -0.2259
s.e.  0.0946   0.0941

sigma^2 estimated as 0.4557:  log likelihood = -110.88,  aic = 227.76
```



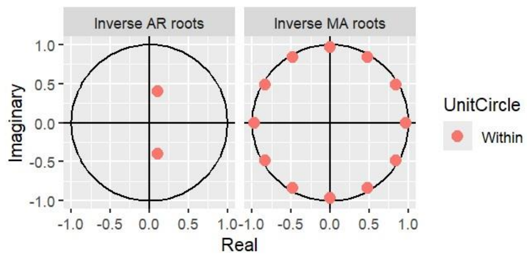


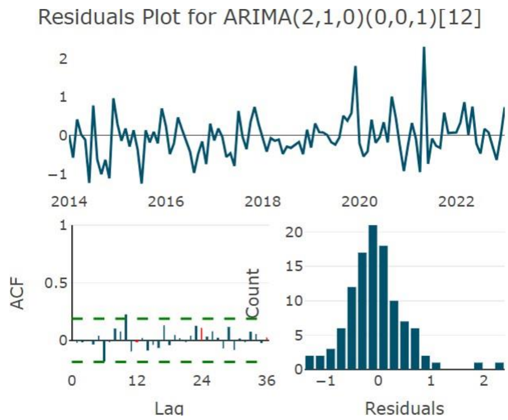
SARIMA(2,1,0)(0,0,1)[12]: This seasonal ARIMA model includes a seasonal component with a lag of 12, suggesting annual seasonality.

```
arima(x = training, order = c(2, 1, 0), seasonal = list(order = c(0, 0, 1)))

Coefficients:
         ar1      ar2     sma1
      0.2183  -0.1702  -0.6652
s.e.  0.0968   0.0971   0.1015

sigma^2 estimated as 0.3066:  log likelihood = -92.95,  aic = 193.9
```





7.1 In-Sample Forecasting and Model Performance

The performance of the models was evaluated using Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE):

ARIMA(2,1,0): Test set RMSE =1.3669817 %, MAPE = 24.32322%

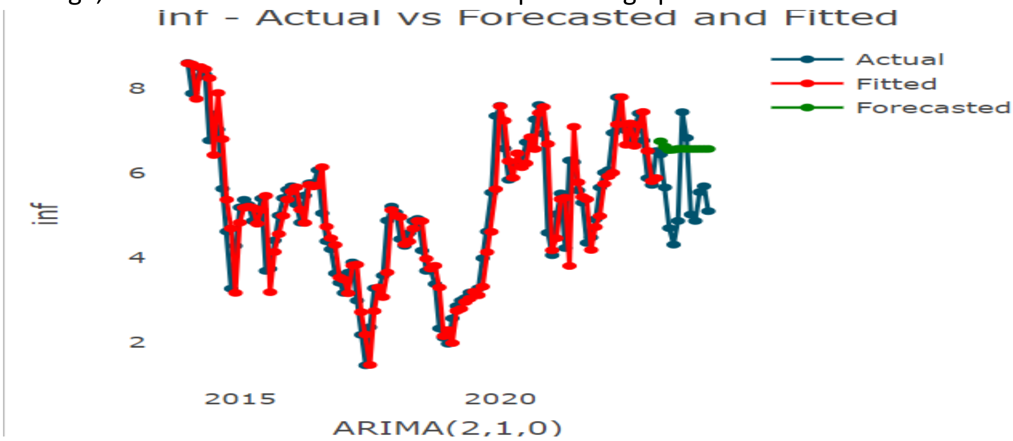
SARIMA(2,1,0)(0,0,1)[12]: Test set RMSE =0.8355113, MAPE = 12.88864

The SARIMA model with an additional seasonal component performed better than the ARIMA model, making it the best model for forecasting.
In this section, we evaluate the performance of the time series models used to forecast inflation rates. The evaluation is based on the Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) calculated on the test set, which consists of the most recent 12 months of data (January 2023 to January 2024).

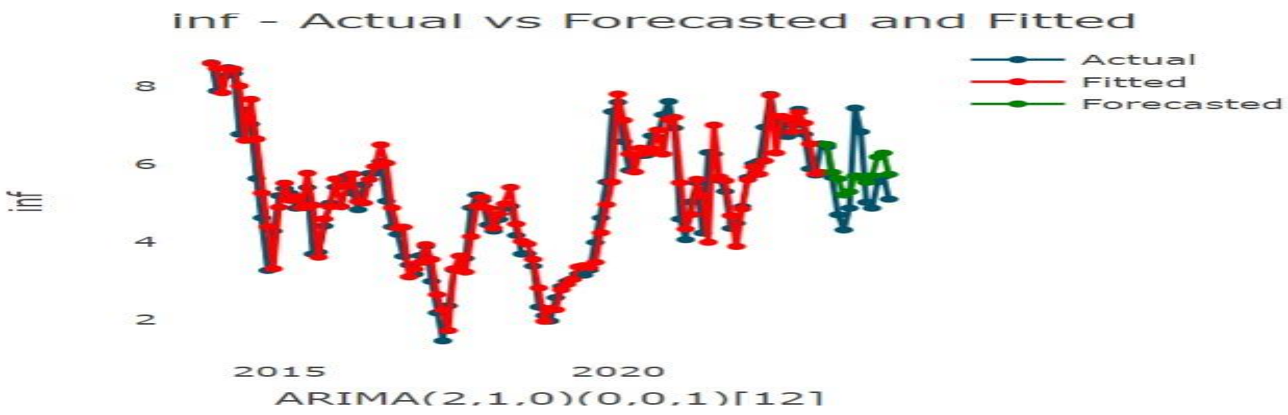
7.1.1 Root Mean Square Error (RMSE)

RMSE measures the average magnitude of the errors between the predicted and actual values. It is computed as the square root of the average of the squared differences between predicted and actual values. A lower RMSE indicates better model performance.

- **ARIMA(2,1,0)**: RMSE =1.3669817
  - This model has the highest RMSE among the models tested, indicating that its forecasts tend to be further from the actual values compared to the other models. The RMSE value suggests that on average, the forecast errors are about 1.36 percentage points.



- **SARIMA(2,1,0)(0,0,1)[12]**: RMSE = 0.8355113
- This model has a lower RMSE than the ARIMA(2,1,0) model, suggesting it provides more accurate forecasts. The RMSE value indicates that its average forecast error is about 0.83 percentage points.



-
-

7.1.2 Mean Absolute Percentage Error (MAPE)

MAPE measures the average percentage error between the predicted and actual values, expressed as a percentage. It provides a clear understanding of the forecast accuracy in relative terms. Lower MAPE values signify better model performance.

- **ARIMA(2,1,0)**: MAPE = 24.32322%
    - This model has the highest MAPE, meaning that its forecasts are off by an average of about 24.32% from the actual values. This high percentage indicates that the model's forecasts are less reliable and less precise compared to the others.

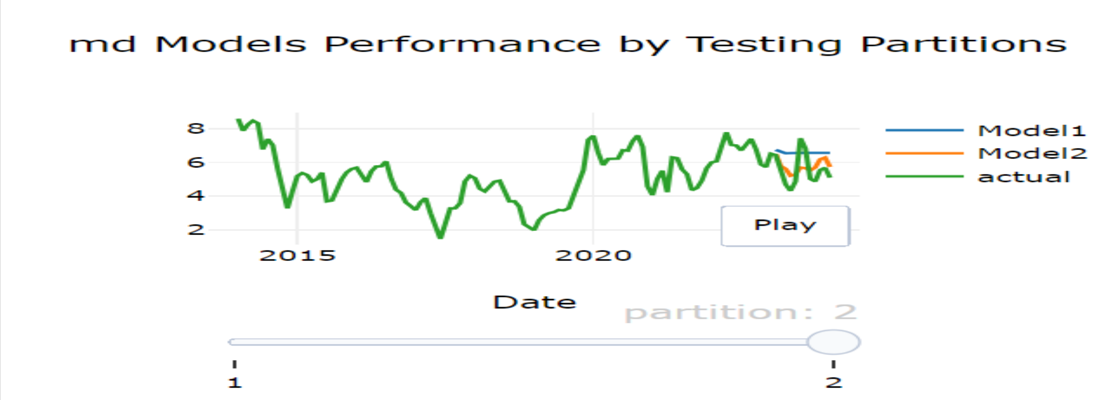
SARIMA(2,1,0)(0,0,1)[12]: MAPE = 12.88864
- This model has a lower MAPE than the ARIMA(2,1,1), showing a better forecast accuracy in relative terms. The average percentage error is about 12.88864 %, suggesting that the forecasts are relatively closer to the actual values.

The **SARIMA(2,1,0)(0,0,1)[12]** model is the most suitable for forecasting inflation rates based on the given data, as it has the lowest RMSE and MAPE. This indicates that it provides the most accurate predictions, with both the smallest absolute errors and the lowest percentage deviations from actual values..

#**Training the models with backtesting**

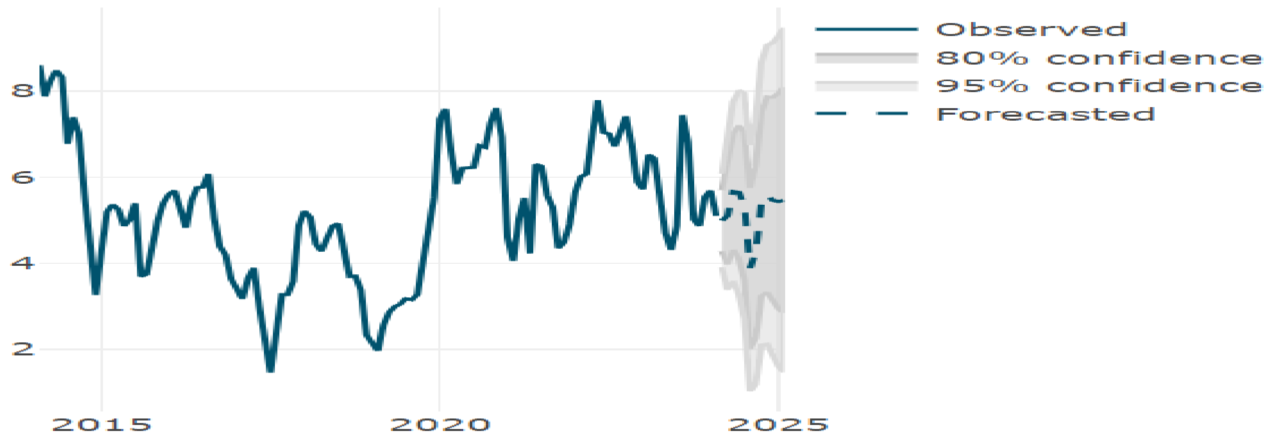
8. Out-of-Sample Forecasting

The final model chosen for out-of-sample forecasting was **SARIMA(2,1,0)(0,0,1)[12]** the The model was used to generate a 7-month forecast from February 2024 to August 2024.



8.1 Forecast Interpretation

The forecast suggests a gradual increase in inflation rates, peaking in May 2024 at approximately 6.18%. This is followed by a decline, with the lowest forecasted inflation rate of 4.05% in July 2024. The prediction intervals indicate some uncertainty, especially in the later months, as reflected by the widening prediction intervals.



```
Forecasts:
         Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
Feb 2024       4.990736 4.281125 5.700347 3.905480 6.075992
Mar 2024       5.101119 3.982681 6.219557 3.390616 6.811623
Apr 2024       5.669223 4.307090 7.031355 3.586021 7.752425
May 2024       5.650890 4.104465 7.197315 3.285837 8.015943
Jun 2024       5.326210 3.613123 7.039298 2.706269 7.946152
Jul 2024       3.885430 2.017022 5.753839 1.027946 6.742915
Aug 2024       4.267167 2.254991 6.279342 1.189810 7.344523
Sep 2024       5.385736 3.239841 7.531631 2.103873 8.667599
Oct 2024       5.597279 3.325673 7.868886 2.123157 9.071401
Nov 2024       5.467381 3.076628 7.858133 1.811040 9.123722
Dec 2024       5.441239 2.936970 7.945509 1.611290 9.271189
Jan 2025       5.468389 2.855531 8.081246 1.472368 9.464409

>
```

orecasting Inflation Rate of India (2014-2024) Using SARIMA (R)

● Applied Box-Jenkins methodology to forecast India's inflation rate from January 2014 to January 2024 using a univariate time series model.

● Performed Exploratory Data Analysis (EDA) with time series plots, ACF, PACF, and decomposition testing to identify trends, seasonality, and autocorrelations.

● Utilized Augmented Dickey-Fuller (ADF), Phillips-Perron (PP), and KPSS tests to check for stationarity, confirming the need for differencing.

● Developed and compared ARIMA(2,1,0) and SARIMA(2,1,0)(0,0,1)[12] models. SARIMA showed better performance with lower RMSE (0.835) and MAPE (12.89%).

● Generated out-of-sample forecasts for February 2024 to August 2024, predicting a peak in inflation in May 2024.


```
Forecasts:
         Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
Feb 2024       4.990736 4.281125 5.700347 3.905480 6.075992
Mar 2024       5.101119 3.982681 6.219557 3.390616 6.811623
Apr 2024       5.669223 4.307090 7.031355 3.586021 7.752425
May 2024       5.650890 4.104465 7.197315 3.285837 8.015943
Jun 2024       5.326210 3.613123 7.039298 2.706269 7.946152
Jul 2024       3.885430 2.017022 5.753839 1.027946 6.742915
Aug 2024       4.267167 2.254991 6.279342 1.189810 7.344523
Sep 2024       5.385736 3.239841 7.531631 2.103873 8.667599
Oct 2024       5.597279 3.325673 7.868886 2.123157 9.071401
Nov 2024       5.467381 3.076628 7.858133 1.811040 9.123722
Dec 2024       5.441239 2.936970 7.945509 1.611290 9.271189
Jan 2025       5.468389 2.855531 8.081246 1.472368 9.464409
```



```
Forecasts:
         Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
Feb 2024       4.990736 4.281125 5.700347 3.905480 6.075992
Mar 2024       5.101119 3.982681 6.219557 3.390616 6.811623
Apr 2024       5.669223 4.307090 7.031355 3.586021 7.752425
May 2024       5.650890 4.104465 7.197315 3.285837 8.015943
Jun 2024       5.326210 3.613123 7.039298 2.706269 7.946152
Jul 2024       3.885430 2.017022 5.753839 1.027946 6.742915
Aug 2024       4.267167 2.254991 6.279342 1.189810 7.344523
```