ANALYZING AND FORECASTING CONSUMER PRICE INDEX TRENDS USING ADVANCED TIME SERIES TECHNIQUES

1. Introduction

The Consumer Price Index, or CPI, measures the change in prices of goods and services typically bought by households. It's a key indicator of inflation and the cost of living. To create the CPI, various data are collected and a range of methods are used, ensuring that the index is accurate and comparable across different regions.

1.1. The CPI index matters because:

- Inflation Tracking: It's a principal measure of inflation, showing how prices change over time.
- Economic Policy: Governments use CPI to shape economic policy and adjust interest rates.
- Wage Adjustments: Employers may base wage increases on CPI to match living cost changes.
- Purchasing Power: It reflects the purchasing power of consumers' money over time.
- Pension Updates: Many pension plans use CPI to adjust benefits to maintain retirees' living standards.
- Investment Insight: Investors use CPI to inform strategies, as it can impact stock and bond markets.
- Business Strategy: Companies use it for pricing strategies and budget planning.
- International Comparisons: CPI allows for comparison of inflation rates between countries.

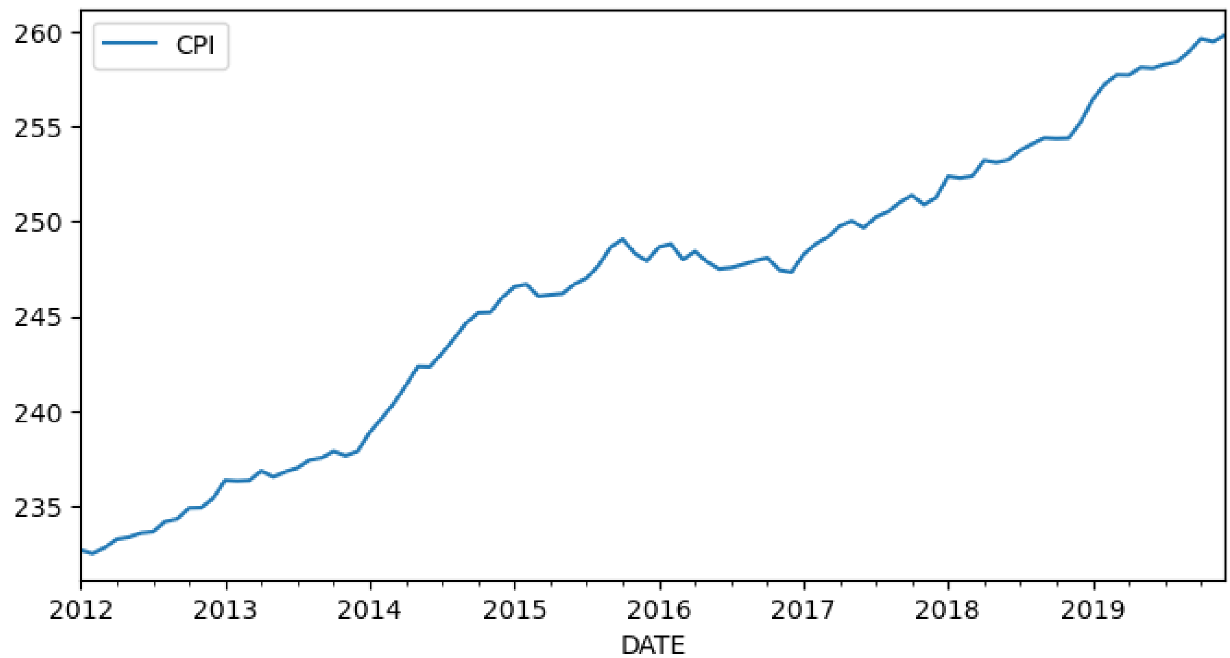
1.2. The Consumer Price Index (CPI) in the USA is influenced by:

- Supply Chain Dynamics: Availability and cost of goods and services.
- Monetary Policy: Interest rates and government spending.
- Consumer Spending: Changes in household buying habits.
- Global Economics: Import prices and global demand.
- Technology: Innovations that change production costs.
- Market Trends: Shifts in consumer preferences and trends.
- Natural Events: Weather impacts on food and energy prices.
- Regulatory Changes: Taxes and government regulations.

2. Data

The series **CPIUFDNS** on the Federal Reserve Bank of St. Louis's FRED website represents the Consumer Price Index for All Urban Consumers: All Items in U.S. City Average, Not Seasonally Adjusted. It tracks the changes in prices paid by urban consumers for a representative basket of goods and services over time. This data is crucial for economic analysis as it provides insight into inflationary trends and cost of living adjustments, helping policymakers, businesses, and consumers make informed decisions.

Link - https://fred.stlouisfed.org/series/CPIUFDNS
Range - January 1, 2012, to December 1, 2019
Frequency: Monthly

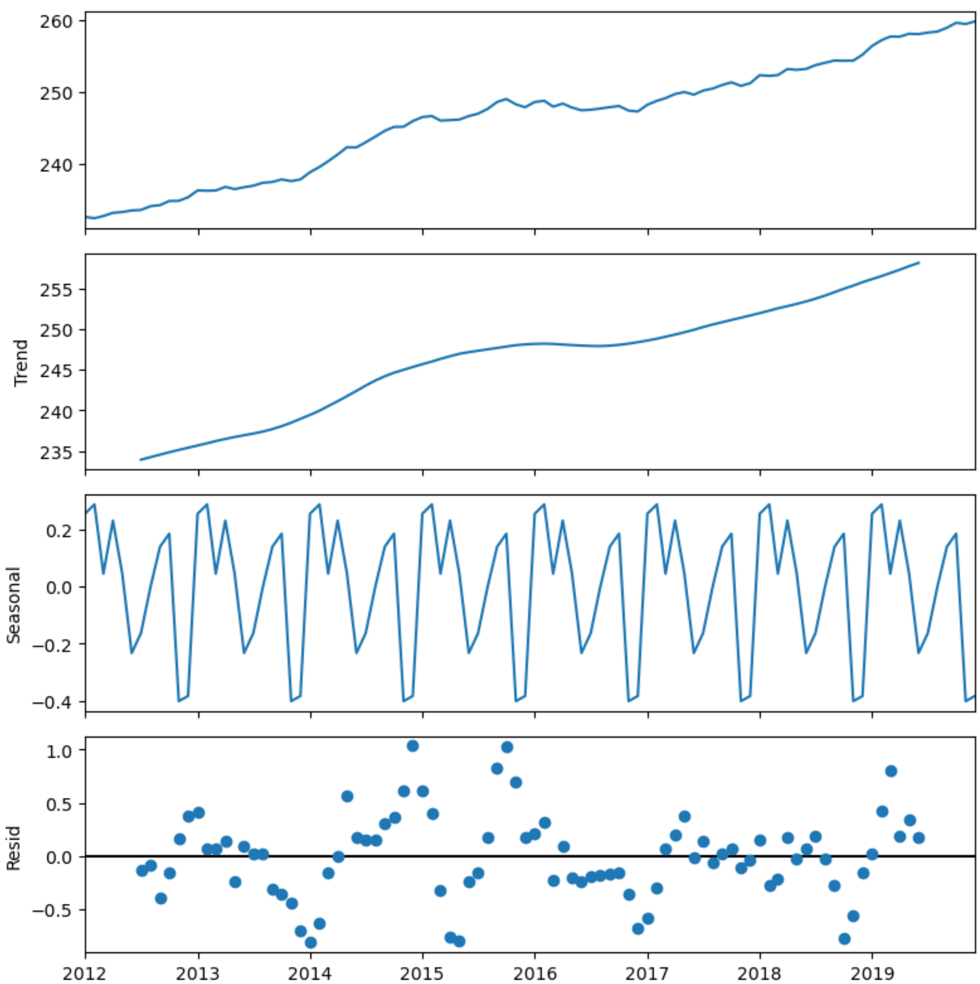


3. Exploratory Analysis

3.1. Seasonal Decomposition

Technique used in time series analysis to separate a time series into three components-

1. **Trend**: This represents the long-term progression of the series, showing movements in the mean over time, which could be due to factors like economic growth, inflation, or increasing population.
2. **Seasonal**: This captures regular patterns that repeat over time, such as monthly or quarterly patterns in the data, which could be influenced by factors like seasonal consumer behavior, holiday effects, or weather patterns.
3. **Residual**: These are the irregular or random fluctuations that cannot be attributed to the trend or seasonal components. It's essentially the "noise" left over after the trend and seasonal components have been removed.



The seasonal decomposition of the CPI data from 2012 to 2019 reveals a consistent upward trend, indicating inflationary pressures in the economy as consumer prices have generally increased. The seasonal component indicates that there are predictable seasonal fluctuations in the CPI, which could be leveraged for anticipatory economic measures. The residuals do not show any apparent pattern, suggesting that the model has effectively captured the trend and seasonality in the CPI data. Overall, the analysis of the CPI through seasonal decomposition provides valuable insights for economic planning and forecasting.

4. Train Test Split

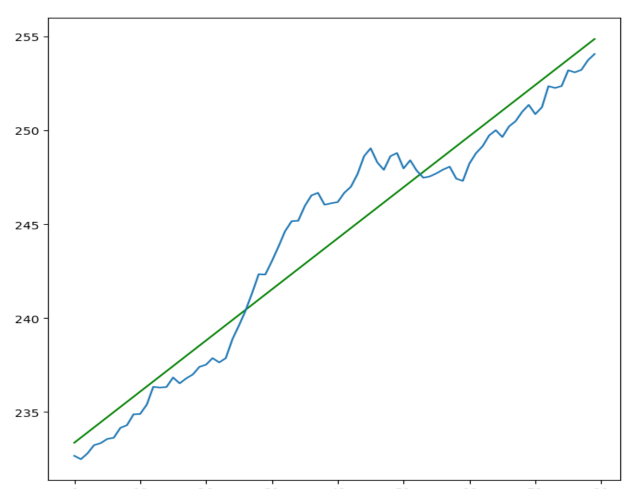
The dataset comprising 95 data points representing the Consumer Price Index (CPI) was strategically divided to facilitate the development and validation of a predictive model. The distribution of the data is as follows:

1. Training Data: The initial 80 data points form the training set, which corresponds to approximately 84% of the total data. This segment will serve as the foundation for training the model, enabling it to learn the underlying patterns and trends in the CPI data.
2. Testing Data: The subsequent 15 data points constitute the test set, making up about 16% of the total data. This portion is reserved for evaluating the model's predictive accuracy, providing an unbiased measure of its performance on new data it has not previously encountered.

5. Clean Data

5.1. Detrending CPI Data Using Linear Regression Analysis

A linear regression model is being used to detrend the Consumer Price Index (CPI) data, a method that helps in understanding the underlying trend by removing the effect of seasonality and irregular fluctuations.



The Consumer Price Index (CPI) data was redefined by removing the time-related trend. This was done by first calculating the trend using a linear model, where the trend equals time (t) multiplied by the model's slope (lr.coef_[0]), plus the intercept (lr.intercept_). We then subtracted this calculated trend from the original CPI values.

The result of this subtraction is the 'Detrend' series. This new series represents the CPI data free from its time-based linear trend. It highlights the true, underlying patterns in the CPI, focusing on cyclical and irregular movements that the original trend might have masked. This detrended data is crucial for a clearer understanding of the CPI's behavior over time.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                    CPI   R-squared:                       0.941
Model:                            OLS   Adj. R-squared:                  0.940
Method:                 Least Squares   F-statistic:                     1236.
Date:                Fri, 15 Dec 2023   Prob (F-statistic):           1.34e-49
Time:                        14:13:14   Log-Likelihood:                -150.08
No. Observations:                  80   AIC:                             304.2
Df Residuals:                      78   BIC:                             308.9
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const        233.3633      0.354    658.572      0.000     232.658     234.069
t              0.2723      0.008     35.163      0.000       0.257       0.288
==============================================================================
Omnibus:                       12.896   Durbin-Watson:                   0.076
Prob(Omnibus):                  0.002   Jarque-Bera (JB):               12.096
Skew:                           0.875   Prob(JB):                      0.00236
Kurtosis:                       2.248   Cond. No.                         90.7
==============================================================================
```

From the OLS regression results, we observe an R-squared value of 0.941, indicating that approximately 94.1% of the variation in the CPI can be explained by the time variable 't'. This high R-squared value suggests a strong fit of the model to the data.

The model summary shows that the slope of the trend line is approximately 0.2723, and the intercept is 233.3633. The slope indicates that, on average, there is a monthly increase of 0.2723 units in the CPI over the period analyzed. The positive slope confirms the inflationary trend observed in the first part of the decomposition analysis.

The F-statistic and its associated p-value indicate that the time variable is statistically significant in predicting the CPI, meaning the relationship observed is not by random chance.

Furthermore, the Durbin-Watson statistic is around 0.076, which is quite low, suggesting that there is a positive autocorrelation in the residuals of our model. Ideally, a Durbin-Watson value close to 2.0 indicates no autocorrelation. This could imply that the model may not be capturing some of the trend-related information, which could be due to other cyclical factors not accounted for by a simple linear trend.

In conclusion, the linear regression model effectively detrends the CPI data, revealing a steady inflationary trend over the years. However, additional investigation into potential autocorrelation and cyclical patterns could further refine the model's accuracy and forecasting ability.

5.2. Analyzing Seasonal Effects and Deseasonalising the CPI Data

Seasonal adjustment is a statistical technique used to remove the seasonal component of a time series that exhibits a systematic, calendar-related movement. It simplifies the underlying trend and other cyclical patterns that may be present in the data. In the context of the CPI data, this seasonal component could represent regular fluctuations due to factors like holidays, weather changes, or fiscal policies.

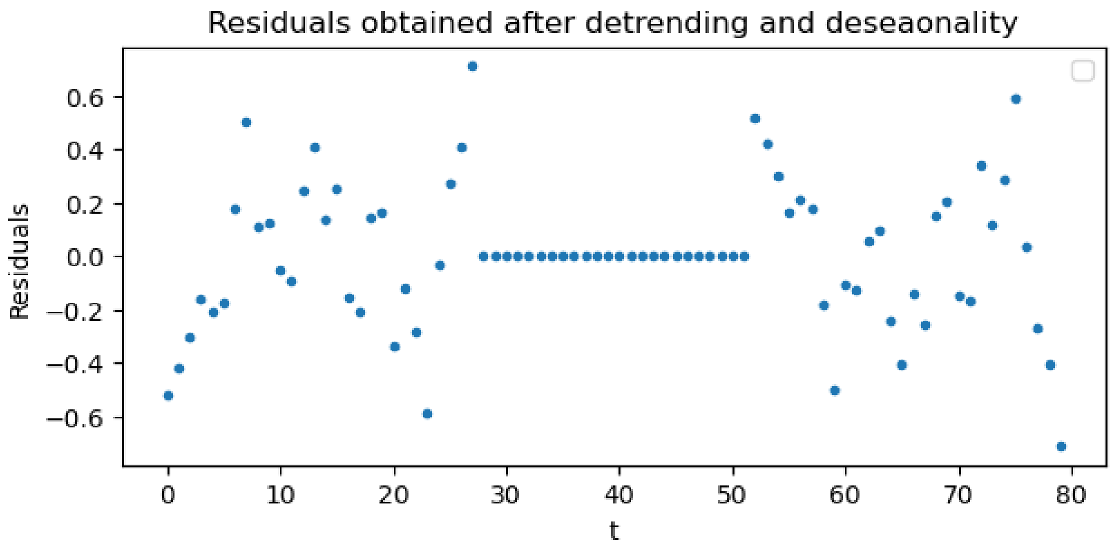
Seasonal Mean Calculation:
To capture the regular patterns that occur at specific intervals within our data, we computed the average values corresponding to each interval in a complete cycle, which is presumed to have 12 intervals — representing months in a year. This computation was carried out by aggregating the data points for each interval and then averaging them. The resulting averages reflect the typical seasonal fluctuations that can be expected for each interval, allowing us to discern and adjust for these regular patterns in our dataset.

The sum of the calculated seasonal averages approached zero (-2.5579538487363607e-12), a strong indication that the seasonal influences across the dataset have been evenly accounted for and offset. This near-zero sum confirms that the adjustment has effectively neutralized the seasonal component, ensuring that it no longer distorts the underlying trends and patterns in the data.

5.3. Calculation of Residuals

After adjusting for seasonality, we further refined our CPI dataset by isolating the random or irregular components. This was achieved by deducting the seasonal averages from the detrended data. The resulting values are known as residuals, which are expected to be random, devoid of any trend or seasonal pattern.

The residuals were plotted against time to assess their distribution and to check for randomness. The ideal outcome is a plot where the residuals are scattered evenly without forming any discernible patterns, suggesting that the significant predictable components have been removed.



Observations from the Plot:

1. Randomness: The scatter plot did not reveal any systematic patterns in the residuals, indicating that the adjustments for trend and seasonality were successful.
2. Consistency: The residuals appeared to be uniformly distributed, reinforcing the effectiveness of the detrending and deseasonalization.
3. Anomalies: We did not observe any significant outliers, which would have suggested remaining anomalies not addressed by the seasonal or trend adjustments.

The numerical and graphical analyses suggest that the process of detrending and deseasonalization has been executed correctly. The absence of structure in the residuals indicates that the primary deterministic components typically influencing the CPI have been adequately neutralized. The residuals can now serve as a clean dataset for advanced analyses, including identifying anomalies or developing stochastic models for future predictions.

5.4. Test for Stationarity

The Dickey-Fuller test was performed to determine the stationarity of the time series data after detrending and deseasonalization. The results of the test are as follows:
The Dickey-Fuller test results indicate that the time series data is stationary. This conclusion is based on two key findings:

- The test statistic of -3.683 is lower than the critical values for 1%, 5%, and 10% significance levels, suggesting a high level of confidence in the result.
- A p-value of 0.0043 reinforces this confidence, showing a low probability that the data is non-stationary.

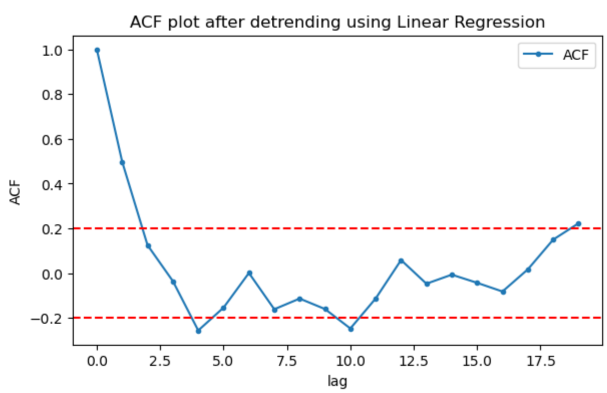
In practical terms, this means that the data does not show patterns of trend or seasonality and is therefore suitable for reliable statistical analysis and forecasting.

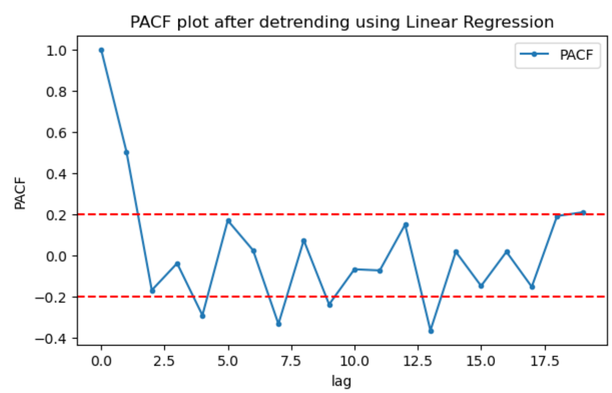
6. ARMA Model Selection Based on ACF and PACF Plots

In the process of forecasting using the ARMA model, we assess the autocorrelation of data points through two diagnostic plots: ACF (Autocorrelation Function) and PACF (Partial Autocorrelation Function). These plots reveal how the data points in a time series are related over time. Specifically:

- ACF Plot: Shows the correlation between the series and its lags, guiding the choice of moving average components.
- PACF Plot: Indicates the direct effect of past data on the current data, assisting in selecting autoregressive components.

Selecting an ARMA model based on these plots ensures that the model captures the significant relationships in the data, which is essential for making reliable predictions.





- ACF and PACF Plots: Both plots show a prominent initial spike at the first lag, which exceeds the threshold value of 0.2, indicating significant correlation at the first lag.
- Threshold Setting: The threshold of 0.2 helps us distinguish between significant and insignificant lags.
- Model Parameters: The strong initial spike in both plots leads us to choose an ARMA(1,1) model. This means the model considers one past data point and the error from the previous forecast in predicting the next value.

The plots guide us to a model that looks just one step back into the past data to predict the future. This approach, supported by the data's patterns, is expected to forecast future values effectively.

7. Training and Performance of the ARMA(1,1) Model

We trained an ARMA(1,1) model on our Consumer Price Index (CPI) dataset. This model was chosen based on the analysis of ACF and PACF plots, which indicated that the first lag of data has a significant influence on the series. The fitted model was then used to generate in-sample and out-sample predictions for the CPI.

A SARIMAX model with ARIMA(1,0,1) specifications was fitted to the residual data obtained from 80 observations. The SARIMAX model extends the ARIMA model by incorporating seasonal trends, which were not needed in this case, given the ARIMA specification.

Model Fit and Statistical Significance:

- Log Likelihood: The positive value of 9.789 indicates that the model has a good fit to the data.
- Akaike Information Criterion (AIC): With a value of -11.577, the AIC suggests the model is parsimonious and fits the data well.
- Bayesian Information Criterion (BIC): The BIC value of -2.049 also supports the model's selection by balancing model complexity and fit.

Coefficient Analysis:

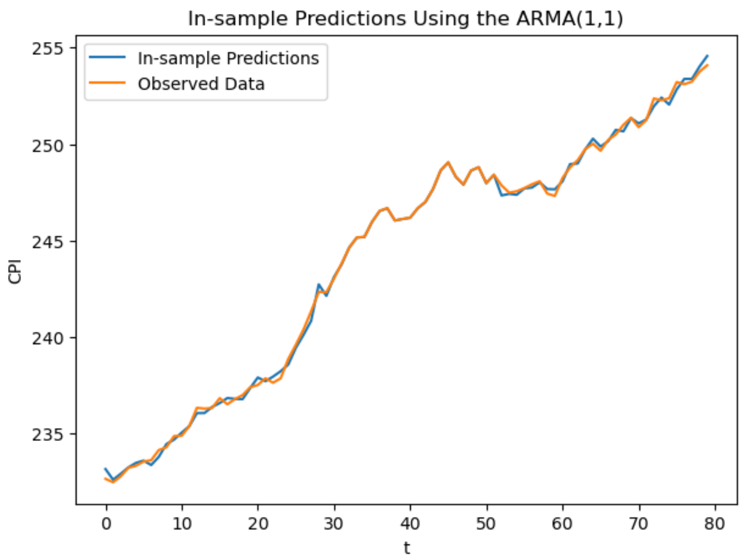
- Constant Term: The model estimates a constant term (const) of -0.0107, but this is not statistically significant (p-value: 0.810), suggesting it does not contribute meaningfully to the model.
- AR Coefficient (ar.L1): The coefficient value of 0.2427, with a p-value of 0.140, indicates a weak autoregressive effect that is not statistically significant.
- MA Coefficient (ma.L1): The moving average term has a coefficient of 0.4575, which is statistically significant (p-value: 0.006), suggesting that past forecast errors are useful in predicting the current value.

Variance (sigma2):
The estimated variance of the residuals is 0.0455, indicating the average magnitude of the fluctuations around the mean of the model.

Diagnostic Tests:

- Ljung-Box Test: The p-value of 0.81 for the Ljung-Box test indicates that there is no significant autocorrelation in the residuals, which is desirable.
- Jarque-Bera Test: With a p-value of 0.93, the Jarque-Bera test suggests that the residuals are normally distributed.
- Heteroskedasticity Test: The p-value of 0.92 implies no heteroskedasticity, meaning the variance of the residuals does not change over time.

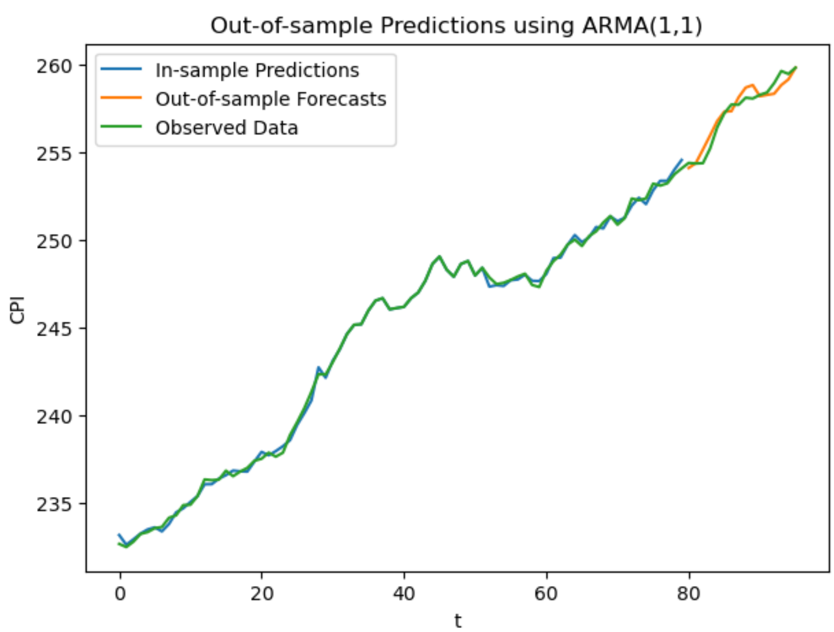
7.1. In sample CPI Forecasting



The ARMA(1,1) model has demonstrated its effectiveness in modeling the CPI data, as evidenced by the close match between the predictions and the observed values. This supports the model's suitability for short-term forecasting and its potential application in economic planning and policy formulation. Further validation on out-of-sample data is recommended to confirm the model's predictive power.

1. In-sample (Linear) RMSE: 0.209702639258
2. In-sample (Linear) MAE: 0.152691230205
3. In-sample (Linear) MSE: 0.043975196911

7.2. Out-of-sample CPI Forecasting



- Consistency: The out-of-sample forecasts continue to track closely with the actual observed CPI data, indicating that the model generalizes well beyond the training dataset.
- Forecast Accuracy: The alignment of the forecasted values with the observed data suggests a high level of accuracy, with the model capturing the key trends in the CPI data effectively.
- Model Validation: The similarity between in-sample predictions and out-of-sample forecasts validates the model's stability and predictive reliability.

1. Out-of-sample (Linear) RMSE: 0.485205840
2. Out-of-sample (Linear) MAE: 0.398817482
3. Out-of-sample (Linear) MSE: 0.23542470

The performance of the ARMA(1,1) model in out-of-sample predictions is a critical indicator of its utility in practical scenarios. The close correspondence between the forecasted and observed data demonstrates the model's effectiveness in anticipating future movements of the CPI. This affirms the model's potential use in economic analysis and policy-making, with the provision that ongoing model assessments are conducted to ensure its continued accuracy over time.

8. Improvement 1 - Change Detrending technique and perform the forecasting

8.1. Concept of Detrending with One-Sided Moving Average

Detrending using a one-sided moving average involves smoothing the data by removing variations around the trend. This method calculates the trend component by averaging a set number of past observations. The resulting smoothed values reveal the underlying trend by minimizing short-term fluctuations.

Observance:

- Smoothness: The moving average trend appears to be smoother than the linear regression, which reflects short-term variations more closely.
- Adaptability: The moving average adapts to changes in the trend more flexibly than the linear regression, which assumes a constant linear relationship over time.

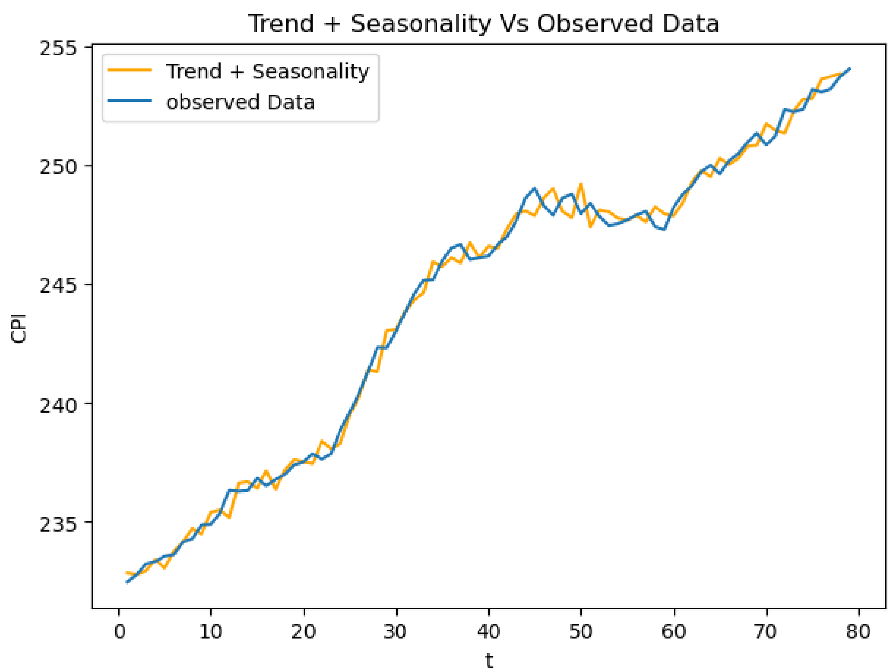
When comparing detrending methods, the one-sided moving average offers a smoother representation of the underlying trend and can adapt to changes more dynamically. However, it may not reflect rapid changes in trend as effectively as linear regression.

8.2. Integrating Trend and Seasonality in CPI Analysis

In this phase of analysis, we aimed to reconstruct the Consumer Price Index (CPI) data by combining the trend component with seasonal fluctuations. The goal was to assess how well the combined model aligns with the observed data.

Methodology:

- Trend: We had previously extracted a trend using a one-sided moving average, which smooths out short-term volatility to reveal the underlying direction of the CPI data.
- Seasonality: We then reintegrated the seasonal component, which captures regular, periodic fluctuations in the data, often related to factors like consumer behavior, fiscal policy changes, or other cyclical events.
- Combination: The trend and seasonal components were summed to form a composite series representing both long-term movements and recurring patterns.



Insights:

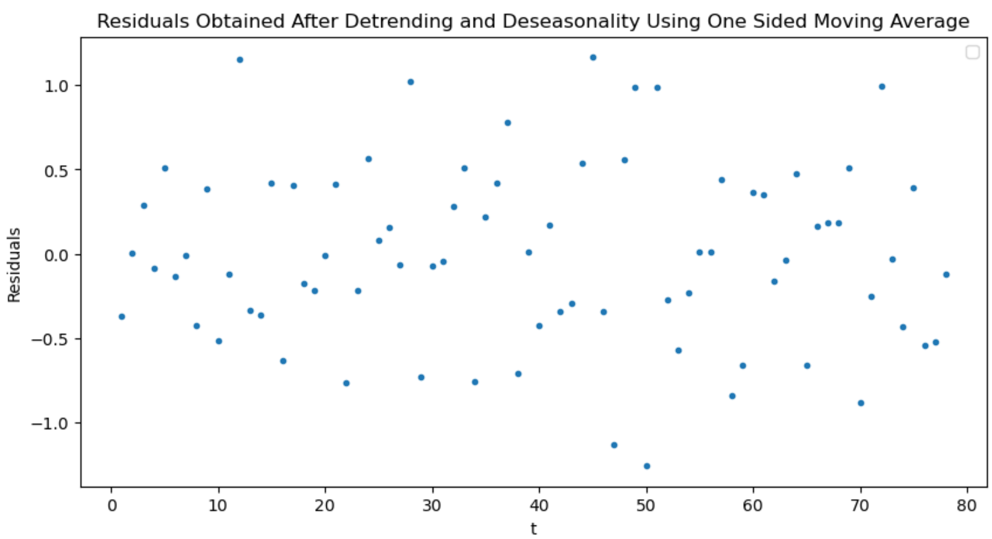
- Alignment with Observed Data: The composite series closely tracks the observed CPI, indicating that the combination of trend and seasonal components effectively captures the data's behavior.
- Model Accuracy: The close match between the composite series and the observed data suggests that the model accurately represents both the gradual changes and the regular seasonal variations in the CPI.
- Implications for Forecasting: The ability to match the observed data with the combined trend and seasonal model suggests it can be a reliable approach for forecasting future CPI values, assuming the trend and seasonal patterns continue.

The analysis demonstrates that accounting for both trend and seasonality provides a comprehensive understanding of the CPI's dynamics. The combined model offers a nuanced view of the data, essential for accurate economic analysis and forecasting.

8.3. Analysis of Residuals After Detrending and Deseasonalizing

The purpose of this analysis was to evaluate the residuals — the differences between the observed CPI data and the combined model of trend and seasonality, derived using a one-sided moving average.

We removed both the trend and seasonal components from the CPI data. This process aims to isolate the irregular or random components that are not explained by the systematic patterns of trend and seasonality. The residuals represent the leftover variations in the data after accounting for the predictable components.



The plot of residuals indicates that the detrending and deseasonalizing process, using a one-sided moving average, has effectively standardized the CPI data, leaving behind what appears to be random noise. This suggests that the systematic components of the data — its trend and seasonality — have been accurately modeled and removed. As a result, any remaining structure in the residuals could be the target of further investigation for anomalies or non-standard patterns.

8.4. Test for Stationarity

Initial Test Using Linear Regression Detrending:

- Test Statistic: -3.683, indicating stationarity.
- p-value: 0.0043, confirming the low likelihood of non-stationarity.
- Conclusion: The data was considered stationary, suitable for analysis and forecasting.

Subsequent Test Using One-Sided Moving Average Detrending:

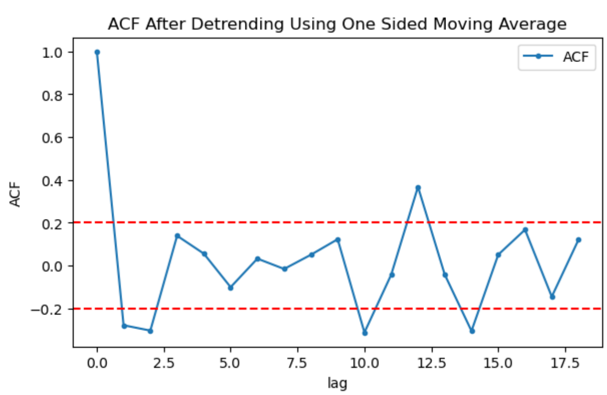
- Test Statistic: -10.64734, significantly more negative, suggesting stronger evidence of stationarity.
- p-value: Approximately 4.76e-19, an extremely low probability of non-stationarity.
- Lags Used: Reduced to 1, implying less complexity in achieving stationarity.

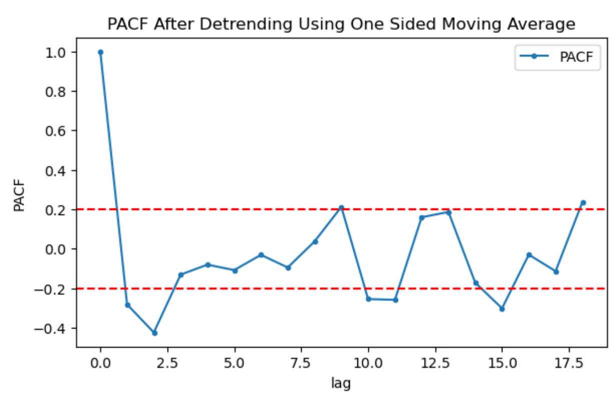
Comparison:

The ADF test following the one-sided moving average detrending provides a more decisive indication of stationarity compared to the linear regression approach. The improved test statistic and reduced p-value suggest that the one-sided moving average method may be more effective in removing trends and seasonal effects from the series. The practical implication is a higher confidence in using the detrended series for further modeling and forecasting tasks.

In summary, the comparison of ADF test results before and after applying different detrending methods demonstrates a substantial improvement in the stationarity of the time series, enhancing the reliability of subsequent analytical steps.

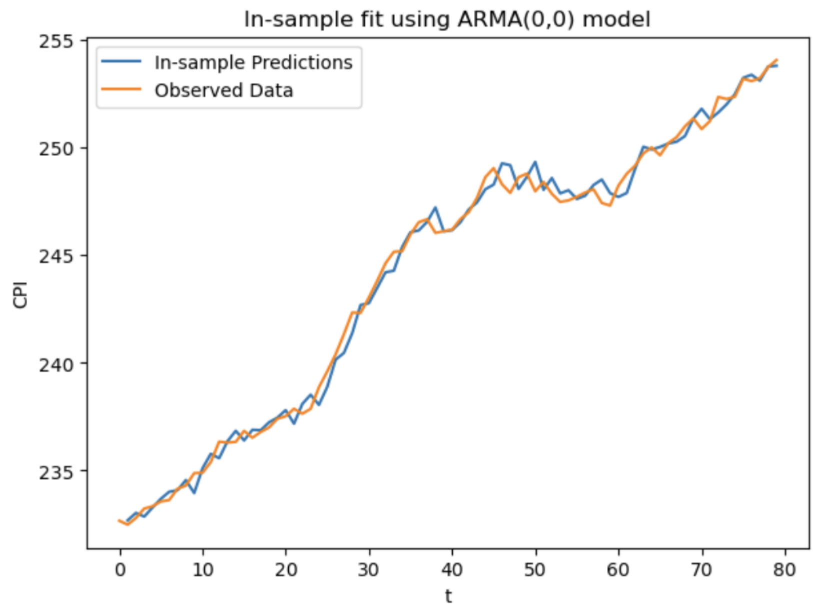
8.5. ACF and PACF Analysis Post-Detrending (One sided moving average)



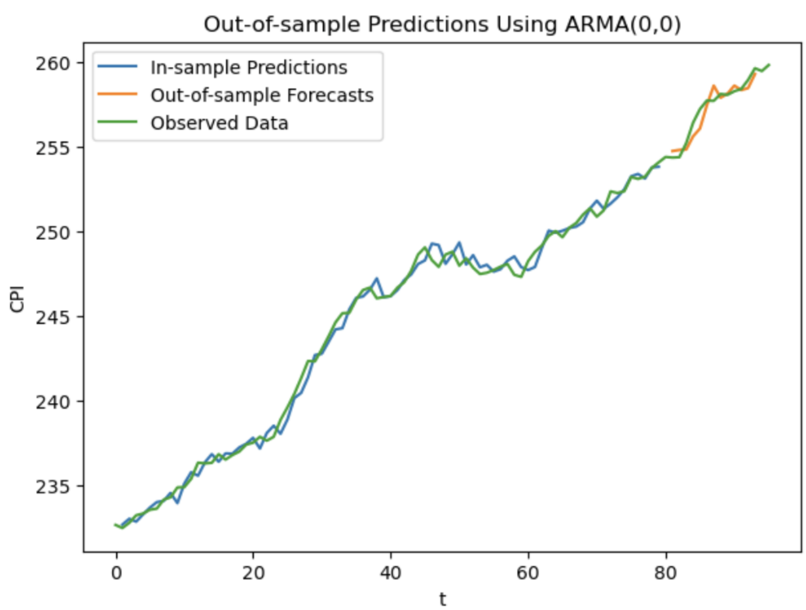


We chose the **ARMA(0,0)** model.

8.6. Prediction



1. In-sample (One Sided Moving Average) RMSE: 0.55686
2. In-sample (One Sided Moving Average) MAE: 0.4390486
3. In-sample (One Sided Moving Average) MSE: 0.3100937

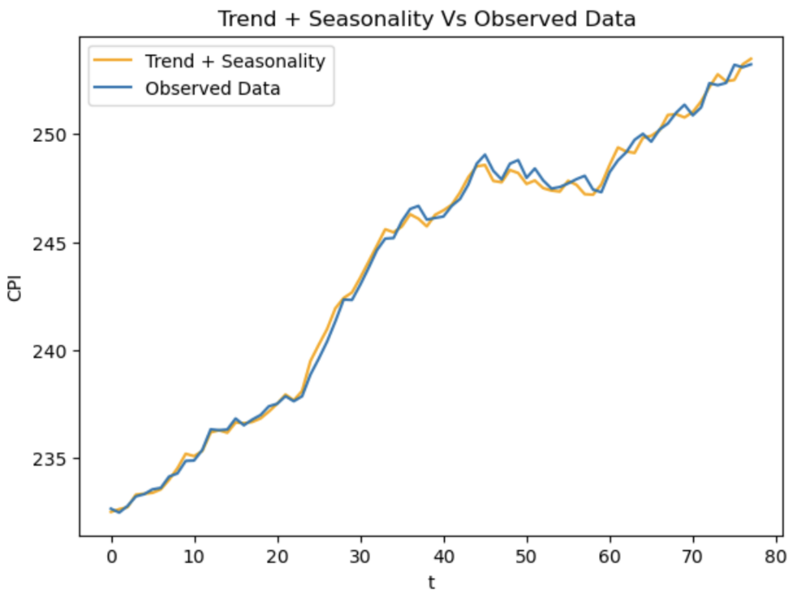


1. Out-of-sample (One Sided Moving Average) RMSE: 0.520398
2. Out-of-sample (One Sided Moving Average) MAE: 0.3986358
3. Out-of-sample (One Sided Moving Average) MSE: 0.2708148

9. Improvement 2 - Right sided moving average

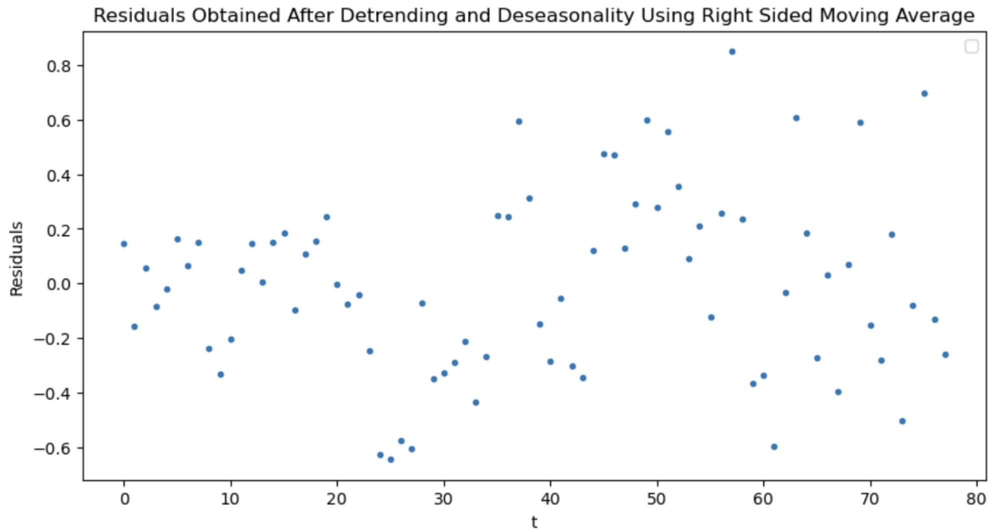
This approach uses a window of 3 data points to calculate a forward-looking moving average for the Consumer Price Index (CPI).Instead of using past data, this method averages each data point with the next two future values, offering a unique 'right-sided' perspective. The original CPI data is adjusted by subtracting this right-sided moving average, effectively removing the trend that includes near-future information. This method allows for retrospective analysis where all data points are known.This technique can reveal patterns and anomalies not easily identified through traditional detrending methods.

9.1. Assessment of Trend and Seasonality



The closely aligned trends of the 'Trend + Seasonality' line with the 'Observed Data' suggest that the model incorporating both trend and seasonal factors provides an accurate representation of the actual CPI movements. The reconstruction captures the overall directional movement and periodic fluctuations, confirming that the combination of these components reflects the true behavior of the CPI over time. This alignment indicates the effectiveness of the right-sided moving average method in modeling the CPI data for this specific period, reinforcing its potential use in economic analysis and forecasting.

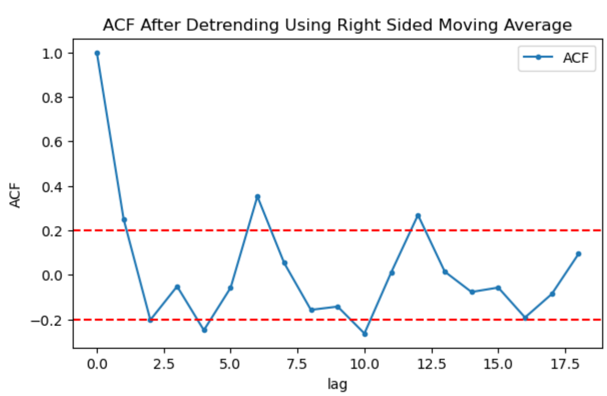
9.2. Analyzing Residuals Post Right-Sided Moving Average Detrending

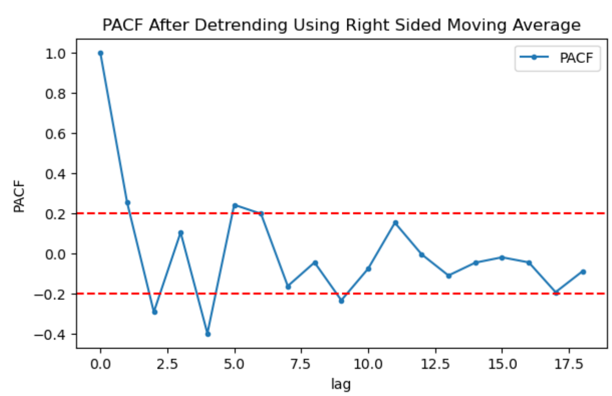


The end portions of the plot show randomness, indicating the trend was captured and removed effectively. The straight line formed by data points in the middle suggests the method didn't fully remove all patterns.

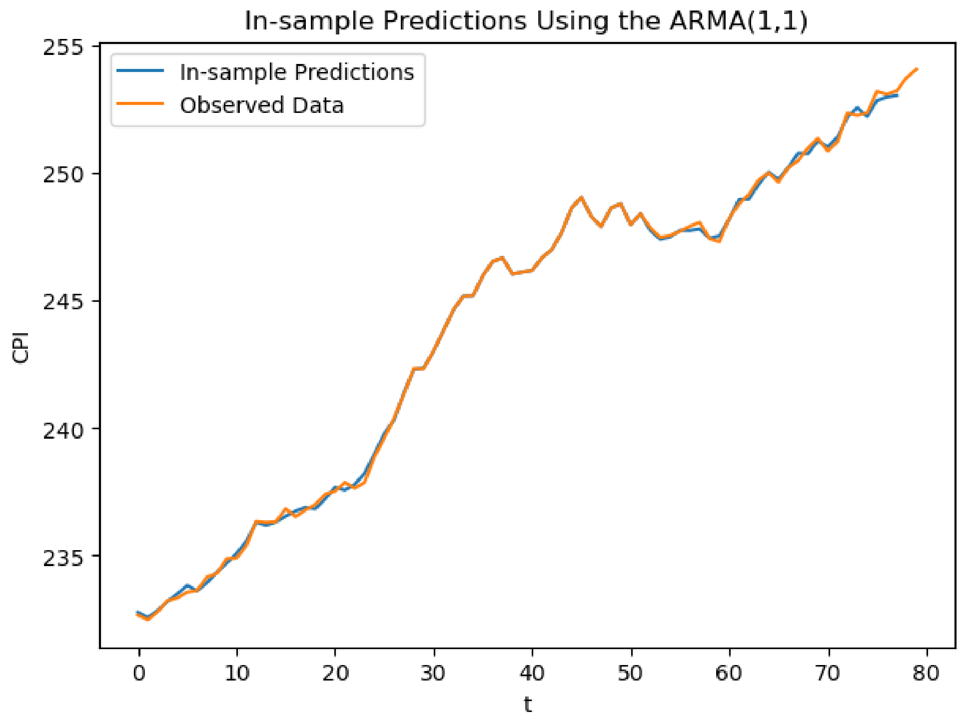
9.3. Choosing ARMA model using ACF/PACF Plots

The scatter plot of residuals, after applying a right-sided moving average for detrending and deseasonalization, indicates an absence of any systematic patterns, suggesting effective isolation of the CPI data's core fluctuations. The strong initial spike in both plots leads us to choose an **ARMA(1,1)** model. This means the model considers one past data point and the error from the previous forecast in predicting the next value.

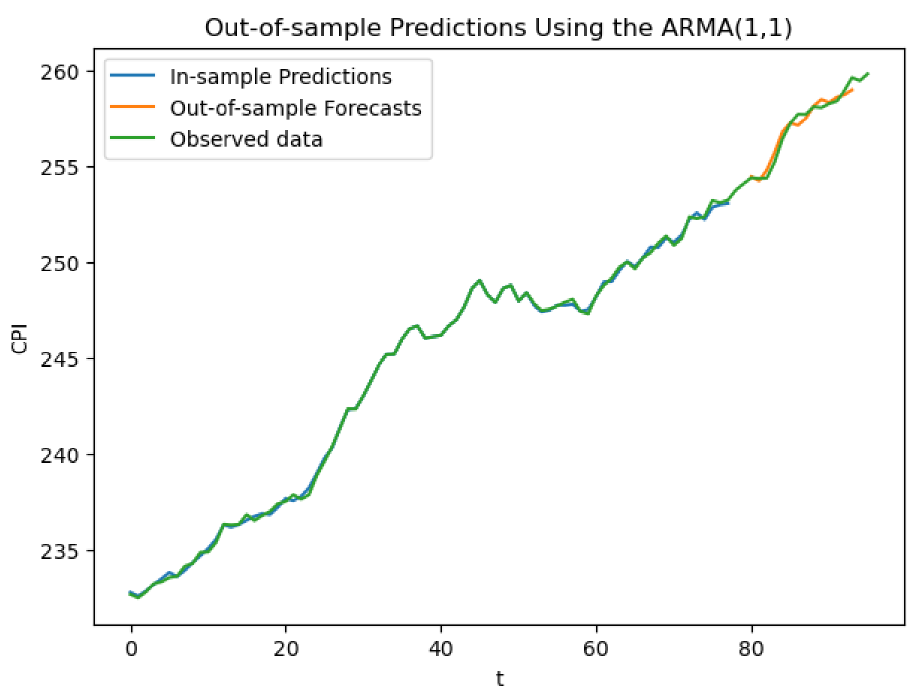




9.4. Predictions



1. In-sample (Right Sided Moving Average) RMSE: 0.1431522329786891
2. In-sample (Right Sided Moving Average) MAE: 0.10351958159065461
3. In-sample (Right Sided Moving Average) MSE: 0.020492561806784885



1. Out-of-sample (Right Sided Moving Average) RMSE: 0.3451876575599079
2. Out-of-sample (Right Sided Moving Average) MAE: 0.2784759450885933
3. Out-of-sample (Right Sided Moving Average) MSE: 0.11915451893169623

10. Conclusion

As the results of three different ways of detrended data, the forecasts obtained by detrending using the right sided moving average are better than those obtained using one sided moving average or linear regression.

IN-SAMPLE

1. In-sample (Linear) RMSE: 0.20970263925812888
2. In-sample (Linear) MAE: 0.15269123020538716
3. In-sample (Linear) MSE: 0.043975196911824936

1. In-sample (One Sided Moving Average) RMSE: 0.558847548702529
2. In-sample (One Sided Moving Average) MAE: 0.43420533545333466
3. In-sample (One Sided Moving Average) MSE: 0.310002976867383

4. In-sample (Right Sided Moving Average) RMSE: 0.1431522329786891
5. In-sample (Right Sided Moving Average) MAE: 0.10351958159065461
6. In-sample (Right Sided Moving Average) MSE: 0.020492561806784885

OUT-OF-SAMPLE

1. Out-of-sample (Linear) RMSE: 0.4852058405790021
2. Out-of-sample (Linear) MAE: 0.39881748227798575
3. Out-of-sample (Linear) MSE: 0.23542470773197602

1. Out-of-sample (One Sided Moving Average) RMSE: 0.52031704450272
2. Out-of-sample (One Sided Moving Average) MAE: 0.3982384941096482
3. Out-of-sample (One Sided Moving Average) MSE: 0.2708023669873176

4. Out-of-sample (Right Sided Moving Average) RMSE: 0.3451876575599079
5. Out-of-sample (Right Sided Moving Average) MAE: 0.2784759450885933
6. Out-of-sample (Right Sided Moving Average) MSE: 0.11915451893169623

Here is a comprehensive approach taken to analyze the Consumer Price Index (CPI) data-

1. Detrend the CPI Data: Apply linear regression analysis to remove the trend, isolating the underlying pattern of the Consumer Price Index.
2. Deseasonalize the Data: Calculate and remove seasonal effects to focus on the non-seasonal fluctuations.

3. Analyze Residuals:Evaluate the residuals obtained after detrending and deseasonalizing to ensure that only random noise remains.
4. Conduct Stationarity Testing:Use the Dickey-Fuller test to confirm that the residual series is stationary, indicating suitability for further analysis.
5. Select an ARMA Model: Based on ACF and PACF plots, identify the appropriate Autoregressive Moving Average (ARMA) model parameters.
6. Train the ARMA Model:Fit the ARMA model to the detrended and deseasonalized data and assess its in-sample forecasting accuracy.
7. Perform Out-of-Sample Forecasting:Generate and evaluate out-of-sample CPI forecasts to test the model's generalizability.
8. Improve with Alternative Detrending: Implement a one-sided moving average approach for a different detrending method and forecast based on this model.
9. Integrate Trend and Seasonality:Reconstruct the CPI data by combining the trend with seasonal components and compare with the observed data for validation.
10. Refine with Right-Sided Moving Average:Explore the right-sided moving average as an alternative detrending technique and assess its forecasting performance.
11. Compare the forecasts from linear regression, one-sided, and right-sided moving average detrending methods to conclude the most effective approach for CPI analysis and forecasting.