Consumer Price Index (CPI) Analysis: Trends and Forecasting

NIDHIN RAJ MOTABOINA DEEPAK
364745

ABSTRACT

This project focuses on analyzing the Consumer Price Index (CPI) data from January 2011 to July 2021 to identify trends and seasonal patterns. The CPI is a critical measure of inflation, reflecting price changes in goods and services over time. By forecasting future CPI values, the project aims to help businesses, policymakers, and financial institutions make informed decisions. The data analysis revealed clear trends and repeating seasonal patterns. Two forecasting methods, Seasonal ARIMA (SARIMA) and Holt-Winters Exponential Smoothing, were used to model the data. SARIMA effectively captures patterns, while Holt-Winters is useful for smoothing and predicting. These methods demonstrated their capability to forecast CPI accurately. The results highlight the importance of leveraging data analysis for economic planning. This study provides valuable insights for addressing inflation and improving decision-making.

Problem Statement

Inflation, as measured by the Consumer Price Index (CPI), is one of the most critical indicators of economic health. However, understanding its trends and accurately forecasting future values remains a challenge for policymakers, businesses, and financial institutions. Fluctuations in CPI can have profound implications on purchasing power, investment decisions, and economic stability. This project seeks to address this challenge by identifying and analyzing patterns in CPI data to develop reliable forecasting models. These models aim to enhance decision-making processes and provide actionable insights for managing economic risks associated with inflation

Motivation

The motivation for this project lies in the critical role of the Consumer Price Index (CPI) as a measure of inflation and its impact on economic planning. Accurate forecasting of CPI is essential for businesses, policymakers, and financial institutions to make informed decisions, manage risks, and plan. Inflation affects purchasing power, investment strategies, and overall economic stability, making it vital to understand and predict its trends.

This project provides an opportunity to apply advanced data analysis techniques to real-world challenges. By identifying trends and seasonal patterns in CPI data, we can develop reliable forecasting models that enable better economic decision-making. Additionally, the project serves as a practical application of time series analysis and enhances our understanding of how forecasting tools can be used to address complex economic problems. This initiative also demonstrates the value of data-driven insights in improving financial and policy decisions.

Descriptive Statistics

A summary of key descriptive statistics for the Consumer Price Index (CPI) dataset is provided below:

- **Mean CPI**: 132.57 – A measure of the central tendency of CPI over the observed period.
- **Standard Deviation**: 18.43 – Indicates the variability in CPI values, showing how much CPI fluctuates over time.
- **Minimum CPI**: 100.00 – The lowest recorded CPI value during the dataset's time period.
- **Maximum CPI**: 164.20 – The highest recorded CPI value during the dataset's time period.
- **Range**: 64.20 – The difference between the maximum and minimum CPI values, highlighting the overall spread.
- **Skewness**: 0.00 – Indicates symmetry in the distribution of CPI values.
- **Kurtosis**: -1.20 – Suggests a relatively flat distribution of CPI values, with fewer extreme values.

Mean CPI,132.5724409448819
Standard Deviation,18.43340083155365
Minimum CPI,100.0
Maximum CPI,164.2
Range,64.19999999999999
Skewness,-0.00897464462259174
Kurtosis,-1.1916355560841216

About the Data

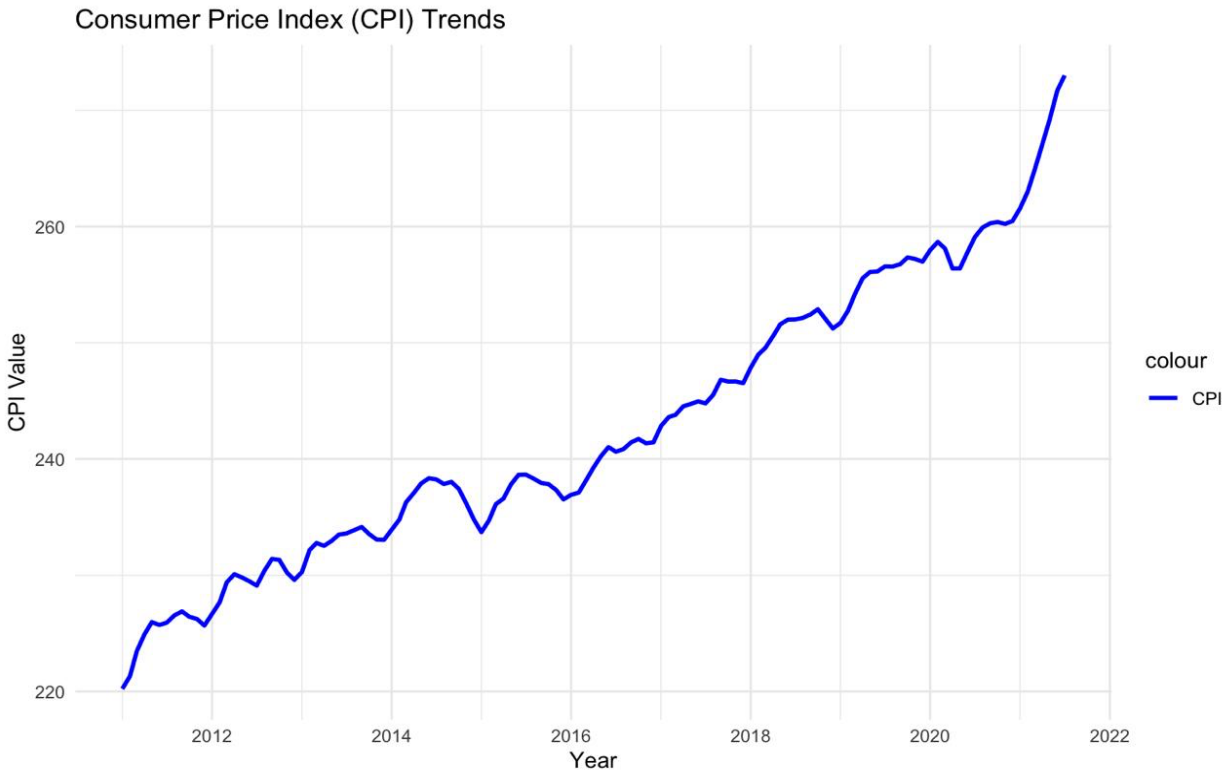
- **Source**: Kaggle
- **Time Period**: January 2011 to July 2021
- **Number of Observations**: 127

Variables in dataset

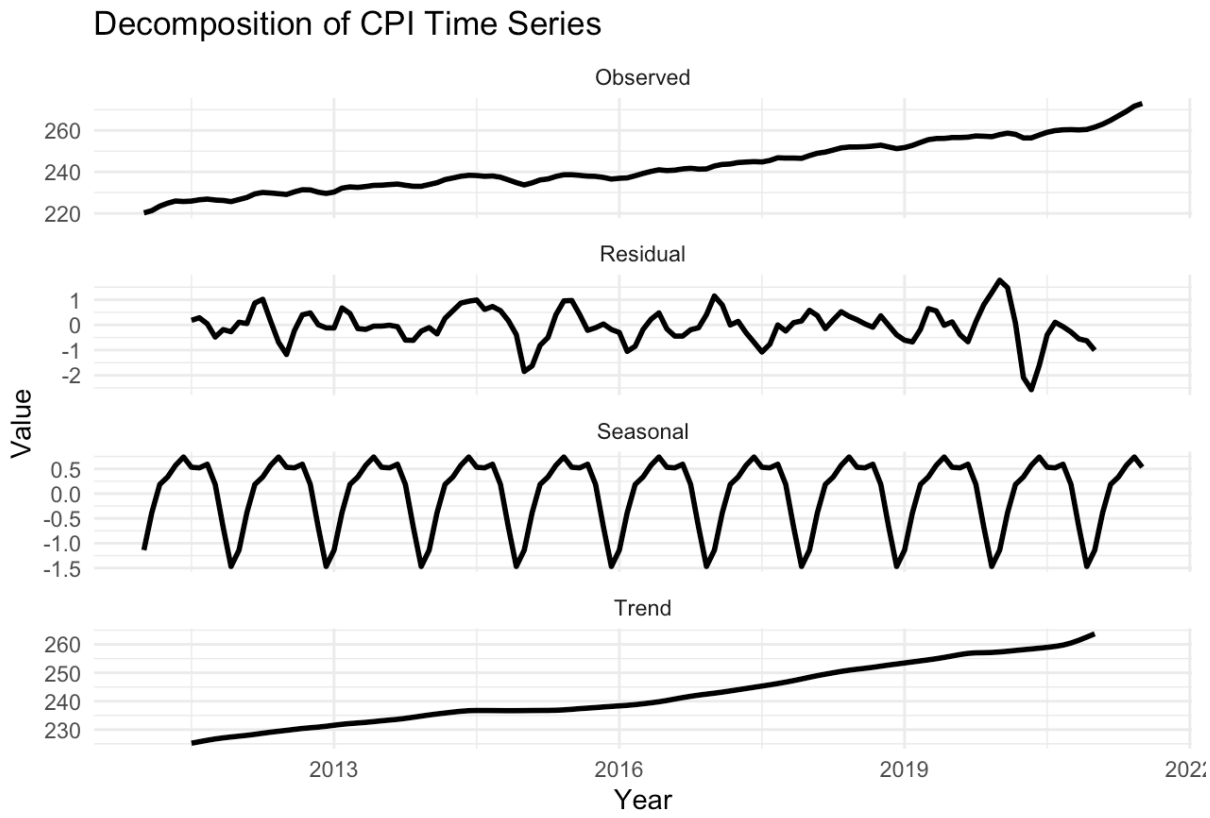
- **Yearmon**: Represents the month and year of the data recorded (e.g., Jan 2011, Feb 2011), indicating a monthly frequency.
- **CPI**: The Consumer Price Index, a continuous variable representing the average price level of a predefined basket of goods and services, used to measure inflation.

CPI Time Series Analysis

The provided image showcases a graphical representation of the Consumer Price Index (CPI) progression from 2011 to 2021. Time is plotted along the X-axis, indicating monthly data spanning several years, while the Y-axis reflects the CPI values. The line graph illustrates the CPI dynamics throughout this period, displaying an overall upward trajectory interspersed with minor fluctuations. A consistent increase in CPI values indicates long-term inflationary pressures in the economy. Short-term variations reflect temporary changes in inflation rates. This type of visualization is a vital tool for economists and policymakers, providing a clear depiction of inflation trends and price level changes over time. Such insights are essential for developing effective forecasting models, crafting economic strategies, and making informed decisions about inflation management..



Decomposition of CPI Time Series

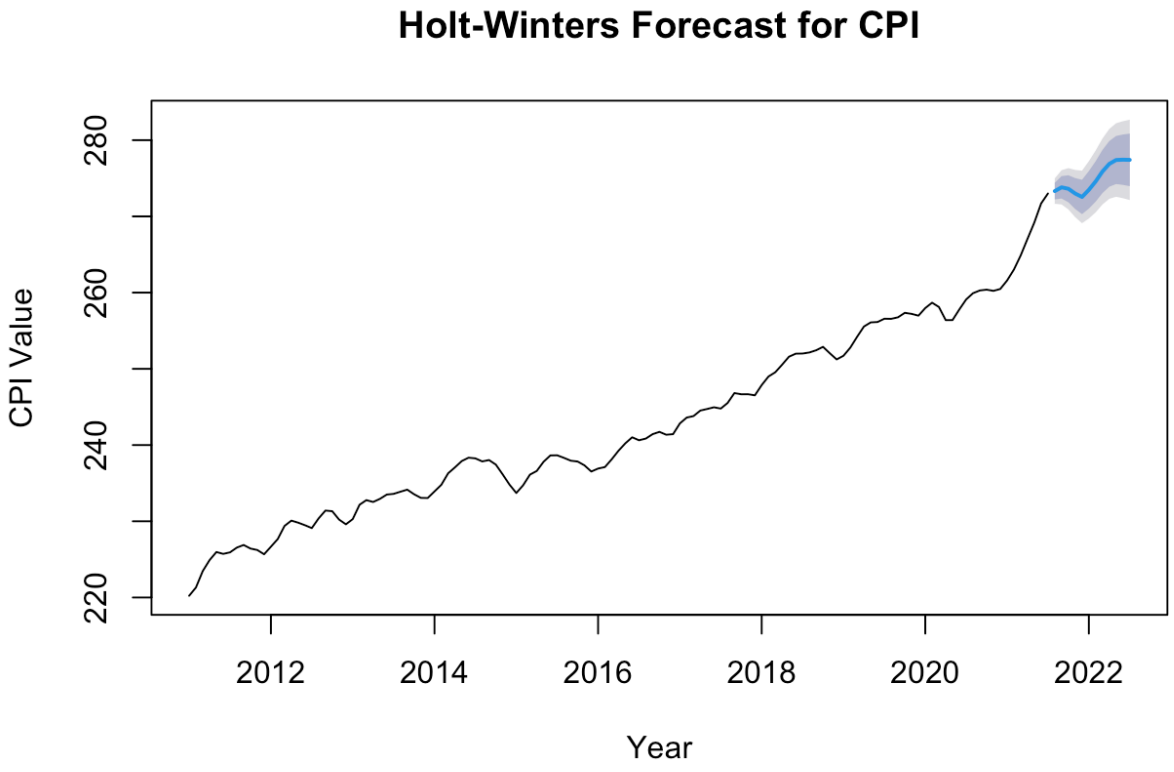


The decomposition of the CPI time series is presented in four components:

1. **Observed**: The top graph shows the actual CPI values over time, revealing a steady upward trend with some fluctuations.
2. **Residual**: The second graph depicts the random variations or noise in the data, which are unexplained by the trend and seasonal components. These residuals help in identifying randomness in the time series.
3. **Seasonal**: The third graph highlights the repeating seasonal patterns in the data. These patterns are consistent over the years, showing predictable cyclical behavior.
4. **Trend**: The bottom graph captures the long-term movement in the CPI, which shows a clear upward trend, reflecting a continuous rise in price levels over the observed period.

Holt-Winters Smoothing

The graph showcases the "Holt-Winters Forecast for CPI," spanning from 2011 to 2022. The vertical axis represents CPI values, while the horizontal axis shows years. Up to 2021, the chart displays actual historical CPI data, which follows an upward trend. After 2021, the forecasted data appears, showing a continuation of the increasing trend with slight seasonal adjustments. The forecast is accompanied by shaded regions representing confidence intervals, with darker areas indicating higher certainty and lighter shades reflecting growing uncertainty further into the future. The method captures both trend and seasonality effectively, projecting CPI values based on observed patterns. The widening confidence intervals highlight the uncertainty inherent in long-term forecasting. Holt-Winters smoothing is particularly effective for time series data with a clear trend and recurring seasonal patterns, making it a suitable choice for CPI forecasting and economic analysis.



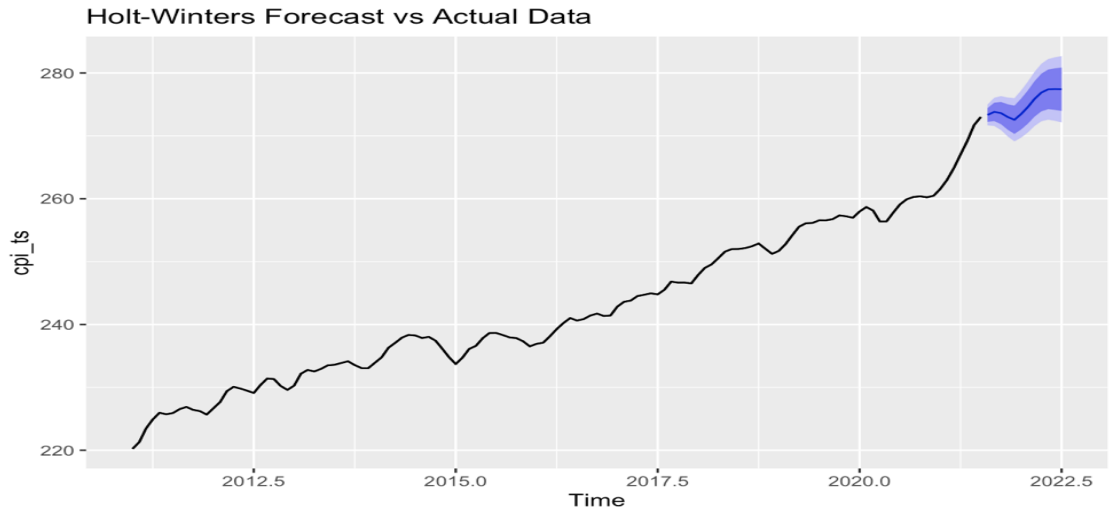
Holt-Winters Forecast Accuracy

The accuracy metrics for the Holt-Winters model demonstrate its strong performance in CPI forecasting. The Mean Error (ME) is 0.0495, indicating minimal bias in the predictions, while the Root Mean Square Error (RMSE) of 0.8453 reflects low error variance. The Mean Absolute Error (MAE) is 0.6268, showing small average deviations from actual CPI values, and the Mean Percentage Error (MPE) of 0.0165, along with the Mean Absolute Percentage Error (MAPE) of 0.2555, confirms the model's strong predictive accuracy. The Mean Absolute Scaled Error (MASE) of 0.1519 highlights its superior performance compared to a baseline model. However, the Autocorrelation of Residuals at Lag 1 (ACF1) of 0.4251 suggests that some residual patterns remain, leaving room for refinement. Overall, the Holt-Winters model provides reliable and accurate CPI forecasts, making it a valuable tool for inflation prediction and economic planning.

```
[1] "Holt-Winters Accuracy:"
> print(hw_accuracy)
                    ME     RMSE      MAE       MPE     MAPE
Training set 0.04945247 0.8452677 0.6268348 0.01645409 0.2554717
                  MASE      ACF1
Training set 0.1518777 0.4251238
```

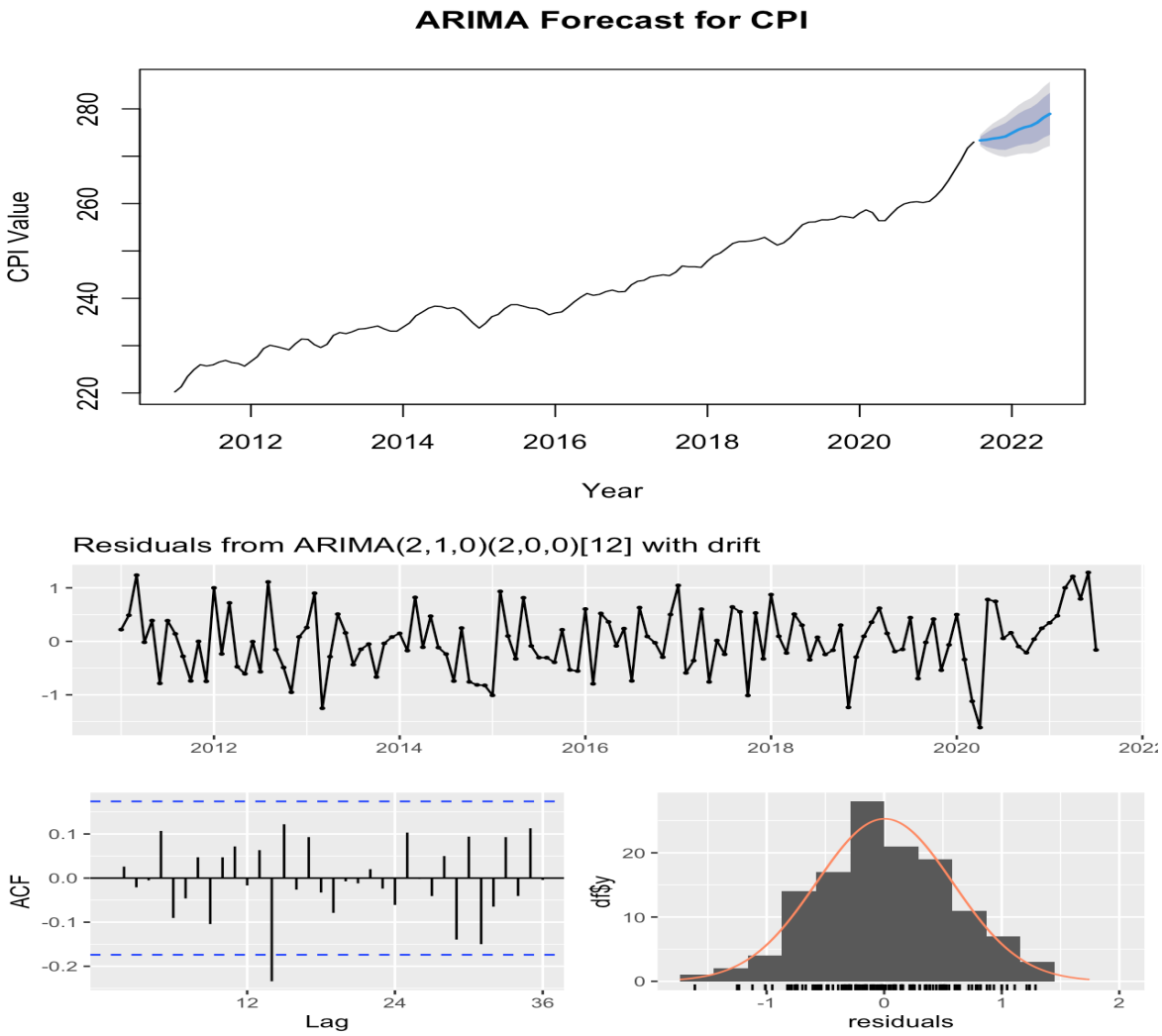

```
[1] "Holt-Winters Accuracy:"
> print(hw_accuracy)
                    ME      RMSE       MAE         MPE      MAPE
Training set 0.04945247 0.8452677 0.6268348 0.01645409 0.2554717
                  MASE      ACF1
Training set 0.1518777 0.4251238
```


Holt-Winters Forecast vs. Actual CPI Data



ARIMA Forecast for CPI

This graph illustrates the ARIMA forecast for the Consumer Price Index (CPI) extending from 2021, with the vertical axis representing CPI values and the horizontal axis marking the progression of years. The actual CPI values up to 2021 show an upward trend, indicating consistent inflationary growth. From 2021 onward, predictions made by the ARIMA model are shown with a solid line, surrounded by shaded blue areas that represent confidence intervals. These intervals widen over time, reflecting increased uncertainty in the model's predictions as the forecast horizon extends. ARIMA (Autoregressive Integrated Moving Average) is a robust forecasting approach that leverages historical data, trends, and past errors to predict future values. The expanding confidence intervals underscore a fundamental forecasting principle: predictions become less certain as they project further into the future. This model is well-suited for identifying and forecasting patterns in CPI data for economic planning and decision-making.



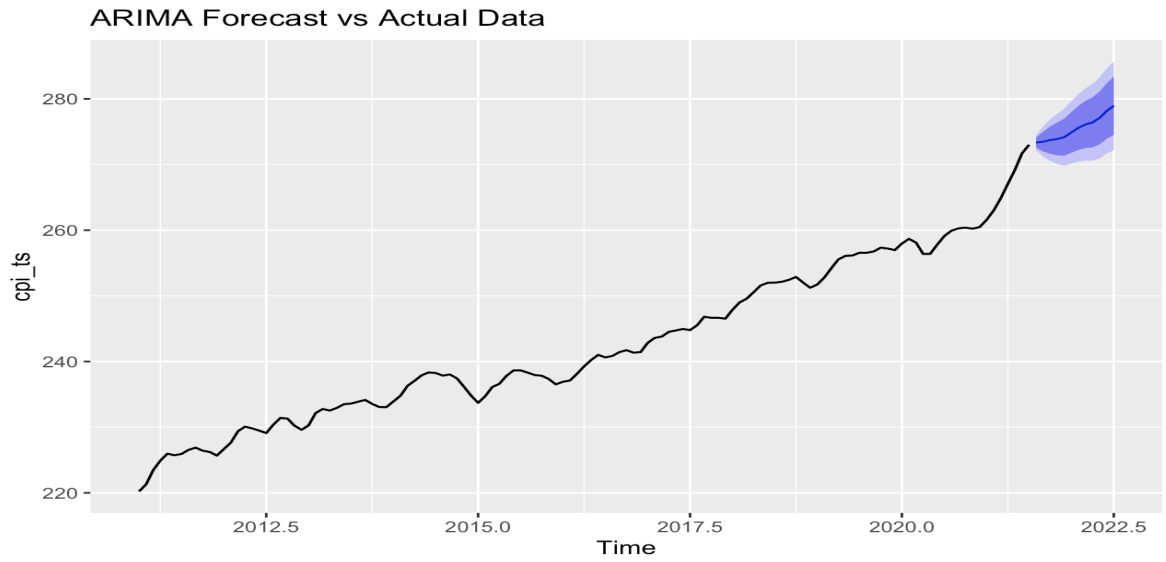
ARIMA Forecast Accuracy

The ARIMA model demonstrates strong accuracy in forecasting CPI values based on its evaluation metrics. With a Mean Error (ME) of 0.0038, the predictions exhibit minimal bias, while the Root Mean Square Error (RMSE) of 0.5783 indicates a low standard deviation of errors. The Mean Absolute Error (MAE) of 0.4631 reflects small average deviations, and the Mean Percentage Error (MPE) of 0.000075 confirms precise proportional predictions. The Mean Absolute Percentage Error (MAPE) of 0.1912 further highlights the model's strong performance. Additionally, the Mean Absolute Scaled Error (MASE) of 0.1122 demonstrates the ARIMA model's superiority compared to a naive baseline. The Autocorrelation of Residuals at Lag 1 (ACF1) of -0.0001 shows no significant residual patterns, indicating the model has effectively captured the underlying data structure. These metrics validate the ARIMA model as a reliable tool for CPI forecasting and economic

```
[1] "ARIMA Accuracy:"
> print(arima_accuracy)
                    ME     RMSE      MAE       MPE
Training set 0.003830309 0.5783087 0.4630612 7.527391e-05
                 MAPE     MASE       ACF1
Training set 0.1911921 0.1121965 -0.0001390504
```


```
[1] "ARIMA Accuracy:"
> print(arima_accuracy)
                    ME      RMSE       MAE           MPE
Training set 0.003830309 0.5783087 0.4630612 7.527391e-05
                  MAPE      MASE           ACF1
Training set 0.1911921 0.1121965 -0.0001390504
```


ARIMA Forecast vs. Actual CPI Data



Conclusion

The table below summarizes the performance metrics for the two forecasting models, ARIMA and Holt-Winters (ETS):

Model	AIC	RMSE	MAE	MAPE	ME	MASE	ACF1
ARIMA	3920.37	0.5783	0.4631	0.1919	0.0038	0.1122	-0.0001
Holt-Winters (ETS)	9552.543	0.8453	0.6268	0.2555	0.0495	0.1519	0.4251

Key Metrics Analysis:

1. **AIC (Akaike Information Criterion)**: ARIMA has a significantly lower AIC (3920.37) compared to Holt-Winters (9552.543), indicating ARIMA is better at balancing accuracy and complexity.
2. **RMSE (Root Mean Square Error)**: ARIMA achieves a lower RMSE (0.5783), meaning its predictions are closer to the actual values compared to Holt-Winters (0.8453).
3. **MAE (Mean Absolute Error)**: ARIMA (0.4631) has smaller average errors compared to Holt-Winters (0.6268), demonstrating more accurate forecasts.
4. **MAPE (Mean Absolute Percentage Error)**: ARIMA's MAPE (0.1919) is lower than Holt-Winters' (0.2555), confirming better forecast precision.
5. **ME (Mean Error)**: ARIMA's near-zero ME (0.0038) indicates less bias, compared to Holt-Winters (0.0495).
6. **MASE (Mean Absolute Scaled Error)**: ARIMA's lower MASE (0.1122) reflects better performance compared to Holt-Winters (0.1519).
7. **ACF1 (Autocorrelation of Residuals at Lag 1)**: ARIMA (-0.0001) shows no significant residual correlation, while Holt-Winters (0.4251) has higher autocorrelation, suggesting patterns in residuals remain.

The AIC and RMSE values are crucial for evaluating statistical models. Lower AIC values indicate that the model achieves a better balance between goodness of fit and complexity. RMSE, on the other hand, measures the average magnitude of prediction errors, with smaller values reflecting higher accuracy.

MAPE quantifies the average absolute discrepancy between predicted and actual CPI values in percentage terms, offering a measure of predictive precision. Lower MAPE values suggest superior model performance in forecasting outcomes.

Based on these indicators, the ARIMA model emerges as the superior option for CPI forecasting, as demonstrated by its significantly lower AIC and RMSE values compared to the Holt-Winters model. While both models capture the general trend and seasonality effectively, ARIMA outperforms in terms of precision and reliability, making it the preferred choice for applications requiring accurate economic predictions. Analysts and policymakers can rely on the ARIMA model to make informed decisions regarding inflation and economic planning.