School of Social Sciences and Philosophy
Assignment Submission Form

Student Name:	**Athena Rodrigues**
Student ID Number:	**24332991**
Programme Title:	**Applied Social Data Science**
Module Title:	**POP77014 - Social Forecasting**
Assessment Title:	**Final Assignment**
Lecturer (s):	**Thomas Chadefaux**
Date Submitted	**26 April 2025**

I have read and I understand the plagiarism provisions in the General Regulations of the University Calendar for the current year, found at: http://www.tcd.ie/calendar

I have also completed the Online Tutorial on avoiding plagiarism 'Ready, Steady, Write', located at http://tcd-ie.libguides.com/plagiarism/ready-steady-write

Signed: _________ABR________________

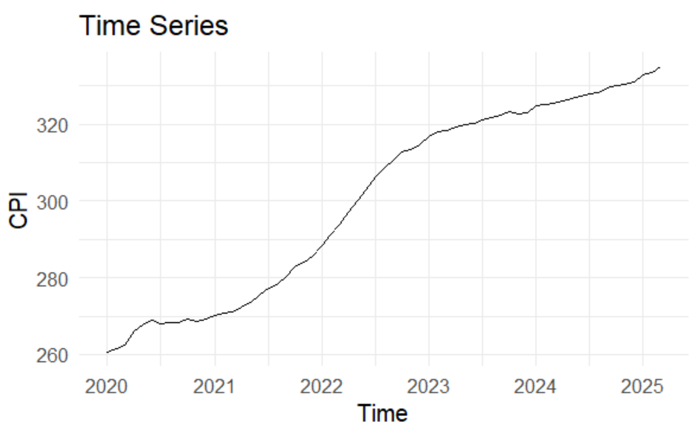
Date: ___________26/04/2______________

Introduction

This report analyses the Food and Beverages Consumer Price Index (CPI) for U.S. urban consumers, as reported by the Bureau of Labour Statistics. The index reflects monthly, non-seasonally adjusted price changes over time, offering insight into inflation trends and pricing dynamics within the food industry. Understanding and forecasting these trends in CPI is critical for both industry stakeholders and policymakers, informing decisions related to pricing strategies, subsidies, and broader economic planning. Accurate CPI forecasts can aid companies in inventory management, product pricing, and cost forecasting for business expenses. This analysis uses ARIMA modelling to generate a 15-month forecast based on CPI data from January 2020 to March 2025. By identifying long-term trends, evaluating seasonality, and assessing model performance, this report aims to provide actionable insights for forward-looking decision-making.

Data Preparation

Figure 1



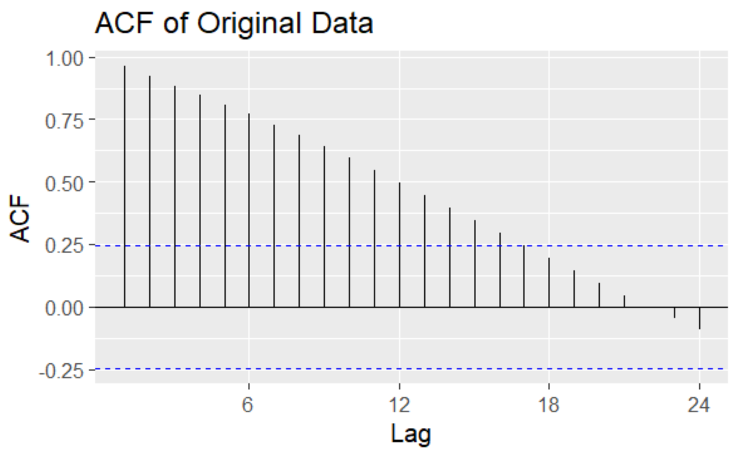
Visualising the time series data (Figure 1) helps identify patterns such as trends, potential seasonality and outliers. The plot reveals a consistent upward trend in the Food and Beverage CPI, reflecting ongoing inflationary pressures. Although some short-term fluctuations are present, no clear seasonal pattern emerges—an observation supported by the autocorrelation function (ACF) analysis. Given these characteristics, an ARIMA (AutoRegressive Integrated Moving Average) model is appropriate, as it effectively captures trends in non-seasonal, non-stationary time series. ARIMA combines autoregressive terms, moving averages, and differencing to model the underlying structure and generate forecasts based on historical values and residuals. This is ideal for datasets where there is a clear trend but no strong seasonal components, as the food and beverage price index shows.

Model Fitting

The ARIMA model was chosen to capture underlying trends in the Food and Beverages price index without introducing explicit seasonal components, which would require a SARIMA approach. Since the data from the Bureau of Labour Statistics is not seasonally adjusted, no additional seasonal transformations were necessary. For modelling, the data was split into training and testing sets: the training set included 80% of the data (January 2020 to December 2023), while the testing set comprised the remaining 20% (January 2024 to March 2025).

To assess seasonality and stationarity, an AutoCorrelation Function (ACF) plot (Figure 2) was generated, which shows the correlation of the series with its past values at various lags. The gradual decay in autocorrelations confirmed non-stationarity, consistent with the observed trend. The absence of strong periodic spikes further supported the classification of the series as non-seasonal and the use of ARIMA.

Figure 2



An ARIMA model was fitted using auto.arima(), which selects optimal parameters based on the training data. The best-fitting model was ARIMA(1,1,0)(1,0,0)[12] with drift.

The corresponding equation is:

$$y_t = 1.1417 + 0.6992y_{t-1} + 0.3669\epsilon_{t-12} + \epsilon_t$$

This equation indicates that the current CPI value depends on the previous month's value (y_{t-1}), a seasonal error component from the same month in the previous year ($\in_{t-12}$), and the current error term ($\in_t$), along with a drift term reflecting the overall upward trend. This structure captures the series' non-stationarity while accommodating any mild residual seasonality.

Model Evaluation

The model's performance was assessed using two metrics: Mean Absolute Percentage Error (MAPE) and Mean Absolute Error (MAE). MAPE measures the average percentage difference between predicted and actual values, while MAE quantifies the average absolute difference. The in-sample MAE of 0.6002 and MAPE of 21% suggest that the model has captured the trend reasonably well, with relatively small errors in both absolute and percentage terms. An examination of the residuals can show if there is patterning that the model has not fully captured and should be addressed in further tuning.

Residual analysis further supports the model's validity. As shown in Figure 3, the residuals fluctuate randomly around zero with no discernible pattern, indicating that the ARIMA model has effectively accounted for the underlying structure in the training data. Figure 4 compares the fitted values to the actual observations in the training set. The ARIMA model's line (in red) closely follows the observed CPI, demonstrating strong in-sample fit and accurate trend representation.

However, in-sample accuracy does not ensure reliable forecasting. The next step involves comparing the model's forecasts to the test set to assess predictive performance.

Figure 3

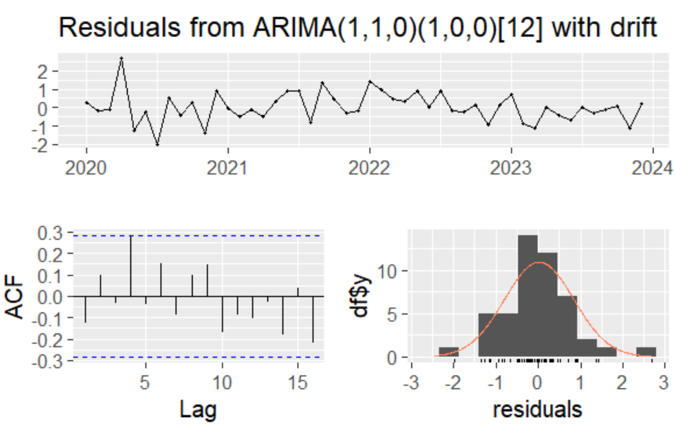
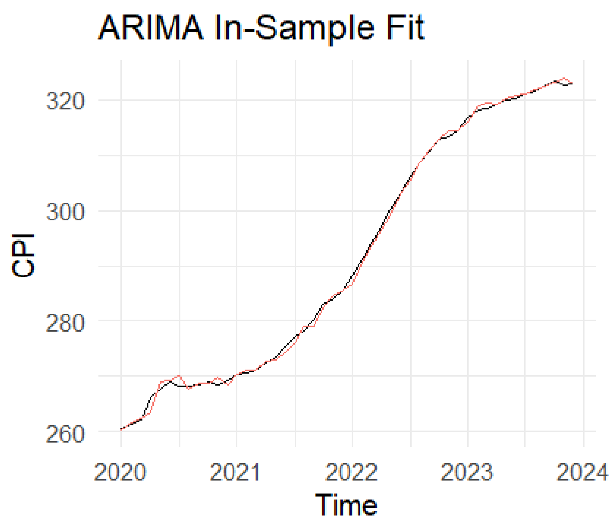


Figure 4

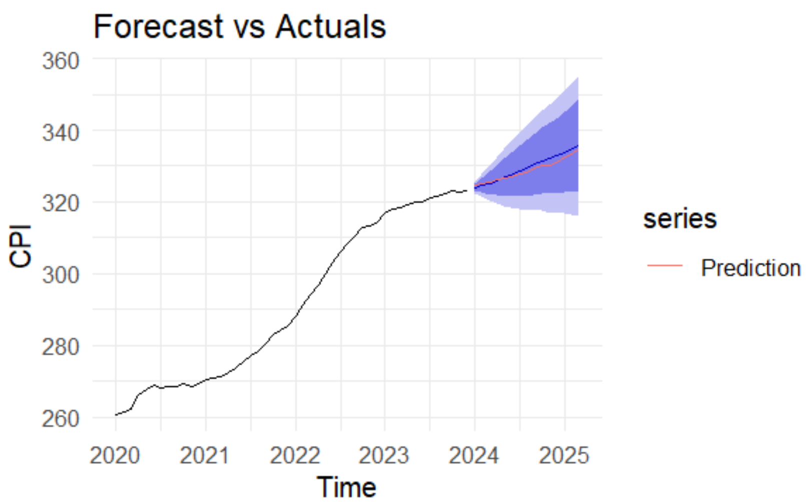


Forecasting

The ARIMA model was used to forecast the CPI for Food and Beverages from January 2024 to March 2025. While the in-sample fit was strong, the model's out-of-sample performance indicates room for improvement. In the test period, the MAE increased to 0.9148 and the MAPE rose to 27.73%, suggesting diminished accuracy when predicting unseen data.

Figure 5 illustrates the model's forecast (in red) alongside the actual CPI values (in black) for the forecast horizon. The blue shaded areas represent the 80% and 95% confidence intervals, which widen over time, reflecting growing uncertainty. Although the model captures the general trend, it consistently underestimates recent values, contributing to the higher forecast error. This highlights that while the model is effective for broad trend estimation, further tuning may be necessary to improve short-term predictive precision.

Figure 5



Conclusion

Using ARIMA modelling, this report analysed and forecasted the Consumer Price Index for food and beverages among U.S. urban consumers. The model effectively captured the strong upward trend observed in recent years, highlighting ongoing inflationary pressures in the sector. While the in-sample fit was solid and residual diagnostics showed no significant autocorrelation, forecasting performance on unseen data revealed reduced accuracy, particularly in short-term predictions.

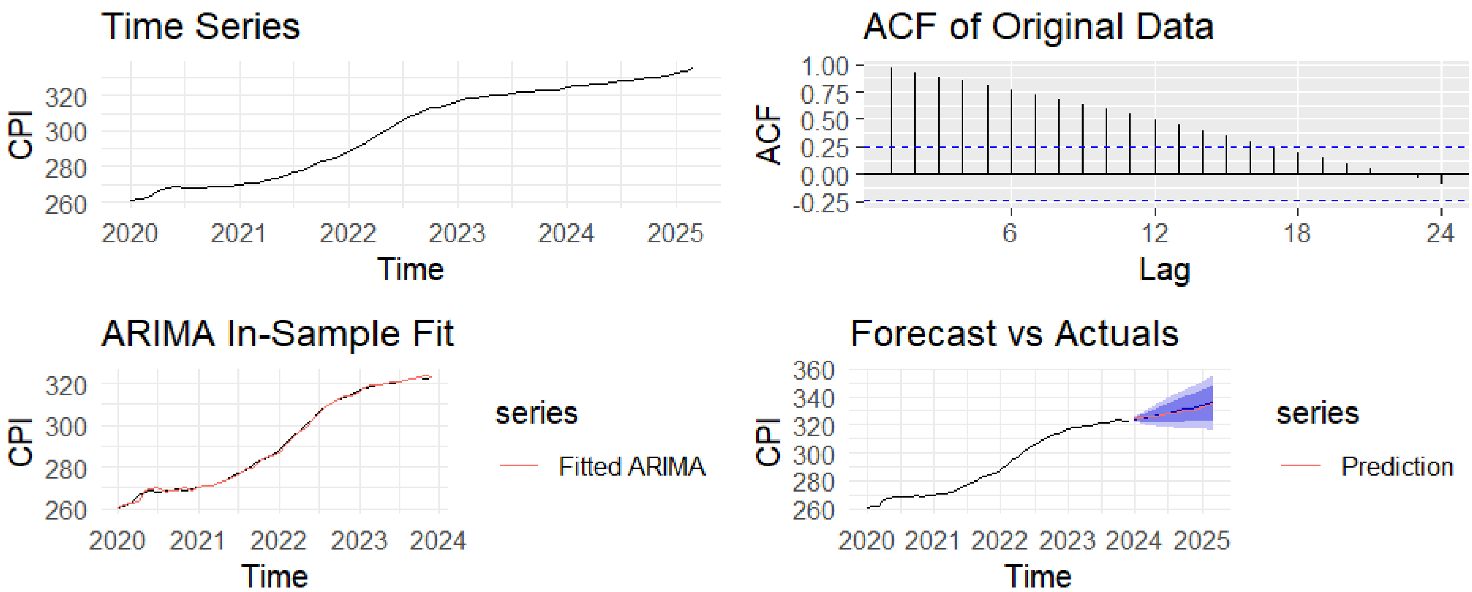
These limitations point to the need for further model tuning and the potential inclusion of external economic variables to better account for the inherent volatility of CPI data. Despite this, the ARIMA model remains a valuable tool for identifying trends and supporting data-driven decisions. With refinements, it can aid businesses and policymakers in anticipating price shifts and managing risks associated with inflation.

Appendix

GitHub Repository with code:

https://github.com/rodriguesathena/SocialForecasting_Spring2025

Combined Graph Panel:



References

U.S. Bureau of Labor Statistics. (n.d.). *Consumer Price Index for All Urban Consumers: Food and Beverages [CPI-U], not seasonally adjusted (Series ID: SUUR0000SAF).* from https://data.bls.gov/dataViewer/view/timeseries/SUUR0000SAF.