A Time Series Analysis of Inflation Based on Consumer Price Index

Author: Murpys Duran Mendez
md256@students.uwf.edu

STA - 6856 Spring 2025
Professor: Tharindu De Alwis

CHAPTER 1: INTRODUCTION

Inflation, the general price increase over time, plays a crucial role in shaping economic policy and financial decision-making. Understanding its behavior through data analysis allows researchers and policymakers to detect patterns, assess stability, and forecast future trends. This report applies time series methods to analyze inflation dynamics, uncovering structural characteristics and exploring potential drivers behind its fluctuations.

The used dataset contains a time series that studies inflation, measured by the consumer price index for different countries worldwide. This indicator reflects the annual percentage change in the cost to the average consumer of acquiring a basket of goods and services. The data originates from the International Monetary Fund and is hosted on data.worldbank.org.

The dataset contains the following variables:

- Time: Annual periodicity. Range (1960 - 2023)
- Country Name: Full country name
- Country Code: Three-letter abbreviation of country name
- Observed variable: Inflation, consumer price index (CPI). (Aggregation method: median)

OBJECTIVES

1- Perform a univariate analysis of inflation in the available period for the United States.

2-Perform a multivariate analysis of the indicator in the United States as influenced by countries that have a strong influence on the U.S economy (e.g., Mexico and Canada)

3-Compare the two models based on their forecasting performance using metrics such as RMSE and MAE, and assess whether incorporating the additional CPIs improves the prediction of the U.S. metric. Additionally, the statistical significance of this influence will be evaluated through Granger causality tests.

LITERATURE REVIEW

Several sources were consulted, covering both theoretical and empirical approaches to economic modeling. While some focus directly on inflation, others provide broader insights into time series methods and relationships that support the analysis in this project.

“ A Model of the Fed’s View on Inflation” models how the relationship between inflation, unemployment, and expectations evolves by allowing the underlying parameters to change gradually. This approach captures shifts in economic dynamics and helps explain how the Federal Reserve’s interpretation of inflation has adapted across different periods of uncertainty.

Benjamin M. Blau, in “ Inflation and Bitcoin: A descriptive time-series analysis”, explores the relationship between inflation and Bitcoin using descriptive time series analysis. The authors examine whether Bitcoin behaves like a hedge against inflation across several countries, focusing on correlations between inflation rates and Bitcoin prices.

The paper “A Time Series Analysis of Determinants of PCE Inflation in the United States”, by Elliot Charron uses multiple linear regression with lagged variables for modeling relationships over time. It includes tests for stationarity (like the ADF test) and uses autoregressive components to account for persistence in inflation.

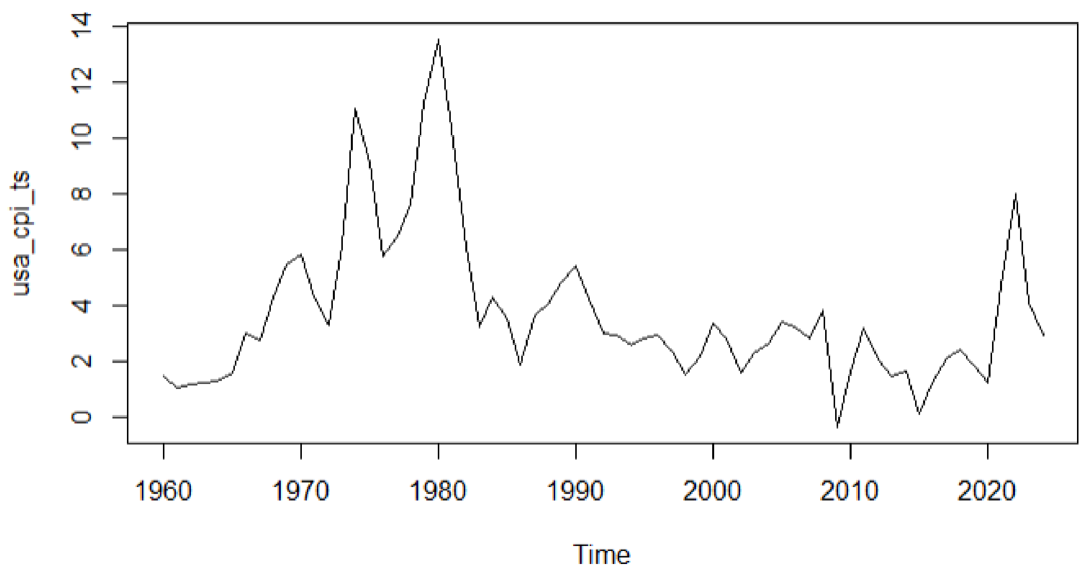
“Predicting the Future: Forecasting Inflation Using Time Series Models in R”, explores how to forecast inflation using various time series models implemented in R. Tigran Danielyan applies models such as ARIMA, ETS (Exponential Smoothing), and Prophet, comparing their performance on U.S. inflation data. The piece emphasizes model selection, forecast accuracy, and the practical implementation of time series forecasting techniques.

“Time-series Modeling for Consumer Price Index Forecasting using Comparison Analysis of AutoRegressive Integrated Moving Average and Artificial Neural Network”. This paper compares the performance of ARIMA and Artificial Neural Network (ANN) models in forecasting the Consumer Price Index (CPI). Using historical CPI data, the authors evaluate each model’s accuracy and ability to capture inflation patterns. The results show that ARIMA handles linear trends well.

CHAPTER 2: PRELIMINARY ANALYSIS

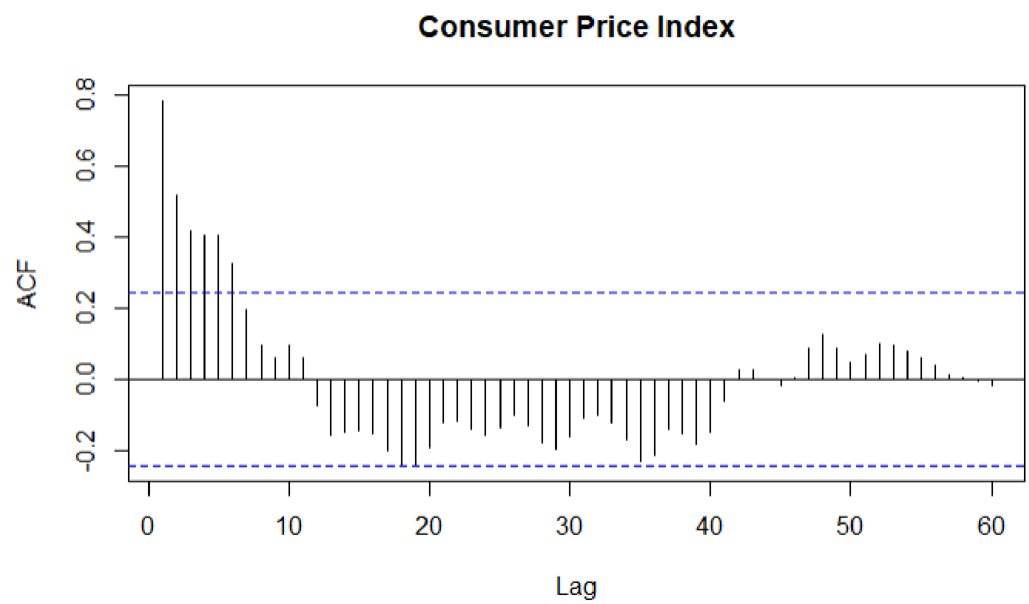
In this initial exploration of the data, we employ various graphical tools to examine, summarize, and gain initial insights into the time series data. The US data is stored in a time-series object called "usa_cpi_ts", and it is of size [1:64], ranging from 1960 to 2023.

USA Time Series Plot:

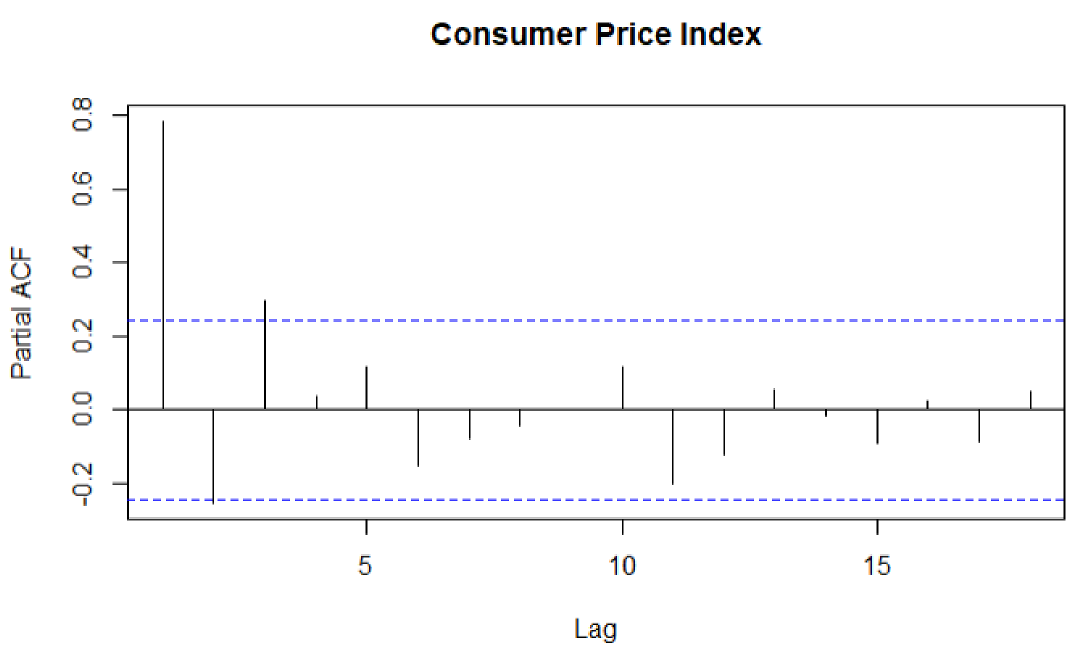


The series does not exhibit a clear trend. We can see periodical variations; however, it is difficult to discern regularity based on a visual inspection of the series plot. Between 1970 and 1980, and again in 2022, the time series shows visible spikes in values, indicating deviations from the general level of the series. These spikes match periods of economic changes and conflicting forces in commerce and society.

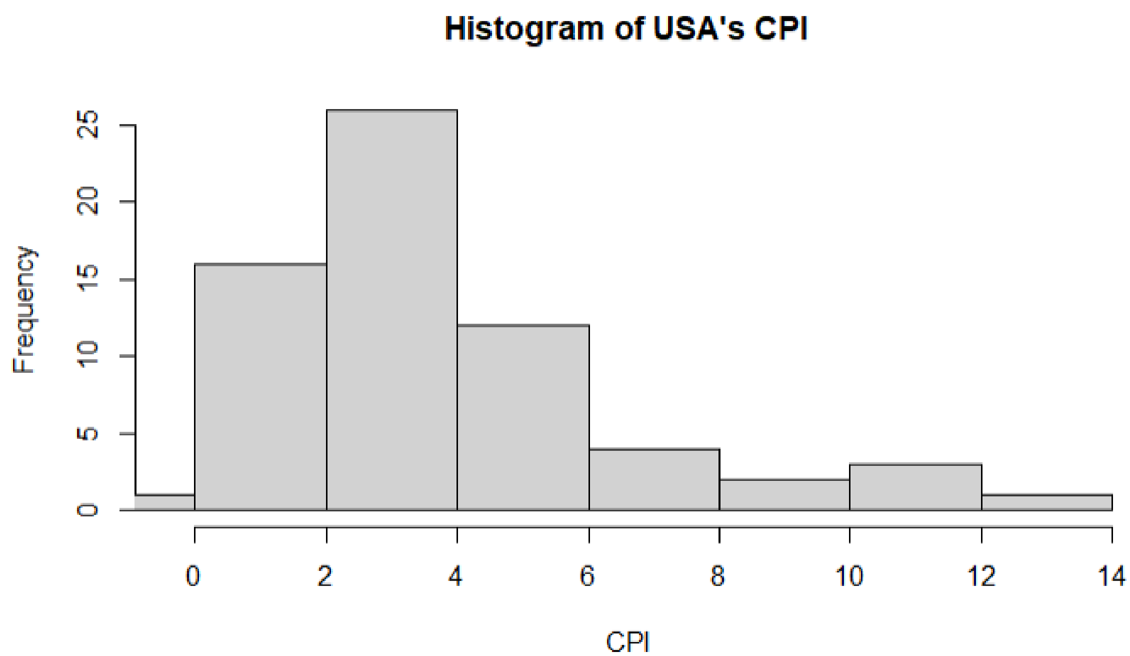
The ACF was plotted at several lag values, revealing no seasonality. The lags are significant until six.

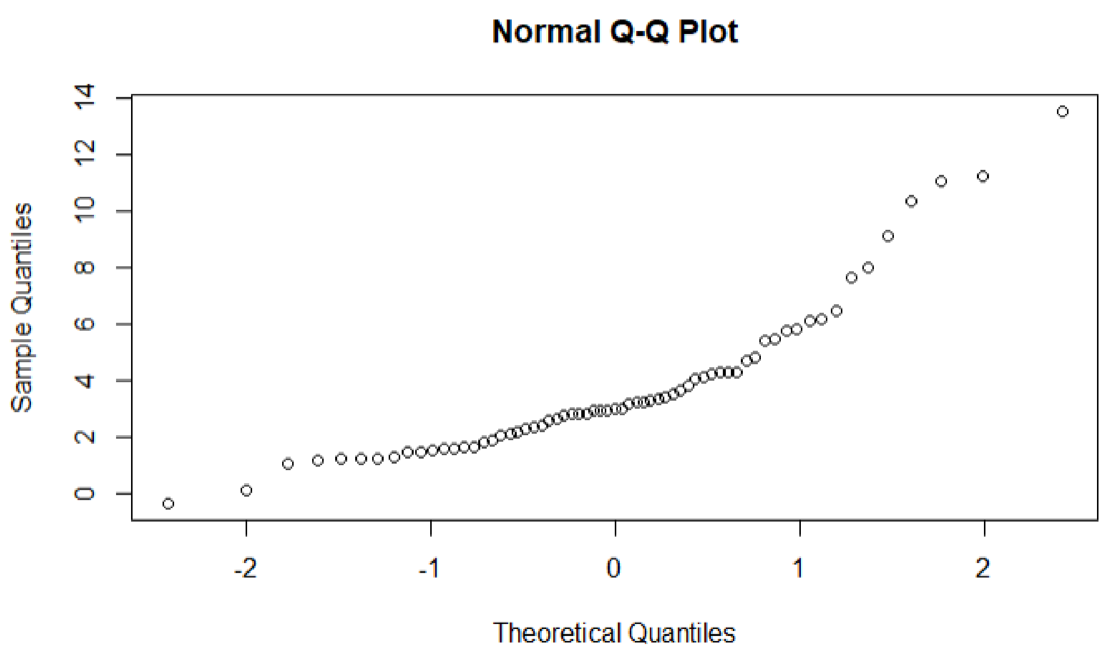


The PACF shows significance at lag one, while lags two and three are close to the significance line.

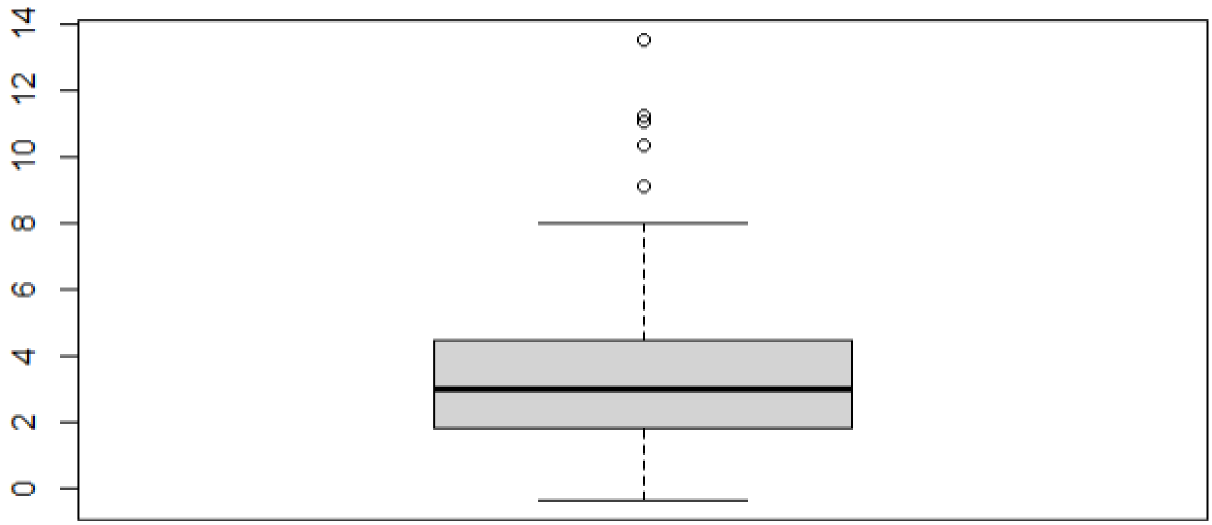


The observed high values impact the data, producing right skewness and pulling the data away from a normal distribution.



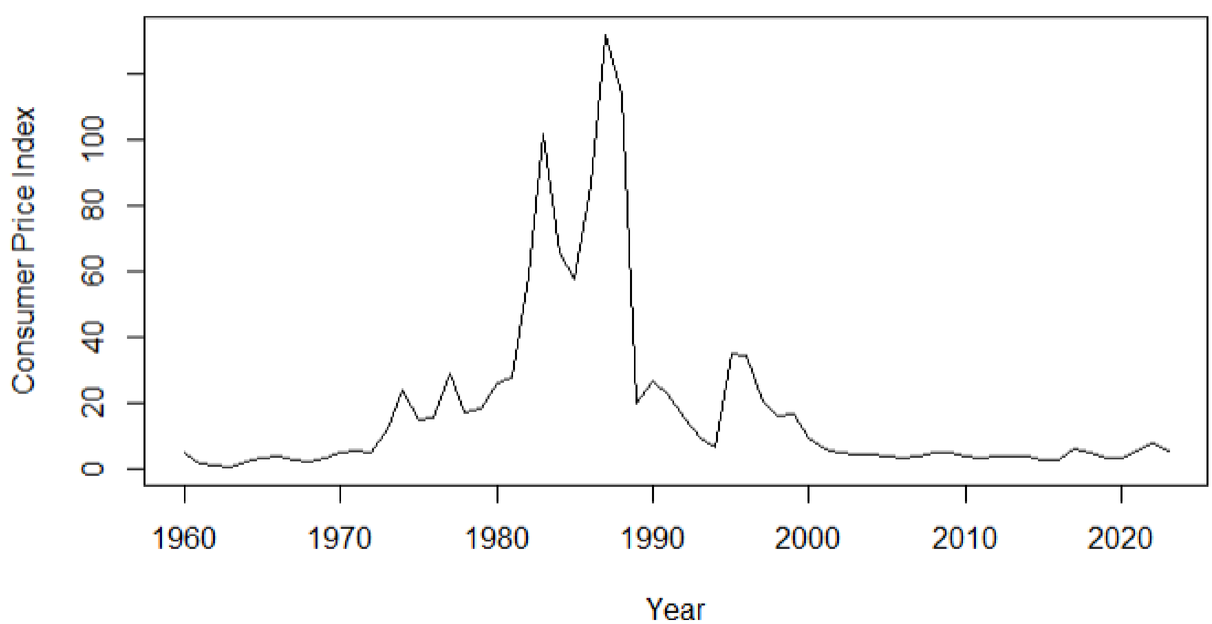


This result is confirmed by the Shapiro-Wilk normality test, with *p-value* = 0.000001616, leading to reject the Null hypothesis (H_0: *the data follows a normal distribution*). The box plot shows those higher values as potential outliers, but we know they are valid datapoints.



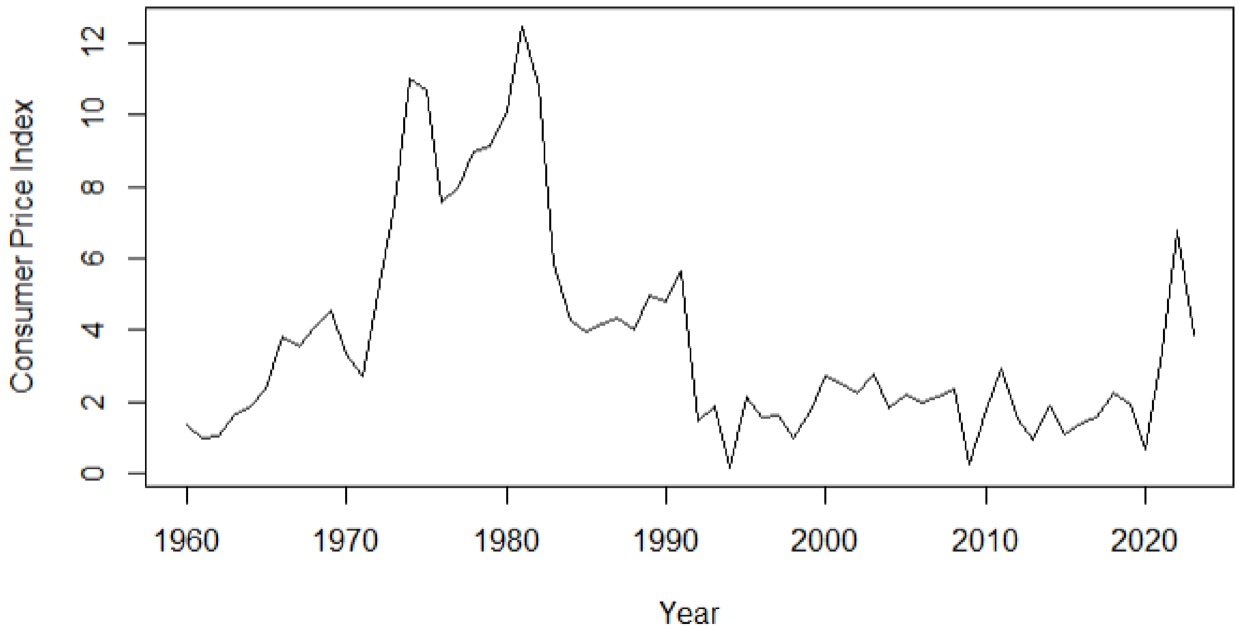
Mexico's time series was plotted, showing that the curve follows a similar spike as the USA series, though the rest of the series has less variation.

Mexico's Time Series:

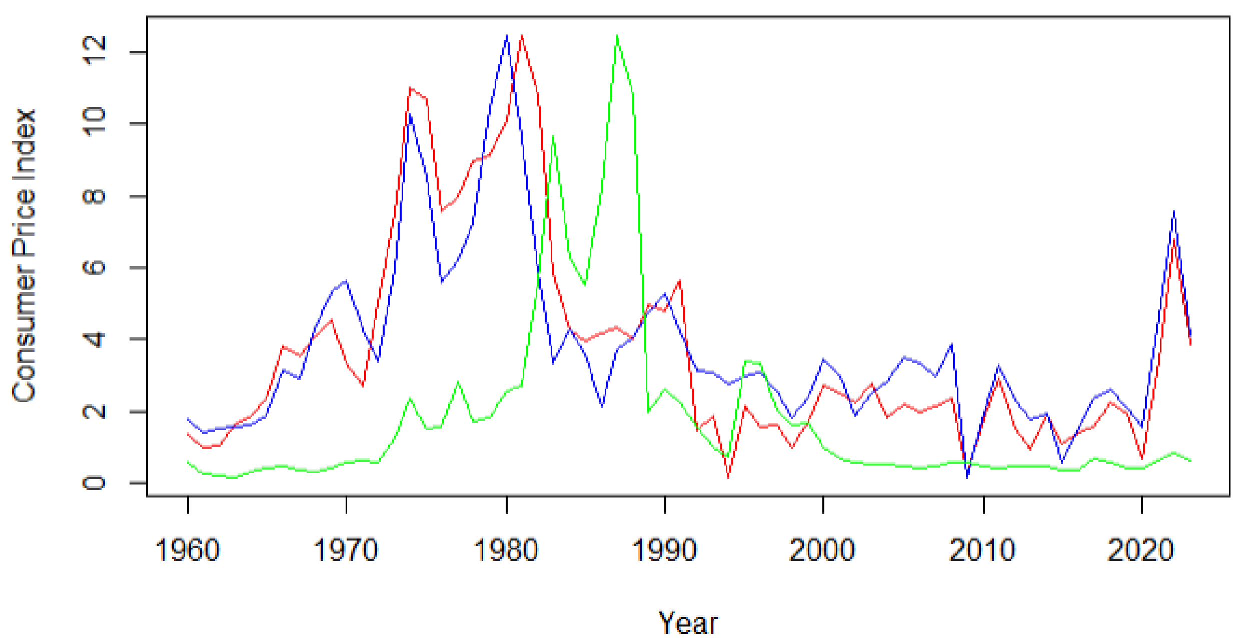


The same procedure was applied to Canada’s data. Given its closer economic ties to the US, the plot shows similar patterns during the same years.

Canada’s Time Series:



USA, Canada, and Mexico’s CPI Time Series:



Canada ----------

USA ----------

Mexico ----------

Interestingly, all three series exhibit similarly shaped and spaced spikes. For Canada and the U.S., the events are nearly synchronized, while in Mexico, a similar spike appears after a few lags.

Testing stationarity through the Augmented Dickey-Fuller Test.

United States
H_0: The US series has a unit root (it is non-stationary)
H_1: The series is stationary.

For the U.S series, a *p-value* = 0.3312 was obtained. Therefore, we fail to reject the Null hypothesis and conclude that the series is non-stationary.

Canada
H_0: The Canada series has a unit root (it is non-stationary)
H_1: The series is stationary.

The Canada series resulted in a *p-value* = 0.4453. We fail to reject H_0, concluding that the series is non-stationary.

Mexico
H_0: The Mexico series has a unit root (it is non-stationary)
H_1: The series is stationary.

The test for the Mexico series produced a *p-value* = 0.586. We conclude that the series is not stationary.

Candidate models include: ARIMA(1,1,0), ARIMA(0,1,1), ARIMA(1,1,2), ARIMA(0,1,2)

CHAPTER 3: SECONDARY ANALYSIS

Univariate Model

Building on the previous results and the plotted series, we fitted different models in an iterative process, aiming to find the model that best fits the series. A hypothesis test was performed for all models to determine the significance of the coefficients.

H_0: The parameter is not different from zero ($\phi = 0$)

H_1: The parameter is different from zero ($\phi \neq 0$)

ARIMA(1,1,0)
An ARIMA(1,1,0) model was applied to the U.S.A. data.
The *p-value* of 0.356 indicates that we fail to reject the null hypothesis, suggesting that the coefficient is not significant.

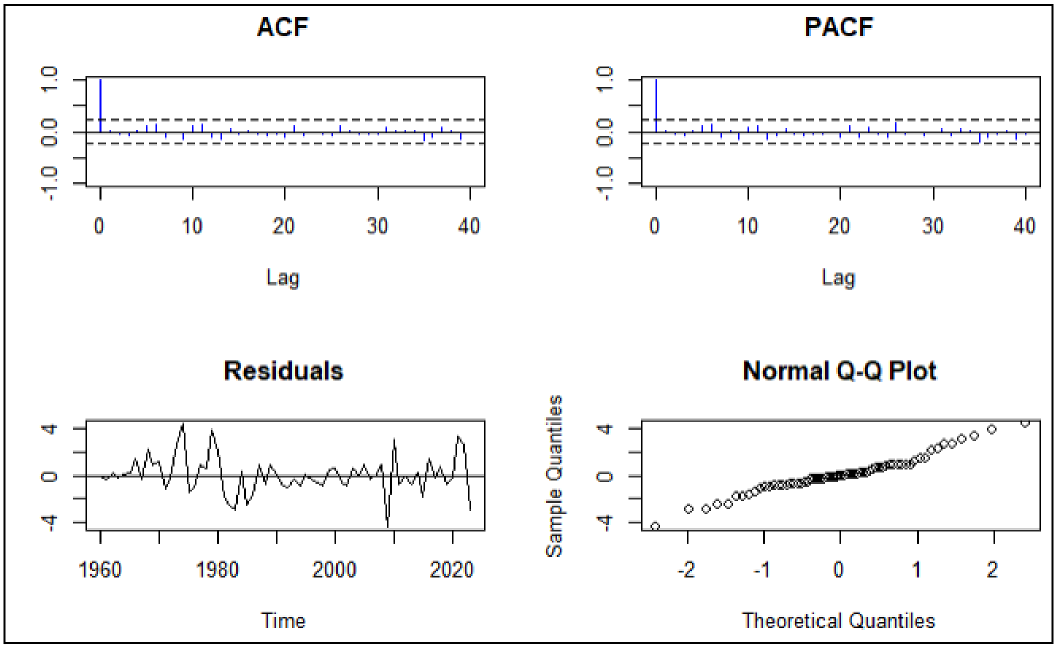
ARIMA(1,1,2)
Based on the EACF we fitted an ARIMA(1,1,2) model.
In this model, the AR coefficient remained non-significant (*p-value* = 0.2546), suggesting that the AR term has little impact on explaining the series. Therefore, we proceeded with fitting models that only include MA components.

ARIMA(0,1,2)
In the ARIMA(0,1,2) model, the MA1 component was found to be non-significant (*p-value* = 0.4998), indicating that it did not contribute meaningfully to explaining the series. Since the non-significance of this component suggested it was not necessary for capturing the underlying patterns in the data, we decided to simplify the model by considering ARIMA(0,1,1), which includes only the MA1 term. This adjustment aimed to improve model parsimony while retaining essential characteristics of the data; however, a comparison of the AIC values showed that ARIMA(0,1,2) has significantly lower digits.

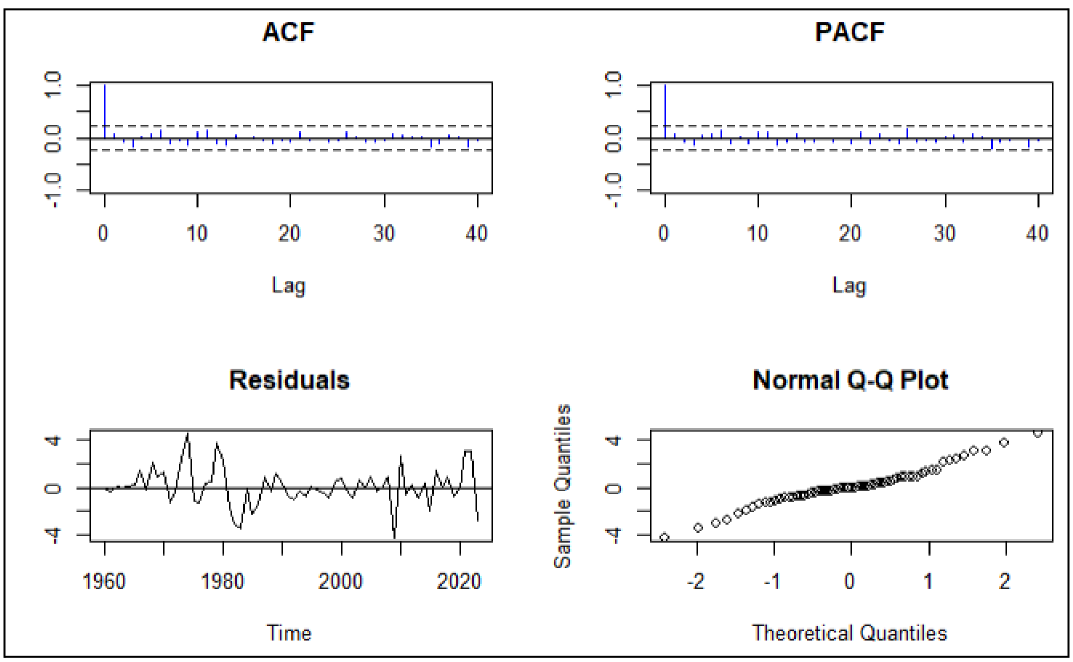
	p-value			AIC
Model \ Term	**ar1**	**ma1**	**ma2**	
ARIMA(1,1,0)	*0.356*	--	--	*255.0239*
ARIMA(0,1,1)	--	*0.0155*		*252.3027*
ARIMA(1,1,2)	*0.2546*	*0.538*	*<0.001*	*244.9541*
ARIMA(0,1,2)	--	*0.4998*	*<0.001*	*244.2753*

Similar results were obtained when the residuals of the best two candidate models (ARIMA(1,1,2) and ARIMA(0,1,2)) were tested. Both residuals are iid noise.

ARIMA(1,1,2) Ljung-Box: *p-value = 0.877*



ARIMA(0,1,2) Ljung-Box: *p-value = 0.814*

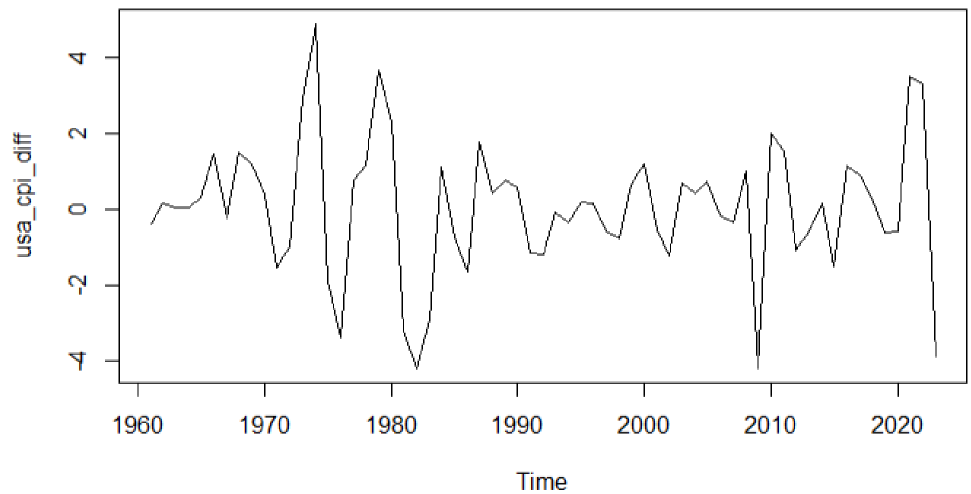


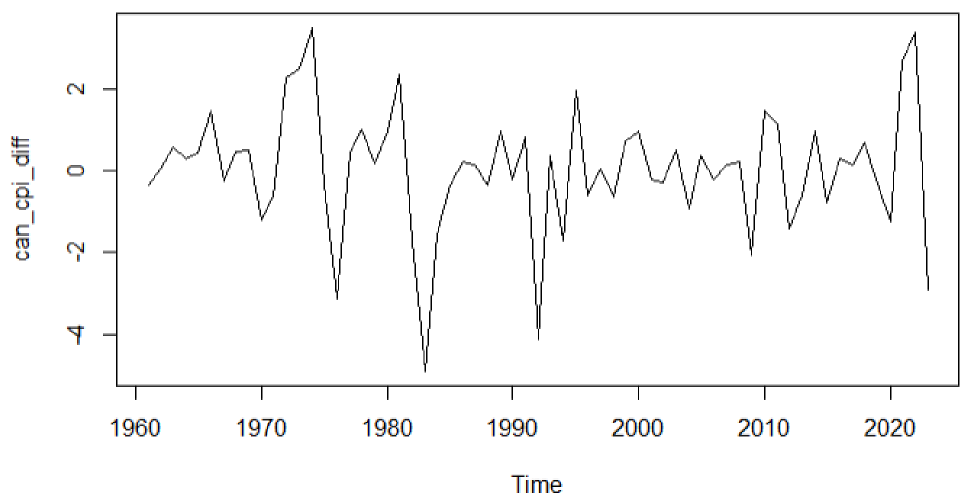
Following the principle of Ockham's razor, we will stick to the ARIMA(0,1,2) model for the univariate case.

Multivariate Model

We decided to model the USA and Canada time series together, dropping Mexico since we observed a shift or displacement in the pattern.

Differencing both series, we obtained stationary data.





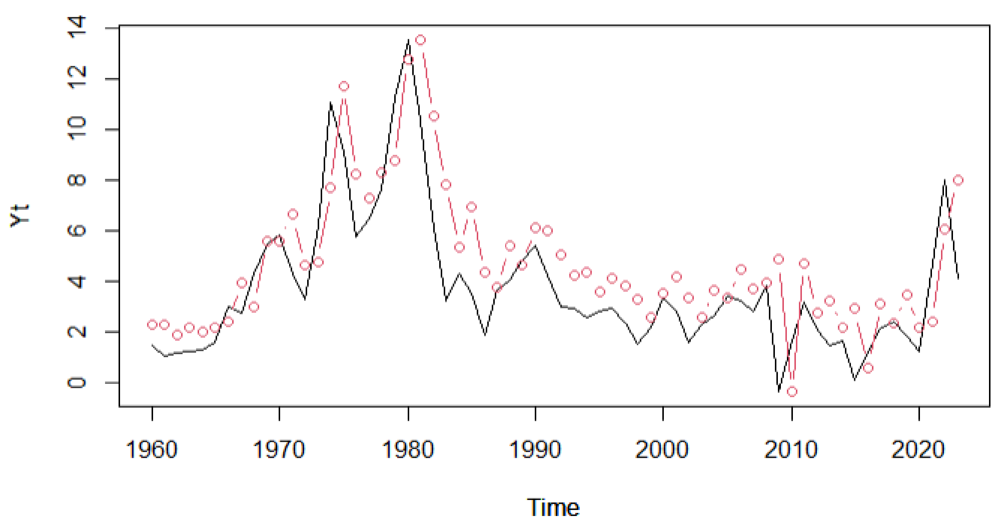
Augmented Dickey-Fuller Test: *p-value* = 0.01 for both series.

Models comparison

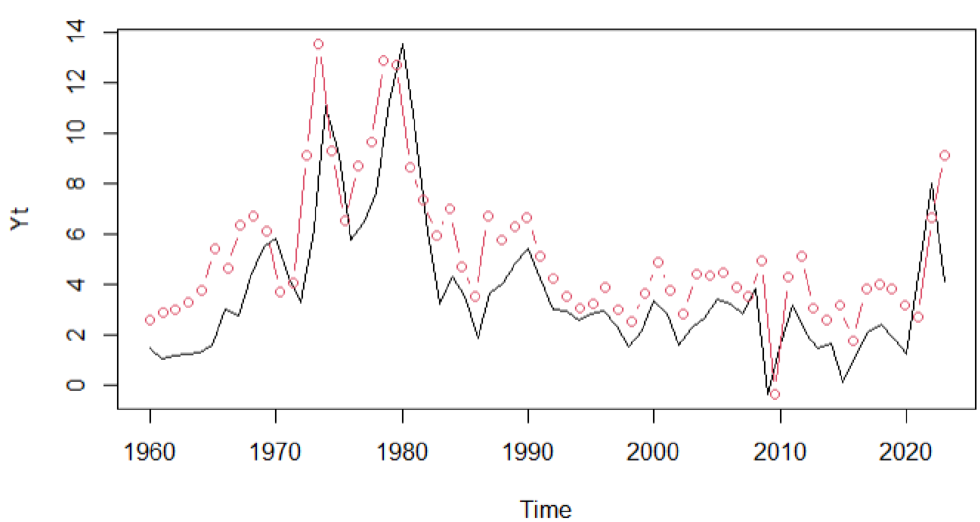
The comparison of the Mean Squared Error (MSE) between the univariate MA(2) model and the multivariate model for predicting the USA CPI shows a small but notable improvement. The MSE for the univariate model is 2.512184, while the multivariate model yields a slightly lower MSE of 2.34531. This indicates that incorporating additional variables, such as the Canadian CPI, provides a modest reduction in prediction error. Although the improvement in MSE is not large, it demonstrates the benefit of considering more data sources for more accurate forecasting.

<u>Models plots vs the original series:</u>

$$\boldsymbol{W_t = 0.077\, e_{t-1} - 0.45\, e_{t-2}}$$



$$usa\_cpi\_ts = 0.856\, usa\_cpi\_ts_{t-1} + 0.175\, can\_cpi\_ts_{t-1} + 0.640\, usa\_cpi\_ts_{t-2} + 0.333\, can\_cpi\_ts_{t-2} + 1.072$$



CONCLUSION

In this project, we analyzed the performance of univariate and multivariate models for forecasting consumer price indices, with a focus on the USA and Canada. By comparing both models' Mean Squared Error (MSE), we observed that the multivariate model generally offered a slightly better fit, highlighting the importance of incorporating multiple variables in forecasting. While the univariate model demonstrated reasonable accuracy, the inclusion of additional predictors in the multivariate model helped refine predictions, resulting in a lower MSE. This analysis emphasizes the utility of multivariate models in capturing more complex relationships and improving forecast accuracy in time series analysis.

A limitation of this analysis is that both the USA and Canada consumer price indices followed similar patterns, which may have influenced the performance of the models. In future work, it would be beneficial to incorporate more diverse and varied data sources to capture a broader range of dynamics and improve the robustness of the forecasting models.

References

1. A Model of the Fed's View on Inflation. Thomas Hasenzagl1 , Filippo Pellegrino , Lucrezia Reichlin , and Giovanni Ricco. December 2017. https://arxiv.org/abs/2006.14110v1

2. Benjamin M. Blau, Todd G. Griffith, Ryan J. Whitby, Inflation and Bitcoin: A descriptive time-series analysis, Economics Letters, Volume 203, 2021, 109848, ISSN 0165-1765, https://doi.org/10.1016/j.econlet.2021.109848.

3. A Time Series Analysis of Determinants of PCE Inflation in the United States. Elliot Charron. EEB Vol 14 (2021). Bryant University

4. Predicting the Future: Forecasting Inflation Using Time Series Models in R. Tigran Danielyan. February 2023. https://medium.com/@tigrandanielyan_46621/predicting-the-future-forecasting-inflation-using-time-series-models-in-r-4b7dc713dbb5

5. Time-series Modeling for Consumer Price Index Forecasting using Comparison Analysis of AutoRegressive Integrated Moving Average and Artificial Neural Network.

   Intan Yuniar Purbasari, Fetty Tri Anggraeny, and Nindy Apsari Ardiningrum. https://www.scitepress.org/Papers/2020/103692/103692.pdf

Appendix I

```
---
title: "STA_6856 Final_Project"
format: html
author: "TDA"
date: "2025-03-28"
output: pdf_document
echo: false
---
```{r set up}
library(TSA)
library(lmtest) #Testing parameters
library(itsmr)
library(tseries)
library(forecast)
library(vars)
options(scipen = 999)
```

```{r import}
data <- read.csv("inflation_cpi.csv", header = FALSE, check.names = FALSE,
stringsAsFactors = FALSE)

```
```{r}
head(data)
```



```{r clean up}
# Promote the 3rd row to column names
colnames(data) <- as.character(unlist(data[3, ]))

# Drop the first 3 rows (metadata + header row)
data <- data[-c(1:3), ]

# Reset row numbering
rownames(data) <- NULL
```

```{r clean_up cont}
data <- data[,-c(3,4)] #drop columns 3 and 4
data <- data[, -ncol(data)] #drop last empty column
data <- na.omit(data)
```

```{r usa ts}
# Create a time series for United States' CPI
us_row_number <- which(data$`Country Name` == "United States")
usa_cpi <- as.numeric(data[us_row_number, 3:ncol(data)])
usa_cpi_ts <- ts(usa_cpi, start = 1960, end = 2023, frequency = 1)

```

```{r mex ts}
```
```

````
# Create a time series for Mexico's CPI
mex_row_number <- which(data$`Country Name` == "Mexico")
mex_cpi <- as.numeric(data[mex_row_number, 3:ncol(data)])
mex_cpi_ts <- ts(mex_cpi, start = 1960, end = 2023, frequency = 1)
```


```{r can ts}
# Create a time series for Canada's CPI
can_row_number <- which(data$`Country Name` == "Canada")
can_cpi <- as.numeric(data[can_row_number, 3:ncol(data)])
can_cpi_ts <- ts(can_cpi, start = 1960, end = 2023, frequency = 1)
```

```{r test 1}
#us_row_number
#class(usa_cpi_ts)
#str(usa_cpi)
#data[us_row_number, ]
#head(usa_cpi_ts)
#usa_cpi[1]
#is.numeric(usa_cpi_ts1)
str(usa_cpi_ts)
#usa_cpi_ts
```

```{r plot ts usa}
plot(usa_cpi_ts, ylab = "Consumer Price Index", xlab = "Year")
hist(usa_cpi_ts, xlim = range(usa_cpi_ts), main = "Histogram of USA's CPI",
xlab = "CPI")
boxplot(usa_cpi_ts)

```
```{r USA acf/pacf}
#acf(usa_cpi_ts, main = "Consumer Price Index")
acf(usa_cpi_ts, main = "Consumer Price Index", lag.max = 60)
pacf(usa_cpi_ts, main = "Consumer Price Index")
```

```{r}
qqnorm(usa_cpi_ts)
```
```{r stationarity test}
#summary(usa_cpi_ts)
adf.test(usa_cpi_ts)
adf.test(can_cpi_ts)
adf.test(mex_cpi_ts)
#shapiro.test(usa_cpi_ts)
```
```{r}

```

```{r plot ts mexico}
````

````
plot(mex_cpi_ts, ylab = "Consumer Price Index", xlab = "Year")
hist(mex_cpi_ts, xlim = range(mex_cpi_ts), main = "Histogram of Mexico's
CPI", xlab = "CPI")
#boxplot(usa_cpi_ts)

```


```{r plot ts canada}
plot(can_cpi_ts, ylab = "Consumer Price Index", xlab = "Year", col = "red")
par(new=TRUE)
plot(usa_cpi_ts, ylab = "", xlab = "", col = "blue", axes = FALSE)
par(new=TRUE)
plot(mex_cpi_ts, ylab = "", xlab = "", col = "green", axes = FALSE)

#hist(can_cpi_ts, xlim = range(can_cpi_ts), main = "Histogram of Canada's
CPI", xlab = "CPI")
#boxplot(usa_cpi_ts)

```


```{r eacf}
eacf(usa_cpi_ts)
```



```{r modeling univariate.AR(1)}
fit.ar1=arima(usa_cpi_ts,order=c(1,1,0),include.mean = FALSE)
coeftest(fit.ar1)
AIC(fit.ar1)
```

```{r modeling univariate.MA(1)}
fit.ma1=arima(usa_cpi_ts,order=c(0,1,1),include.mean = FALSE)
coeftest(fit.ma1)
AIC(fit.ma1)
```

```{r modeling univariate.ARMA(1,2)}
fit.arma12=arima(usa_cpi_ts,order=c(1,1,2),include.mean = FALSE)
coeftest(fit.arma12)
AIC(fit.arma12)
```
```{r modeling univariate.MA(2)}
fit.ma2=arima(usa_cpi_ts,order=c(0,1,2),include.mean = FALSE)
coeftest(fit.ma2)
AIC(fit.ma2)
#Wt = 0.077*Wt-1 - 0.45*Wt-2
```
$$
W_t = 0.077\ e_{t-1} - 0.45\ e_{t-2}
$$

```{r modeling univariate.MA(6)}
#we were just studying what happens if we go to the last significant lag in
````

````
afc
fit.ma6=arima(usa_cpi_ts,order=c(0,1,6),include.mean = FALSE)
coeftest(fit.ma6)
AIC(fit.ma6)
```




```{r auto arima}
fit.auto <- auto.arima(usa_cpi_ts)
summary(fit.auto)
coeftest(fit.auto)
```


```{r residuals arima(1,1,2)}
residuals.12 = residuals(fit.arma12)
test(residuals.12)
shapiro.test(residuals.12)

```


```{r residuals arima(0,1,2)}
residuals.02 = residuals(fit.ma2)
test(residuals.02)
shapiro.test(residuals.02)
```

```{r fitted values plot}
resid_vals <- residuals(fit.ma2)
fitted_vals <- usa_cpi_ts - resid_vals
plot(usa_cpi_ts,col=1,ylab="Yt")
par(new=TRUE)
plot(fitted_vals,type="b",col=2, axes = FALSE, ylab="Yt")

```



```{r individual series prep}
usa_cpi_diff <- diff(usa_cpi_ts)
can_cpi_diff <- diff(can_cpi_ts)

```

```{r usa_cpi_diff tests}
acf(usa_cpi_diff)
plot(usa_cpi_diff)
adf.test(usa_cpi_diff)
```


```{r can_cpi_diff tests}
acf(can_cpi_diff)
````

````
plot(can_cpi_diff)
adf.test(can_cpi_diff)
```

```{r multivariate model}
#combine the three series into a multivariate matrix
multi_data <- cbind(usa_cpi_ts, can_cpi_ts)
VARselect(multi_data, lag.max=12, type = "const")

```
```{r model fitting}
mult_cpi <- VAR(multi_data, p=2, type = "const")
summary(mult_cpi)
```

$$
usa_cpi_ts = 0.856 \ usa_cpi_ts_{t-1} + 0.175 \ can_cpi_ts_{t-1} +
0.640 \ usa_cpi_ts_{t-2} + 0.333 \ can_cpi_ts_{t-2} + 1.072
$$

```{r tests}
irf_fit <- irf(mult_cpi, impulse = "usa_cpi_ts", response = "usa_cpi_ts",
n.ahead = 10, boot = TRUE)

```


```{r fitted values plot multivariate}
fitted_var_usa <- fitted(mult_cpi)[, "usa_cpi_ts"]
plot(usa_cpi_ts,col=1,ylab="Yt")
par(new=TRUE)
plot(fitted_var_usa,type="b",col=2, axes = FALSE, xlab="", ylab="")

```




```{r test block}
#summary(fit.arma12)
#fvs <- fitted.values(fit.ar1)
#summary(fvs)
#sum(is.na(usa_cpi_ts))
#str(usa_cpi_ts)
#length(usa_cpi_ts)
#summary(fit.ma2)
```

```{r comparison}
#arima(0,1,2)
mse <- mean(residuals(fit.ma2)^2)
mse


#var(2)
mse_var <- mean(residuals(mult_cpi)[, "usa_cpi_ts"]^2)
````

```
mse_var

```