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


```