Project

Hortencia J. Hernandez* University of Texas at El Paso (UTEP)

14 December, 2023

About Data

Throughout the years, it is common for women to become unemployed due to being pregnant or having kids. Although the factors vary it is important to see the change and conclude if there are any percautionary measures to take. The data set can be found here[1] and comes from the U.S. Bureau of Labor Statistics. This is a subset that contains 840 observations from January 1950 to December 2019. It is calculated based on the amount of women in the work force with the total amount of people. More details are found at the link.

Goal

The goal of this project is find a model that would best fit for predicting women's unemployment rate. Although the world was hit with a global pandemic, if that event did not occur how would have the unemployment rate change? What does the spectral density tell?

Plot

Differencing was employed as a technique to transform the non-stationary data into a (mostly) stationary form, which can be see from Figure 1.

*hjhernandez4@miners.utep.edu

[1] https://fred.stlouisfed.org/series/LNS14000002

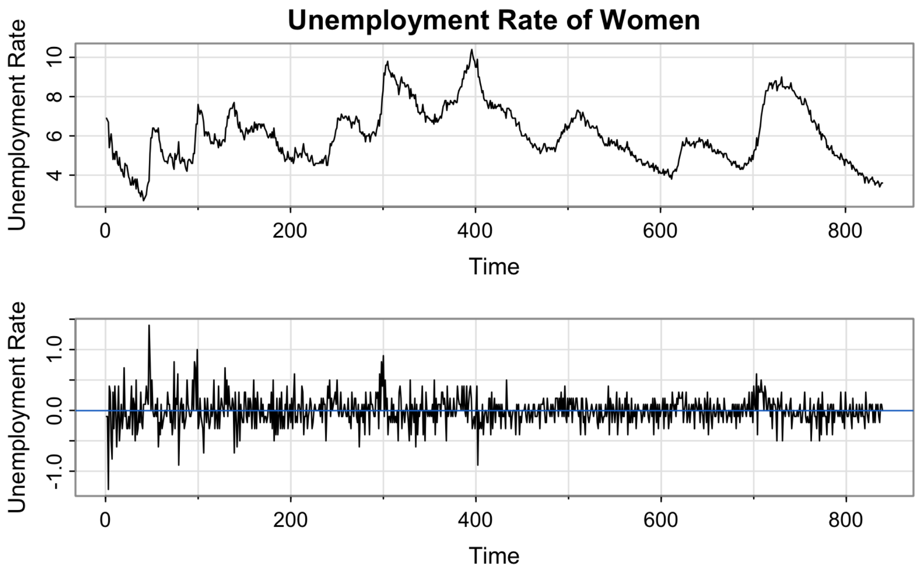

Figure 1: Time Series Plot of Unemployment-Women

Estimated ACF/PACF

The ACF of the data implies a differentiation thus, Figure 2 shows the estimated ACF and PACF plots. The ACF appears to cut off after lag 1 and the PACF seems to be tailing off, which could imply an MA(1) . The plots could also suggest an ARMA(1,1) model.

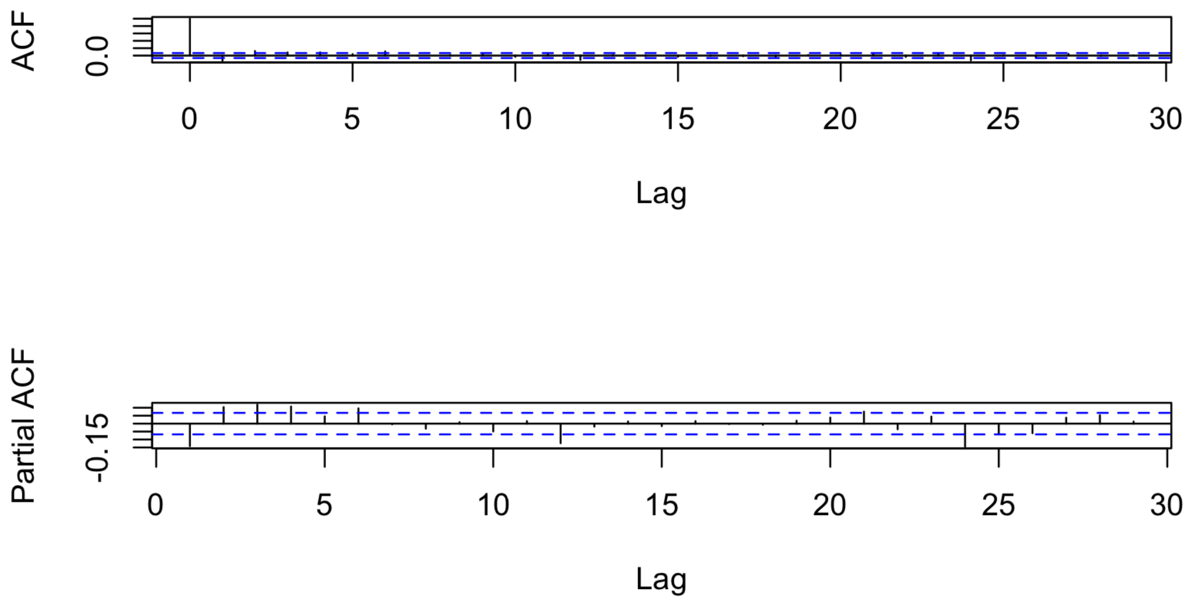

Figure 2: Estimated ACF/PACF

Estimated Spectral Density

The estimated Spectral Density can be seen in Figure 3. The peaks are 0.0011574, 0.0023148 which are found at periods 864, 432. From the image it could be potentially concluded that there is a MA factor due to the shape of the graph. Further, the confidence interval is large which could be due to some uncertainities in the estimated spectral density.

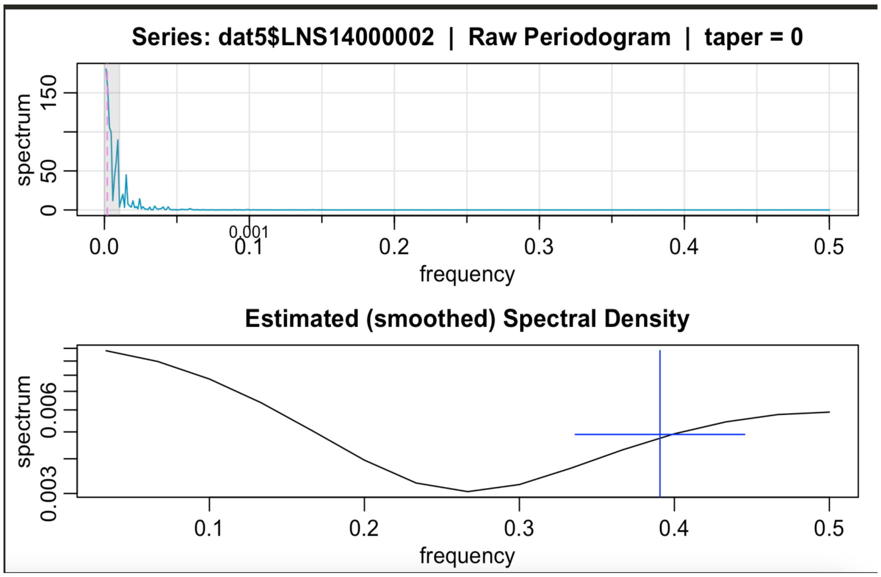

Figure 3: Estimated Spectral Density

(S)ARIMA Model

Based on Figure 2, the following models: MA(1), MA(2), ARMA(1,1), and ARMA(1,2) were tested. Figure 4 shows the model that was believed to be the best. The normal Q-Q plot suggests that the assumption of normality is not unreasonable, with very few outliers. Compared to the other models, for ARMA(1,2) most of the p-values for the Q - statistics exceed or are at 0.05.

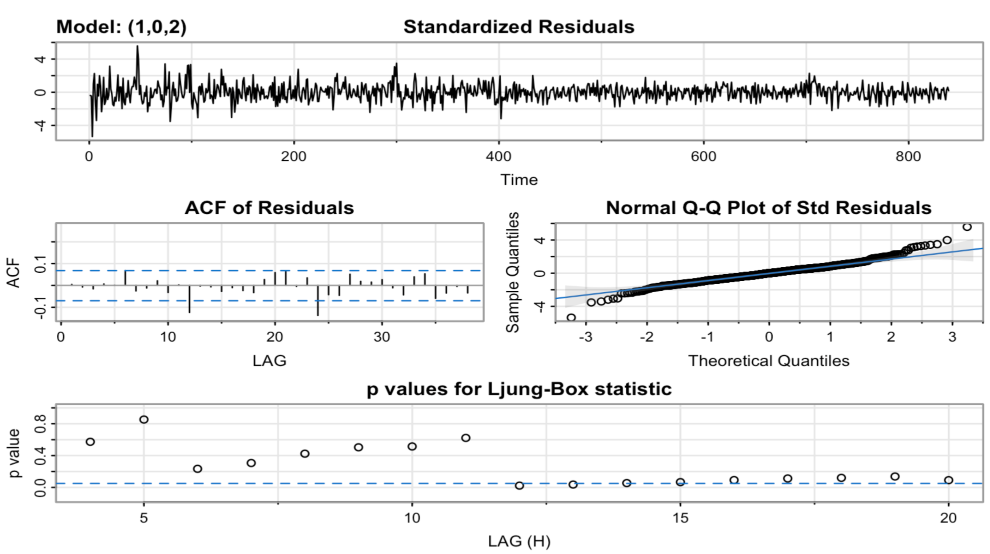

Figure 4: Model

Forecasting

From Figure 5, it can be noted that the unemployment rate for women appears to be gradually decreasing.

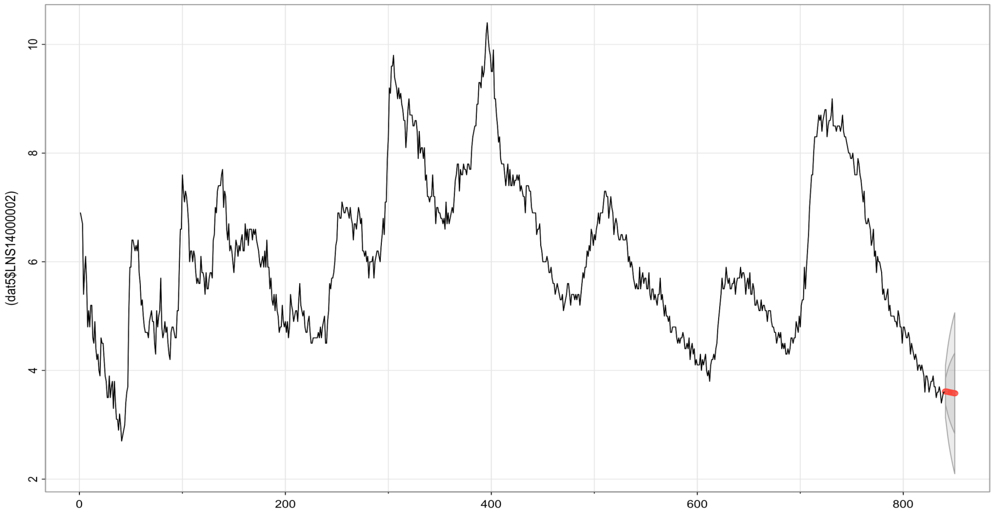

Figure 5: Forecasting - 10 Years

Discussion

With the aid of Figure 1 and the ACF of the original data, to transform the original data into a stationary form, differencing was employed. After performing some diagnostics the model selected was an ARMA(1,2) on the differenced data (or ARIMA(1,1,2) on the original data). From the forecasting we can see that the unemployment rate for women is gradually decreasing.

Future considerations might include seeing the affect of COVID and how post-pandemic might have affected the outcomes. Would it still be showing a gradual decrease at the same rate post-pandemic?

Appendix

```
knitr::opts_chunk$set(echo = FALSE, message = FALSE,
                      warning = FALSE, dev = 'pdf')

library(dplyr)
library(astsa)
library(xts)
library("forecast")

dat5 <- read.csv("LNS14000002.csv")
#Plotting the Data
par(mfrow= c(2,1))
tsplot(dat5$LNS14000002, ylab = "Unemployment Rate",
       main = "Unemployment Rate of Women")
tsplot(diff(dat5$LNS14000002), ylab = "Unemployment Rate")
abline(a = mean(diff(dat5$LNS14000002)), b =0 , col = 4)
#See original data acf2
acf2((dat5$LNS14000002), main = " ")
#Est. P/ACF
par(mfrow= c(2,1))
dat3_acf <- acf(diff(dat5$LNS14000002), main = "")
dat3_pacf <- pacf(diff(dat5$LNS14000002), main = "")


#Est. Spectral Density Plot
dat3_density <- spec.pgram(dat3_acf$acf, spans = c(2, 4, 6, 8),  taper = 0.1,
                           main = "Estimated (smoothed) Spectral Density")
par(mfrow= c(2,1))
mvspec(dat5$LNS14000002, col = rgb(.05,.6,.75))
  rect(0, -1e5, 1/95, 1e5, density = NA, col = gray(.5, .2))
  abline(v = 0.002, lty = 2, col = "violet")
  mtext("0.001", side = 1, line = 0, at =.1, cex = .75)

spec.pgram(dat3_acf$acf, spans = c(2, 4, 6, 8), taper = 0.1,
                           main = "Estimated (smoothed) Spectral Density")

#Periodogram of the original data
spec <- mvspec(dat5$LNS14000002, col = rgb(.05,.6,.75))
  rect(0, -1e5, 1/95, 1e5, density = NA, col = gray(.5, .2))
  abline(v = 0.002, lty = 2, col = "violet")
  mtext("0.001", side = 1, line = 0, at =.1, cex = .75)

max_spec <- spec$spec[which.max(spec$spec)]

# Find frequencies corresponding to peaks in the spectrum
peaks <- spec$freq[spec$spec > 0.8 * max_spec]
```

```
# Display the predominant periods
periods <- 1 / peaks
#Models
#ARMA(1,2)
sarima(diff(dat5$LNS14000002), 1, 0, 2)
#AIC = -0.01740778; BIC = 0.01079372
#sarima(diff(dat5$LNS14000002), 1, 0, 1) # pvals did not exceed 0.05
#AIC = 0.01962643 ; BIC = 0.04218762
#sarima(diff(dat5$LNS14000002), 0, 0, 1)# pvals did not exceed 0.05
#AIC = 0.02603459 : BIC = 0.04295549
#sarima(diff(dat5$LNS14000002), 0, 0, 2)# pvals did not exceed 0.05
# AIC = 0.008629878 ; BIC = 0.03119107
#Forecasting - 10 years
sarima.for((dat5$LNS14000002),n.ahead = 10, 0,1,2, plot.all = TRUE)
#abline(v = 2020, lty = 2, col = 4)
lines(dat5$LNS14000002)
```