Joel Cabrera

Economic Forecasting and Big Data (01:220:423)

Professor Berisha

September 25, 2019

**Project #1**

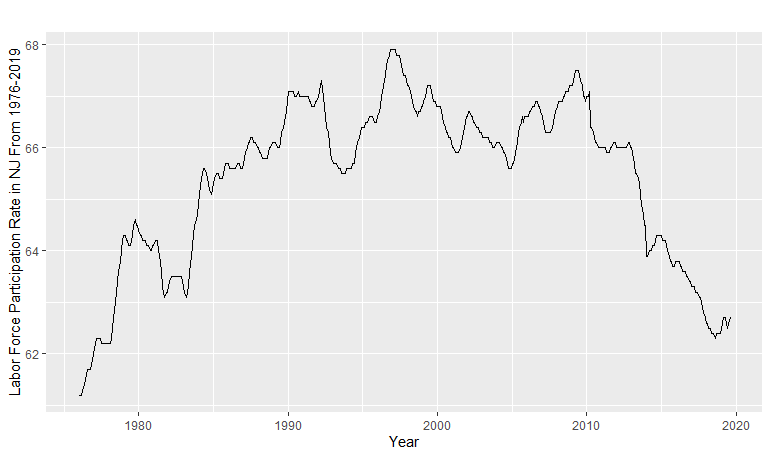
1. **Please identify the two variables you have chosen. Show the series graphically. Show the autocorrelation coefficients for the two series. Looking at the autocorrelation plots, do you think the original series have any trend, seasonal pattern, or cyclical pattern?**

The two variables that I have chosen are: the Labor Force Participation Rate in New Jersey (NJ) since 1976 and the Unemployment Rate in New Jersey (NJ) since 1976. They both collectively have 524 observations and are based on monthly frequencies (starting and ending on January 1976 and August 2019, respectively). The datasets that contain these two variables, which are coded in R/hereafter referred to as LFPR and UR, respectively, can be found from these two websites, also respectively: <https://fred.stlouisfed.org/series/LBSSA34> and <https://fred.stlouisfed.org/series/NJURN>

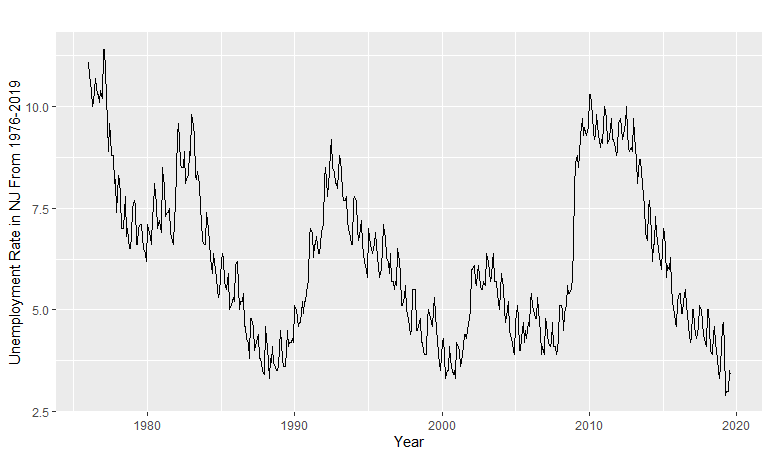
Below are graphs that illustrate the variables in two types of statistical matters. Graphs 1.1a and 1.2a depict the LFPR and UR variables in a time-series fashion, respectively. Graphs 1.1b and 1.2b depict the autocorrelation plots and coefficients of said two variables, also respectively.

Based on their respective autocorrelation plots, the time plot for the labor force participation force in NJ since 1976 seems to have a cyclical pattern (peaks and troughs are not at fixed frequencies; every 20 years), whereas the time plot for the unemployment rate in NJ since 1976 seems to have a seasonal pattern (and possibly a trend, if looking at the major peaks and troughs, but would need much earlier data to claim that).

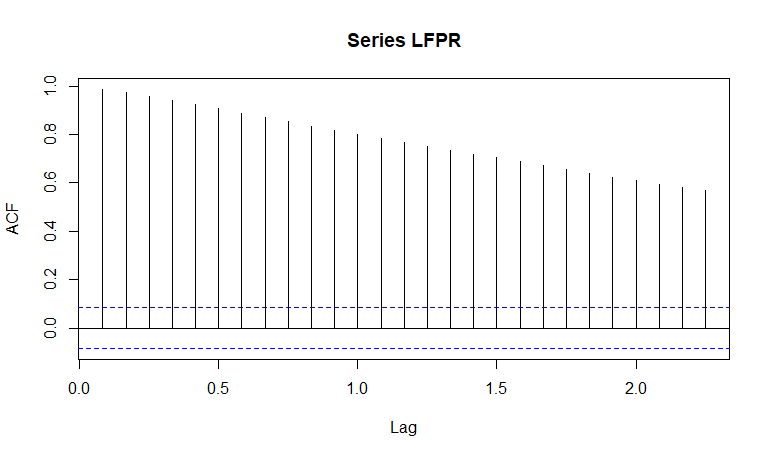
Graph 1.1a:



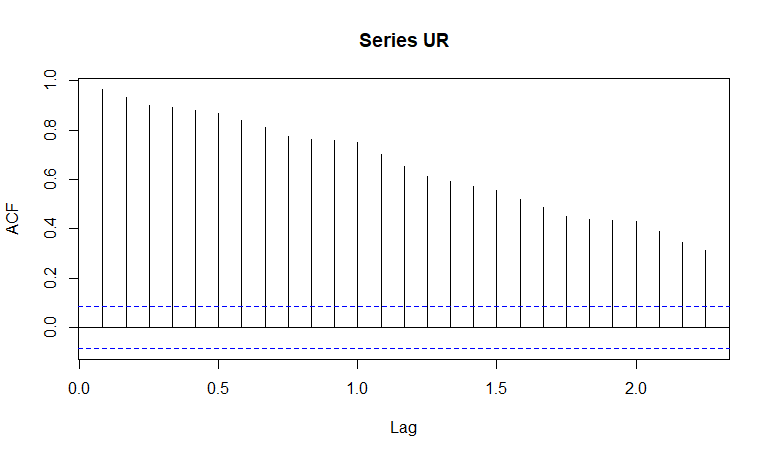
Graph 1.2a:



Graph 1.1b:



Graph 1.2b:



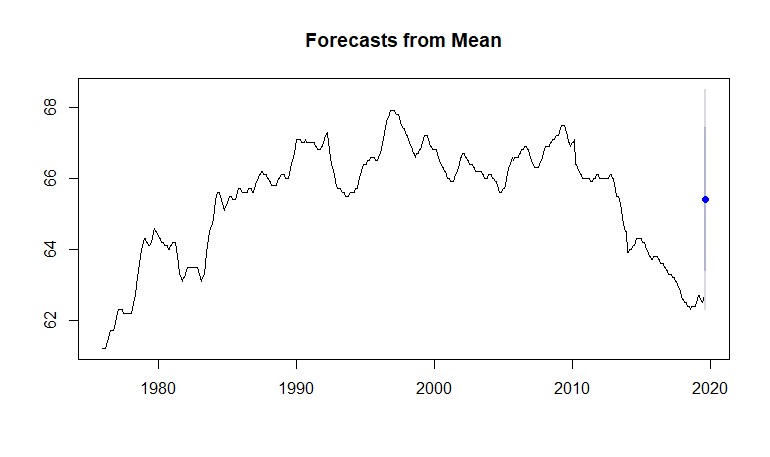
1. **Estimate the four simple forecasting models to forecast the next period values for both variables:**
2. **Average Method**
3. **Naïve Method**
4. **Drift Method**
5. **Seasonal Naïve Method**

*Note: Below each question are time-series graphs that depict the expected forecasts of the LFPR and UR variables based on each method. Also, all forecasts are rounded to two decimal places.*

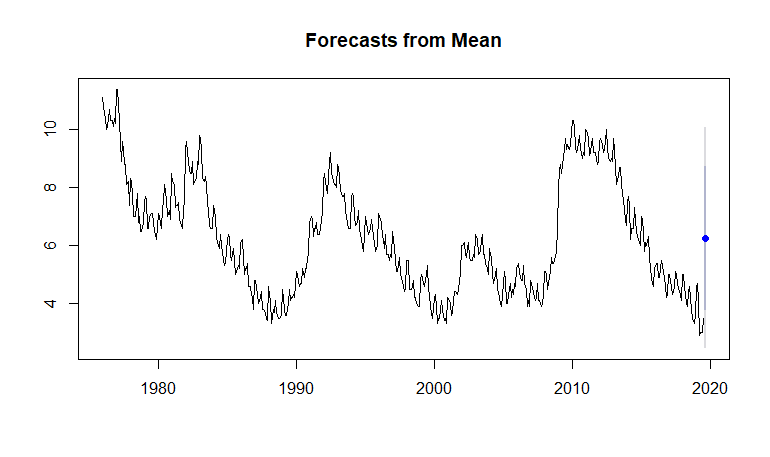
2a. Using the average method, the labor force participation rate in NJ is expected to be approximately 65.40% in September 2019 (as shown in Graph 3.1a).

Using the average method, the unemployment rate in NJ is expected to be approximately 6.26% in September 2019 (as shown in Graph 3.1b).

Graph 3.1a:



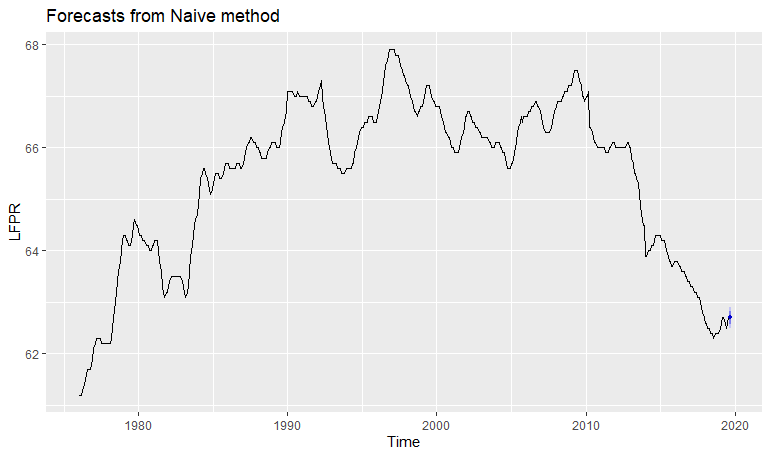
Graph 3.1b:



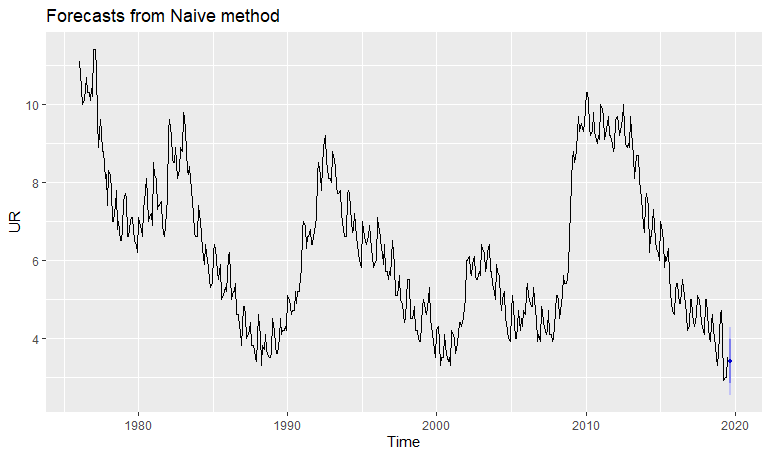
2b. Using the naïve method, the labor force participation rate in NJ is expected to be approximately 62.70% in September 2019 (as shown in Graph 3.2a).

Using the naïve method, the unemployment rate in NJ is expected to be approximately 3.40% in September 2019 (as shown in Graph 3.2b).

Graph 3.2a:



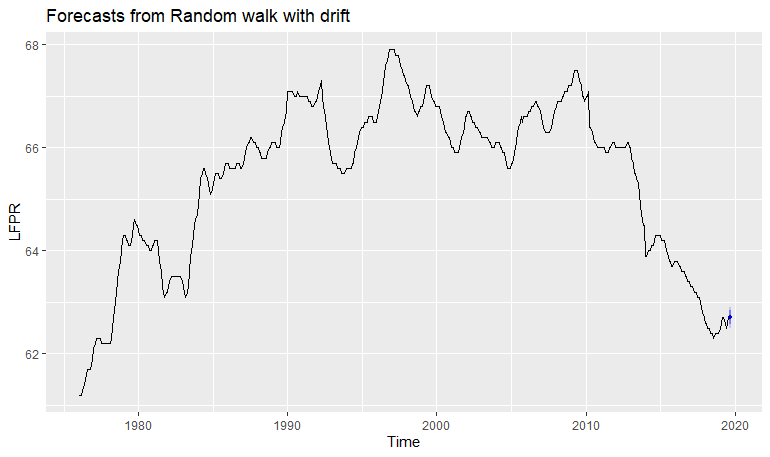
Graph 3.2b:



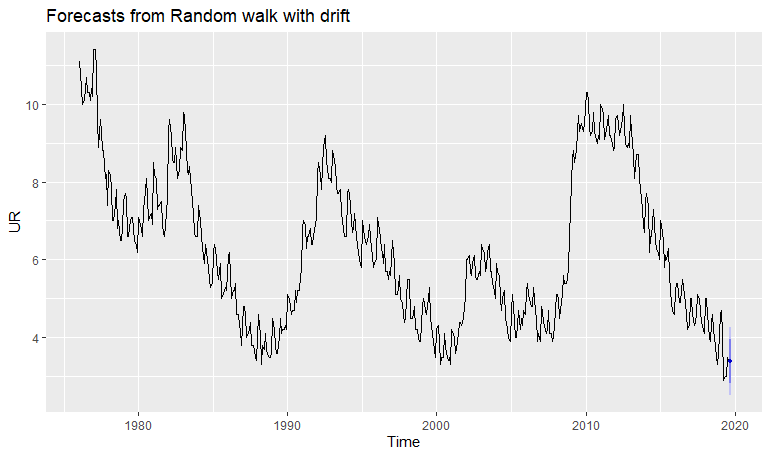
2c. Using the drift method, the labor force participation rate in NJ is expected to be approximately 62.70% in September 2019 (as shown in Graph 3.3a).

Using the drift method, the unemployment rate in NJ is expected to be approximately 3.39% in September 2019 (as shown in Graph 3.3b).

Graph 3.3a:



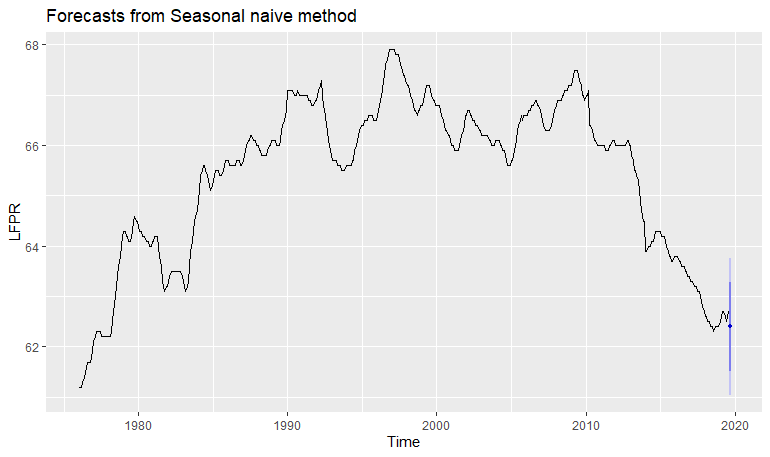
Graph 3.3b:



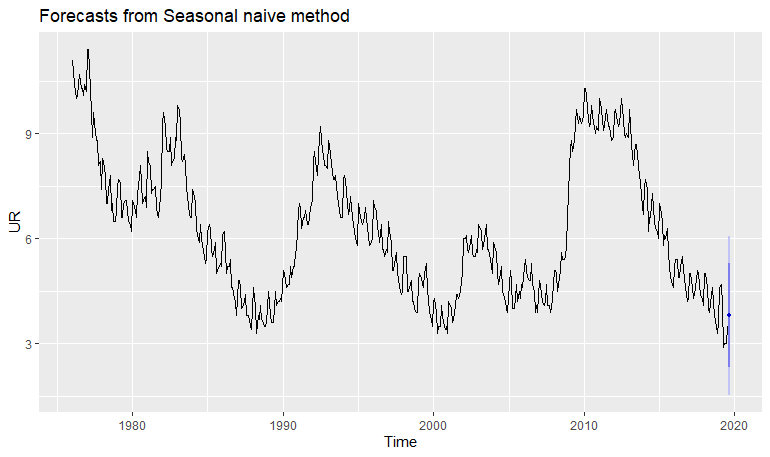
2d. Using the seasonal naïve method, the labor force participation rate in NJ is expected to be approximately 62.40% in September 2019 (as shown in Graph 3.4a).

Using the seasonal naïve method, the unemployment rate in NJ is expected to be approximately 3.80% in September 2019 (as shown in Graph 3.4b).

Graph 3.4a:



Graph 3.4b:

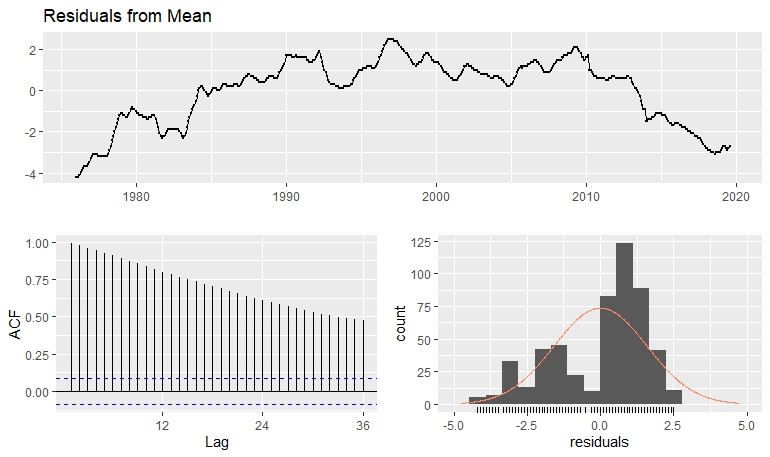


1. **For each variable, which model produces forecasts that are more reliable? Please explain how you arrived at the conclusion for choosing the right model.**

While all four of the methods provide forecast estimates for each variable, only one method for each variable can provide the most reliable estimates. To do this, we must check the residuals of each method performed on each variable. Below are time-series graphs that depict the residuals each in a different statistical fashion of each variables’ models.

Graphs 4.1a, 4.1b, 4.1c, and 4.1d all depict the residuals and lags of the average, naïve, drift, and seasonal naïve methods performed on LFPR. Below the graphs is R output that shows statistics based on the Ljung-Box test and an explanation of whether the model in question is reliable. Every test is different since it is based on the residuals of each corresponding method for the LFPR variable.

Graphs 4.1a:



Ljung-Box test

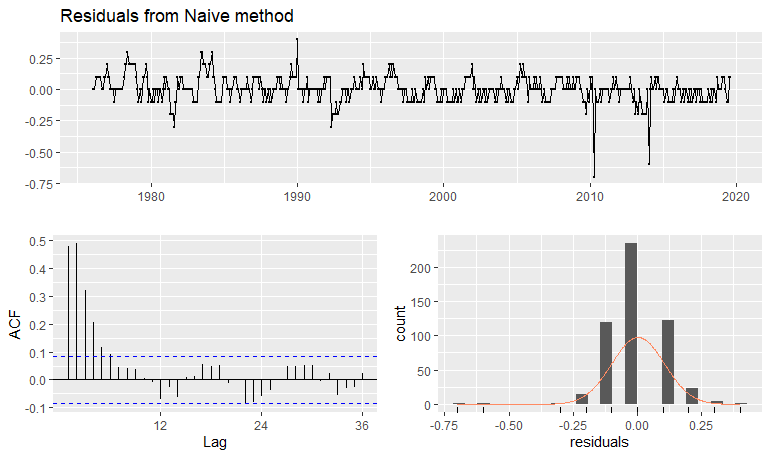
data: Residuals from Mean

Q\* = 8340.6, df = 23, p-value < 2.2e-16

Model df: 1. Total lags used: 24

Examining Graphs 4.1a, we see that the residuals do not seem to have a mean zero, the lags consistently pass the upper blue significance threshold, residuals are not normally distributed, and the p-value is less than 0.05. By these statistics, we can reject the H0 – there is no serial correlation between the residuals. Because we can reject the H0, we can also assert that the average method’s forecasts for the labor force participation rate in NJ are not reliable.

Graps 4.1b:



Ljung-Box test

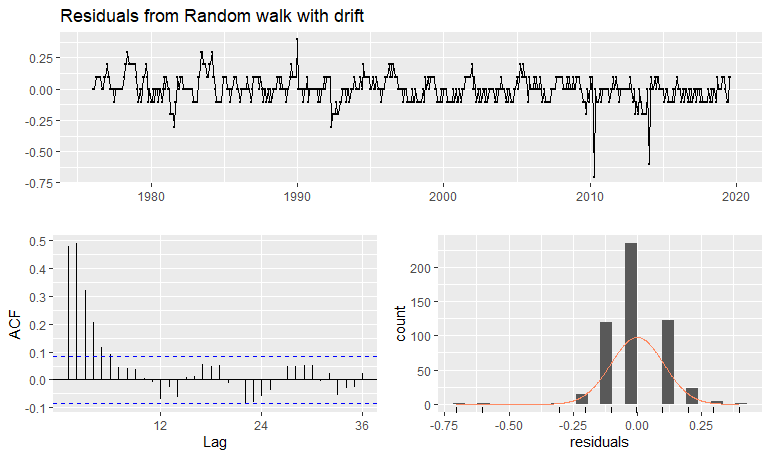
data: Residuals from Naive method

Q\* = 358.45, df = 24, p-value < 2.2e-16

Model df: 0. Total lags used: 24

Examining Graphs 4.1b, we see that the residuals seem to almost have a mean zero, some lags pass the upper blue significance threshold, residuals are not normally distributed, and the p-value is less than 0.05. By these statistics, we can reject the H0 – there is no serial correlation between the residuals. Because we can reject the H0, we can also assert that the naïve method’s forecasts for the labor force participation rate in NJ are not reliable.

Graphs 4.1c:



Ljung-Box test

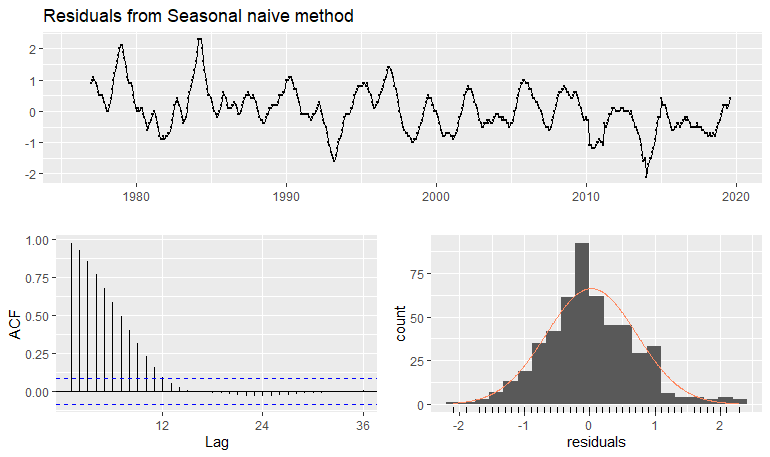
data: Residuals from Random walk with drift

Q\* = 358.45, df = 23, p-value < 2.2e-16

Model df: 1. Total lags used: 24

Examining Graphs 4.1c, we see that the residuals seem to almost have a mean zero, some lags pass the upper blue significance threshold, residuals are not normally distributed, and the p-value is less than 0.05. By these statistics, we can reject the H0 – there is no serial correlation between the residuals. Because we can reject the H0, we can also assert that the drift method’s forecasts for the labor force participation rate in NJ are not reliable.

Graphs 4.1d:



Ljung-Box test

data: Residuals from Seasonal naive method

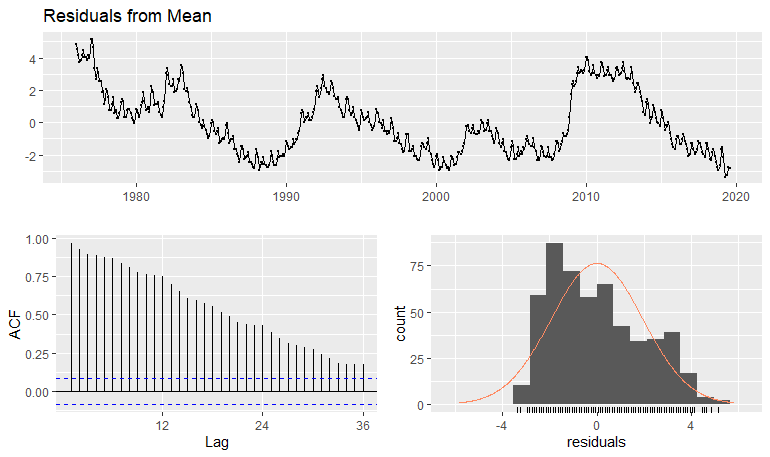
Q\* = 2361.2, df = 24, p-value < 2.2e-16

Model df: 0. Total lags used: 24

Examining Graphs 4.1d, we see that the residuals seem to almost have a mean zero, the lags consistently pass the upper blue significance threshold, residuals are not normally distributed (but is closer compared to the other graphs), and the p-value is less than 0.05. By these statistics, we can reject the H0 – there is no serial correlation between the residuals. Because we can reject the H0, we can also assert that the naïve seasonal method’s forecasts for the labor force participation rate in NJ are not reliable.

Graphs 4.2a, 4.2b, 4.2c, and 4.2d all depict the residuals and lags of the average, naïve, drift, and seasonal naïve methods performed on UR. Below each graph is R output showing statistics based on the Ljung-Box test and an explanation of whether the model in question is reliable. Every test is different since it is based on the residuals of each corresponding method for the UR variable.

Graphs 4.2a:



Ljung-Box test

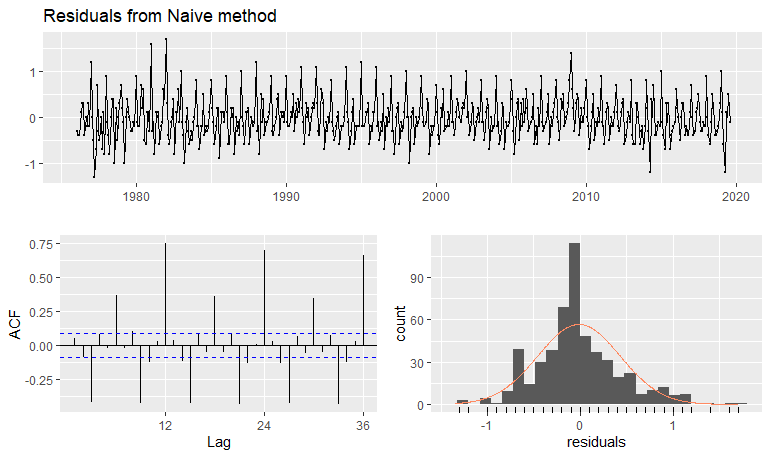
data: Residuals from Mean

Q\* = 6506.8, df = 23, p-value < 2.2e-16

Model df: 1. Total lags used: 24

Examining Graphs 4.2a, we see that the residuals seem to almost have a mean zero, the lags consistently pass the upper blue significance threshold, residuals are not normally distributed, and the p-value is less than 0.05. By these statistics, we can reject the H0 – there is no serial correlation between the residuals. Because we can reject the H0, we can also assert that the average method’s forecasts for the unemployment rate in NJ are not reliable.

Graphs 4.2b:



Ljung-Box test

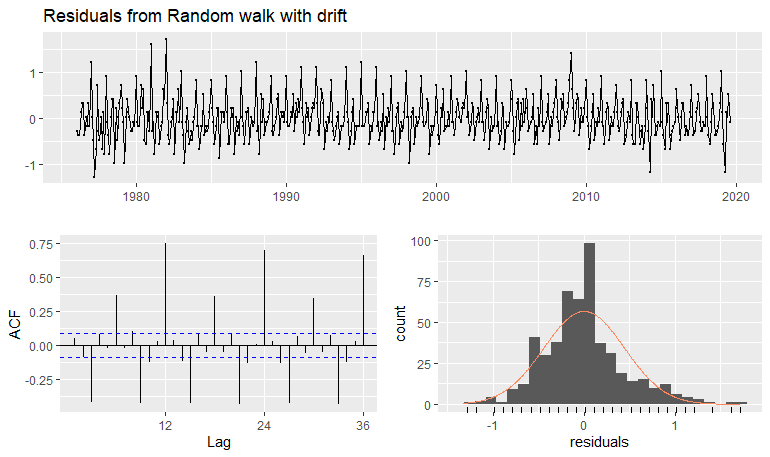
data: Residuals from Naive method

Q\* = 1154.4, df = 24, p-value < 2.2e-16

Model df: 0. Total lags used: 24

Examining Graphs 4.2b, we see that the residuals seem to almost have a mean zero, the lags consistently pass the upper and lower blue significance thresholds, residuals are not normally distributed, and the p-value is less than 0.05. By these statistics, we can reject the H0 – there is no serial correlation between the residuals. Because we can reject the H0, we can also assert that the naïve method’s forecasts for the unemployment rate in NJ are not reliable.

Graphs 4.2c:



Ljung-Box test

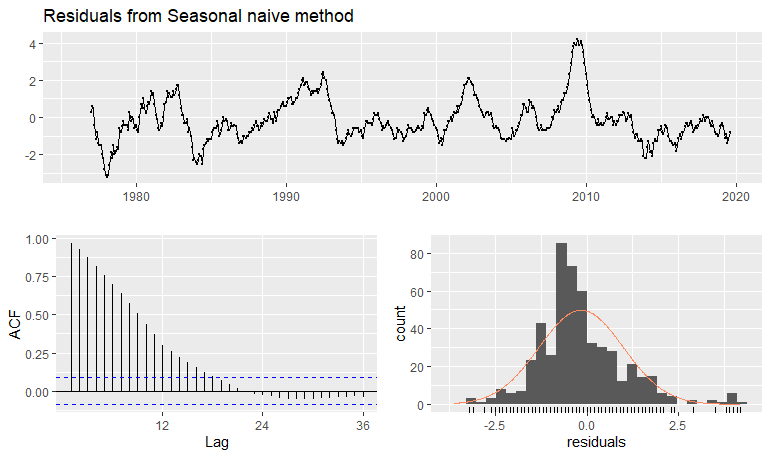
data: Residuals from Random walk with drift

Q\* = 1154.4, df = 23, p-value < 2.2e-16

Model df: 1. Total lags used: 24

Examining Graphs 4.2c, we see that the residuals seem to almost have a mean zero, the lags consistently pass the upper and lower blue significance thresholds, residuals are not normally distributed, and the p-value is less than 0.05. By these statistics, we can reject the H0 – there is no serial correlation between the residuals. Because we can reject the H0, we can also assert that the drift method’s forecasts for the unemployment rate in NJ are not reliable.

Graphs 4.2d:



Ljung-Box test

data: Residuals from Seasonal naive method

Q\* = 3079.9, df = 24, p-value < 2.2e-16

Model df: 0. Total lags used: 24

Examining Graphs 4.2d, we see that the residuals do not seem to have a mean zero, lags consistently pass the upper blue significance threshold, residuals are not normally distributed, and the p-value is less than 0.05. By these statistics, we can reject the H0 – there is no serial correlation between the residuals. Because we can reject the H0, we can also assert that the seasonal naïve method’s forecasts for the unemployment rate in NJ are not reliable.

Overall, based on the descriptive and inferential statistics shown above for all four methods performed on each variable, it can be said that no method is reliable in forecasting the labor force participation rate (LFPR) and unemployment rate (UR) in NJ next month (September 2019).

**Appendix A – R Code**

# Joel Cabrera

# Economic Forecasting and Big Data (01:220:423)

# Professor Berisha

# September 24, 2019

#Loading data

data=read.csv(file.choose(), header=TRUE)

summary(data)

#libraries MUST BE RAN FIRST

library(TSA)

library(dplyr)

library(ggplot2)

library(forecast)

library(tseries)

#variables becoming time series

#1

LFPR=ts(data$LBSSA34, frequency=12, start = c(1976,1))

UR=ts(data$NJURN, frequency=12, start = c(1976,1))

autoplot(LFPR)+xlab("Year") + ylab("Labor Force Participation Rate in NJ From 1976-2019")

autoplot(UR)+xlab("Year") + ylab("Unemployment Rate in NJ From 1976-2019")

#Note: can also use plot(), has different style

acf(LFPR)

acf(UR)

#ggAcf(LFPR)

#ggAcf(UR)

#Note: can also use ggAcf(), has different style

#2: Forecast Methods - for each variable

#2a. Average

plot(meanf(LFPR, h=1)) #forecastss next period #Note: blue dot = mean/most likely will be; 80% CI is dark blue; 95% CI is lighter bar

meanf(LFPR, h=1) #h = 1 -> forecasts one period forward

plot(meanf(UR, h=1)) #forecastss next period

meanf(UR, h=1) #h = 1 -> forecasts one period forward

#2b. Naive

LFPR1=naive(LFPR, h=1) #LFPR1 = renaming for naive

LFPR1

autoplot(naive(LFPR, h=1)) #can do plot, too, to check

UR1=naive(UR, h=1) #UR1 = renaming for naive

UR1

autoplot(naive(UR, h=1)) #can do plot, too, to check

#2c. Drift

LFPR2=rwf(LFPR, h=1, drift = TRUE)

LFPR2 #different name and gives dif. forecast #

autoplot(LFPR2)

UR2=rwf(UR, h=1, drift = TRUE)

UR2 #different name and gives dif. forecast #

autoplot(UR2)

#2d. Seasonal Naive

LFPR3=snaive(LFPR, h=1)

LFPR3

autoplot(LFPR3)

UR3=snaive(UR, h=1)

UR3

autoplot(UR3)

#3.Which model is most reliable?

#Average

checkresiduals(meanf(LFPR))

checkresiduals(meanf(UR))

#Naive

checkresiduals(naive(LFPR))

checkresiduals(naive(UR))

#Drift

checkresiduals(rwf(LFPR, drift = TRUE))

checkresiduals(rwf(UR, drift = TRUE))

#Seasonal Naive

checkresiduals(snaive(LFPR))

checkresiduals(snaive(UR))