Pallavi Sethi

Assignment 1 : Forecasting Methods

Time Series Analysis

*Unemployment Rate of College Graduates with Master's Degree for 25 years and over*

***INDEX***

|  |  |
| --- | --- |
| ***Topics*** | ***#Page*** |
| ***Introduction*** | ***2*** |
| ***Plotting Time Series*** | ***3*** |
| ***Seasonality*** | ***4*** |
| ***Sub Series Plot*** | ***5*** |
| ***Lag Plot*** | ***6*** |
| ***Acf Plot*** | ***7*** |
| ***Conclusion*** | ***8*** |
| ***Appendix*** | ***9*** |

***Introduction:***

United States has been witnessing an unprecedented spike in jobless claims in the last few weeks. In addition to a health and economic crisis, US is now approaching the devastating period of massive unemployment, one that – we are told by decades of research – will leave deep financial and emotional scars for the workers and families trapped within it. Currently unemployment rate in US is at an all-time high, representing the amount of people taking positions lower than their education or experience level. These people are stuck in roles related to gig economy, temporary work or lower-end positions.

The lack of jobs available, and employers' desired skills, are beginning to prove to be major cause of U.S. graduate unemployment. Graduates complete school with a degree and a knowledge-filled head, but still lack work experience to impress employers with white collars. This is emotionally exhausting. I am currently pursuing my Master’s degree in US and this topic is especially concerning for me. This issue saps the energy and self-confidence and stirs fears for the future employment a graduate student hoped for.

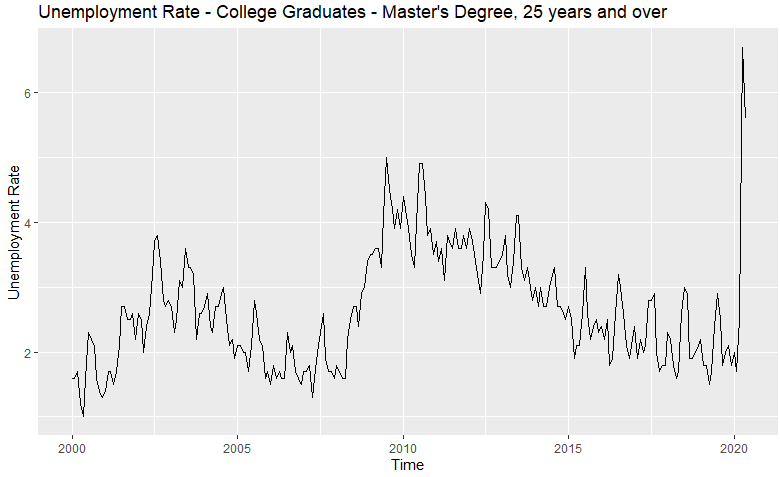
In order to shed more light on this issue. I got a time series dataset “Unemployment Rate.csv“ on Unemployment Rate of College Graduates with Master's Degree for 25 years and over, from the below link : <https://fred.stlouisfed.org/series/CGMD25O>.

The data set I selected has two columns, viz Date and URate. The date range is from Jan/2000 till May/2020. In order to load the data into R environment run and for setting the working directory run ***steps 1-3*** from the appendix .

For doing time series analysis we need to convert the csv data into ts() object. The data I collected is the monthly data so I stored URate as a numerical vector Rate and then converted it to a ts object tsRate by running ***step 4***,also check the class of the tsRate using Class() and str() function to check the start and end time of the data.

***Plotting Time Series***

Once the time series data is loaded into R, the next step is to make a plot of the time series. Run ***step 5*** from the appendix. Below time series graph shows the data collected for the Unemployment Rate from 2000 to 2020.



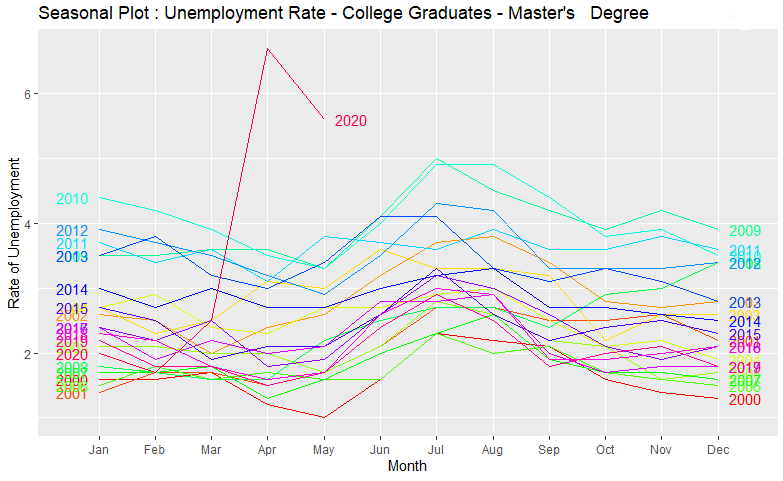
*Findings from the Time plot*:

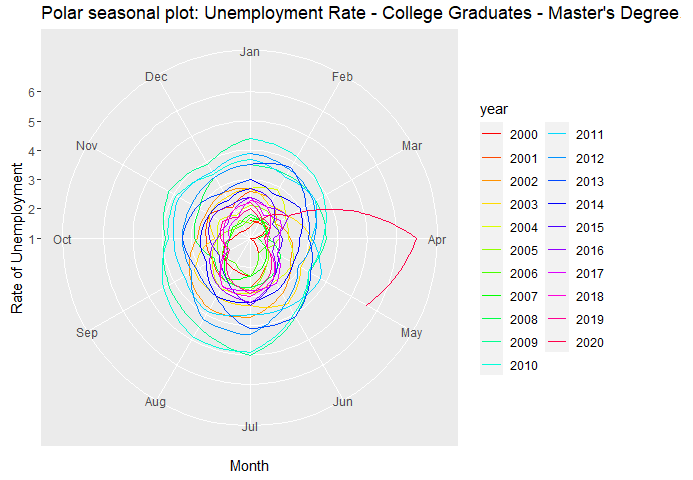
1. Unemployment rate is gradually increasing among graduates every 10 year.
2. Unemployment rate shows some strong cyclic behavior with a period of about 6–10 years
3. The 2001 recession and the economic slowdown that lasted until 2003 altered outcomes for all workers. Unemployment rates increased and the recovery began in the middle of 2003, which brought down the Unemployment rate considerably low.
4. Then again in 2008 recession there was a sudden spike in rate, which resulted in around 5%, the highest by 2010.
5. The recent change in Unemployment rate came in 2020, soars to 6.7% by May2020.

***Seasonality***:

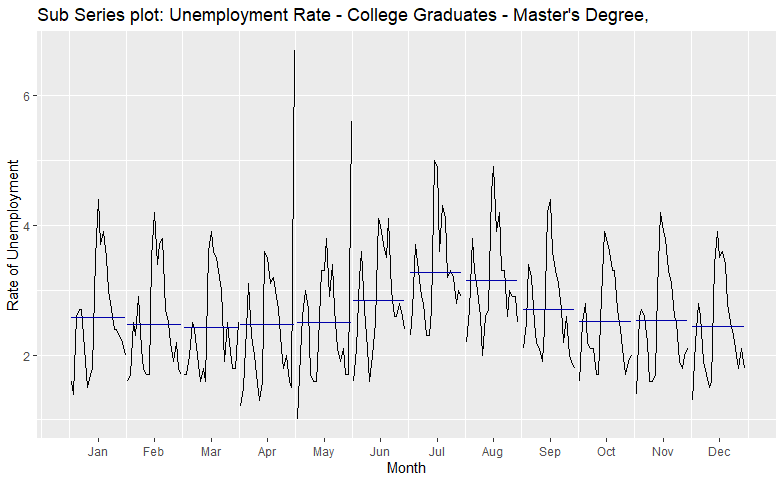
Many time series display seasonality. Seasonality mean periodic fluctuations. A seasonal plot allows the underlying seasonal pattern to be seen more clearly and is especially useful in identifying years in which the pattern changes.

Run ***step-6*** from the appendix for plotting Seasonal plot and Polar Seasonal plot.

**



***Seasonal Subseries Plot:***



The seasonal subseries plot containing monthly data of Unemployment Rate and reveals a strong seasonality pattern.

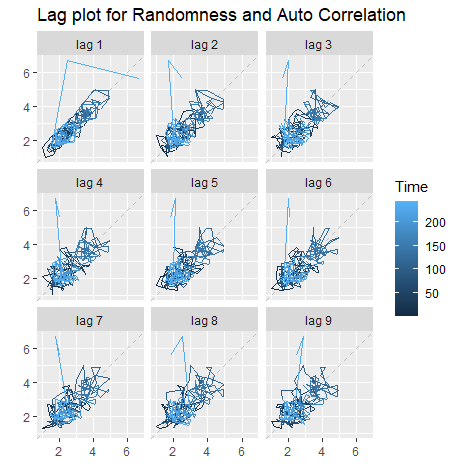
Sub Series Plot reveals below points:

1. From the plot we can see that July has the highest average Unemployment rate.
2. From January till July there is an increase in the Unemployment rate and then after July the rate tends to decrease slowly.
3. This form of plot highlighted the underlying seasonal pattern and also shows the changes in seasonality over time.

***Lag-Plot***

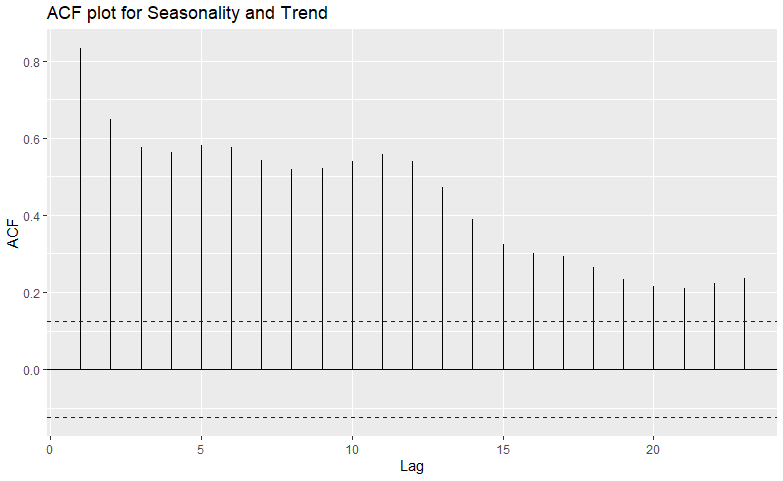
A lag plot checks whether the time series is random or not. Random data should not exhibit any identifiable structure in the lag plot. Non-random structure in the lag plot indicates that the underlying data are not random.

Run ***step-7*** for creating the lag Plot.



### ***ACF plots*** *also known as* correlogram.

In this plot the autocorrelation coefficients are plotted against the lag. Run **step 7** for creating ACF plot



The above ACF plot highlighted below points:

1. From the Seasonal plot we can say that the data shows seasonality and strong cyclicity, and the ACF plot confirms the autocorrelations seems to be large and positive.
2. The decrease in the ACF as the lags increase is due to the trend, while the “cave” shape is due the seasonality.
3. From the ACF plot we can confer that the time series for the Unemployment Rate doesn’t contain White Noise, since all the ACF spikes are outside these bounds of ±2/√T(dotted blue line) , where T is the length of the Time Series.

***Conclusion***

Historically, each time the global economy endured a crisis, unemployment has also risen. Going by the analysis we can say that the most of the recent unemployment rate rise can be attributed to cyclical causes and the statistics from the previous Unemployment surge rate says that, the rate might typically peaks about 15 months after the crisis begins, or 4 months after the crisis ends, and then begins to gradually fall over time as the economy recovers.

***Appendix:***

***Step 1: Setting working directory***

setwd("C:/………./Summer2020/Forecasting/Assignments/1")

***Step 2: Loading the required library***

library(ggplot2)

library(fpp2)

library(forecast)

library(corrplot)

library(ggcorrplot)

***Step 3: Loading the csv file into R environment***

Rawdata = read.csv("Unemployment Rate.csv", header= TRUE)

***Step 4: Converting dataframe into ts() object***

*Rate* <- data[,"URate"] *# This creates Rate vector from Rawdata dataframe*

tsRate = ts(Rate,start=2000, frequency = 12) *# This convert Rate vector to ts() object*

class(tsRate) *# tsRate is of class “ts”*

str(tsRate) *# to see the structure of the tsRate*

*>> Time-Series [1:245] from 2000 to 2020: 1.6 1.6 1.7 1.2 1 1.6 2.3 2.2 2.1 1.6….*

***Step 5: Creating Time Plot***

autoplot(tsRate) +

xlab("Time") + ylab("Unemployment Rate") +

ggtitle("Unemployment Rate - College Graduates - Master's Degree, 25 years and over")

***Step 6: Creating Seasonal Plot***

ggseasonplot(tsRate, col=rainbow(21),year.labels = TRUE,year.labels.left = TRUE) +

ylab("Rate of Unemployment") +

ggtitle("Seasonal Plot : Unemployment Rate - College Graduates - Master's Degree, 25 years and over")

*#For creating Polar Seasonal Plot*

ggseasonplot(tsRate, col=rainbow(21),polar= TRUE) +

ylab("Rate of Unemployment") +

ggtitle("Polar seasonal plot: Unemployment Rate - College Graduates - Master's Degree, 25 years and over")

*# For creating sub series plot*

ggsubseriesplot(tsRate) +

ylab("Rate of Unemployment") +

ggtitle("Unemployment Rate - College Graduates - Master's Degree, 25 years and over")

***Step 7: Creating Lag Plot & checking White Noice***

window <- ts(tsRate,start=2008)

gglagplot(window)

ggAcf(window)