Report

Rui Xing Wang

7/27/2021

Contents


```
if(!require(tseries)) install.packages("tseries", repos = "http://cran.us.r-project.org")
if(!require(forecast)) install.packages("forecast", repos = "http://cran.us.r-project.org")

library(tseries)
library(forecast)
```

1 Introduction

Sea ice is frozen seawater that floats on the ocean surface. Blanketing millions of square kilometers, sea ice forms and melts with the polar seasons, affecting both human activity and biological habitat. The monthly Sea Ice Index provides a quick look at Arctic-wide changes in sea ice. In this project, I am going to use ARIMA model to forecast the sea ice index in the future based on the data from the past.

2 Data

The data could be found online at: https://nsidc.org/data/seaice_index

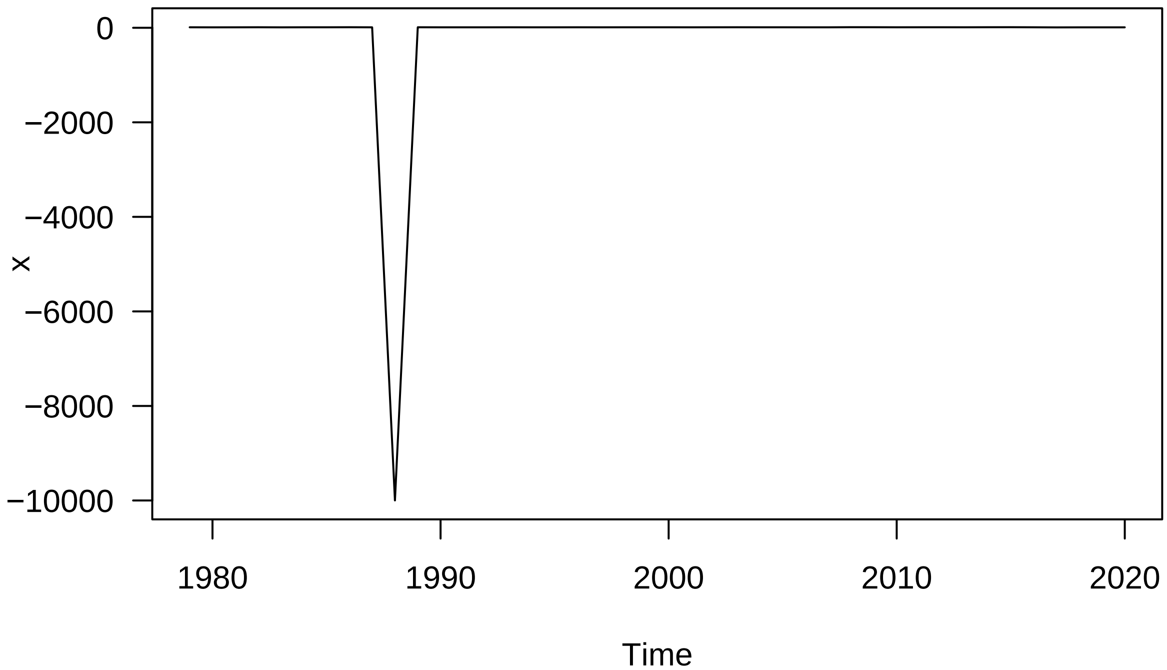
This data set have 6 columns and 43 rows, which includes "year", "mo", "data-type", "region", "extent", "area". And they have same "month number".

```
S12<- read.csv("./S_12_extent_v3.0.csv")
```

2.1 Preprocessing

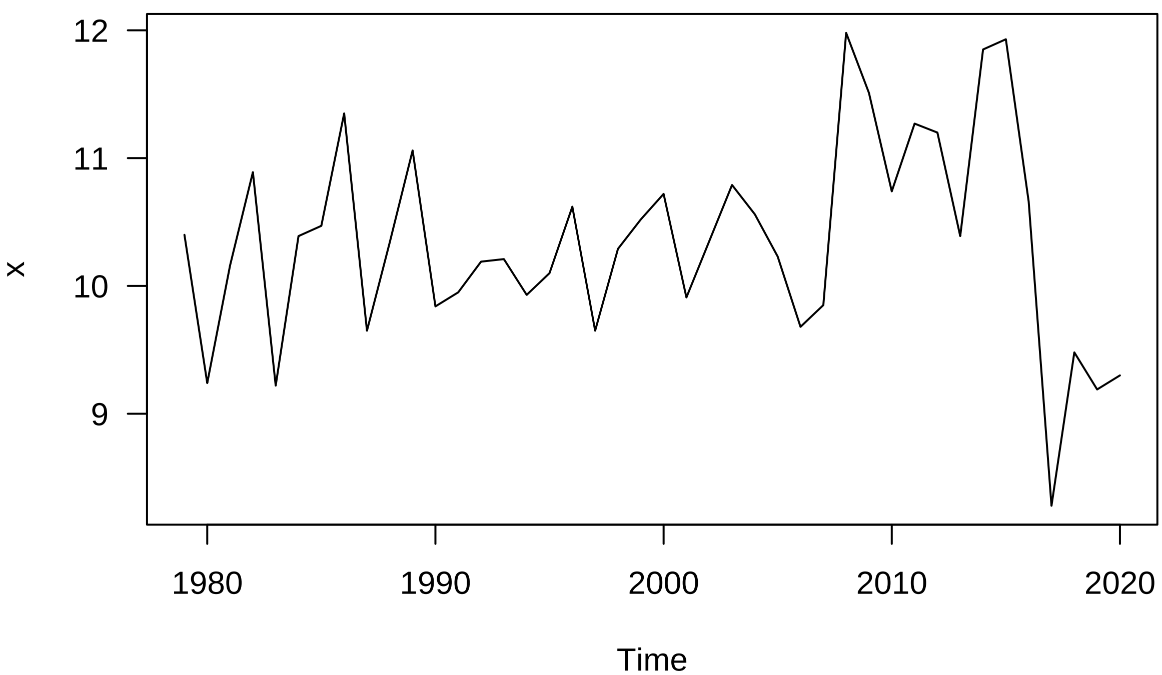
Before we build the model, we need to reduce the data to stationary Step1: firstly, I set extent as x and then plot it.

```
x<-ts(S12$extent,start=1979,frequency=1)
plot(x,las=1)
```



Step2: I find that there is a particular point, which can be seen as an outlier. So I set that value to be the mean value of the data, (see that plot now).

```
x[10]=NA
x[10]=mean(x,na.rm = T)
plot(x,las=1)
```



Step3: through "adf.test(x)" this code, I find that p-value of this data is 0.8016, which is bigger than 0.05, hence I got conclusion that this data set is not stationary.

```
adf.test(x)
```

```
##
##  Augmented Dickey-Fuller Test
```

```
##
## data:  x
## Dickey-Fuller = -1.4216, Lag order = 3, p-value = 0.8016
## alternative hypothesis: stationary
```

```r
diff(x)
```

```
## Time Series:
## Start = 1980
## End = 2020
## Frequency = 1
##  [1] -1.1600000  0.9200000  0.7300000 -1.6700000  1.1700000  0.0800000
##  [7]  0.8800000 -1.7000000  0.6914634  0.7185366 -1.2200000  0.1100000
## [13]  0.2400000  0.0200000 -0.2800000  0.1700000  0.5200000 -0.9700000
## [19]  0.6400000  0.2300000  0.2000000 -0.8100000  0.4400000  0.4400000
## [25] -0.2300000 -0.3300000 -0.5500000  0.1700000  2.1300000 -0.4700000
## [31] -0.7700000  0.5300000 -0.0700000 -0.8100000  1.4600000  0.0800000
## [37] -1.2700000 -2.3800000  1.2000000 -0.2900000  0.1100000
```

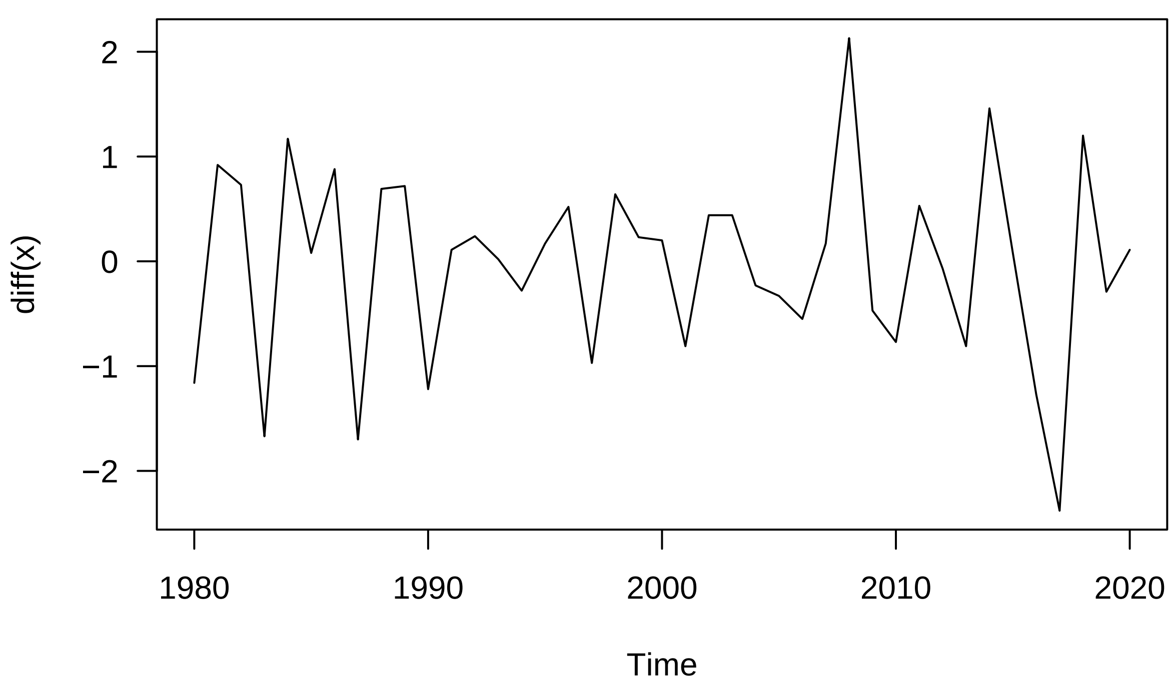
Step4: through differencing x, and then use "adf.test(diff(x))", I can see that if this data after differencing can be stationary. The result is that p-value is 0.02163, which is less than 0.05.

```r
adf.test(diff(x))
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  diff(x)
## Dickey-Fuller = -3.9458, Lag order = 3, p-value = 0.02163
## alternative hypothesis: stationary
```

Step5: then I got conclusion that diffenencing this data can make it to stationarity.

```r
plot(diff(x),las=1)
```

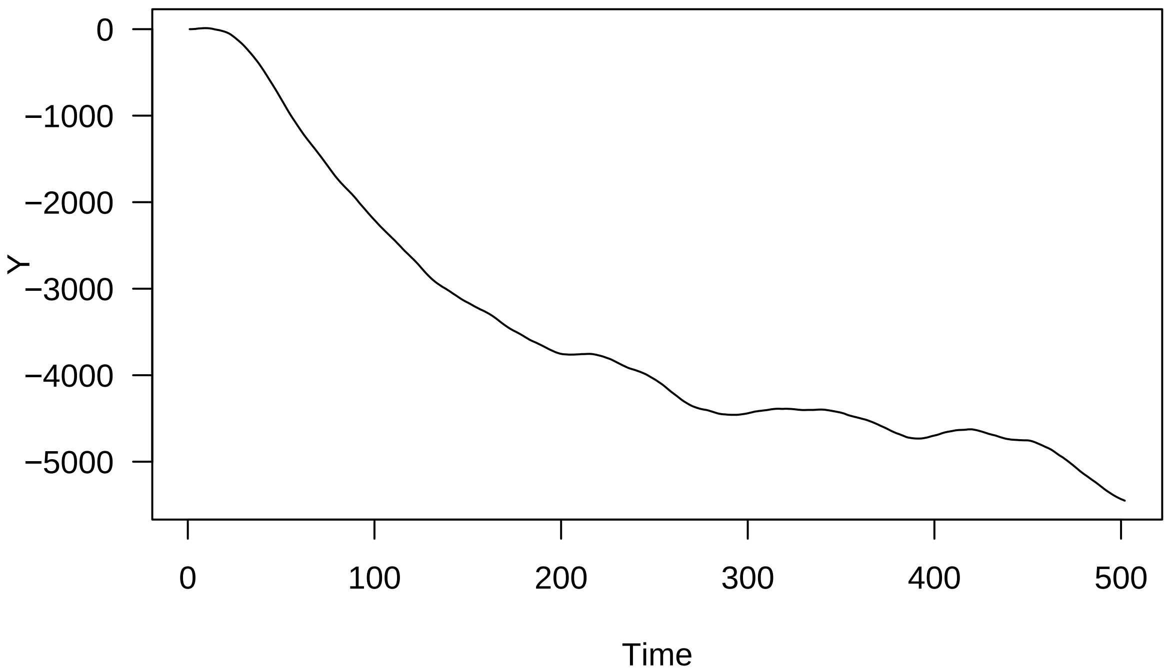


3 Methods

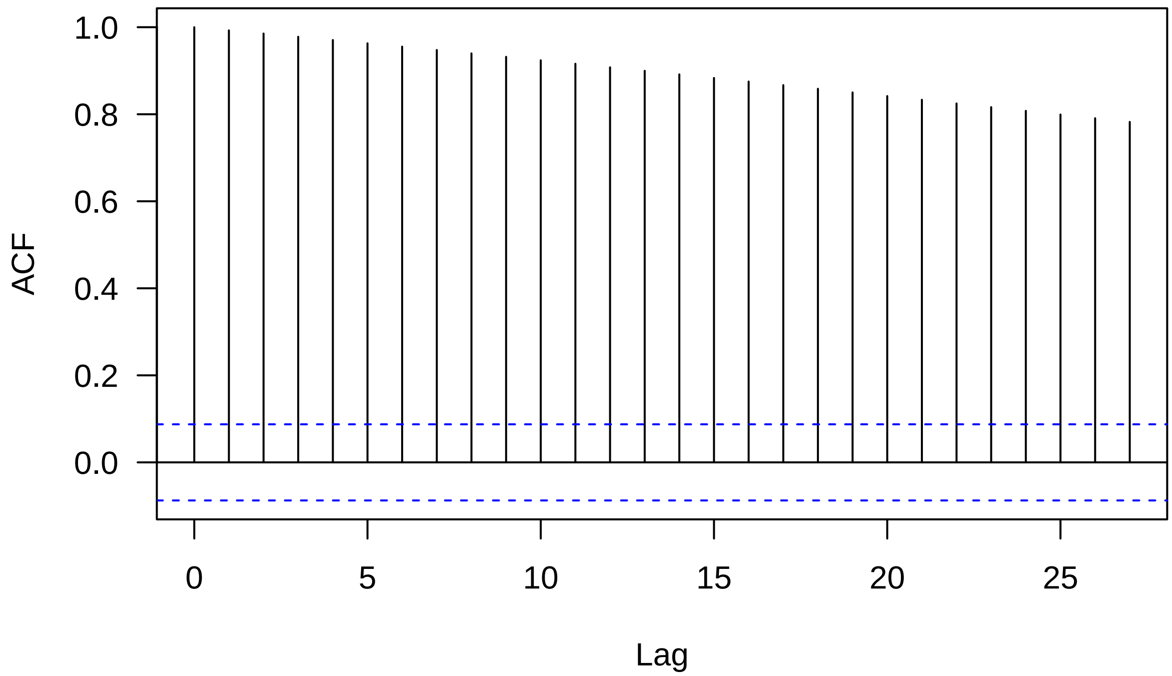
3.1 ARIMA

Autoregressive integrated moving average (ARIMA) model is a model that uses into time series data and predicts future points in the series (forecasting).

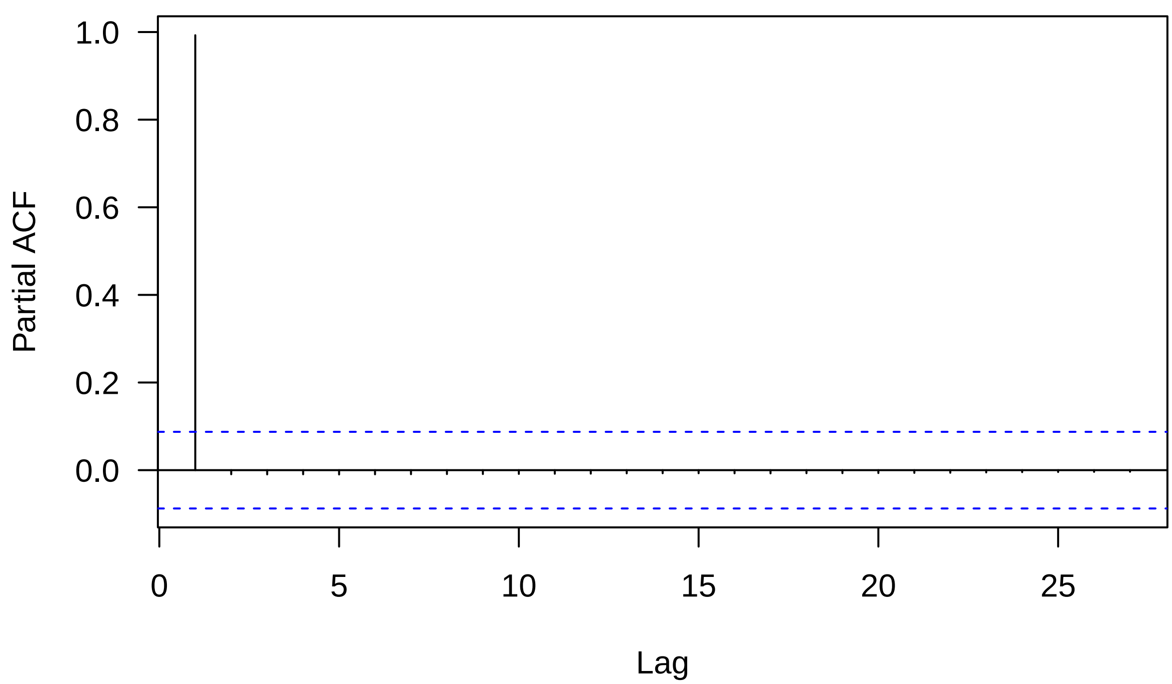
```
set.seed(116101521)
Y<-arima.sim(list(order=c(1,2,1),ar=c(0.7),ma=c(-0.5)),n=500)

plot(Y,las=1,main='')
```



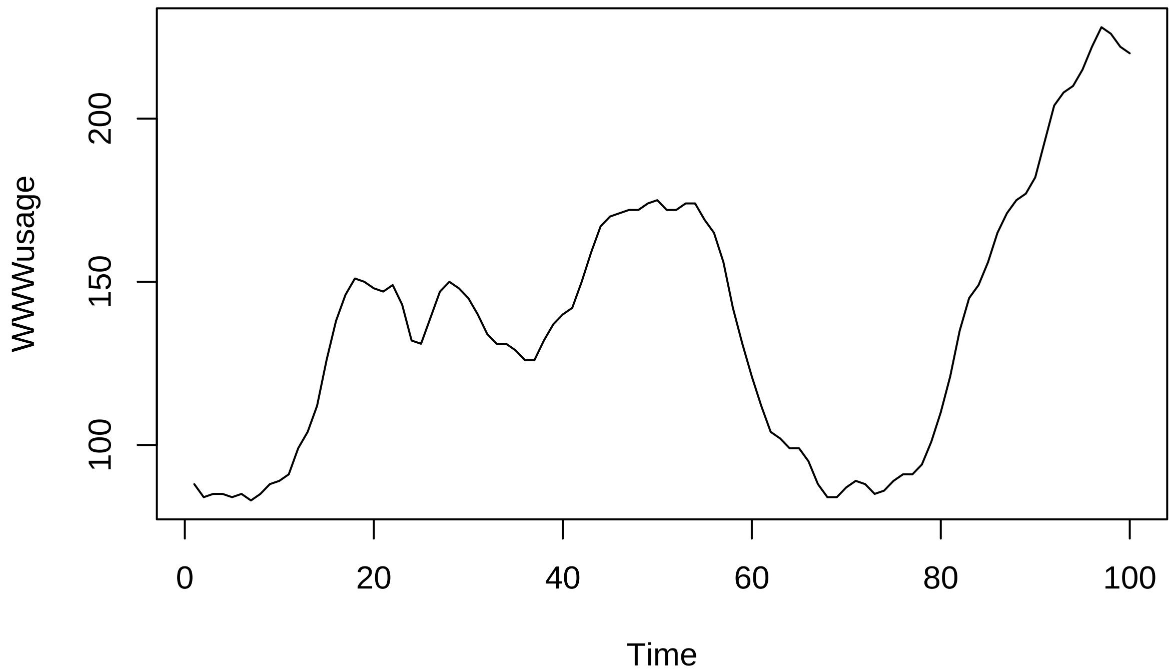

```
acf(Y,las=1,main='')
```




```
pacf(Y,las=1,main='')
```




```
plot(WWWusage)
```




```
aics <- matrix(, 6, 6, dimnames = list(p = 0:5, q = 0:5))
```

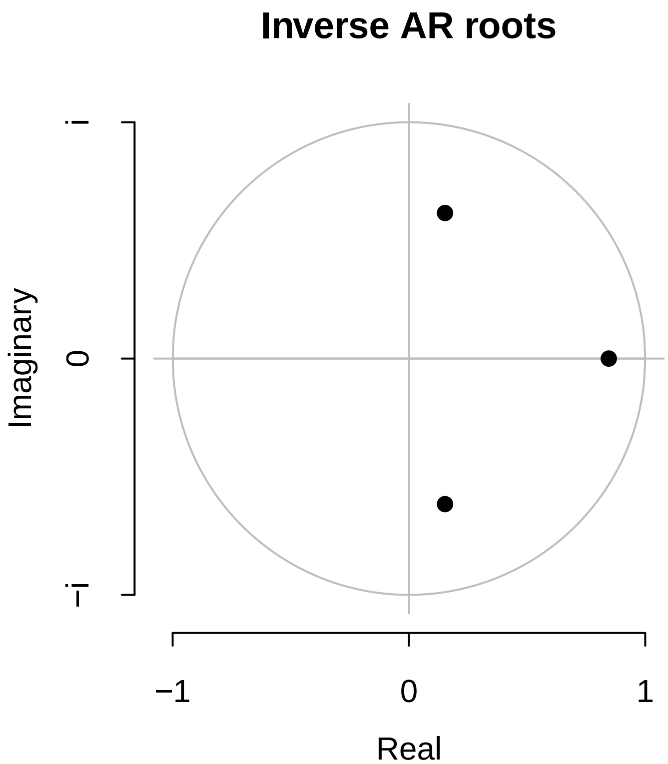
For ARIMA, the choices of p and q are important Through a “for loop”, The possible p and q values are brought into the for loop to obtain several possible models. Then the aic method is used to obtain the optimal p and q values that minimize aic.

```
for(q in 0:1) aics[1, 1+q] <- arima(WWWusage, c(0, 1, q),
                                    optim.control = list(maxit = 500))$aic


for(p in 1:3)
  for(q in 0:1) aics[1+p, 1+q] <- arima(WWWusage, c(p, 1, q),
                                        optim.control = list(maxit = 500))$aic
(round(aics - min(aics, na.rm = TRUE), 2))
```

```
##    q
## p        0     1  2  3  4  5
##   0 119.00 37.81 NA NA NA NA
##   1  17.24  2.31 NA NA NA NA
##   2  10.18  4.30 NA NA NA NA
##   3   0.00  1.94 NA NA NA NA
##   4     NA    NA NA NA NA NA
##   5     NA    NA NA NA NA NA
```

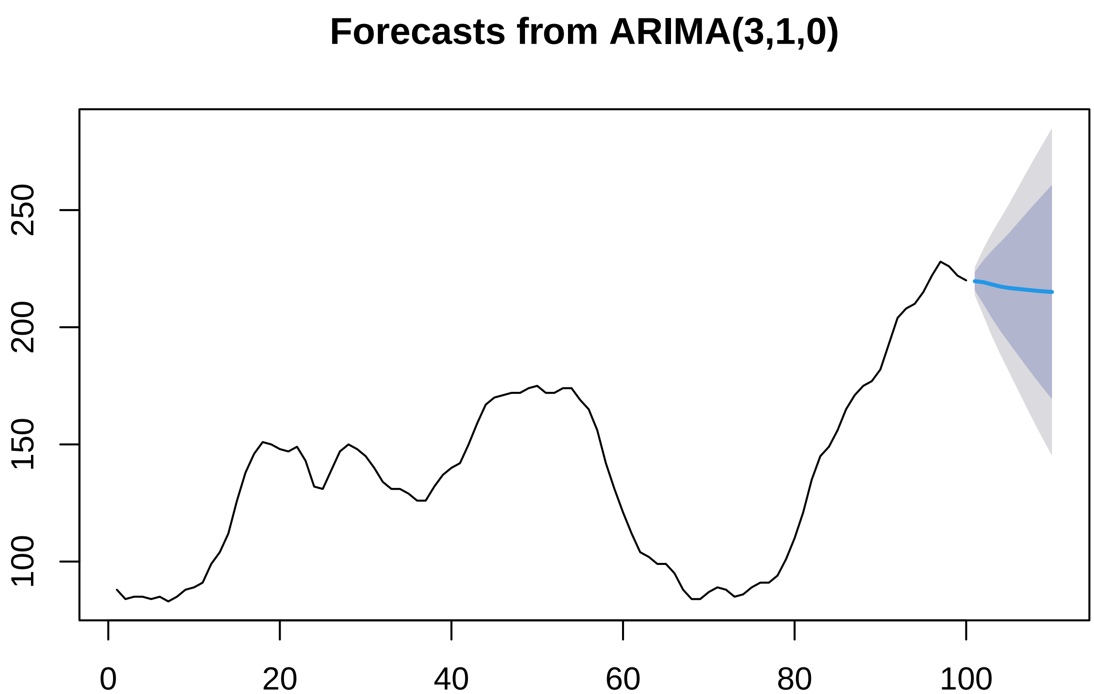
I can find the best p and q by seeking the min aic they’ve made. Then I find that when p=3, q=0, I can get the least aic.

```
Best<-arima(WWWusage, c(3, 1, 0),optim.control = list(maxit = 500))
plot(Best)
```



Use the best p and q I've find to establish the arima model. After establish that model, I plot that model and make forcast for it.

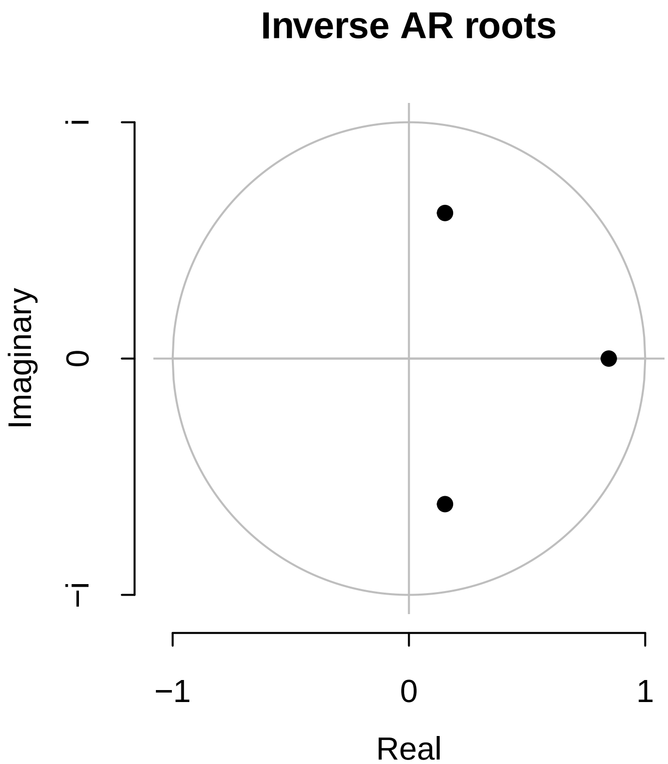
```
plot(forecast(Best),col = "black")
```



Results

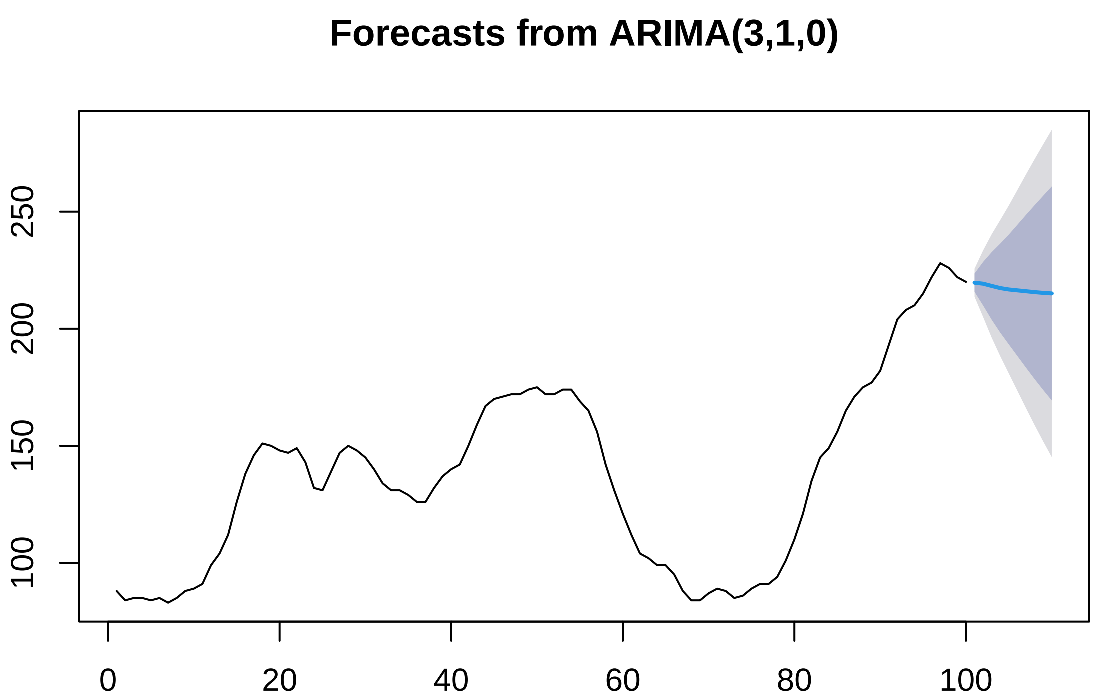
This result shows this AR model is not stationary since the plot that the three points all inside the unit circle.

```
plot(Best)
```



This diagram shows that in different confident interval, I can get the forecast shows in blue.

```r
plot(forecast(Best),col = "black")
```



4 Conclusion

In conclusion, I fit an ARIMA model into the Time Series data and forecast the upcoming data.