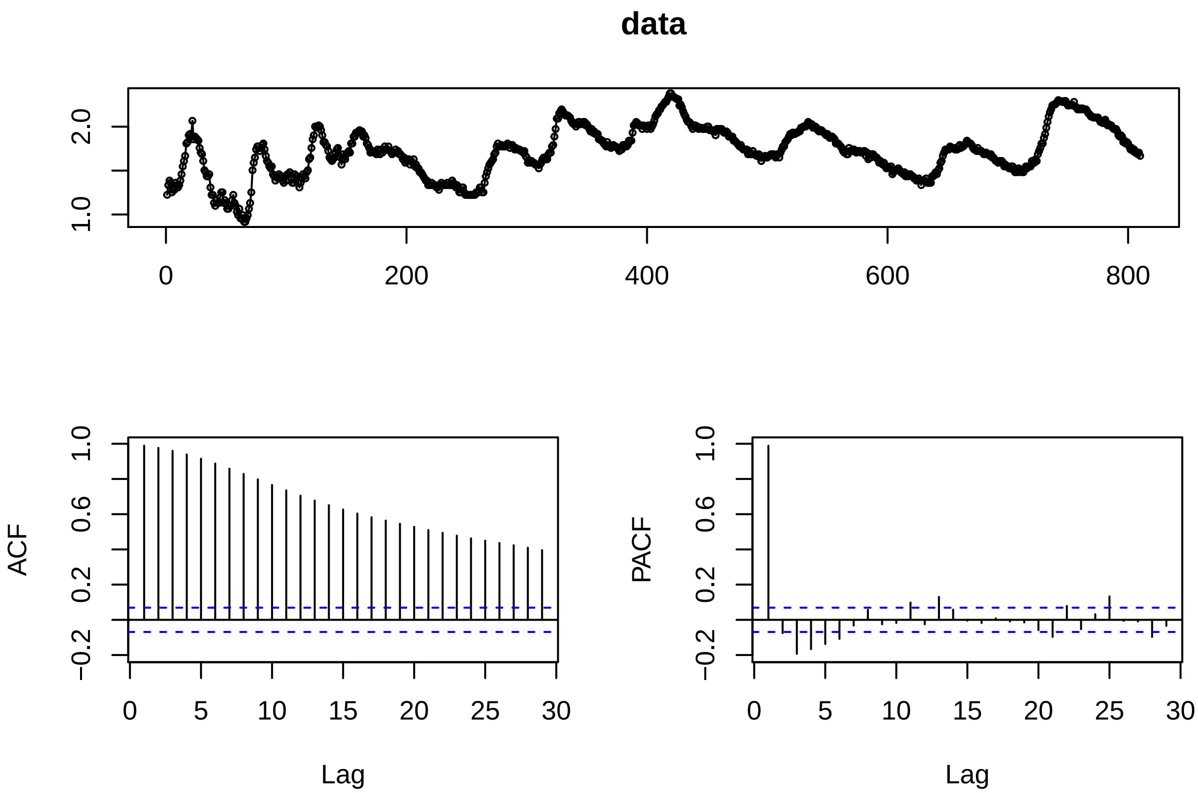
Forecasting Homework 5

Yuhao Zhao, Yz3085

October 28, 2015

A)

```r
library('itsmr')
library('forecast')
data = log(
  read.csv('http://people.stern.nyu.edu/churvich/Forecasting/Data/UNEMPLOYMENT.CSV')[,2])
tsdisplay(data)
```

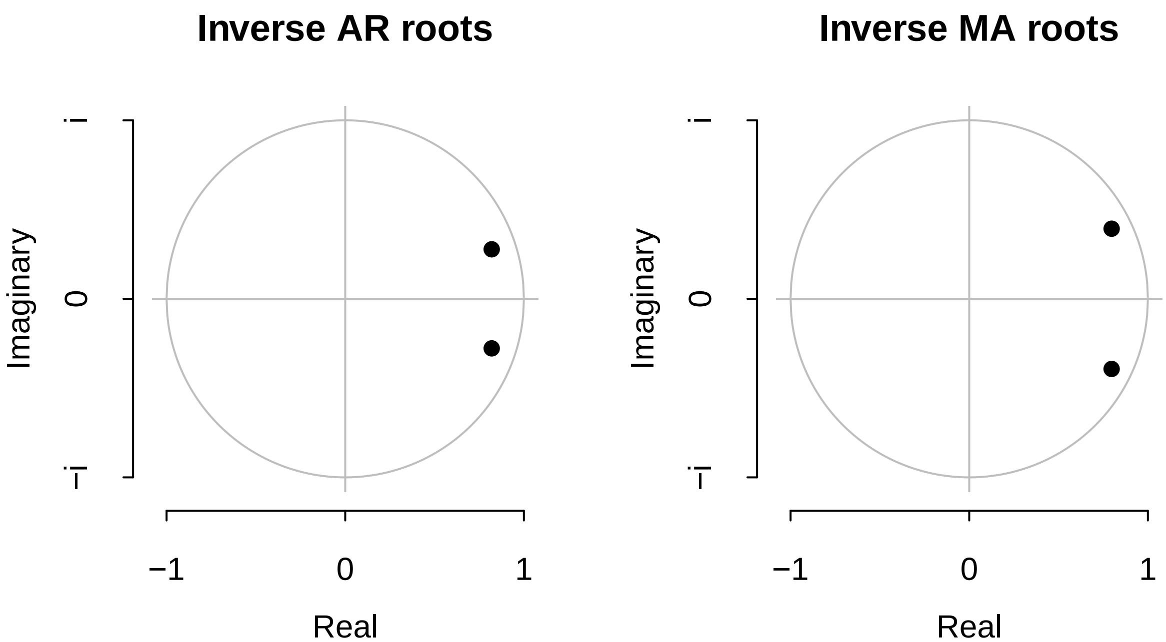



```r
model= auto.arima(data,parallel = F,trace = T,ic = "aicc",stepwise = T,allowmean = TRUE
                  ,allowdrift = T,seasonal =FALSE,max.p = 5,max.q=5 ,max.d = 3)
```

```
##
##  ARIMA(2,1,2) with drift         : -3066.565
##  ARIMA(0,1,0) with drift         : -2966.929
##  ARIMA(1,1,0) with drift         : -2986.767
##  ARIMA(0,1,1) with drift         : -2974.754
##  ARIMA(0,1,0)                    : -2968.774
##  ARIMA(1,1,2) with drift         : -3051.661
##  ARIMA(3,1,2) with drift         : -3062.786
```

```
##  ARIMA(2,1,1) with drift         : -3053.222
##  ARIMA(2,1,3) with drift         : -3062.507
##  ARIMA(1,1,1) with drift         : -3029.326
##  ARIMA(3,1,3) with drift         : -3060.944
##  ARIMA(2,1,2)                    : -3068.46
##  ARIMA(1,1,2)                    : -3053.685
##  ARIMA(3,1,2)                    : -3064.768
##  ARIMA(2,1,1)                    : -3055.244
##  ARIMA(2,1,3)                    : -3064.54
##  ARIMA(1,1,1)                    : -3031.346
##  ARIMA(3,1,3)                    : -3062.923
##
##  Best model: ARIMA(2,1,2)
```

```r
plot.Arima(model)
```



The ACF and PACF plot shows that the simple AR or MA model is not sufficient to model the log of unemployment data. Therefore, we need to use AICc to identify the ARIMA model. I tried p from 0 to 5, q from 1 to 5, and d from 0 to 3.The stationary model with lowest AICc correspond to the model ARIMA(2,1,2) without constant. The characteristic roots of the model shows that the model is stationary and invertible.

B)

```r
pvalue = (1-pnorm(abs(model$coef)/sqrt(diag(model$var.coef))))*2;pvalue
```

```
## ar1 ar2 ma1 ma2
##   0   0   0   0
```

We assusume the dimension is high enough for the z value to aprroximate t value. From the p-value we get from the model, all parameters are significantly different from 0. The model with constant has larger AICc than the model without constant, therefore, we should not include the constant.

C)

```r
model$coef
```

```
##        ar1        ar2        ma1        ma2 
##  1.6398038 -0.7492623 -1.5941598  0.7895618
```

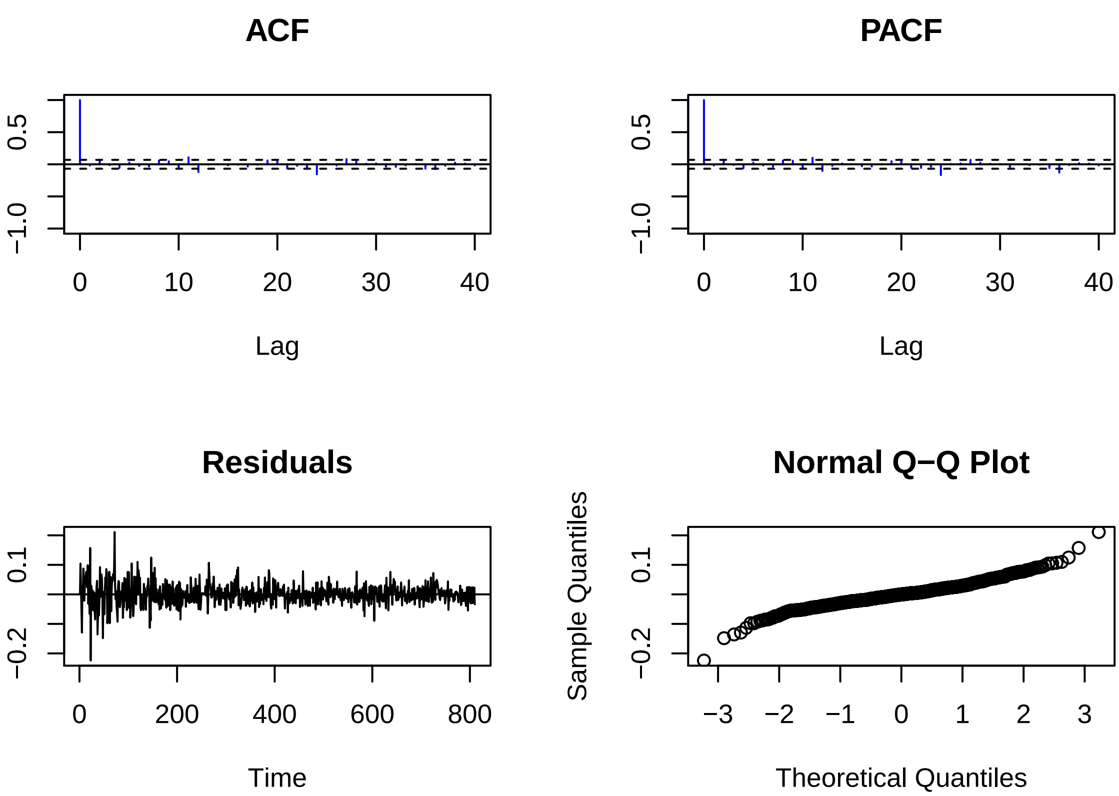
The complete model is

$$X_t = 1.6398038X_{t-1} - 0.7492623X_{t-2} + \epsilon_t \ - 1.5941598\epsilon_{t-1} + 0.7895618\epsilon_{t-2}$$

D)

```r
test(model$residuals)
```

```
## Null hypothesis: Residuals are iid noise.
## Test                        Distribution Statistic   p-value
## Ljung-Box Q                Q ~ chisq(20)     44.95    0.0011 *
## McLeod-Li Q                Q ~ chisq(20)    273.34         0 *
## Turning points T   (T-538.7)/12 ~ N(0,1)       545    0.5972
## Diff signs S      (S-404.5)/8.2 ~ N(0,1)       387    0.0333 *
## Rank P      (P-163822.5)/3845.7 ~ N(0,1)    162184    0.6701
```



From the Ljung-Box statistics, the p-value is $0.0011 < 0.05$, we reject the Null hypothesis that the Residuals are iid noise. Therefore the model is not adequate, however it's the best model we can fit so far.

E)

From the ACf and PACF, it shows that the there are some lag that is outside the bar. The Residuals plot also shows non-constant variance. Therefore, the residual is not a white noise series. This shows that the model is inadequate.

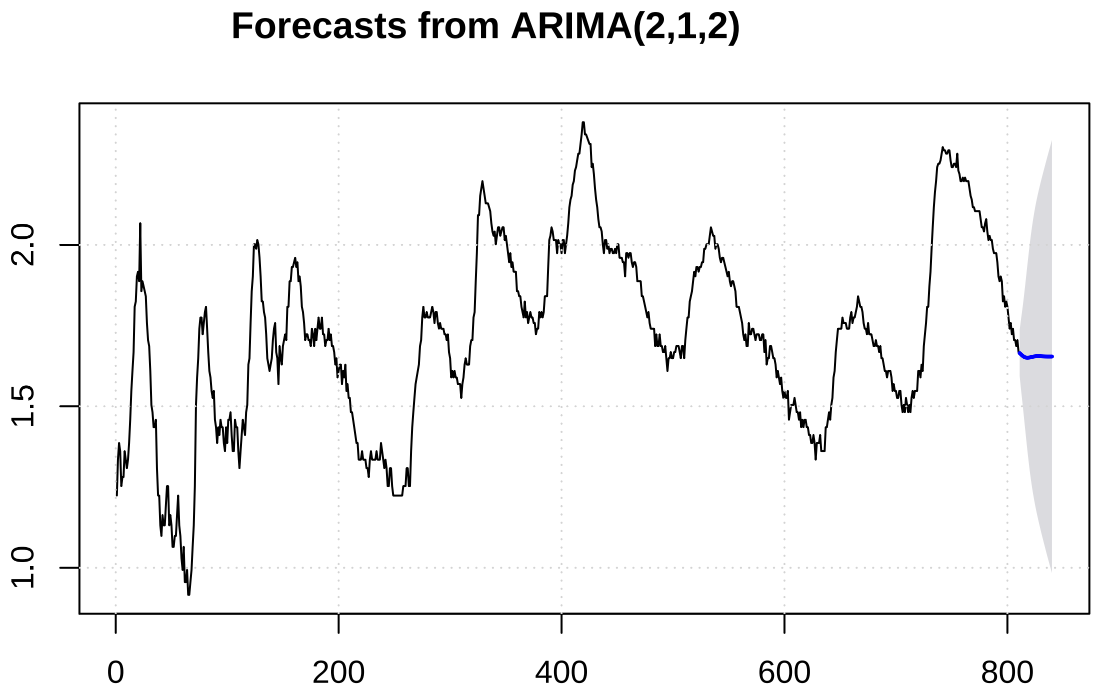
F)

```
fcast <- forecast(model,h=30,level = 95);fcast
```

```
##     Point Forecast     Lo 95    Hi 95
## 811       1.664831 1.5940884 1.735574
## 812       1.661529 1.5591748 1.763883
## 813       1.658268 1.5270446 1.789492
## 814       1.655396 1.4945263 1.816265
## 815       1.653129 1.4610854 1.845172
## 816       1.651563 1.4271548 1.875971
## 817       1.650695 1.3935234 1.907866
## 818       1.650444 1.3610023 1.939885
## 819       1.650683 1.3302437 1.971122
## 820       1.651263 1.3016650 2.000860
## 821       1.652035 1.2754433 2.028626
## 822       1.652866 1.2515511 2.054181
## 823       1.653651 1.2298116 2.077490
## 824       1.654315 1.2099572 2.098673
## 825       1.654816 1.1916804 2.117952
## 826       1.655140 1.1746737 2.135606
## 827       1.655296 1.1586557 2.151935
## 828       1.655308 1.1433858 2.167231
## 829       1.655212 1.1286704 2.181754
## 830       1.655046 1.1143622 2.195729
## 831       1.654844 1.1003570 2.209331
## 832       1.654639 1.0865872 2.222690
## 833       1.654452 1.0730154 2.235889
## 834       1.654301 1.0596269 2.248976
## 835       1.654193 1.0464226 2.261963
## 836       1.654128 1.0334123 2.274845
## 837       1.654104 1.0206097 2.287598
## 838       1.654112 1.0080275 2.300196
## 839       1.654144 0.9956754 2.312612
## 840       1.654189 0.9835578 2.324821
```

G)

```
plot(fcast)
grid()
```



H)

The Forecast doesn't seem reasonable, because there is a clear seasonality of the data. However the model doesn't deal with the seasonality and it gives forecast closed to the mean. The forecast inervals seems ok to me, because it tend to be wide for steps further and covers the whole range of the dataset.