PSTAT174_HW5(1-2)

Celeste Herrera

5/5/2021

1. Create a glossary of R-commands for time series. It should contain all commands that you learned so far in the labs, doing homework, and reviewing posted lecture slides. At the minimum, the glossary should include commands that allow

– define working directory; getwd()

setwd("/Users/celesteherrera/Documents/PSTAT174/Homework/Homework5")

– read and plot data;

read.csv("iowa.csv")

– simulate and plot ARMA models;

x = arima.sim(n,model) ts.plot(x)

– add trend and mean line to the original time series plot;

plot(x ~ as.numeric(1:length(x))) abline(lm(x ~ as.numeric(1:length(x)))) abline(h=mean(x), col = "red")

– calculate and plot theoretical acf/pacf for ARMA models;

plot(0:100, ARMAacf(model, lag.max =),xlim=c(),ylab="r",type='h') abline(h=0)

plot(ARMAacf(model, lag.max =,pacf = TRUE),type= "h",xlab="lag",ylim=c())

(abline(h=0)

– calculate and plot sample acf/pacf;

plot(acf(x,lag.max=)) plot(pacf(x,lag.max=))

– check whether a particular model is causal/invertible (R commands to find and plot roots of polynomials)

source("poly.roots.R") plot.roots(NULL,polyroot(c()))

– perform Box-Cox transforms;

t = 1:length(x) fit = lm(x ~ y) bcTransform = boxcox(x ~ y,plotit = FALSE) lambda = bcTransform **x.bc = (1/lambda)*(wine^lambda-1)**

– perform differencing data at lags 1 and 12;

diff(x, 1) diff(x, 12)

– perform Yule-Walker estimation and find standard deviations of the estimates.

ar(x = x, aic = TRUE, order.max = NULL, method = c("yule-walker"))

– perform MLE and check AICC associated with the model.

x_fit = arima(x,order = c(), method = "ML") library(qpcR) AICc(x_fit)

2. Choose a dataset that you will be interested to analyze for your class final project. URLs of time series libraries are posted on Gaucho Space. Provide the following information about the project:

(a) Data set description: briefly describe the data set you plan to use in your project.

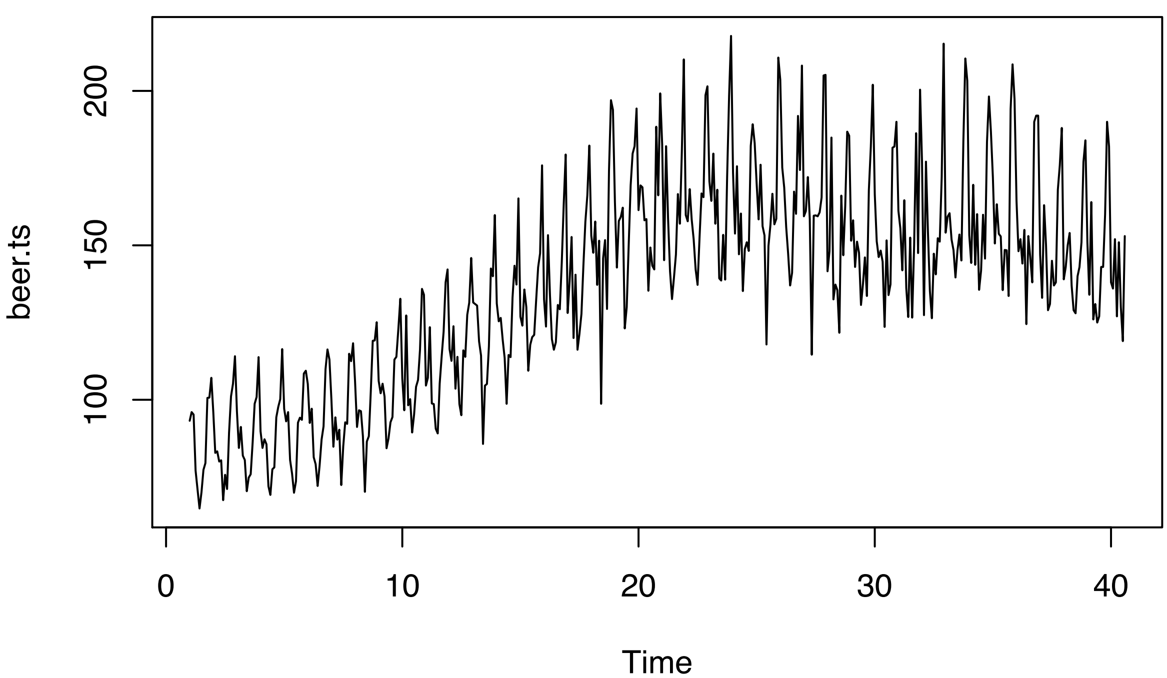
The dataset I plan to use for my project is a time series dataset the breifly describes the amount of production from alcohol in Australlia in the year of 1956 - 1995

(b) Motivation and objectives: briefly explain why this data set is interesting or important. Provide a clear description of the problem you plan to address using this dataset (for example to forecast).

The data is important because it can help forcast the possibility of how much production of beer will be in Australlia.

(c) Plot and examine the main features of the graph, checking in particular whether there is (i) a trend; (ii) a seasonal component, (iii) any apparent sharp changes in behavior. Explain in detail.

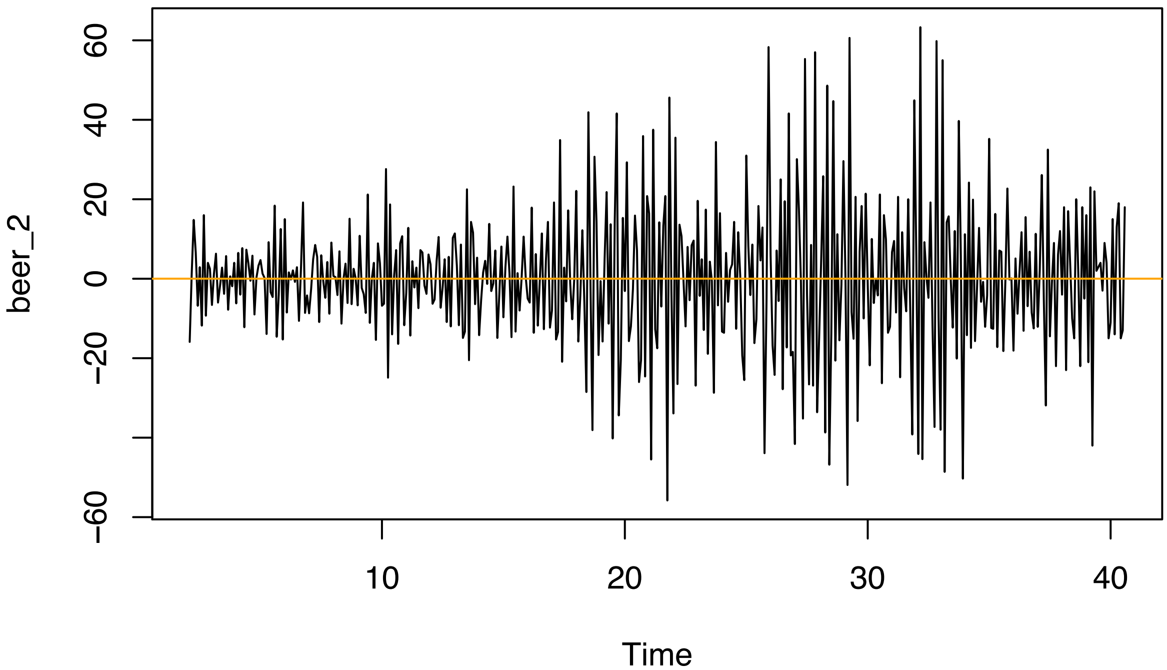
```
beer <- read.csv("beer_production.csv")
beer.ts<-ts(beer$Monthly.beer.production, frequency = 12)
ts.plot(beer.ts)
```



There is a trend that is going upwards then is slightly going back down at the end. There is no clear indication as to whether it is a seasonal componet. There is no sharp changes within the behavior above.

(d) Use any necessary transformations to get stationary series. Give a detailed explanation to justify your choice of a particular procedure. If you have used transformation, justify why. If you have used differencing, what lag did you use? Why? Is your series stationary now?

```
beer_1 <- diff(beer.ts,12)
beer_2 <- diff(beer_1, 1)
ts.plot(beer_2)
abline(h = mean(beer_2), col = "orange")
```




```
var(beer_2)
```

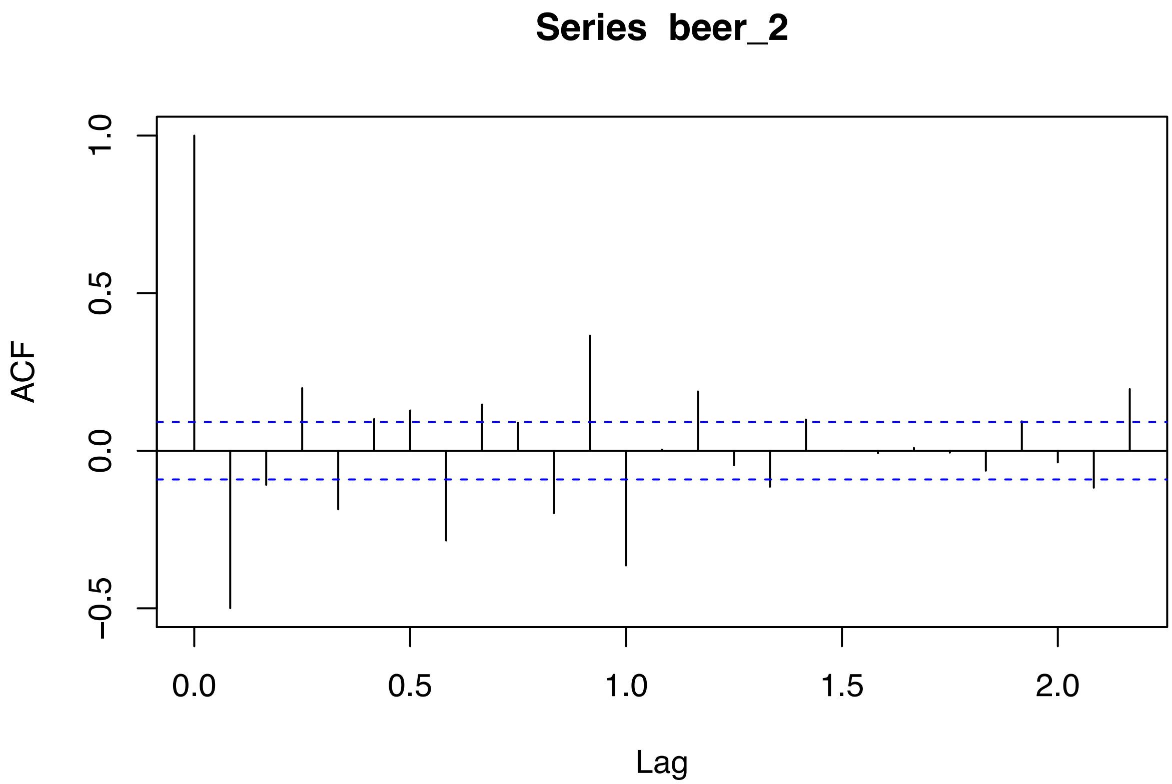
```
## [1] 337.6726
```

```
var(beer.ts)
```

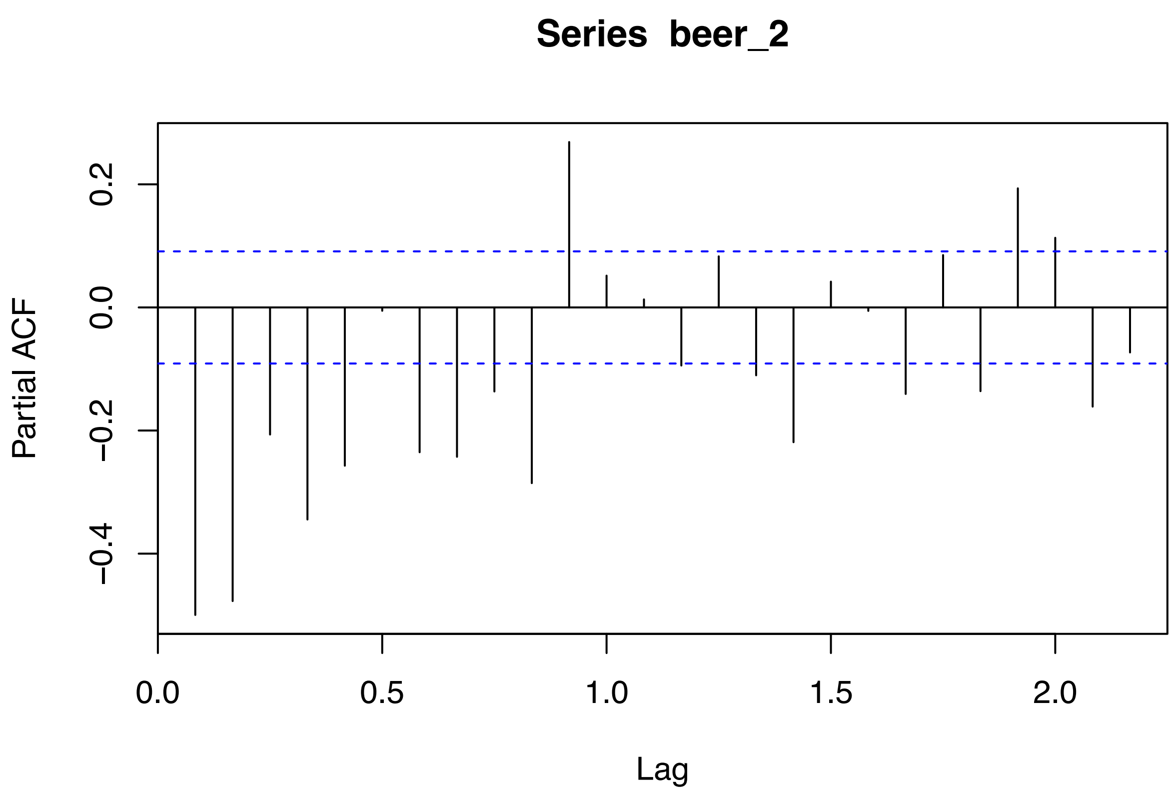
```
## [1] 1138.302
```

(e) Plot and analyze the ACF and PACF to preliminary identify your model(s): Plot ACF/PACF. What model(s) do they suggest? Explain your choice of p and q here.

```
acf(beer_2)
```




```r
pacf(beer_2)
```



As shown above we can see there is non stationary variance. It is not known to be a good fit for the project therefore another dataset will be chosen.

Celeste Herrera

PSTAT 174 HW #5

may 5, 2021

3 ARMA (3,0) model fit to quarterly time series

$X_t - 2.637 = .252(X_{t-1} - 2.637) + .061(X_{t-2} - 2.637) - .202(X_{t-3} - 2.637) + Z_t$

$2.637 + .252(2.93 - 2.637) + .061(4.62 - 2.637) - .202(2.12 - 2.637)$

$\Rightarrow 2.93623$

2.9363 is known to be smaller than 3 therefore we can conclude Choice A best fits our statement

4 AR(1) mean 0: $\rho(2) = .215$ $\rho(3) = -.100$

$x_T = -.431$

So $\hat{x}_{T+1}$ = Value of x_{T+1}

$= \hat{\phi} x_T$

$\hat{\phi}^2 = \rho(2) \Rightarrow \hat{\phi} = -\sqrt{.215} = -.463809$

$\hat{x}_{T+1} = (-.463809)(-.431)$

$= .1998465$

5. AR(1) ARMA(1,1) ARMA(1,2) ARMA(2,3) ARMA(4,3)

AIC = -2x log-likallihood +2x(p+q+2)

$$AR(1) \overset{AIC}{=} 12 \times (-650) + 2 \times (1+0+2) = 1306$$

$$ARMA(1,1) \overset{AIC}{=} -2(-641) + 2 \times (1+1+2) = 1209$$

$$ARMA(1,2) \overset{AIC}{=} -2 \times (-636) + 2(1+2+2) = 1282$$

$$ARMA(2,3) \overset{AIC}{=} -2 \times (-630) + 2 \times (2+3+2) = 1274$$

$$ARMA(4,3) \overset{AIC}{=} -2 \times (-629) + 2 \times (4+3+2) = 1276$$

ARMA(2,3) has the best known model for time series