Name : Hee Chul Jeong

Student ID : 2016314895

Assignment Title : Introduction to Time Series Analysis Midterm 1

1 – a) The time plot strongly indicates there is an increasing trend and suggests it is highly likely that the data has seasonality. In terms of a trend, the data certainly shows an increasing trend with either linear or quadratic trend. In terms of a seasonality, it seems that each t is related to each t+50 or higher. There are about 9 tipping points, which is related to the fact that the data is accumulated for approximately 9 years, and one year is about 52 weeks.

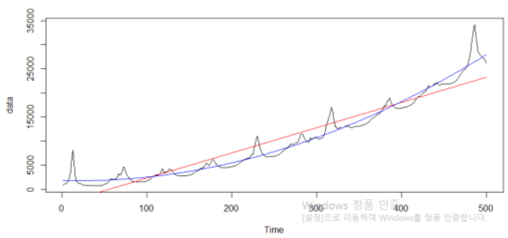
1 – b) Method of Removing trend : Quadratic Regression

Method of Removing seasonality : Exponential Moving Average

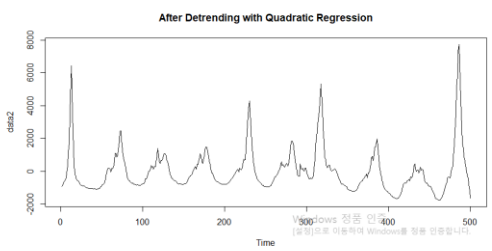
1 – c) The data shows an obvious increasing trend so detrending was needed. It seemed the data is either a linear or quadratic so the regression method was appropriate. To specify a model to choose, the linear regression and quadratic regression were compared. The multiple R-squared, adjusted R-squared, and graphics all indicated that the quadratic regression line is more suitable in this case. The statistics used for comparisons are given in (1) and (2), and the graphic is given in Pic 1. Therefore, Pic 2 is the visual representation of the detrending.

(1) (Linear Regression : Multiple R-Squared - 0.9018, Adjusted R-Squared - 0.9016)

(2) (Quadratic Regression : Multiple R-Squared - 0. 0.9718, Adjusted R-Squared - 0. 0.9717)



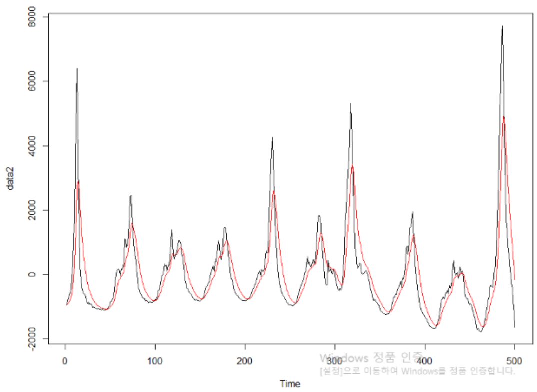
Pic 1



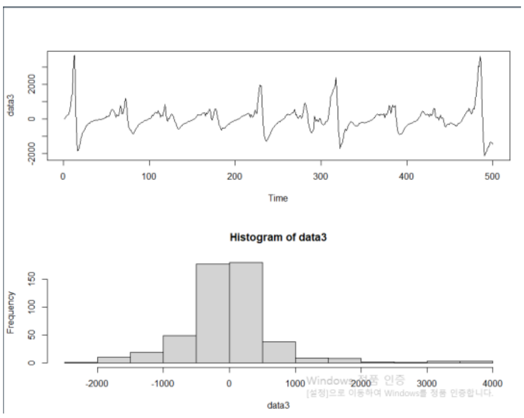
Pic 2

After detrending, there were 4 candidates to choose for removing seasonality : harmonic regression, differencing, seasonal averaging, and exponential moving averaging. The harmonic regression, differencing, seasonal averaging were too invariant with the changes of the data, and the residuals were rather right-skewed. The exponential moving averaging, on the other hand, was able to mitigate and apply the variability of the data, and the residuals followed normal distribution with mean 0 and a standard deviation (see pic 3 and pic 4).

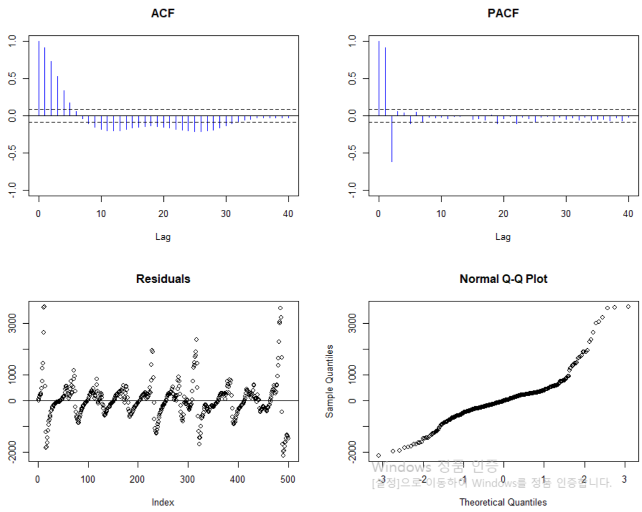
The remaining residual is stationary because it has a constant mean, which is 0 in this case, and the variance that does not depend on t, but only h. We can verify it from pic 4. However, it cannot be claimed that the residual is IID sequence because out of 5 tests, Ljung-Box Q, McLeod-Li Q, Turning points T, and Diff signs S all indicate the residual is not IID sequence, and the residual plot clearly shows a periodic pattern (pic 5). Additionally, ACF graphics shows that closer observations are correlated.



Pic 3



Pic 4



Pic 5

1 – d) The given data showed an increasing trend and seasonal pattern. After comparing statistics and graphics of numerous methods for detrending and removing seasonality, the quadratic regression method was selected to detrend, and the exponential moving average was selected to remove seasonal pattern. After the application of both methods, the remaining residual meets the criteria of stationarity; it has a constant mean, 0, and its variance does not depend on t, but only h. However, the residual could not meet the condition of IID sequence since closer observations have high correlations and the residual plot clearly imply there is a seasonal pattern.

```r
library(itsmr)
library(nortest)
library(tseries)
library(robustHD)
library(QuantTools)


# 1
#1-a
data <- as.vector(read.csv('c:/users/rogan/desktop/2021practice1.csv', header = FALSE)[,1])
source('c:/users/rogan/desktop/TS-library.R')
par(mfrow=c(2,1))
plot.ts(data)
acf2(data,499)
par(mfrow=c(1,1))

#1-b
x <- 1:length(data)
x2 <- x^2
lmod <- lm(data~1+x)
lmod2 <- lm(data~1+x+x2)
summary(lmod)
summary(lmod2)
lines(lmod$fitted.values, col='red')
lines(lmod2$fitted.values, col='blue')
data2 <- data - lmod2$fitted.values
plot.ts(data2)
title('After Detrending with Quadratic Regression')

#1-c
# candidate 1 : harmonic regression
n <- length(data2)
t = 1:n
f1 = 9.615385   # f1 = n/d, fi = i*f1, lambda i = fi*(2*pi/n) = i*f1*(2*pi/n)
f2 = (9.615385)*2
f3 = (9.615385)*3
f4 = (9.615385)*4
costerm1 = cos(f1*2*pi/n*t)
sinterm1 = sin(f1*2*pi/n*t)
costerm2 = cos(f2*2*pi/n*t)
sinterm2 = sin(f2*2*pi/n*t)
costerm3 = cos(f3*2*pi/n*t)
sinterm3 = sin(f3*2*pi/n*t)
costerm4 = cos(f4*2*pi/n*t)
sinterm4 = sin(f4*2*pi/n*t)
```



```r
harmonic.model = lm(data2 ~ 1 + costerm1 + sinterm1)
summary(harmonic.model)

harmonic.model2 = lm(data2 ~ 1 + costerm1 + sinterm1 + costerm2 + sinterm2)
summary(harmonic.model2)

harmonic.model3 = lm(data2 ~ 1 + costerm1 + sinterm1 + costerm2 + sinterm2+ costerm3 + sinterm3 + costerm4 + sinterm4)

plot.ts(data2)
lines(x, harmonic.model$fitted, col="blue")
lines(x, harmonic.model2$fitted, col="red")
lines(x, harmonic.model3$fitted, col="green")
legend(0, 500, lty=c(2,2), col=c("blue","red",'green'), c("k=1", "k=2",'k=3'))

data3 <- data2 - harmonic.model$fitted
plot.ts(data3)

data3 <- data2 - harmonic.model2$fitted
plot.ts(data3)

data3 <- data2 - harmonic.model3$fitted
plot.ts(data3)
hist(data3)
#not successful

# candidate 2 : lag differencing
diff12 = diff(data2, lag=10)
par(mfrow=c(1,2))
plot.ts(data2)
plot(x[11:500], diff12, type="l", col="red")
plot.ts(data2 - diff12)

#not successful

# candidate 3 : seasonal smoothing
season.avg = season(data2, d=71)
plot.ts(data2)
lines(x, season.avg + mean(data2), col="red")
data3 <- data2 - (season.avg + mean(data2))
plot.ts(data3)
```

```r
# candidate 4 : exponential moving average
par(mfrow=c(1,1))
ema = smooth.exp(data2, .2)
plot.ts(data2)
lines(x,ema, col="red")
par(mfrow=c(2,1))
data3 <- data2 - ema
plot.ts(data3)
hist(data3)  # ~ normal distribution

test(data3)
```