Homework III

Xingye Ke

February 13, 2018

## Homework III

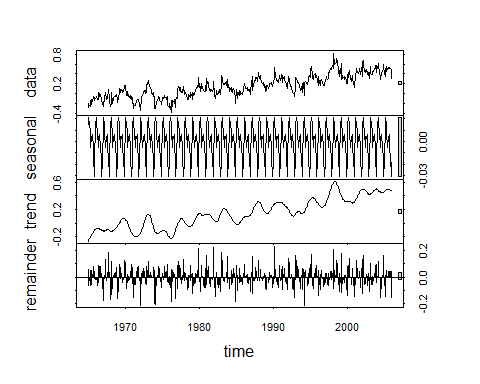
Extract the records from January 1965 to December 2005 from global temperature series. Use this extracted series to answer the following questions.

### Q(a)

Decompose the series using stl command. Plot the components trend, seasonal effect and remainders. Note that stl command may not decompose the series of ts object without some adjustment as example.

The R code and the decomposed graphic are shown below:

setwd("C:/Users/n886528/Desktop/OnUse/WithTracu/Introductory\_Time\_Series\_with\_R\_datasets-master")  
  
dat <- read.table("global.dat")  
  
# reshape  
  
datl <- as.data.frame(matrix(as.matrix(t(dat)),byrow = T, ncol =1 ))  
ts.obj <- ts(datl,start = c(1856,1),end = c(2005,12), frequency = 12)  
  
#extracted  
  
cutted.ts <- window(ts.obj, start = c(1965,1),end = c(2005,12))  
stl.obj <- stl(cutted.ts[,1], s.window = 'periodic')  
plot(stl(cutted.ts[,1],s.window = 'periodic'))



### Q(b)

As we can see from the right side of the graphic:

1.The grey rectangles of trend and the data series have the similar size, so the trend actually dominates the series;

2.If we look into the seasonal’s grey rectangle, we could find the relative large size, which means variation in the seasonality is a much smaller component of the variation exhibited in the original series;

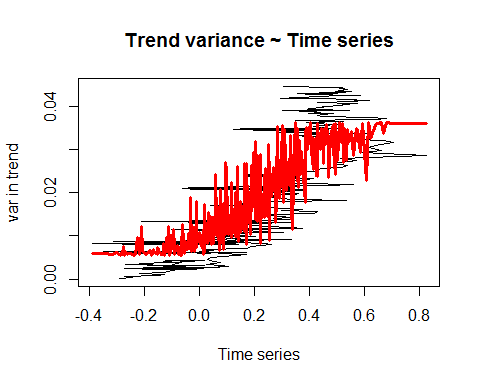
3.We could see from the residual part: the relative small size of grey rectangle means a large part of variation of the orginal series is left unexplained after adjusting for the trend and seasonality, which implys the additive model could not capture most part of the series and might not be a good fit for the series.

### Q(c)

As we know, the additive model will be most appropriate if:

1. The magnitude of seasonality’s fluctuations does not vary with the level of the series.
2. The variation around trend does not vary with the series.  
    Any violation to these two points can be the proof to show a multiplicative model would be prefered. for the seasonality, it is very obvious that the variation in seasonal pattern is not proportional to the level of the time series. for the trend part, we could draw the graphic to show the relationship between variation in trend and time series.

trend.ts <- stl.obj$time.series[,2]  
mag <- 1  
cum\_var <- c()  
for(i in seq(1,length(trend.ts),mag)){  
 cum\_var <- c(cum\_var, var(trend.ts[1:i]))  
}  
  
plot(cum\_var~cutted.ts[seq(1,length(trend.ts),mag)], type = 'l',  
 main = 'Trend variance ~ Time series',  
 xlab = 'Time series',  
 ylab = 'var in trend')  
loess.fit <- loess(cum\_var~cutted.ts[seq(1,length(trend.ts),mag)])  
ord <- order(cutted.ts[seq(1,length(trend.ts),mag)])  
lines(cutted.ts[seq(1,length(trend.ts),mag)][ord],loess.fit$fitted[ord],col = 'red' , lwd = 3)



Note that there is an obvious increasing trend in the plot, which means a multiplicative model can be a good subsitude.

### Q(d)

The negative value in the original series is the biggest obstacle towards log transformation. If log transformation is still needed, we need to do some transformation to keep the information contained in the original series as much as possible. The first very straightforward method is to move the series to upper x-axis, which mins to add a constant to the whole series to make every number postive, then do the log-transformation and the relative time series analysis.

### Q(e)

The column named Point Forecast is the values forecasted by the additive model.

library(forecast)

## Warning: package 'forecast' was built under R version 3.4.3

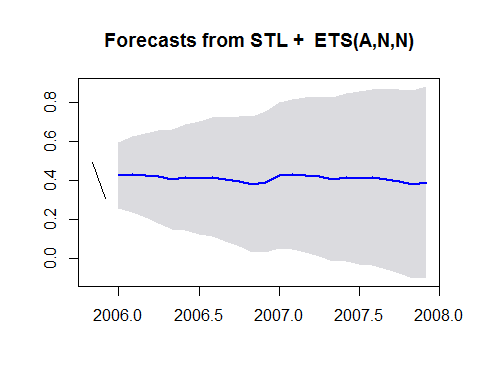
fcst <- forecast(stl.obj, h = 24, level = 95 )  
print(fcst)

## Point Forecast Lo 95 Hi 95  
## Jan 2006 0.4251945 0.256323086 0.5940659  
## Feb 2006 0.4312782 0.236874215 0.6256823  
## Mar 2006 0.4258497 0.208897471 0.6428020  
## Apr 2006 0.4185828 0.181214653 0.6559510  
## May 2006 0.4045110 0.148348942 0.6606731  
## Jun 2006 0.4142124 0.140544067 0.6878808  
## Jul 2006 0.4125724 0.122452177 0.7026926  
## Aug 2006 0.4163472 0.110659240 0.7220351  
## Sep 2006 0.4044878 0.083987444 0.7249882  
## Oct 2006 0.3956951 0.061037316 0.7303529  
## Nov 2006 0.3788537 0.030613515 0.7270938  
## Dec 2006 0.3916235 0.030311163 0.7529357  
## Jan 2007 0.4251945 0.051266799 0.7991222  
## Feb 2007 0.4312782 0.045147040 0.8174094  
## Mar 2007 0.4258497 0.027889110 0.8238104  
## Apr 2007 0.4185828 0.009134388 0.8280313  
## May 2007 0.4045110 -0.016111593 0.8251337  
## Jun 2007 0.4142124 -0.017295095 0.8457200  
## Jul 2007 0.4125724 -0.029552152 0.8546970  
## Aug 2007 0.4163472 -0.036145357 0.8688397  
## Sep 2007 0.4044878 -0.058140404 0.8671160  
## Oct 2007 0.3956951 -0.076851412 0.8682416  
## Nov 2007 0.3788537 -0.103407237 0.8611146  
## Dec 2007 0.3916235 -0.100159983 0.8834069

### Q(f)

The forecasted values are the Blue line, and the shadow area is the corresponding confidence interval.

values <- fcst$mean  
upper <- fcst$upper  
lower <- fcst$lower  
plot(fcst, include =2, PI=TRUE)



### Q(2)

The RMSE is the sqrted mean of sum sqaure error.

dat <- read.csv("qjnj.csv",header = TRUE)  
log.dat <- ts(log(dat$x),start = c(1960,1),end = c(1980,4),frequency = 4)  
  
decomposed <- decompose(log.dat,type = 'additive')  
  
rmse = sqrt(mean(decomposed$random^2,na.rm = TRUE))  
  
print(rmse)

## [1] 0.08154302