ISYE 6501 Intro Analytics Modeling – HW4

Question 7.1 Describe a situation or problem from your job, everyday life, current events, etc., for which exponential smoothing would be appropriate. What data would you need? Would you expect the value of α (the first smoothing parameter) to be closer to 0 or 1, and why?

The World Health Organization has declared the growing coronavirus outbreak in China to be a global health emergency. I think exponential smoothing would be used to forecast the total cases of infections. To forecast the outbreak, we will need the dates and number of confirmed cases of each date. I would expect the value of α to be closer to 0. Because the number of confirmed cases (observed infected cases X_t) is very sensitive to the diagnostic and medical resources. For example, when the number of patients has reached the hospitals' limitation or when the diagnostic kits have been launched will cause big valleys and peaks for the confirmed cases, in the case we should trust S_{t-1} more than X_t.

Question 7.2 Using the 20 years of daily high temperature data for Atlanta (July through October) from Question 6.2 (file `temps.txt`), build and use an exponential smoothing model to help make a judgment of whether the unofficial end of summer has gotten later over the 20 years. (Part of the point of this assignment is for you to think about how you might use exponential smoothing to answer this question. Feel free to combine it with other models if you'd like to. There's certainly more than one reasonable approach.)

The first step is to read in the dataset and do some simple data summary. From the code, I get to know this dataset contains **123 data points** and **21 columns**. After converting the data table into time series format, I plot the temperature in July through October 1996 to 2015. (Fig.1)

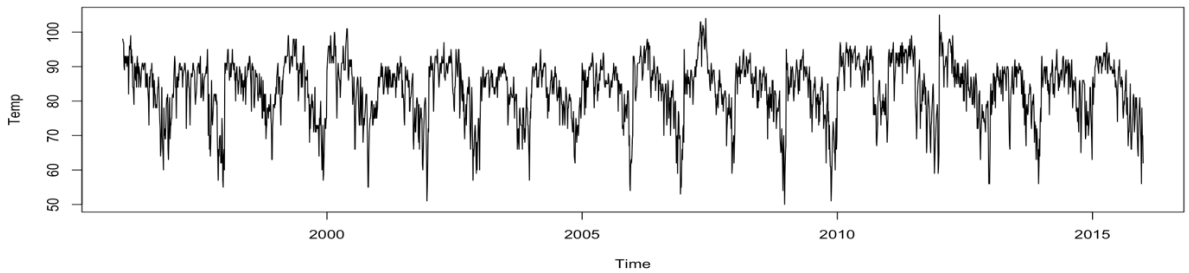

Fig.1

Firstly, I assumed there is no trend or seasonal pattern and tried simple exponential smoothing. I used `HoltWinters` function and set beta and gamma to FALSE. From the plot we can see S_t overlapped pretty well with X_t, also from detail, we can see the big alpha value 0.84, which means that the model thinks there is randomness but not much so trust X_t more. The sum of squared errors is 56,198. (Fig.2)

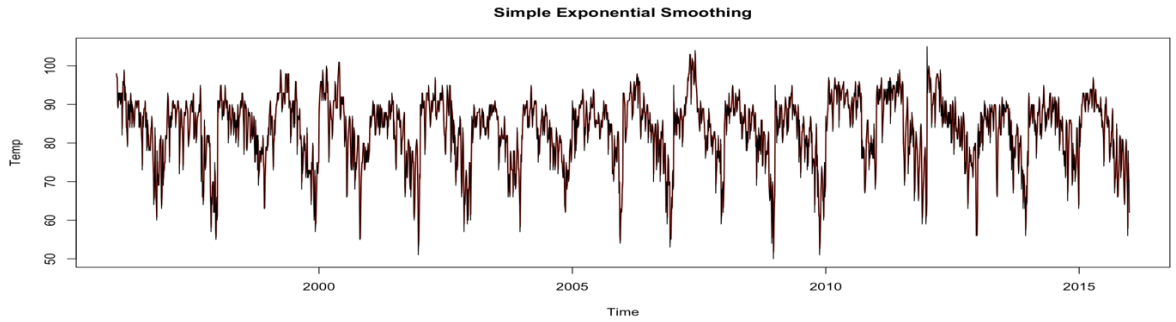

Fig.2

```
Holt-Winters exponential smoothing without trend and without seasonal component.
Call: HoltWinters(x = ts, beta = FALSE, gamma = FALSE)
Smoothing parameters:
 alpha: 0.8388021
 beta : FALSE
 gamma: FALSE
```

To get a better smoothing effect, I set the alpha to 0.2 manually, from the plot below, it does smooth some peaks and valleys and the sum of squared errors is also bigger: 78,768 (Fig.3)

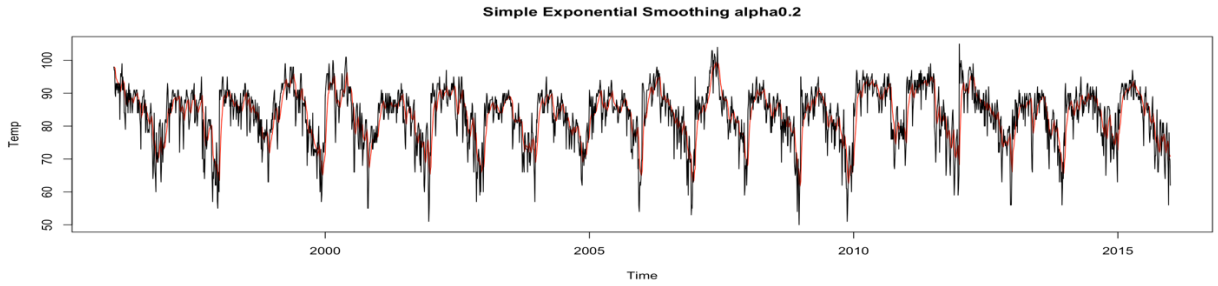

Fig.3

Then I tried two double exponential smoothing by adding trend and seasonal separately. The trend exponential smoothing has alpha: 0.845, beta: 0.004, so the trend is only increasing at the beginning when the T_{t-1} is big and close to 0. The sum of squared error is 66,245. The seasonal exponential smoothing has alpha: 0.661, gamma: 0.625, the sum of squared error: 172,748 (Fig.4).

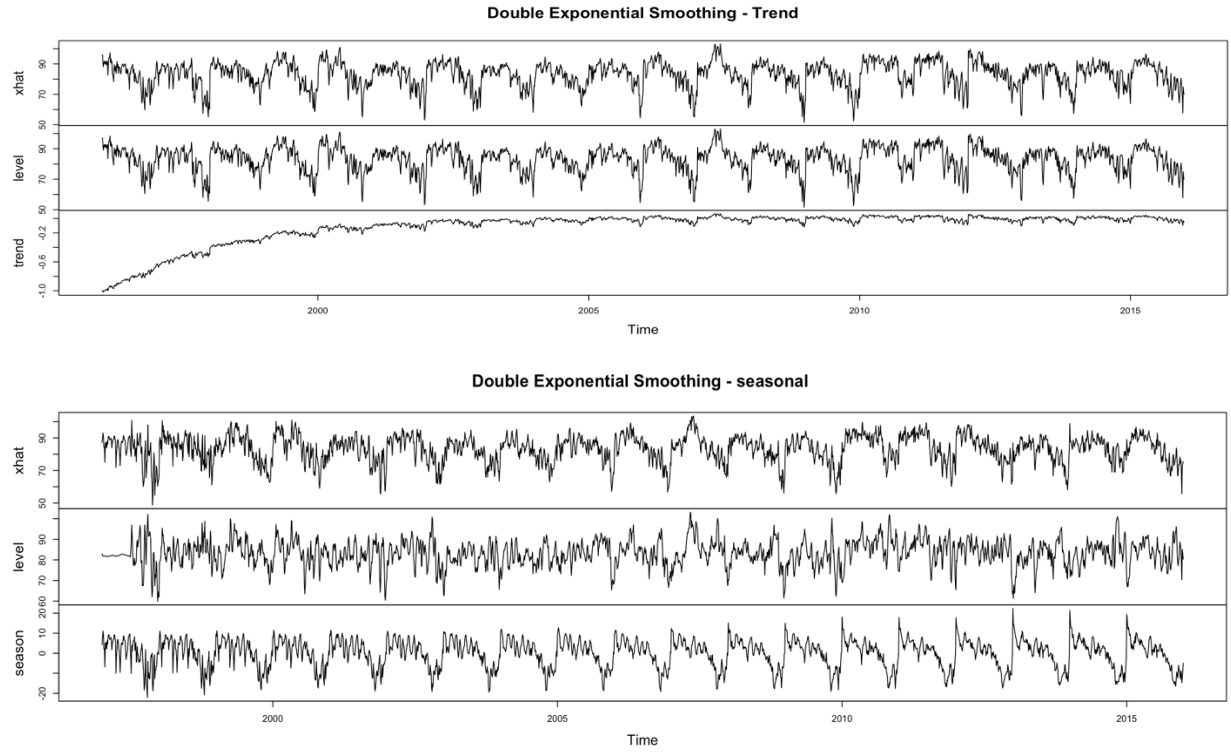

Fig.4

Then I tried triple exponential smoothing, I got alpha: 0.661, beta: 0, gamma: 0.625, error: 66,245. From fitting the model, there is no trend found. I collected the coefficients of seasonal component to create a heatmap. we can see from 1996 to 2015 the seasonal coefficients' pattern is shifting. It looks like both the unofficial start and end of summer have gotten later over the 20 years in Atlanta.

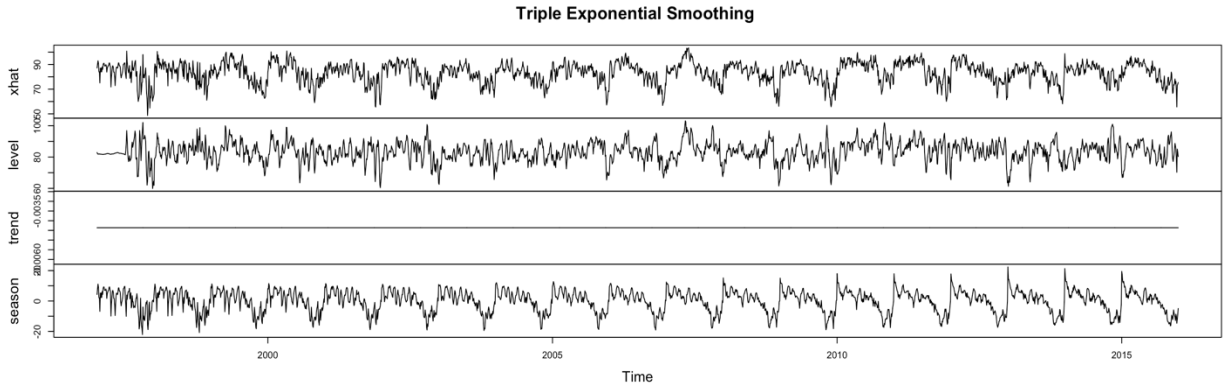

Fig 5

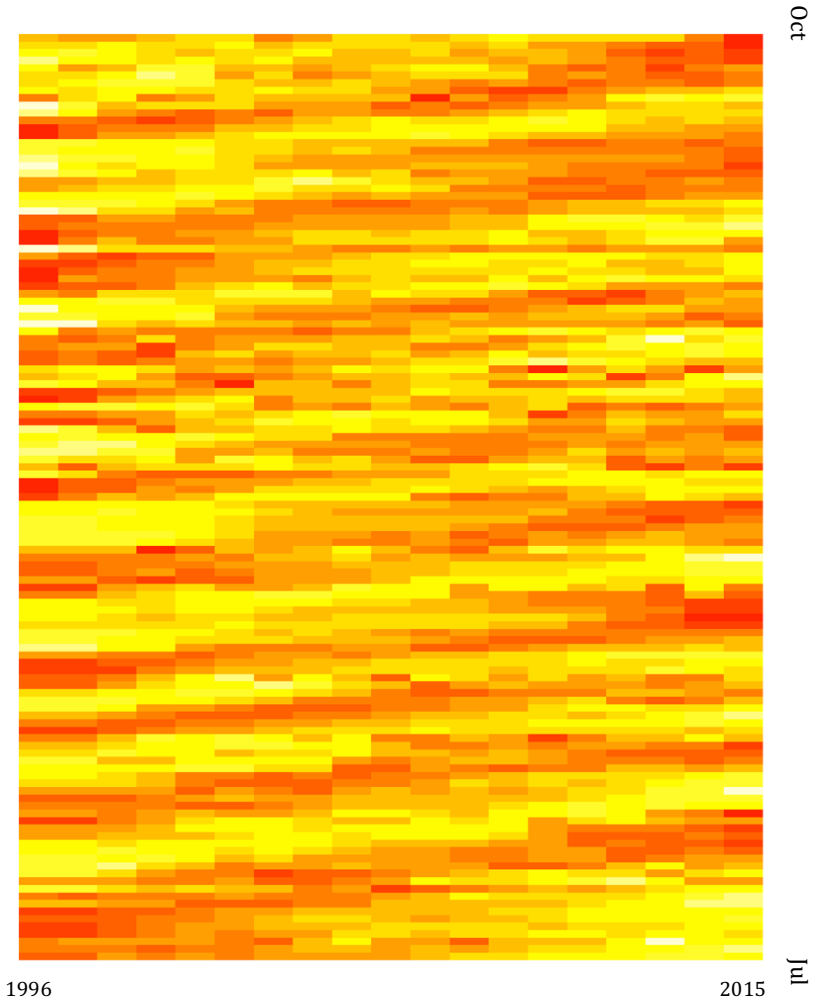

Fig 6

Code

```
# ISYE 6501 Intro Analytics Modeling - HW4
# IP uscrime.txt

# Loading and examining data
df<-read.delim("temps.txt", header = TRUE, sep = "\t")
#### Display Head Lines ####
head(df,2)
#### Show summary, number of row of last column##
summary(df$Crime)
nrow(df)
ncol(df)
for (i in attributes(df)$names ){
  print(paste0('Name:',i,'   Class:',class(df[[i]]),'   nlevels:',length(uniq
ue(df[[i]]))))
  print(summary(df[[i]]))
}

# Question 7.2 Using the 20 years of daily high temperature data for Atlanta
(July through October) ,
# build and use an exponential smoothing model to help make a judgment of
# whether the unofficial end of summer has gotten later over the 20 years.
# (Part of the point of this assignment is for you to think about how you mig
ht use exponential smoothing to answer this question.
# Feel free to combine it with other models if you'd like to. There's certain
ly more than one reasonable approach.)

ts_v = as.vector(unlist(df[,2:21]))                #create time series
ts  <-ts(ts_v, start=1996, frequency = nrow(df))
plot.ts(ts,xlab="Time",ylab="Temp")

#simple exponential smoothing
single_exp_hw<-HoltWinters(ts, beta = FALSE, gamma = FALSE)
single_exp_hw
plot(single_exp_hw,xlab="Time",ylab="Temp",main="Simple Exponential Smoothing
")
single_exp_hw$SSE


single_exp_hw2<-HoltWinters(ts,alpha=0.2, beta = FALSE, gamma = FALSE)
single_exp_hw2
plot(single_exp_hw2,xlab="Time",ylab="Temp",main="Simple Exponential Smoothin
g alpha0.2")
single_exp_hw2$SSE

#double exponential smoothing
double_exp_hw <- HoltWinters(ts, gamma = FALSE)
double_exp_hw
```

```r
plot(double_exp_hw,main="Double Exponential Smoothing - Trend")
seasonal_temp = double_exp_hw$fitted
plot( seasonal_temp, main="Double Exponential Smoothing - Trend" )
double_exp_hw$SSE


#double exponential smoothing2
double_exp_hw2 <- HoltWinters(ts, beta = FALSE)
double_exp_hw2
plot(double_exp_hw2)
seasonal_temp2 = double_exp_hw2$fitted
plot( seasonal_temp2,main="Double Exponential Smoothing - seasonal "  )
double_exp_hw2$SSE

#triple exponential smoothing
triple_exp_hw <- HoltWinters(ts,seasonal = 'additive')
triple_exp_hw
plot(triple_exp_hw)
seasonal_temp = triple_exp_hw$fitted
plot( seasonal_temp,main="Triple Exponential Smoothing " )
triple_exp_hw$SSE

#heatmap
seasonal<- matrix(triple_exp_hw$fitted[,4], nrow = 123)
seasonal_est<-as.data.frame(seasonal,col.names=name(c(1995:2005)))
heatmap(seasonal, Rowv = NA, Colv = NA)
```