ISyE 6501 hw4

9/15/2019

Group Members:

Jingyu, Li | alanli@gatech.edu | GTID: 903520148
Jinyu, Huang | jhuang472@gatech.edu | GTID: 903522245
Yuefan, Hu | yuefanhu@gatech.edu | GTID:903543027
Chengqi, Huang | chengqihuang@gatech.edu | GTID: 903534690

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 (the first smoothing parameter) to be closer to 0 or 1, and why?

Answer

If we would like to detect how popular a web portal is, we probably need the daily number of views of this website. However, this type of data can be misleading, as it must contain a lot of randomness like breaking news, holidays or just users' random behavior. We can use exponential smoothing to deal with the randomness in daily page views. Our model will be: (PV=page views)

$$PVS_t = \alpha PVX_t + (1-\alpha)PVS_{t-1}$$

We propose that the baseline volume of page views is more meaningful for us to get an impression of how popular the web portal is, as the daily page views in a certain day may varies much, because of the randomness listed above. So in order to diminish such randomness, we expect that α should be closer to 0 so that we can lower the weight of observed number of page views on Day t. Thus we are more likely to have a precise estimate of the daily page views, which definitely helps us to better understand about the popularity of this web portal.

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.)

Note: in R, you can use either HoltWinters (simpler to use) or the smooth package's es function (harder to use, but more general). If you use es, the Holt-Winters model uses model="AAM" in the function call (the first and second constants are used "A"dditively, and the third (seasonality) is used "M"ultiplicatively; the documentation doesn't make that clear).

Answer

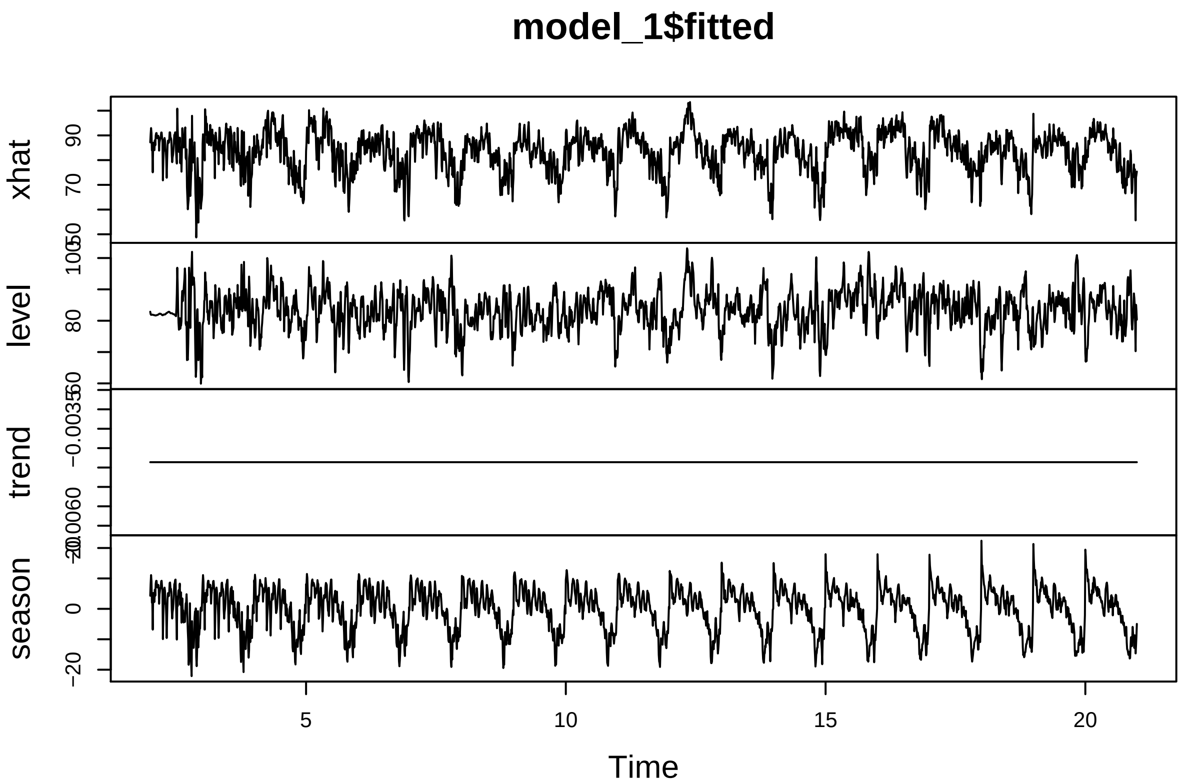
At the very beginning, we load the data needed for exponential smoothing by importing the temperature data and convert it into time series data:

```
#load data, transform into time series data
temp_data <- read.table("temps.txt",
header = TRUE)
data_2 <- vector(mode = "numeric", length = 0)
for (i in seq(2,21)){
data_2 <- c(data_2,temp_data[,i])
}
data_3 <- ts(data_2, frequency = 123, start = 1)
str(data_3)
```

```
##  Time-Series [1:2460] from 1 to 21: 98 97 97 90 89 93 93 91 93 93 ...
```

To follow up, we apply HoltWinters function in order to apply triple exponential smoothing method to the daily high temperature through July to October from 1996 to 2015. The seasonal argument is defined as "additive", because in this case we believe the temperature in each year is not changing proportionally. After running the model, we retrieve the value of seasonal, which enables us to execute further CUSUM analysis to the seasonal factor data set, with the purposes of identifying the date of summer end for each year and exploring whether there is significant change in dates of summer end through the sample years.

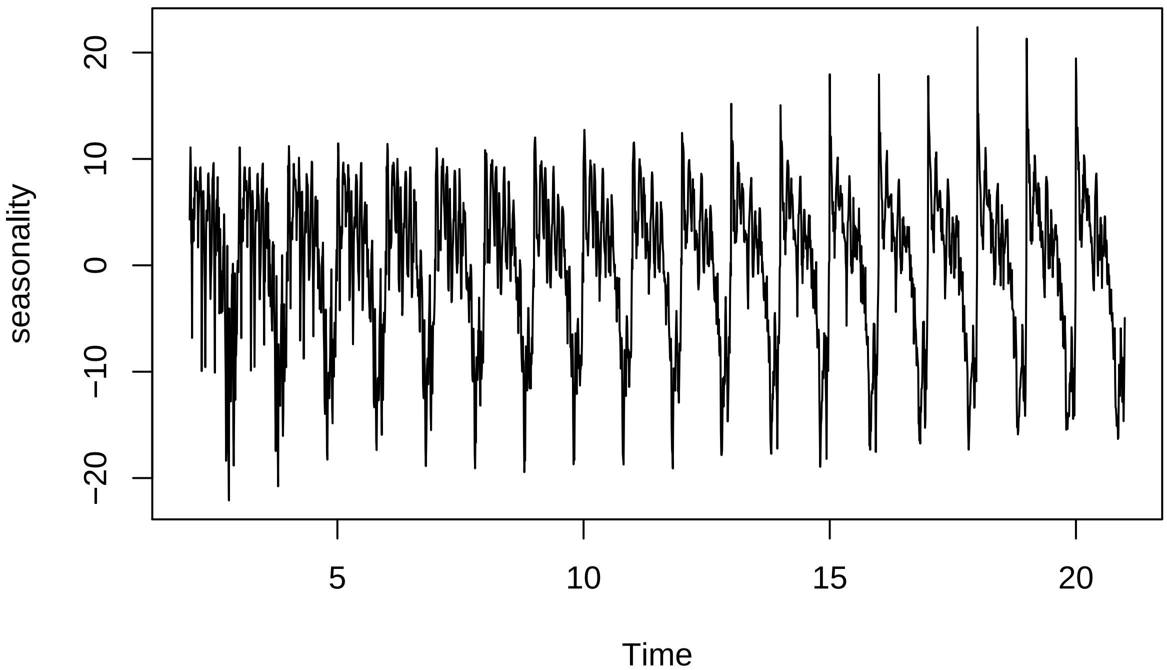
```
#apply HolWinters function
model_1 <- HoltWinters(data_3)
plot(model_1$fitted)
```




```r
#extract seasonal data from model_1
fit <- model_1$fitted
seasonality <- fit[,4]
str(seasonality)
```

```
##  Time-Series [1:2337] from 2 to 21: 4.3 8.24 11.09 9.04 2.07 ...
```

```r
plot(seasonality)
```



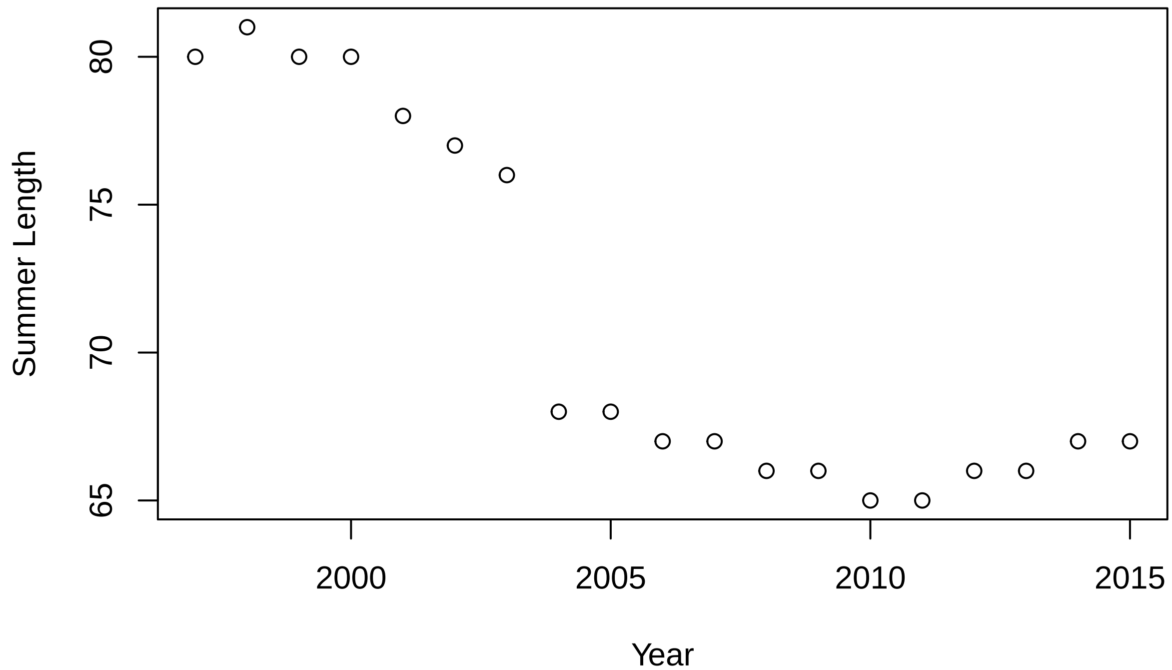
We implement the "detectCUSUMMean" function in the "ffstream" package to detect significant change in seasonality from 1997 to 2015 (the 1996 is excluded because this year is automatically set as the base for the HoltWinters model to generate the data of seasonality). The output value of "detectCUSUMMean" model is "tauhat", which stands for the place where the change point occupies in a vector. The "tauhat" value has another interpretation in this case, namely the length of summer (we assume for each year summer starts from 1st Jul. for the convenience of further study). Finally, we plot out all the values of "tauhat". The graph shows that from 1997 to 2015, the distance between 1st July and the corresponding change points extends. From ths observation, we tend to believe that unofficial end of summer hasn't gotten later over the 20 years. On the opposite, our study may support that unofficial end of summer tended to get earlier from 1996 to 2015.

```
#load package
library(ffstream)
```

```
## Loading required package: Rcpp
```

```
#use detectCUSUMMean fuction to detect the change points each year
change_matrix <- matrix(0,nrow = 19, ncol = 2)
for (i in seq(1,19)){
sea_cusum <- detectCUSUMMean(seasonality[((i-1)*123+1):(i*123)],
k=0.25,
h=8.00,
BL=50,
multiple=FALSE)
change_matrix[i,2] <- sea_cusum$tauhat
}
change_matrix[,1] <- c(seq(1997,2015))

#to see if the summer length has a certain trend in each year
plot(change_matrix,xlab = "Year",ylab = "Summer Length",)
```


Summer Length
80
75
70
65
2000
2005
2010
2015
Year