q7.2.R

Muhammad

2021-09-22

library(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5 v purrr 0.3.4  
## v tibble 3.1.4 v dplyr 1.0.7  
## v tidyr 1.1.3 v stringr 1.4.0  
## v readr 2.0.1 v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(stats)  
library(smooth)

## Loading required package: greybox

## Package "greybox", v1.0.0 loaded.

##   
## Attaching package: 'greybox'

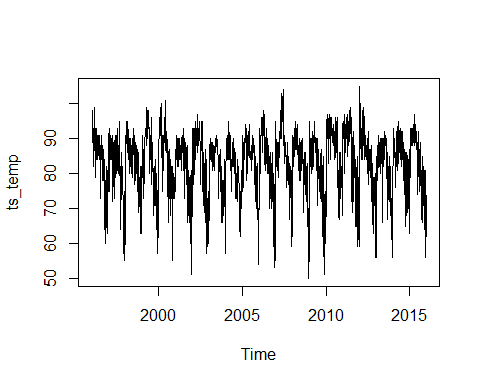
## The following object is masked from 'package:tidyr':  
##   
## spread

## This is package "smooth", v3.1.2

#Q7.2  
set.seed(1)  
  
setwd("C:/Users/Muhammad/ISYE/Hw5")  
  
temp\_data<- read.table("temps.txt", header = TRUE)  
names(temp\_data) = gsub(pattern = "X", replacement = "", x = names(temp\_data))  
temp\_data <- data.frame(unlist(temp\_data[,2:21]))  
ts\_temp= ts(temp\_data, frequency = 123, start=1996)  
summary(ts\_temp)

## unlist.temp\_data...2.21..  
## Min. : 50.00   
## 1st Qu.: 79.00   
## Median : 85.00   
## Mean : 83.34   
## 3rd Qu.: 90.00   
## Max. :105.00

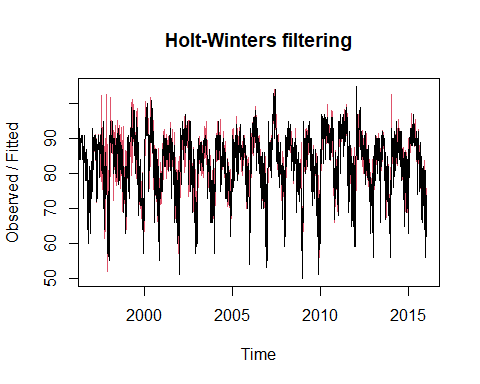
ts.plot(ts\_temp)



Model\_HW\_Mult<-HoltWinters(ts\_temp, seasonal="multiplicative")  
summary(Model\_HW\_Mult)

## Length Class Mode   
## fitted 9348 mts numeric   
## x 2460 ts numeric   
## alpha 1 -none- numeric   
## beta 1 -none- numeric   
## gamma 1 -none- numeric   
## coefficients 125 -none- numeric   
## seasonal 1 -none- character  
## SSE 1 -none- numeric   
## call 3 -none- call

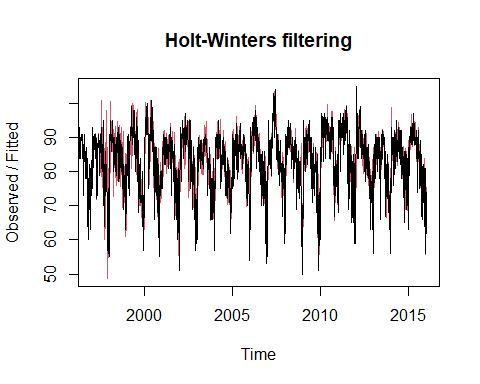
plot(Model\_HW\_Mult)



Model\_HW\_ADD<-HoltWinters(ts\_temp, seasonal="additive")  
summary(Model\_HW\_ADD)

## Length Class Mode   
## fitted 9348 mts numeric   
## x 2460 ts numeric   
## alpha 1 -none- numeric   
## beta 1 -none- numeric   
## gamma 1 -none- numeric   
## coefficients 125 -none- numeric   
## seasonal 1 -none- character  
## SSE 1 -none- numeric   
## call 3 -none- call

plot(Model\_HW\_ADD)



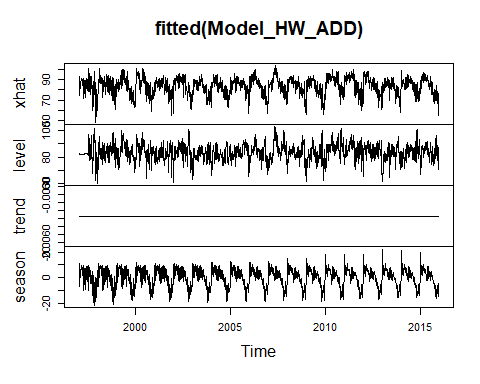
cat("Sum of squared Errors in Multiplicative model",Model\_HW\_Mult$SSE,"\n")

## Sum of squared Errors in Multiplicative model 68904.57

cat("Sum of squared Errors in Addititve model",Model\_HW\_ADD$SSE,"\n")

## Sum of squared Errors in Addititve model 66244.25

#Additive model has smaller SSE hence I am choosing additive seasonality, better way to do it would be using cross-validation but for now I am choosing seasonality as additive  
#Lets visualize the trend to understand if the temperature is increasing   
plot(fitted(Model\_HW\_ADD))



# We can observe that there is no visible change in trend over time in the temperature   
# I think one of the reason why it's hard to see the change over time in temperature is because the temperature is increasing very slowly  
# and the changes would be very minute, additional our data is too small to capture that since there are random fluctuation in yearly temperature as well such as heatwaves etc.  
  
#Let us further use cum-sum to determine if there is change is seasonality over the years  
  
Seasonality\_data<- matrix(Model\_HW\_ADD$fitted[,4],nrow = 123)  
  
temp\_data1<- read.table("temps.txt", header = TRUE)  
names(temp\_data1) = gsub(pattern = "X", replacement = "", x = names(temp\_data1))  
  
dim(Seasonality\_data)

## [1] 123 19

dim(temp\_data1)

## [1] 123 21

colnames(Seasonality\_data)<-c(1997:2015)  
days <- temp\_data1[,1]  
rownames(Seasonality\_data)<-days  
  
cumsum\_data <- data.frame(matrix(0, nrow=123, ncol=19))  
rownames(cumsum\_data) <-days  
colnames(cumsum\_data)<-c(1997:2015)  
  
mean\_year1= mean(Seasonality\_data[,1])  
var=sd(Seasonality\_data[,1])  
day\_counter=1  
threshold= var\*3 #(Using threshold for change as 3 times SD, as I want to give some room for variation)  
  
  
for(i in 1:ncol(cumsum\_data)){  
 date\_change=NULL  
 for (j in 1:nrow(cumsum\_data)){  
 val=Seasonality\_data[j,i]  
 cumsum\_data[j,i] <- max(0, cumsum\_data[j-1,i] + (mean\_year1 - val- var\*0.5))  
   
 if (cumsum\_data[j,i]>=threshold){  
   
 cat("Year",colnames(cumsum\_data[i]),"Day winter started is:",  
 rownames(cumsum\_data)[j],"Day","\n")  
 break  
 }  
   
   
 }  
   
}

## Year 1997 Day winter started is: 30-Sep Day   
## Year 1998 Day winter started is: 1-Oct Day   
## Year 1999 Day winter started is: 1-Oct Day   
## Year 2000 Day winter started is: 1-Oct Day   
## Year 2001 Day winter started is: 2-Oct Day   
## Year 2002 Day winter started is: 2-Oct Day   
## Year 2003 Day winter started is: 4-Oct Day   
## Year 2004 Day winter started is: 3-Oct Day   
## Year 2005 Day winter started is: 4-Oct Day   
## Year 2006 Day winter started is: 5-Oct Day   
## Year 2007 Day winter started is: 6-Oct Day   
## Year 2008 Day winter started is: 6-Oct Day   
## Year 2009 Day winter started is: 6-Oct Day   
## Year 2010 Day winter started is: 5-Oct Day   
## Year 2011 Day winter started is: 5-Oct Day   
## Year 2012 Day winter started is: 5-Oct Day   
## Year 2013 Day winter started is: 5-Oct Day   
## Year 2014 Day winter started is: 6-Oct Day   
## Year 2015 Day winter started is: 6-Oct Day

#According to my CUSUM model with C of 0.5 SD and threshold of 3 times SD, on seasonality measure from our Holt Winters model, it appears that summer end is delayed slowly over the years,it ends with 30th September in 1997   
#but for 2015 it ends in 6th October  
#Note that we can't say if summer is longer or shorter because we don't have data to know if summer started earlier or delayed  
#but according to our model it is gradually delayed