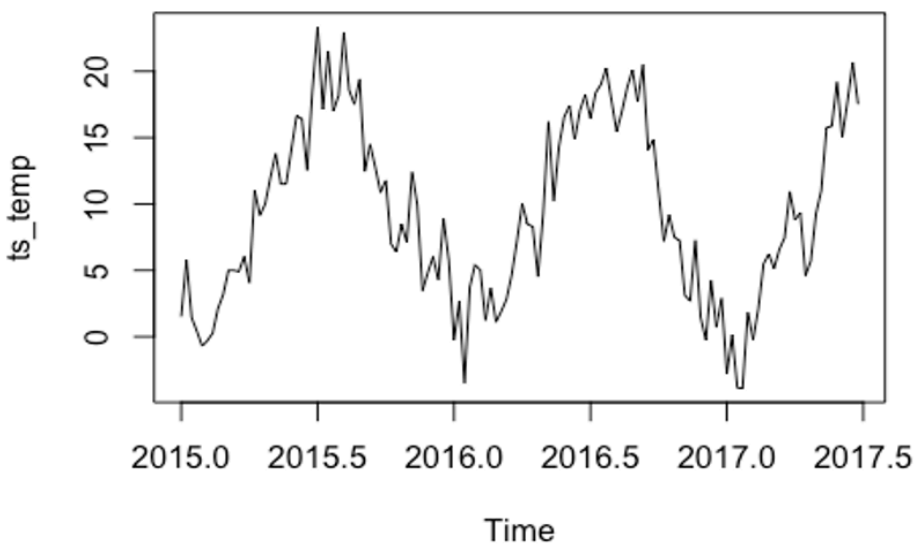
BAX 493A HW#1
Team F: Charles Wang, Qinyi Qiu, Richard Liu, Jie Zhu, Yuxin Yi

Our choice of weather data:

For this homework, our group chose the average temperature from weather to conduct the time series analysis.

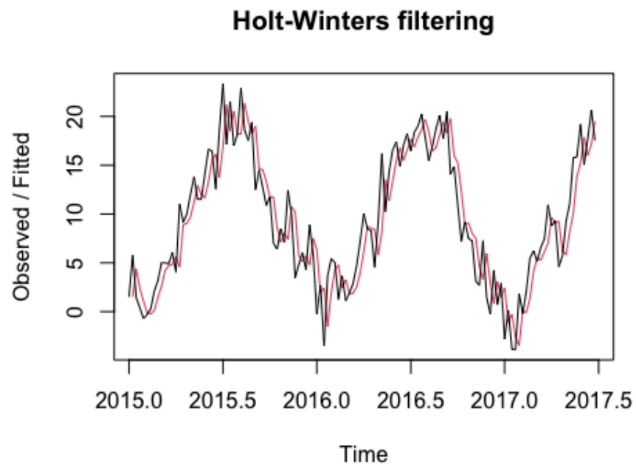
First, we plot a time series graph of the average temperature. In this time series graph, the temperature is plotted against time, where the x-axis represents the time from the beginning of 2015 to 2017. It appears to exhibit a seasonal pattern, as the temperature rises in summer and falls in winter. To forecast the temperature, we are going to use Holt-Winters forecasting.



Try the four combinations of beta (on/off) and gamma (on/off)

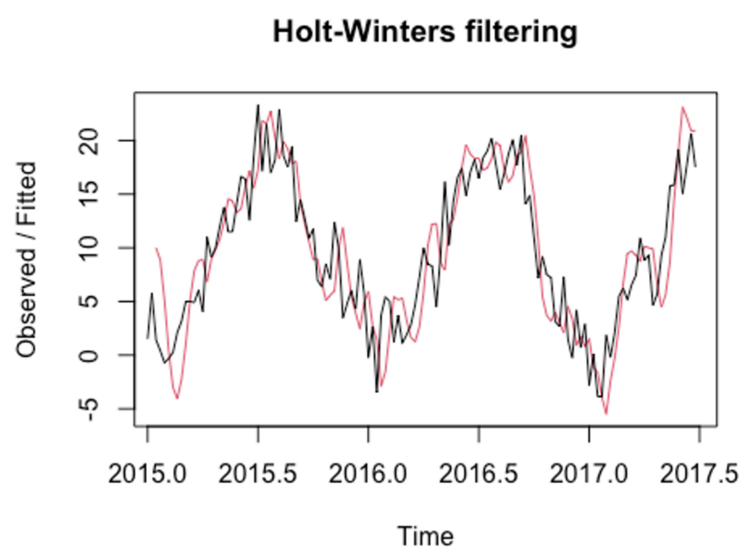
First, we tried the four combinations and here are our interpretations for these models.

Combination 1: Beta off, Gamma off



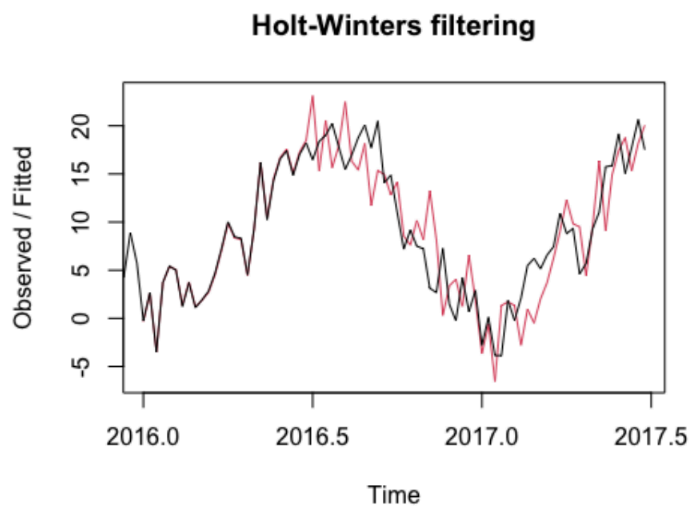
With beta off and gamma off combination, this plot showed us the simple model that only adjusted for the level of the time series. As we can see from the red line, the model didn't capture all of the peaks and troughs as it shows the seasonality of the temperature. Thus, it might be a little less accurate in forecasting.

Combination 2: Beta on, Gamma off



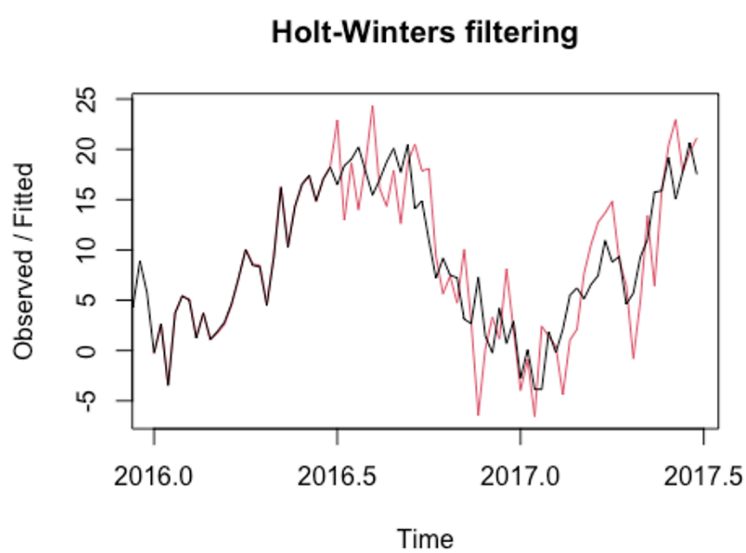
For combination 2, with a beta "on" and gamma "off", this model captures the trends in the data but ignores seasonality. Hence, we can see from the plot that the fitted line briefly follows the direction of the data but usually cannot capture the seasonal peak or low points.

Combination 3: Beta off, Gamma on



For combination 3, The fitted data (red curve), which represents the estimated temperatures using the Holt-Winters method, follows this seasonal pattern closely, rising and falling in sync with the observed data. There is a noticeable lag at some points where the fitted data adjusts to the observed data, but overall it tracks the observed trend fairly well.

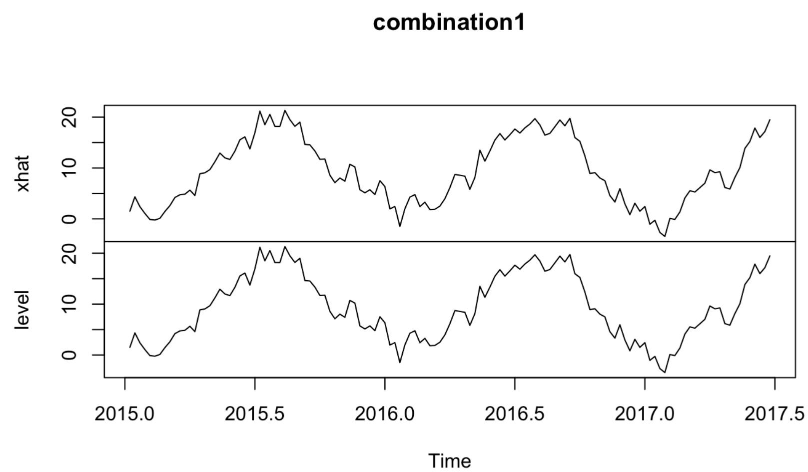
Combination 4: Beta on, Gamma on



With beta “on” and gamma “on” is a full Holt-Winters model that can capture both trend and seasonality in our data. The red curve does overshoot and undershoot at various points compared to the black curve. This suggests that while the model captures the overall seasonal pattern, it does not always accurately estimate the magnitude of the temperature changes.

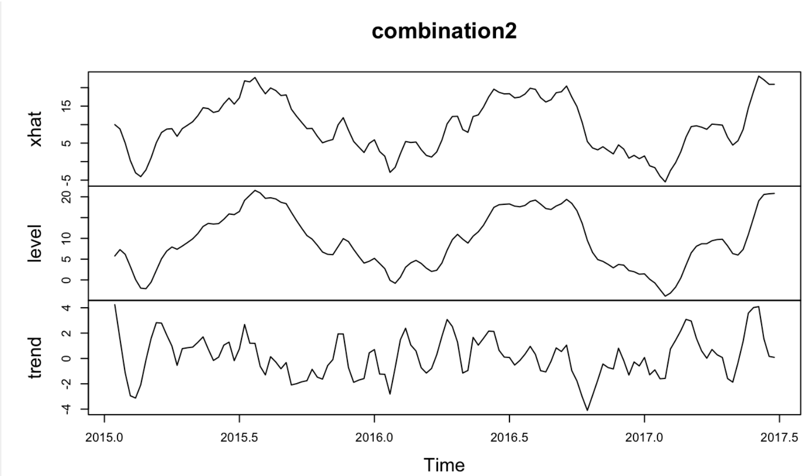
3. Present the plots of data series, trend, and seasonal components for each of the four combinations

<u>Combination 1: Beta off, Gamma off</u>



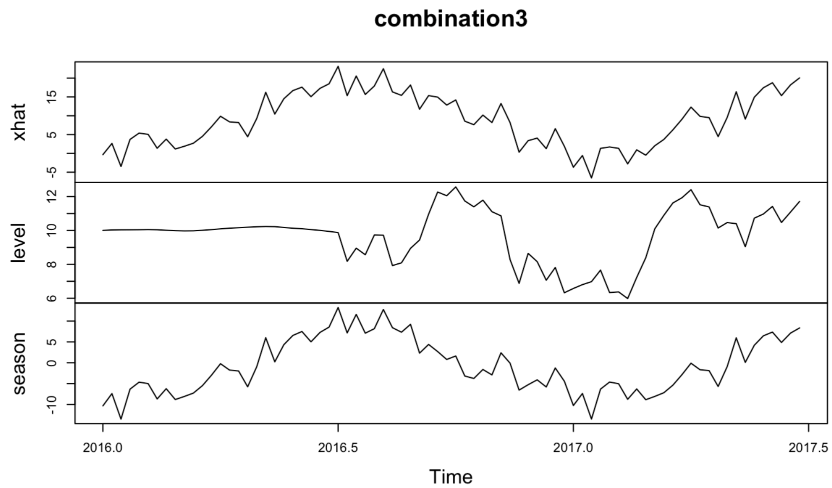
This graph shows the components of a time series forecasting model without any adjustments for both trend and seasonality. The first plot shows the observed data and the second plot shows the average value of the series captured by the model. Since the gamma and beta are all off, the model doesn't capture the trend. We cannot account for the trend through the plot.

<u>Combination 2: Beta on, Gamma off</u>



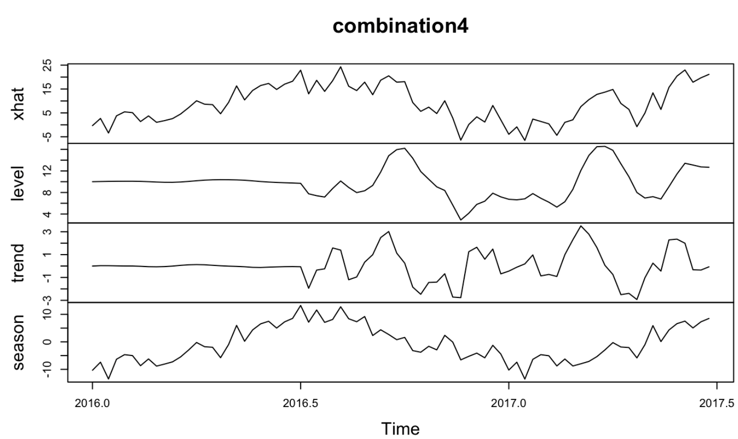
This graph adds a trend component, showing us the clear directional movement over time without the seasonal effects. The level plot shows the level component, which is now being adjusted for the trend in the data. The trend shows the trend component of the model, indicating how the level component's slope is changing over time.

Combination 3: Beta off, Gamma on



This plot displays three components derived from a time series analysis using the Holt-Winters method, with the configuration of beta off (no trend adjustment) and gamma on (seasonal adjustment). The xhat line fluctuates, indicating the model's attempt to predict future values based on the underlying seasonal pattern. The level line appears relatively stable, reflecting the absence of a trend (since beta is off), which means the model does not account for any increasing or decreasing trend over time. The season pattern shows the cyclic fluctuations that repeat over a fixed period.

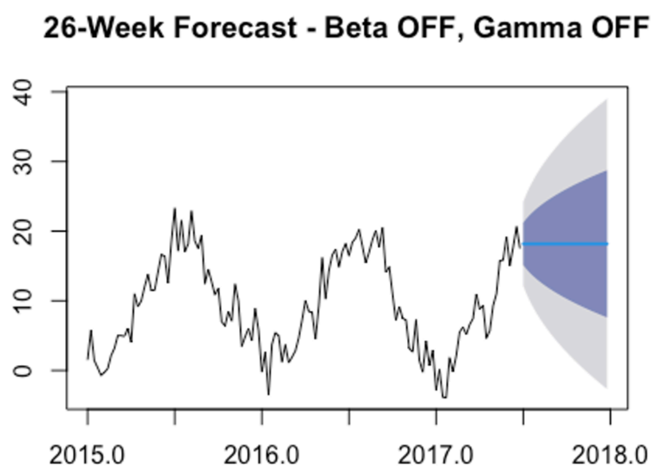
Combination 4: Beta on, Gamma on



This is the most complex model, showing both trend and seasonality. The xhat plot presents the fitted values that incorporate both trend and seasonality adjustments. The level plot shows the baseline level of the series with both trend and seasonality accounted for. The trend plot shows the direction and strength of the trend over time while the season plot captures the regular seasonal patterns.

4. Make out-of-sample forecast for 26 weeks with confidence bands.

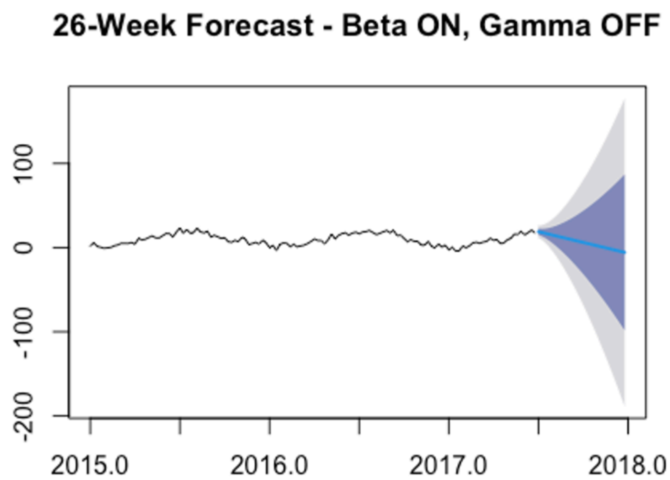
<u>Combination 1: Beta off, Gamma off</u>



The black line represents the historical data from temperature index from 2015 to mid 2017, while the blue line represents the forecast, with the gray and blue shaded areas as its prediction confidence interval.

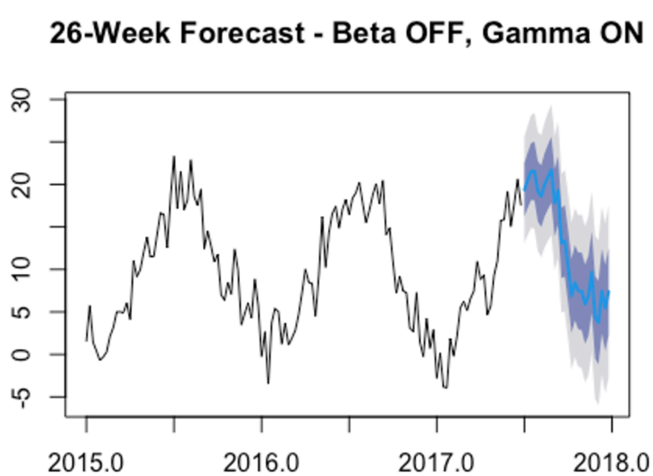
The first combination shows no changes in the forecasting of temperature. The wide area of confidence interval reveals the uncertainty of the forecast, especially when it gradually gets larger as time goes on.

Combination 2: Beta on, Gamma off



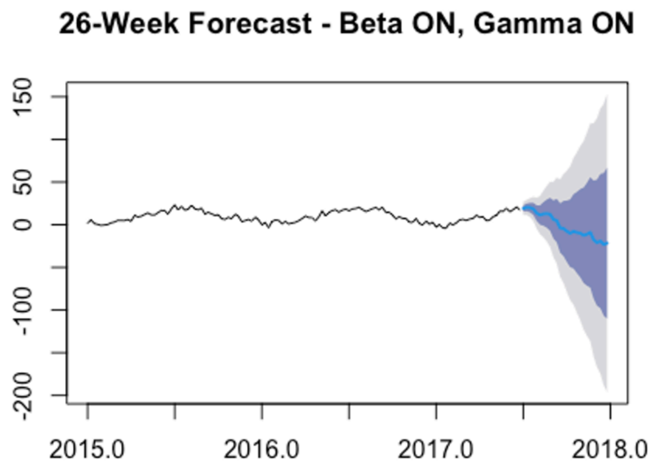
The second combination shows no significant changes in the forecast of the temperature index. The confidence interval is narrow at first but rapidly increases as time goes on, showing the uncertainty of the forecast.

Combination 3: Beta off, Gamma on



The third combination shows continuity from the historical data, which seems to be reasonable as a forecast. The confidence interval is narrow, showing the accuracy of the forecast.

Combination 4: Beta on, Gamma on



The fourth combination shows a relatively flat line in both historical data and forecast, while only downwarding a little bit as time reaches 2018. The confidence interval increases rapidly which suggests the uncertainty of the forecast.

5. Recommend one forecasting model -- give your justification

We recommend model 3: Beta off, Gamma on

As we can see from Holt-Winters filtering, each combination performs a different approach of fitted and observed. In Holt-Winter Filter, beta is for trends and gamma is for seasonality. In our group analysis for avg_temp time series analysis, as we can see in the plot in Q2, the observed value shows a seasonality since it's the average temperature data. Therefore, we need to turn on the gamma for the forecasting model.

From Holt-Winters filtering combination 1 (beta off, gamma off),we can see the fitted graph is following the trends of the observed value. However, for weather forecasting, gamma is a significant factor that needs to be turned on. Therefore, the combination 1 is not the suggested model. Also, from the Q4 26 weeks forecasting, the prediction confidence interval is relatively big, which indicates that the forecasting model has uncertainty in predictions for the average temperature times series with seasonality.

From Holt-Winters filtering combination 2 (beta on, gamma off), we can see the fitted line is somehow not following the trends of the observed value. However, the gamma is off, which is the same as the first forecasting model in which gamma is a significant factor to use when forecasting seasonality. However, as we see in Q4 for this combination of 26 weeks forecasting, the forecasting model has a large prediction confidence interval, which indicates this forecasting model has uncertainty in predictions for the average temperature time series with seasonality.

From Holt-Winters filtering combination 3 (beta off, gamma on), the seasonality gamma is turned on, which is the significance factor for the use of seasonality forecasting. We can see the fitted line in Q2 following exactly the observed value in the graph at the beginning. Also, from the graph for this combination in Q4 26 weeks forecasting, we can see the forecasting model has the smallest prediction of the confidence interval. and it's following the prediction trends. Therefore, this model provides the most accurate forecasting for average temperature time series with seasonality.

From Holt-Winters filtering combination 4 (beta on, gamma on), the seasonality gamma is turned on, that's the significance factor when doing seasonality forecasting for average temperature. In Q2, the fitted line is following exactly the observed value at the beginning of the graph. However, from the graph in Q4 26 weeks forecasting, we can see the prediction of the confidence interval is much wider compared to the combination 3 model, which provides uncertainty in forecasting for the average temperature times series with seasonality.

Therefore, as the justification stated above, the recommended model is the combination 3 with beta off and gamma on. This forecasting model has the narrowest prediction confidence interval and with the gamma (seasonality factor turn on), which provide the most accurate forecasting model among the 4 combinations.

Appendix

```
#install.packages("dplyr")
##install.packages("readr")
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

#install.packages("tidyr")
library(tidyr)
library(psych)
library(readr)
##install.packages("ggplot2")
##install.packages("readxl")
##install.packages("corrplot")
library(readxl)
library(ggplot2)

##
## Attaching package: 'ggplot2'

## The following objects are masked from 'package:psych':
##
##     %+%, alpha

library(corrplot)

## corrplot 0.92 loaded

#install.packages("stringr")
library(stringr)
#install.packages("EnvStats")
library(EnvStats)

##
## Attaching package: 'EnvStats'
```

```
## The following objects are masked from 'package:stats':
##
##     predict, predict.lm
```

```r
#install.packages("PASWR2")
library(PASWR2)
```

```
## Loading required package: lattice
```

```r
#install.packages("pwr")
library(pwr)
#install.packages("car")
library(car)
```

```
## Loading required package: carData

##
## Attaching package: 'car'

## The following object is masked from 'package:EnvStats':
##
##     qqPlot

## The following object is masked from 'package:psych':
##
##     logit

## The following object is masked from 'package:dplyr':
##
##     recode
```

```r
library(forecast)
```

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

```r
#install.packages("forecast")

dataweek <- read_xlsx("data_week.xlsx")
```

Q1

```r
print("For the time series variable, our group choose the avg_temp weather time series in the file to do our analysis.")
```

```
## [1] "For the time series variable, our group choose the avg_temp weather time series in the file to do our analysis."
```

Q2

```r
#Holt-Winters Filter
#Choose avg_temp weather time series
ts_temp <- ts(dataweek$avg_temp, frequency = 52, start = c(2015,1))
plot.ts(ts_temp)

#Combination 1: Beta OFF, Gamma OFF
hw1 <- HoltWinters(ts_temp, beta = FALSE, gamma = FALSE)
plot(hw1)

#Combination 2: Beta ON, Gamma OFF
hw2 <- HoltWinters(ts_temp, beta = TRUE, gamma = FALSE)
plot(hw2)

#Combination 3: Beta OFF, Gamma ON
hw3 <- HoltWinters(ts_temp, beta = FALSE, gamma = TRUE)
plot(hw3)

#Combination 4: Beta ON, Gamma ON
hw4 <- HoltWinters(ts_temp, beta = TRUE, gamma = TRUE)
plot(hw4)
```

\## Q3

```r
#Combination 1: Beta OFF, Gamma OFF
combination1 <- hw1$fitted
plot(combination1)

#Combination 2: Beta ON, Gamma OFF
combination2 <- hw2$fitted
plot(combination2)
```