Homework 4

Richard Albright
ISYE6501
Spring 2018

2/2/2019

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

I work for a Fintech Company analyzing insider trading activity. I have built several exponential smoothing models in the past. One particular exponential smoothing model I built was to construct sector and industry level buy and sell signals. We aggregated up the values of insider's buying and selling activity (using various components) for all companies within an industry or sector, log transformed the data, then applied an arima model accounting for seasonal trends as insider's typically hove blackout period's prior to quarterly earnings which restrict open market transactions. Rather than using a CUSUM method, we used standard deviations from the local mean of our log transformed data to determine buy and sell signals for a given industry or sector. The model can reliably pick market bottoms for a givin sector or industry within 10 days (the average being within 5), while the sell signals were not nearly as reliable (while insider's usually buy due to perceived weakness in their company's stock, they sell for a multitude of reasons). Our alpha in our model was close to 0, as there is a lot of randomness in insider trading data.

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.)

Read in the CSV

```
data <-
  read.table(
    "/Users/ralbright/Dropbox/ISYE6501/week4/homework/temps.txt",
    header=TRUE,
    stringsAsFactors = FALSE,
    sep="\t"
  )
```

Then transform the data into a time series.

```
temps_vec <- as.vector(unlist(data[,2:21]))
temps <- ts(temps_vec, start=1996, frequency=123)
```

Let's see the Head, Tail, and Summary Data of our Atlanta high tempuratures time series.

Head:

```r
table <- xtable(t(as.matrix(head(temps))))
print(table, type='latex', comment=FALSE, scalebox='1.0')
```

	1	2	3	4	5	6
1	98	97	97	90	89	93

Tail:

```r
table <- xtable(t(as.matrix(tail(temps))))
print(table, type='latex', comment=FALSE, scalebox='1.0')
```

	1	2	3	4	5	6
1	67	56	78	70	70	62

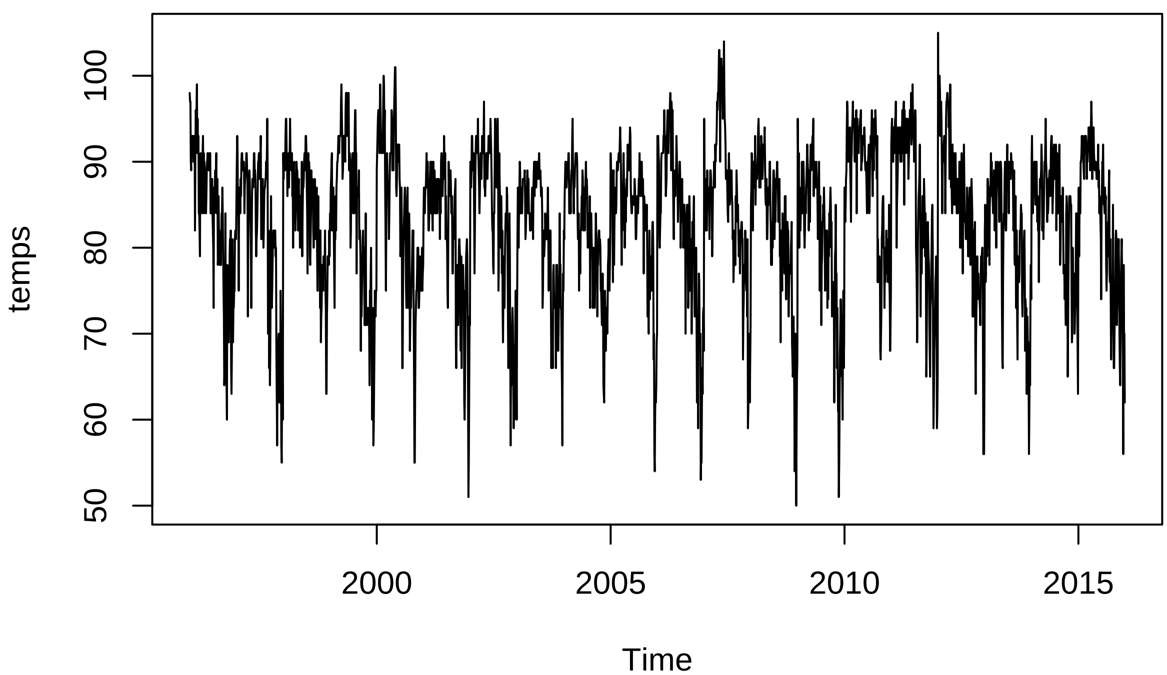
Summary:

```r
table <- xtable(as.matrix(summary(temps)))
print(table, type='latex', comment=FALSE, scalebox='1.0')
```

	x
Min.	50.00
1st Qu.	79.00
Median	85.00
Mean	83.34
3rd Qu.	90.00
Max.	105.00

Then lets plot our time series tempurature data.

```r
plot.ts(temps)
```



We'll then apply the Holt-Winters triple exponential smoothing model. We will leave our alpha, beta, and gamma parameters as null so the model will choose the best values. We will also use a multiplicative seasonal model.

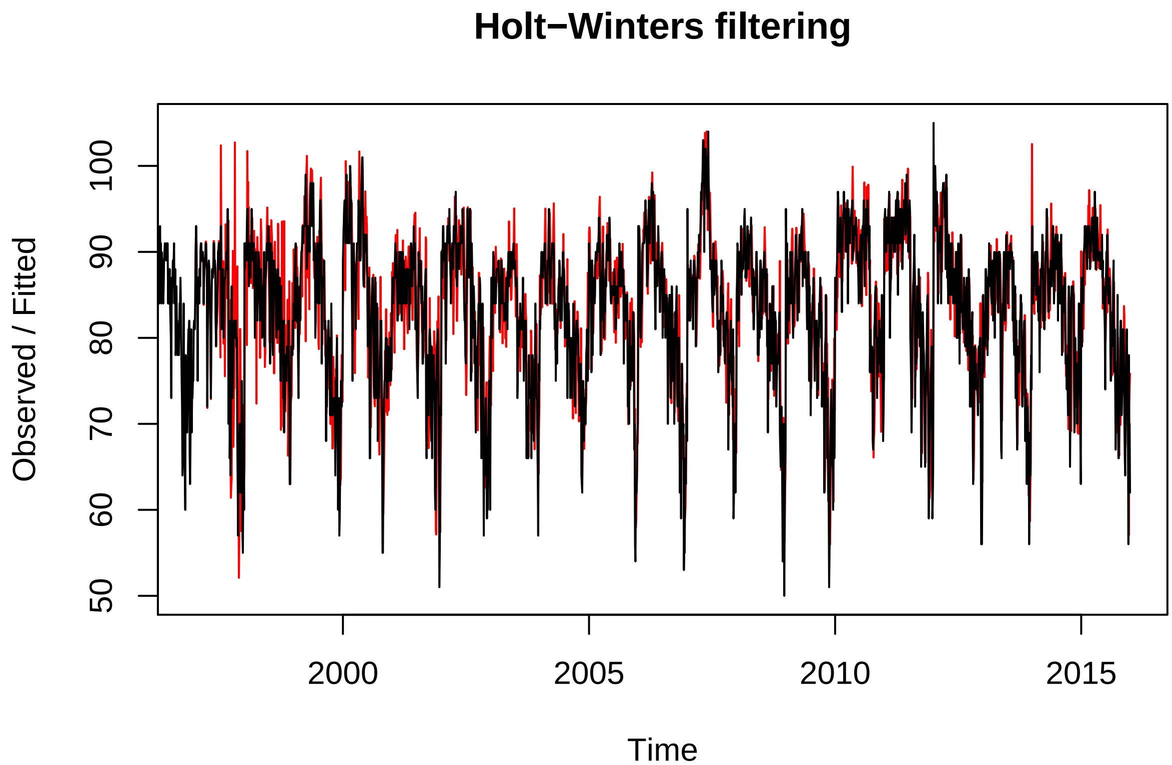
```r
hwModel <- HoltWinters(temps, alpha = NULL, beta = NULL, gamma = NULL, seasonal = "multiplicative")
```

Our model's alpha, beta, and gamma parameters are as follows:

Alpha : 0.615003, Beta: 0, and Gamma: 0.5495256

First let's plot our model.

```r
plot(hwModel)
```



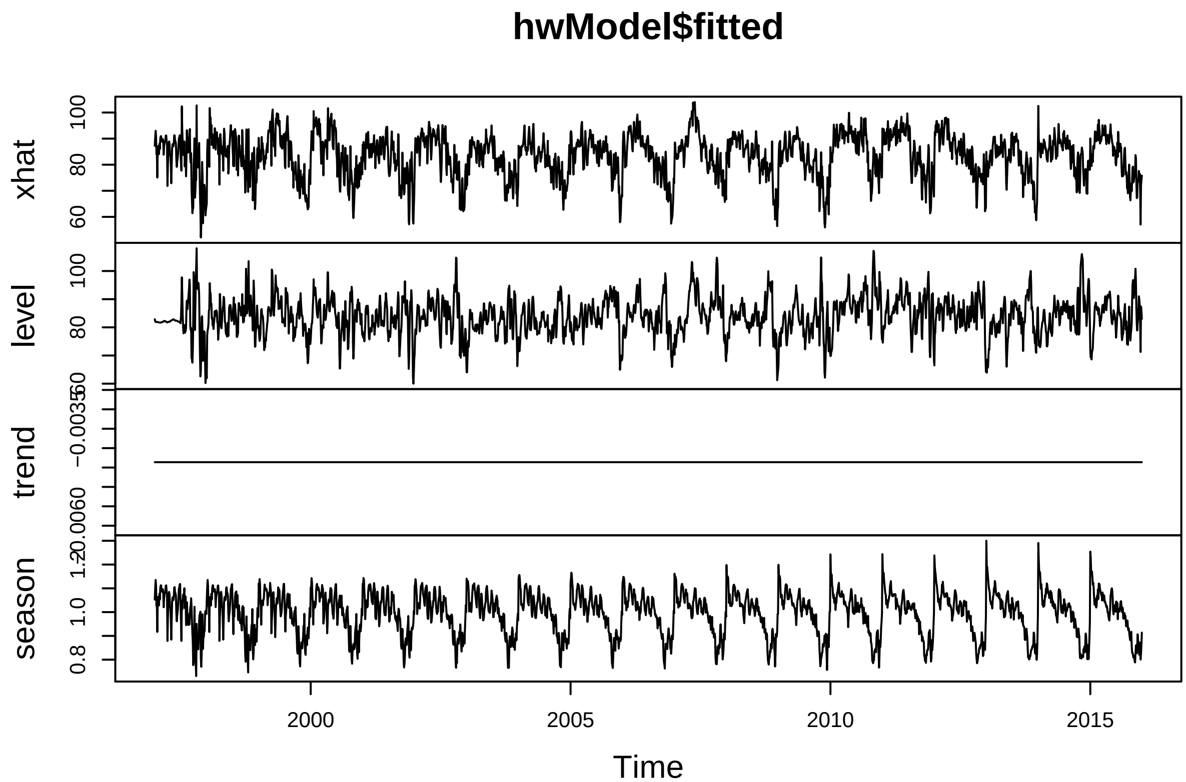
Notice how our fitted model (in red) starts in 1997, and uses the 1st 2 years (1995, 1996) as training data.

With our resulting model has an alpha of 0.615003, it sits on the higher end of 0 < alpha < 1, indicatting that the level of the model is not that random. This implicitly makes sense as is tempurature data taken from the summer going into fall, one would expect the days preceding and following an observation to be quite similar.

The beta for our resulting model is 0. While it is only indicative in the Holts-Winter Model that the initial beta it estimated as correct, a look of at the model's plot also indicates that our model is stationary. Or Gamma parameter of 0.5495256 indicates that there is some seasonality, which makes sense, considering our data set is of high temperatures going from the summer into fall.

We will then plot our fitted model.

```r
plot(hwModel$fitted)
```



If we wanted to forecast 2016's values we could use the R forecast package. We are forecasting July's values for 2016.

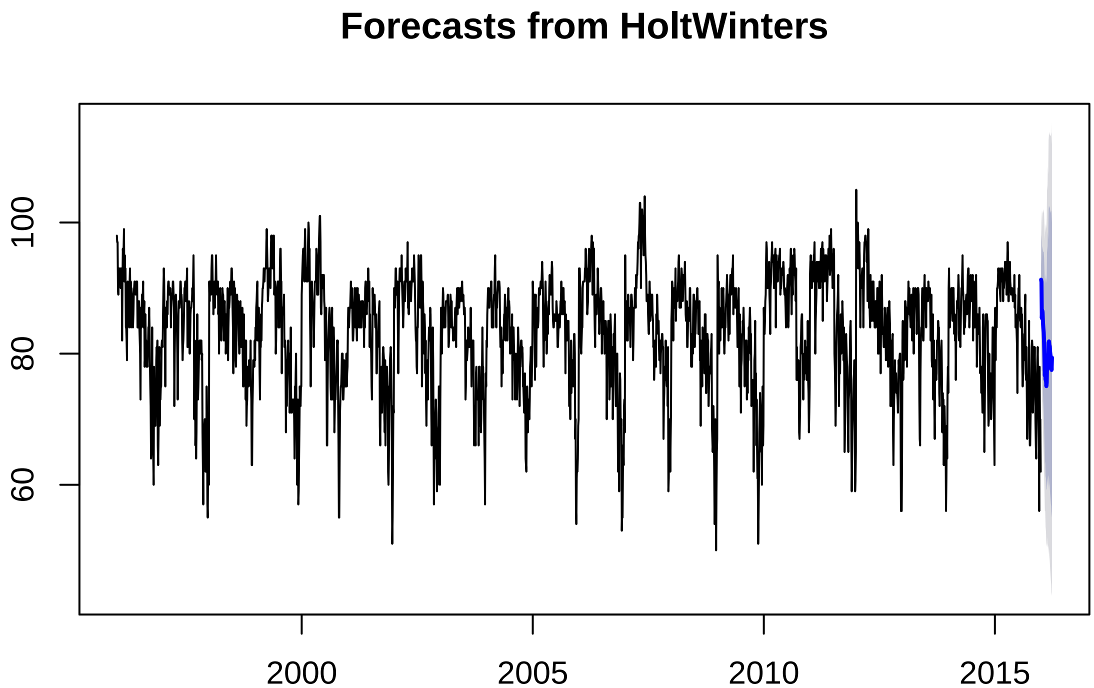
```
forecast <- forecast(hwModel, h=30)
forecast
```

```
##          Point Forecast    Lo 80     Hi 80    Lo 95     Hi 95
## 2016.000       91.28516 85.53219  97.03813 82.48676 100.08357
## 2016.008       90.93510 83.77404  98.09617 79.98320 101.88701
## 2016.016       85.41693 77.33542  93.49844 73.05732  97.77653
## 2016.024       86.57116 77.35084  95.79148 72.46990 100.67242
## 2016.033       86.27852 76.13600  96.42105 70.76687 101.79018
## 2016.041       84.77782 73.90744  95.64821 68.15301 101.40264
## 2016.049       83.91440 72.32111  95.50769 66.18400 101.64480
## 2016.057       83.25410 70.97137  95.53682 64.46929 102.03890
## 2016.065       81.77658 68.95411  94.59904 62.16631 101.38685
## 2016.073       79.25010 66.07243  92.42777 59.09659  99.40361
## 2016.081       76.65370 63.17219  90.13522 56.03550  97.27190
## 2016.089       77.90779 63.56895  92.24663 55.97843  99.83716
## 2016.098       76.01528 61.35964  90.67092 53.60142  98.42915
## 2016.106       76.28765 60.97198  91.60333 52.86435  99.71096
## 2016.114       75.03842 59.36111  90.71572 51.06205  99.01479
## 2016.122       75.57908 59.22460  91.93355 50.56707 100.59108
## 2016.130       78.84387 61.29033  96.39740 51.99806 105.68968
## 2016.138       77.63576 59.81550  95.45601 50.38203 104.88949
## 2016.146       79.80801 61.01698  98.59904 51.06961 108.54641
## 2016.154       78.34675 59.39450  97.29900 49.36178 107.33172
## 2016.163       81.66986 61.48593 101.85379 50.80120 112.53851
## 2016.171       81.87420 61.19263 102.55577 50.24447 113.50393
## 2016.179       81.22923 60.26218 102.19629 49.16289 113.29557
```

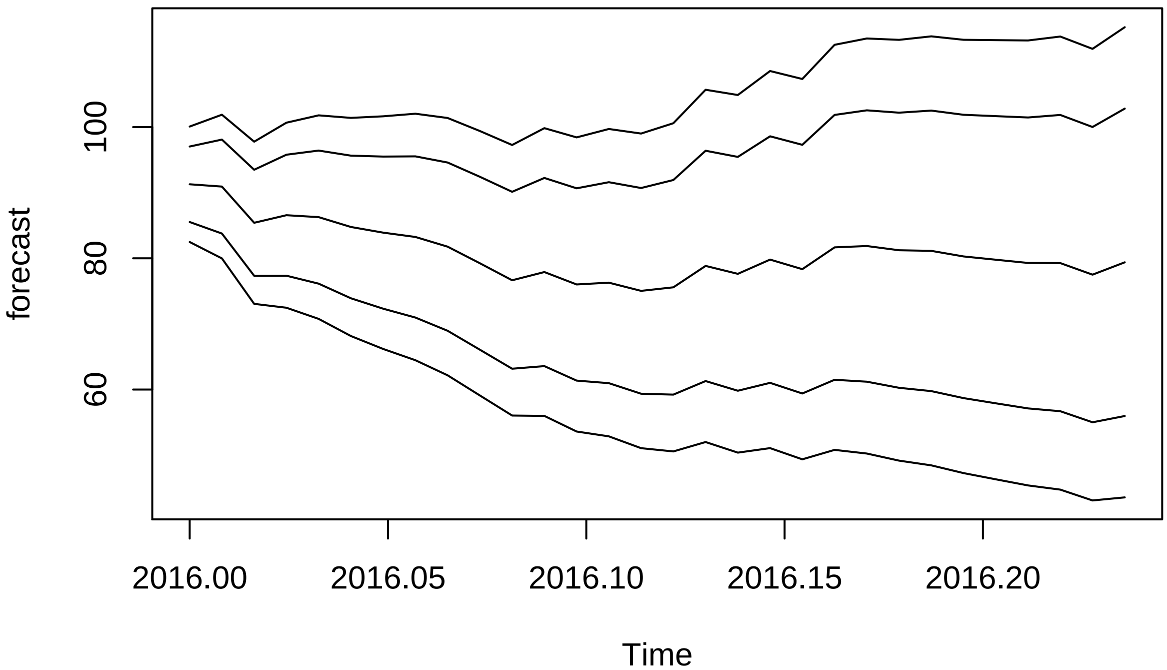
```
## 2016.187       81.13618 59.75822 102.51414 48.44141 113.83094
## 2016.195       80.28483 58.69412 101.87554 47.26469 113.30497
## 2016.203       79.78376 57.89959 101.66794 46.31480 113.25272
## 2016.211       79.29325 57.12110 101.46541 45.38387 113.20264
## 2016.220       79.27248 56.69520 101.84975 44.74352 113.80144
## 2016.228       77.50932 55.00734 100.01131 43.09551 111.92314
## 2016.236       79.39238 55.96382 102.82095 43.56149 115.22328
```

Lets plot our forecast, both on the end of our model, and then just the forecast itself.

```r
plot(forecast(hwModel, h=30))
```




```r
plot.ts(forecast)
```



The forecasted model shows the points of the forecast, and the 80% and 95% bands of certainty around our forecast. Notice how the bands are increasing the further in the future we go. Since the forecast is a stochastic process, it follows a brownian motion. The bands are indicative of more uncertainty of our models forecasting ability as we go further out into the future.

Typically when building such models, one would want to know how significant the autocorrelation is over time. Autocorrelation can best be described as the similarity between observations as a function of time. We can test for autocorrelation with either the acf function or the Box.test function. You would use the residual errors to determine autocorrelation.

Let's look at the autocorrelation over the past 10 days.

```
Box.test(forecast$residuals, lag=10, type="Ljung-Box")
```

```
##
##  Box-Ljung test
##
## data:  forecast$residuals
## X-squared = 332.61, df = 10, p-value < 0.00000000000000022
```

With a p-value of 0.00000000000000022, Our model is highly autocorrelated. Which reiterates what we implicitly know base on our model's alpha.

Since the question posits whether or not summers in Atlanta have been getting longer, we want to analyse the seasonal component. From an initial inspection of our fitted model in the plot above, It appears that the seaonal variation has been increasing since about 2007. So let's dig a little deeper.

Lets transform our the seasonal component back into a matrix for further analysis. Below is the head of our model's seasonal component.

```
hwModel.season <- matrix(hwModel$fitted[,4], nrow=123)
table <- xtable(head(hwModel.season))
print(table, type='latex', comment=FALSE, scalebox='0.75')
```

In order to see if the seasonal component is stationary, the CUSUM method would be ideal for this analysis. We also want to determine when each year has been crossing some threshold to determine if the days' highs

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	1.05	1.05	1.12	1.10	1.12	1.11	1.14	1.14	1.13	1.12	1.16	1.20	1.20	1.24	1.24	1.24	1.30	1.29	1.25
2	1.10	1.10	1.11	1.10	1.11	1.12	1.13	1.15	1.14	1.13	1.14	1.13	1.15	1.17	1.17	1.19	1.19	1.22	1.23
3	1.14	1.14	1.14	1.14	1.14	1.14	1.13	1.16	1.17	1.15	1.15	1.14	1.15	1.16	1.16	1.17	1.19	1.17	1.17
4	1.11	1.11	1.12	1.13	1.13	1.13	1.13	1.14	1.15	1.15	1.14	1.15	1.15	1.16	1.16	1.16	1.17	1.17	1.16
5	1.03	1.03	1.04	1.07	1.08	1.10	1.12	1.10	1.12	1.13	1.13	1.14	1.14	1.11	1.13	1.13	1.15	1.17	1.17
6	1.03	1.03	1.03	1.04	1.05	1.07	1.08	1.09	1.10	1.09	1.08	1.09	1.08	1.10	1.12	1.12	1.12	1.13	1.15

are staying hotter for longer periods or time. Since we will be doing this analysis in excel, we will need to export our seasonal data component into an excel file.

```
excel_file <- '/Users/ralbright/Dropbox/ISYE6501/week4/homework/seasons.xlsx'
wb <-  createWorkbook()
addWorksheet(wb, "Seasons", gridLines = FALSE)
writeData(wb, sheet = 1, hwModel.season, rowNames = TRUE, colNames=TRUE)
saveWorkbook(wb, excel_file, overwrite = TRUE)
```

In order to determine if the summers in Atlanta are getting longer. I performed a CUSUM analysis on the high tempuratures for each year. I used the month of July to calculate my initial mean for each year. I used a 0.5 standard deviation for my C value, and 5 standard deviations for my threshold for each year. After calculating a CUSUM for downward change detection on each year. I then calculated out the 1st date that the CUSUM crossed the threshold value for each year. I then alo calculate out the last date that the CUSUM crossed the threshold value for each year. I plotted both sets of dates to determine if the summers were getting longer in Atlanta over time.

The day that the CUSUM initially crosses the downward threshold is trending further away from date of first initial cross. The day that the CUSUM permanently crosses the downward threshold has a decreasing trend ptior to 2004, then appears stable afterwards. So while the initial cool offs year over year are happening later, the permanent cool offs are happening earlier from 2004 on. We would need to more data on periods preceding 1995 in order to see if there really is a trend of summer's getting longer in Alanta.

Below are the plots used to make that determination.

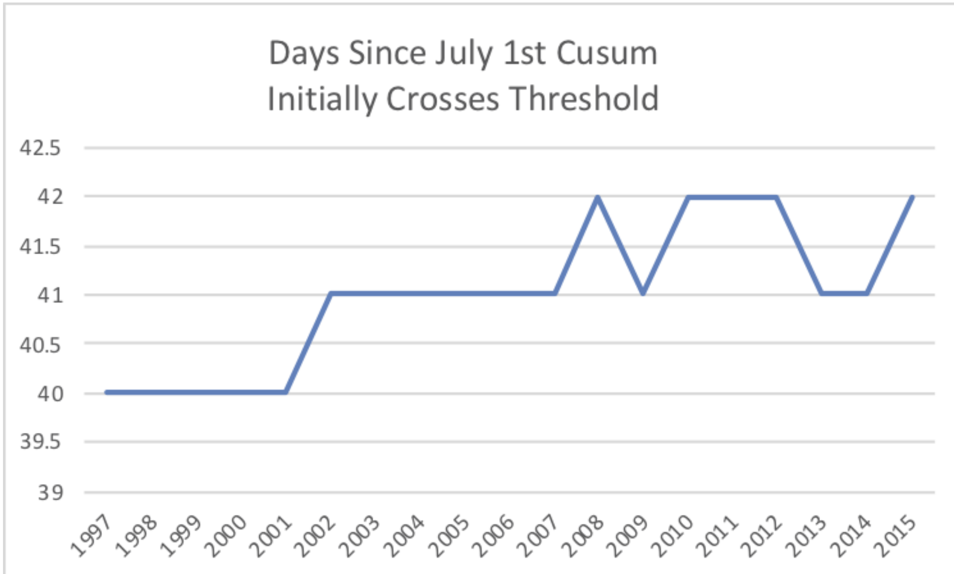


Figure 1: Holt Winters Seasonal Component Yearly Initial Crossing of Threshold

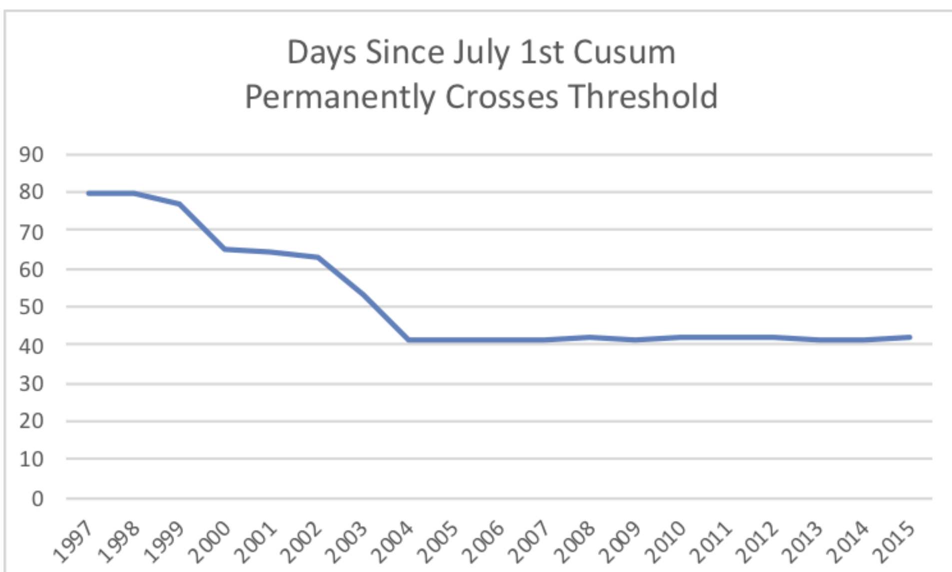


Figure 2: Holt Winters Seasonal Component Yearly Permanent Crossing of Threshold