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?

Response

I believe that exponential smoothing models could have many applications within finance. For example, when analyzing securities, an exponential smoothing model could be used to filter out high frequency daily noise and focus on longer term trends and low frequency seasonal patterns. The information gained by using an exponential smoothing could be used to make trading decisions. The appropriate alpha value would likely vary based upon the security being analyzed, but generally to determine long term trends, I would expect the smoothing parameter to be closer to *0* than *1* since securities often exhibit a high level of high frequency noise.

Question 7.2

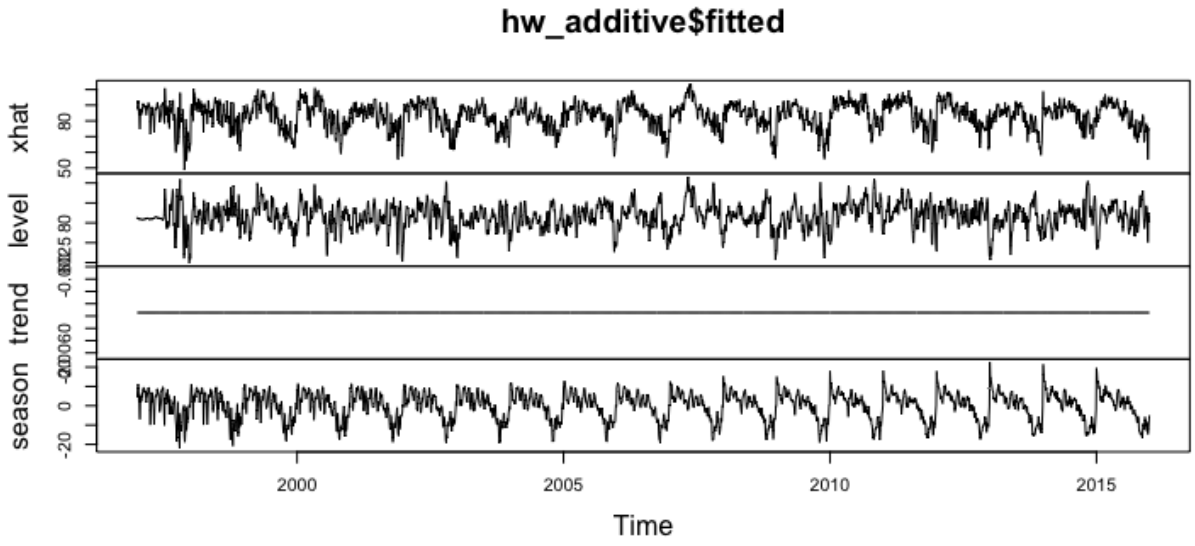
Using the 20 years of daily high temperature data for Atlanta (July through October) from Question 6.2, build and use an exponential smoothing model to help make a judgment of whether the unofficial end of summer has gotten later over the 20 years. (Part of the point of this assignment is for you to 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.)

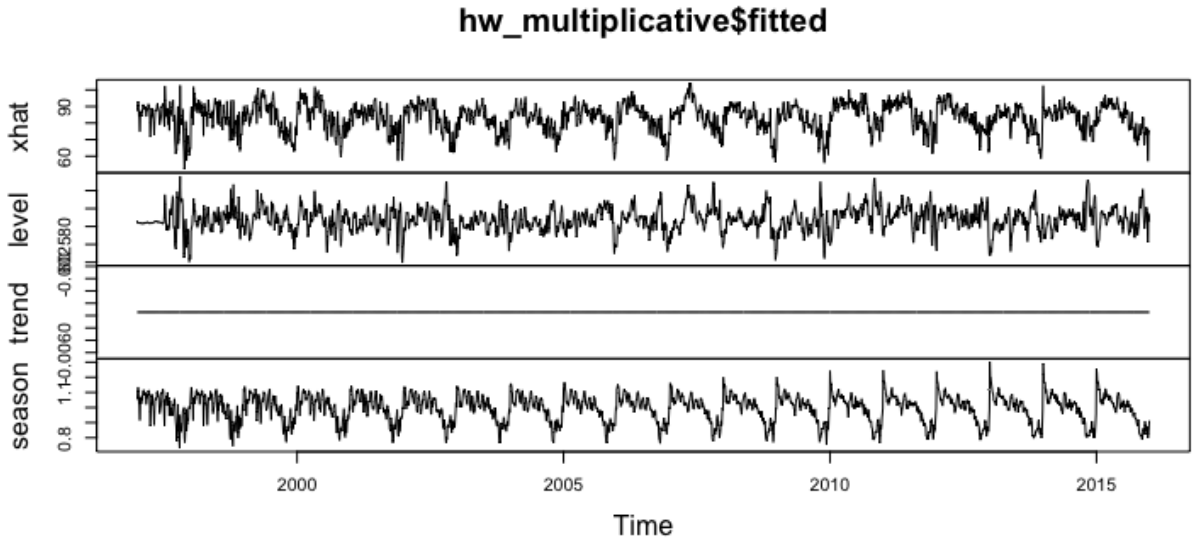
(Please note that my R code for this question is contained in the file 'homework_4_Q7.2.R', and if you wish to run the code, you will need to change line 5 of my R code to your local directory which contains the 'temps.txt' data)

Response

To solve this problem, I implemented *exponential smoothing* in conjunction with the *cumulative sum* algorithm. I implemented *exponential smoothing* in R, using the built-in *HoltWinters* function. To begin, I read in the *temps.txt* data, and converted it into a time series format so that it could be used an input parameter in the function *HoltWinters*.

My next step was to determine if I should use a multiplicative or additive seasonality parameter in the *HoltWinters* function. To do this, I ran the *HoltWinters* function twice, once setting the seasonality parameter to *additive* and once setting the seasonality parameter to *multiplicative*. In both cases, I used the time series temperature data as the only other input in the function. The results may be observed in the figures below:



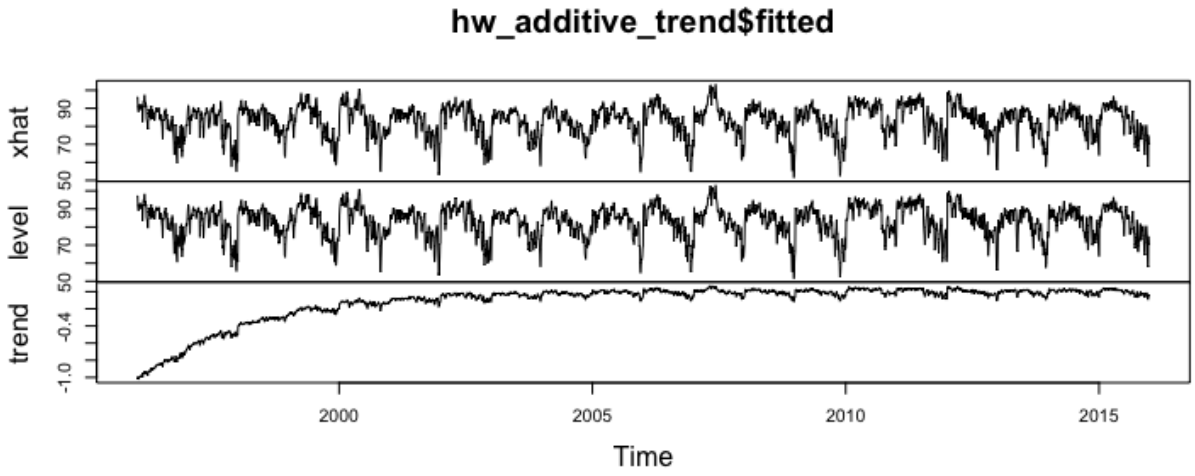


As it may be observed in the two figures above, the results of running the *HoltWinters* function with additive and multiplicative seasonality types both appear to produce very similar results. As mentioned, seasonality and the temperature data were the only input parameters I used in the *HoltWinters* function. This means that *HoltWinters* function automatically determined the optimal *alpha* (α), *beta* (β), and *gamma* (γ) parameters. In the case in which I set *seasonality* to be *additive*, the following parameter values were determined by the function: $\alpha = 0.6611$, $\beta = 0$, and $\gamma = 0.6248$. In the case in which I let set *seasonality* to be *multiplicative*, the following parameter values were determined by the function: $\alpha = 0.6150$, $\beta = 0$, and $\gamma = 0.5495$. In both cases, a similar set of parameters were determined by the *HoltWinters* function. The similar set of parameters produced in

conjunction with the similar plots produced under both conditions indicates that in either scenario, a similar result will likely be determined.

So, which parameter should ultimately be chosen? If the amplitude of *seasonality* is either increasing or decreasing, then *multiplicative* seasonality is the appropriate choice. If the amplitude *seasonality* remains constant, then *additive seasonality* is the appropriate choice. Based on the two figures above, it appears as though we have a somewhat borderline case when observing the amplitude of the seasonality. However, I don't feel that we can definitively say that amplitude is increasing (or decreasing), so I chose to set *seasonality* to *additive* in the rest of my analysis. As mentioned however, either choice will yield similar results.

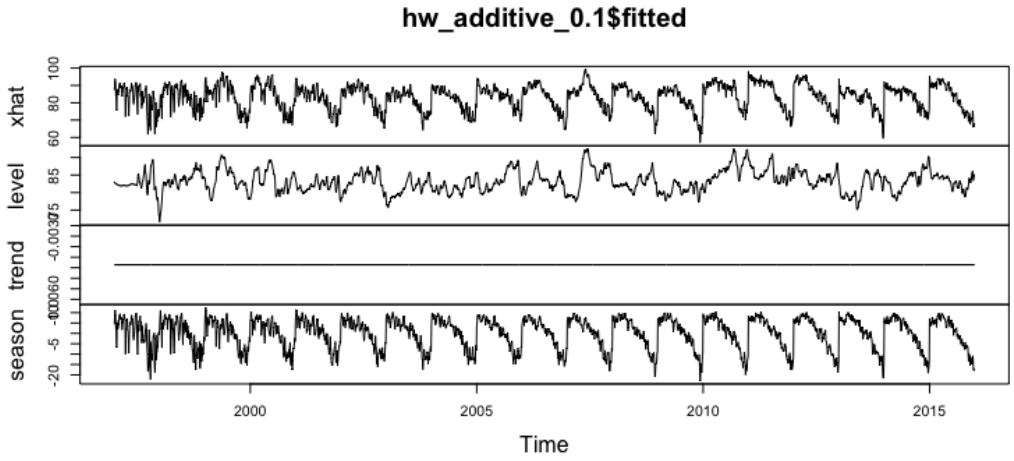
My next step was to again use the *HoltWinters* function with the temperature data as an input, and *seasonality* set to *additive*. However, this time I specified that the parameter *gamma* should be set to FALSE. *gamma* is a seasonality parameter and setting *gamma* to FALSE fits a non-seasonal model. Therefore, by setting *gamma* to false I was able to observe the *trend* of the model. The results of running the *HoltWinters* function with the above parameters may be observed below:

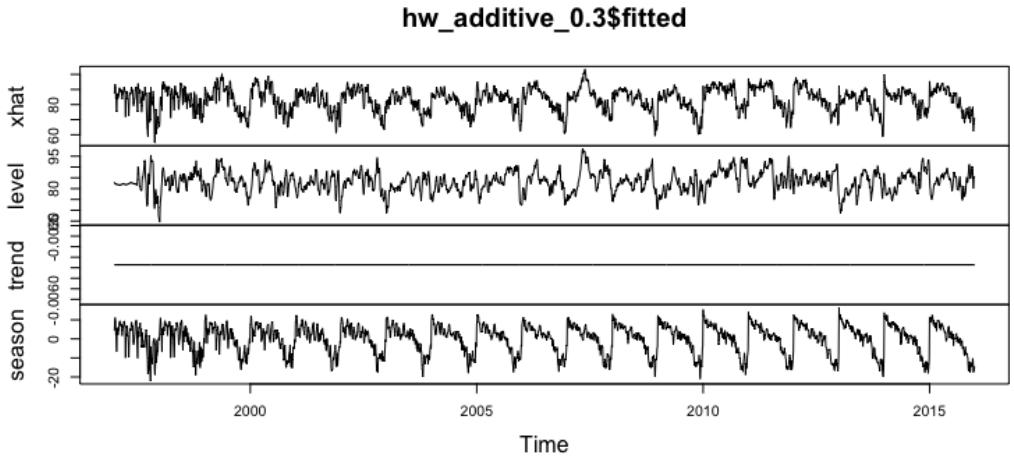


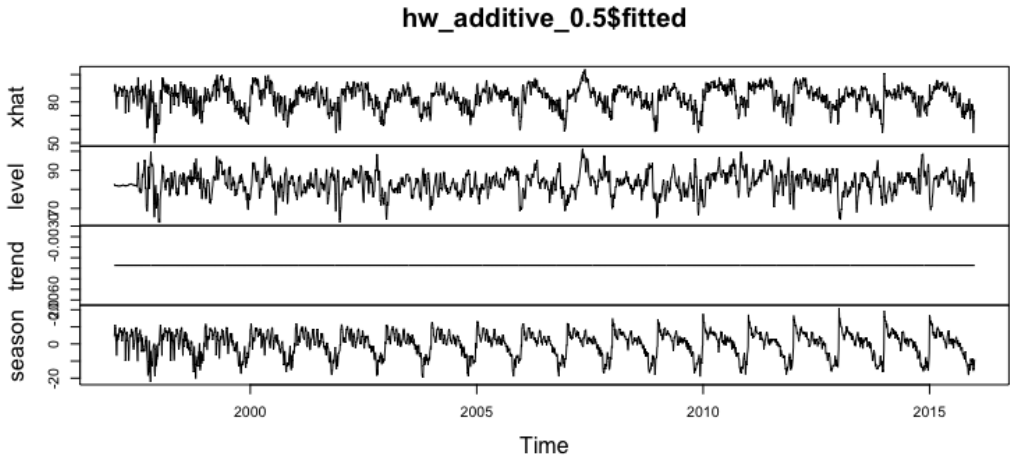
As may be observed by looking at the above figure, the trend does indeed imply a slight increase in temperature, particularly prior to the year 2000. An increasing trend in the summer temperature is indicative of idea that the unofficial end of summer may, on average, be occurring later each year over the twenty-year period we are analyzing. However, this is not definitive proof.

To analyze the question of whether summer is unofficially ending later in a more rigorous manner, I chose to apply the *cumulative sum*, or *cusum* function to the fitted smoothed temperature values (labeled xhat in the figures above) produced by the *HoltWinters* exponential smoothing function.

However, prior to implementing the cumulative sum algorithm, I first wanted to be sure that I had generated the correct exponential smoothing model with regard to the choice of the parameter *alpha*. If no user input is given, then R's *HoltWinters* function automatically assigns a value to *alpha.* However, I wanted to see how the model is affected by altering the parameter *alpha.* Since we expect to see seasonal fluctuations in our data from one year to the next, our ideal choice of *alpha* is the one which will preserve the low-frequency yearly seasonal changes but reduce the "noise" of the high-frequency daily temperature fluctuations. In the first figure above, with the *HoltWinters* determined *alpha* value of $\alpha = 0.6150$, there appears to be an excessive amount of daily noise contained in the model produced. To reduce the daily noise, I ran the *HoltWinters* function three more times and tried setting the parameter *alpha* to the values *0.1 , 0.3,* and *0.5* . The only other input parameters I included in the *HoltWinters* function were the Georgia summer temperature data and the additive seasonality type. The results may be seen in the figures below:







As may be observed in the figures above, when letting $\alpha = 0.1$, we see the greatest smoothing effect, and when letting $\alpha = 0.5$ we see the smallest smoothing effect as expected. By observing the above figures, it appears as though when we let $\alpha = 0.1$ we see a reduction in the clarity of the low-frequency yearly seasonal fluctuations, which means that the model will likely be inaccurate. When we let $\alpha = 0.5$ the seasonal fluctuations are preserved, but there appears to be an excessive amount of daily noise in the model. When we let $\alpha = 0.3$, it appears as though this model does the best job of reducing the high-frequency noise while simultaneously retaining the low-frequency seasonal changes.

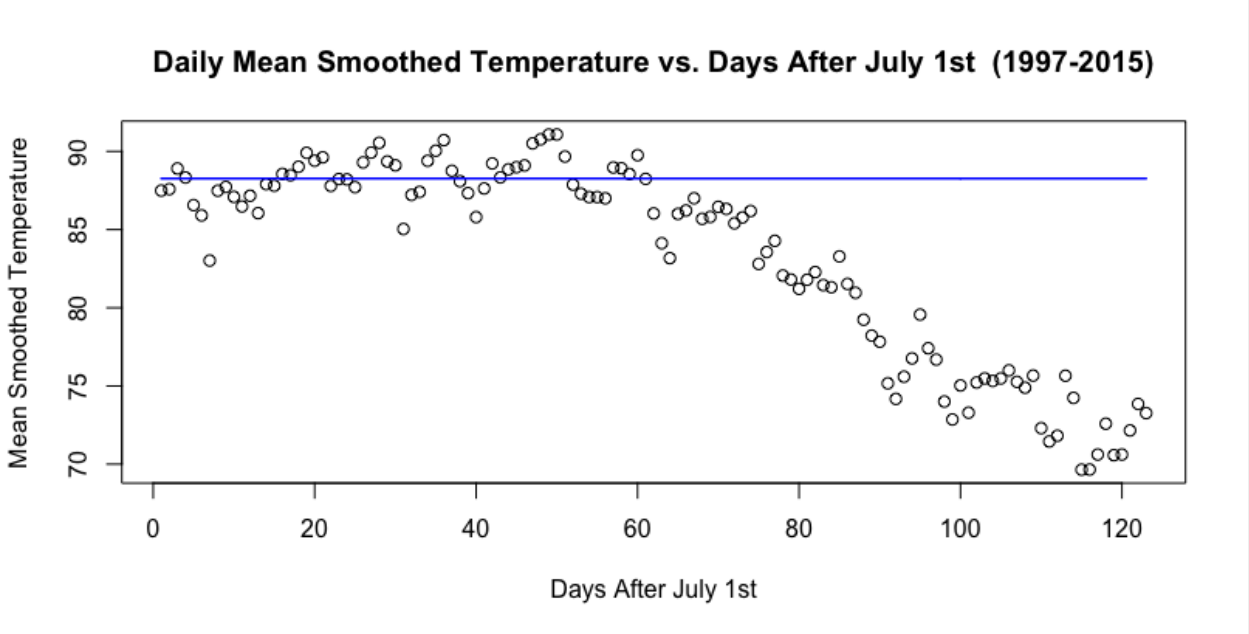
As a final step, I used the *cusum* algorithm on the smoothed temperature component of the model generated by the *HoltWinters exponential smoothing* function. When creating exponential smoothing model, I let $\alpha = 0.3$, and I set the seasonality parameter to *addtive*, since, as described above, these seem to (arguably) be the best parameter choices. To examine if the unofficial end of summer was occurring later each year, I applied the *cusum* algorithm to each year (123 day period) of the fitted temperature values, and detected when a sufficiently large decrease occurred each year, as defined by the *cusum* function and the paramters parameters I supplied, to say that summer had ended. The *cusum* formula used to measure a decrease may be defined as follows:

$$S_t = max\{0, S_{t-1} + (\mu - x_t - c)\}$$

Where S_t = the value returned by the *cusum* function after analyzing point t in the data set, S_{t-1} = the *cusum* value associated with the data point $t-1$ (the previous value returned by the *cusum* function), μ = the average value of the data set being analyzed, if there were no change present, x_t = data point t in the in the dataset being analyzed, and c is a user defined *critical value* parameter that determines how quickly changes accumulate.

To begin my *cusum* analysis, I started by visualizing the seasonal data. I found the mean fitted temperature value on each day over the 19-year period from 1997 to 2015. Additionally, I found the mean summer seasonal data for the combined months of July and August. At this point, it should be pointed out that the model produced by the *HoltWinters*

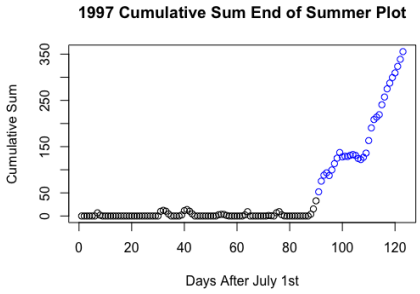
function is a prediction of each day after the first year. The first year, in our case 1996, is used to initialize the model, thus there are only 19 years of fitted data returned, in comparison to the 20 years of input data contained in the Georgia summer temperature data. The results of the plot may be seen below:

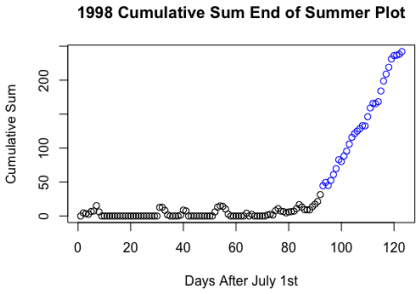


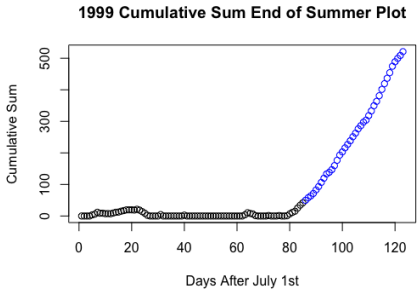
I chose to include the mean of the July and August seasonal values in this figure because it serves as a proxy for the mean summer temperature since we know that the summer does not end before August. Therefore, we can say that when the daily fitted temperature is consistently below the mean of the fitted summer temperature, then summer has unofficially ended, based on our definition above. Based upon observing the above figure, we can see that the daily mean fitted temperature drops below the mean summer temperature at approximately day 60, which corresponds August, 29th.

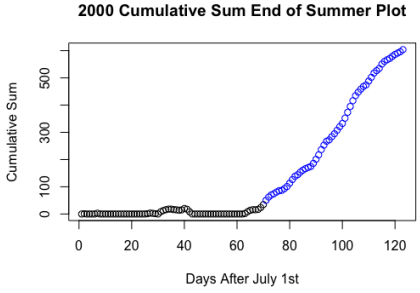
I next analyzed the above hypothesis that summer is unofficially ending at a later date each year on average by using the *cusum* function. I created a function called *cumulative_sum* in R which works in accordance with the formula of *cusum* defined above. To determine if the summer is ending at a later date each year, I used the *cusum* function to find the unofficial summer end date each year. I then used the *cusum* function a second time on the values returned by the first round of *cusum* analysis (the days determined to be the end of summer for each year) to determine if the summer was ending later each year.

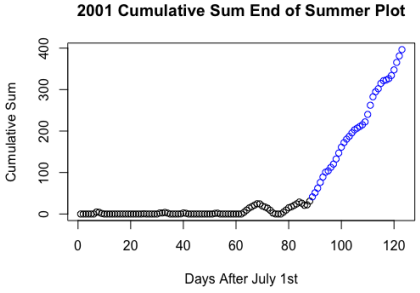
When determining the end date of the summer each year, I let μ = the mean July and August temperature for the given year in the *cusum* algorithm. I let $c = 0.5$ *the standard deviation of daily temperatures in July and August for the given year. And in order to determine, based upon the result of the *cusum* function, when summer had ended, I again included a threshold parameter T as described above. I let $T = 10$ *the standard deviation of daily temperatures in July and August for the given year. The results of this analysis may be observed in the figures below:

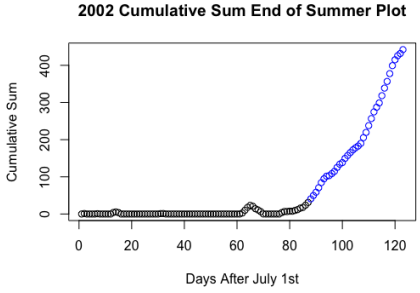

1997 Cumulative Sum End of Summer Plot
Cumulative Sum
0
50
150
250
350
0
20
40
60
80
100
120
Days After July 1st

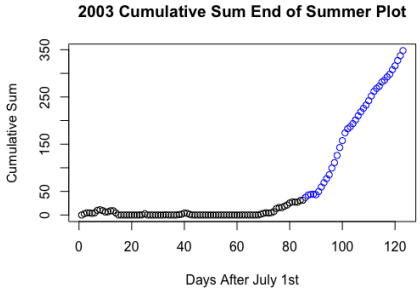

1998 Cumulative Sum End of Summer Plot
Cumulative Sum
0
50
100
200
0
20
40
60
80
100
120
Days After July 1st

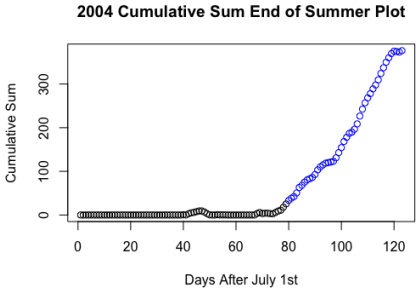

1999 Cumulative Sum End of Summer Plot
Cumulative Sum
0
100
300
500
0
20
40
60
80
100
120
Days After July 1st

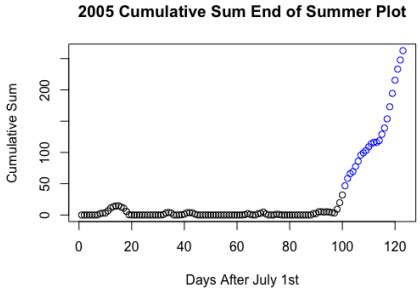

2000 Cumulative Sum End of Summer Plot
Cumulative Sum
0
100
300
500
0
20
40
60
80
100
120
Days After July 1st

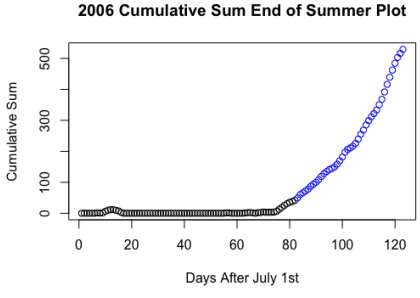

2001 Cumulative Sum End of Summer Plot
Cumulative Sum
0
100
200
300
400
0
20
40
60
80
100
120
Days After July 1st

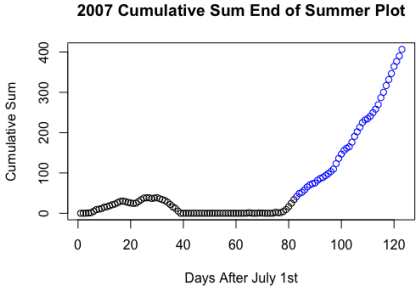

2002 Cumulative Sum End of Summer Plot
Cumulative Sum
0
100
200
300
400
0
20
40
60
80
100
120
Days After July 1st

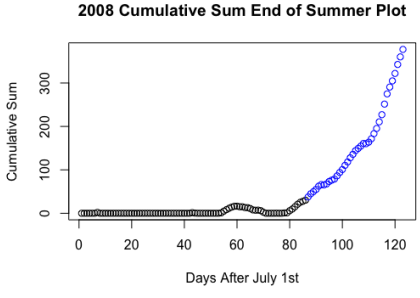

2003 Cumulative Sum End of Summer Plot
Cumulative Sum
0
50
150
250
350
0
20
40
60
80
100
120
Days After July 1st

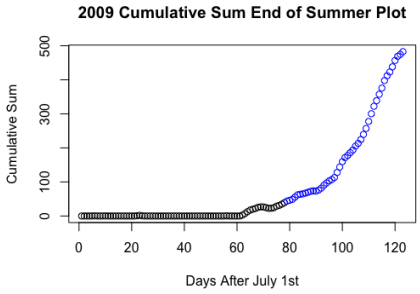

2004 Cumulative Sum End of Summer Plot
Cumulative Sum
0
100
200
300
0
20
40
60
80
100
120
Days After July 1st

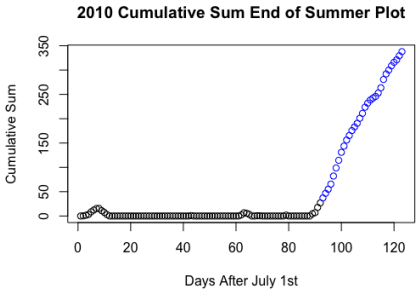

2005 Cumulative Sum End of Summer Plot
Cumulative Sum
0
50
100
200
0
20
40
60
80
100
120
Days After July 1st

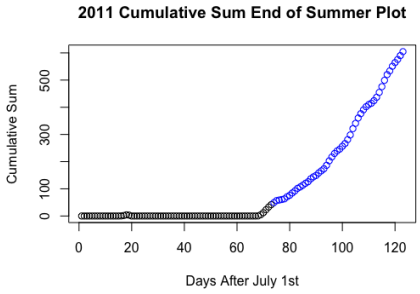

2006 Cumulative Sum End of Summer Plot
Cumulative Sum
0
100
300
500
0
20
40
60
80
100
120
Days After July 1st

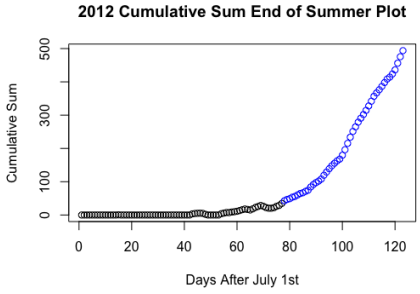

2007 Cumulative Sum End of Summer Plot
Cumulative Sum
0
100
200
300
400
0
20
40
60
80
100
120
Days After July 1st

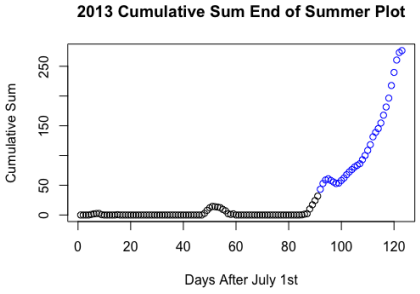

2008 Cumulative Sum End of Summer Plot
Cumulative Sum
0
100
200
300
0
20
40
60
80
100
120
Days After July 1st

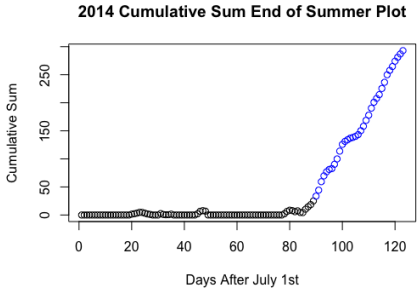


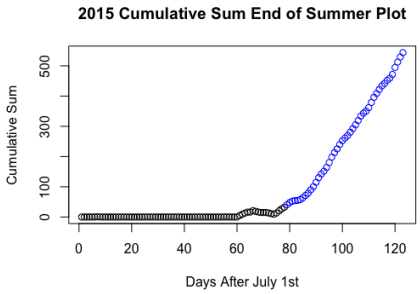




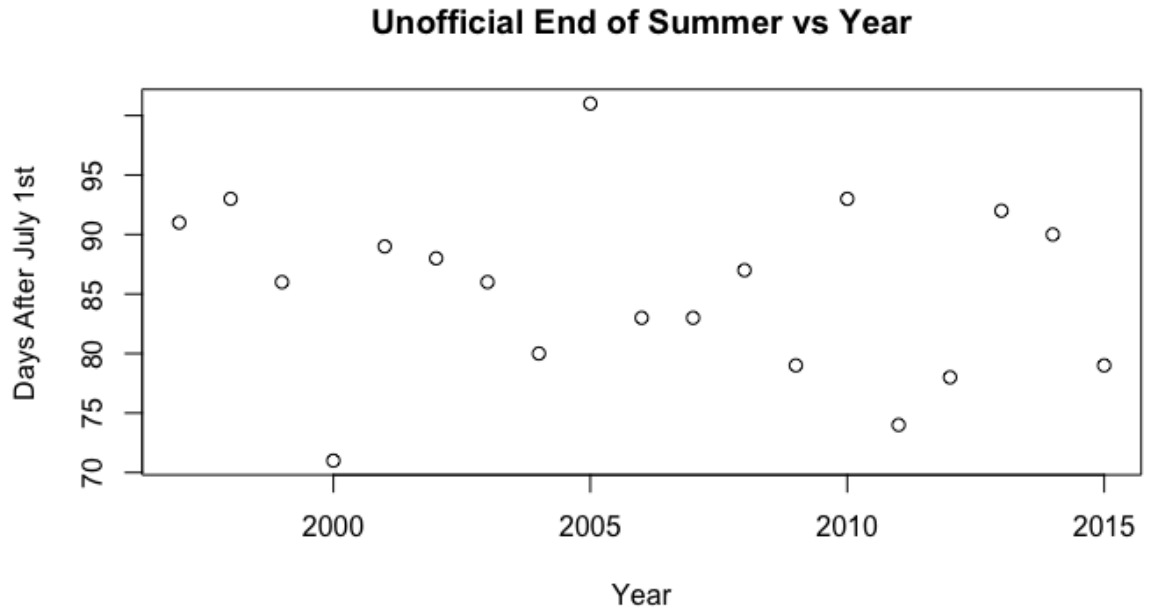








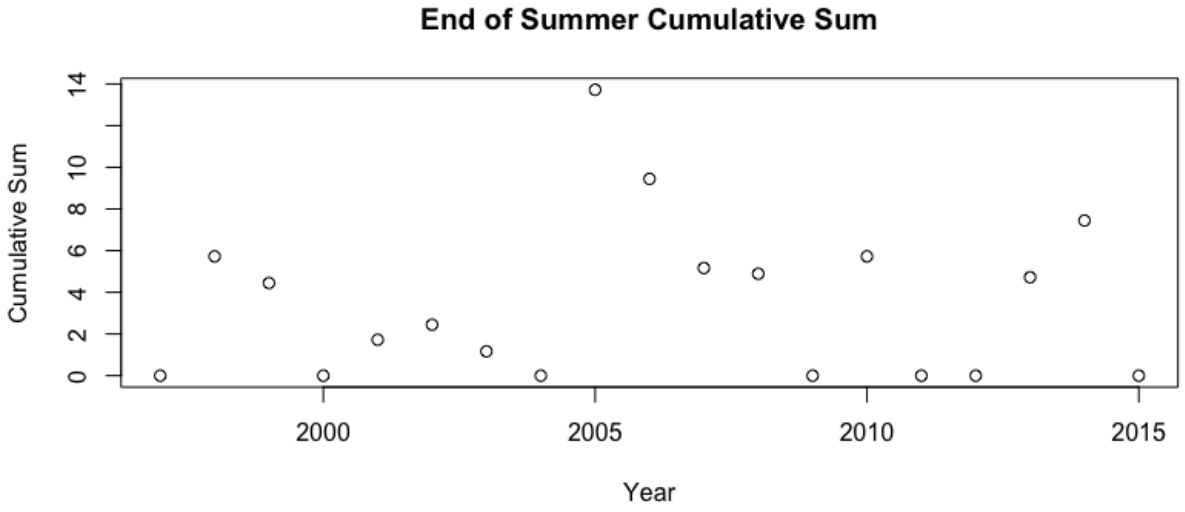
In the figures above the blue points represent the points at which the cumulative sum has surpassed the threshold *T*, that is, when summer has ended. As may be observed in the above figures, there does not appear to be a clear pattern which indicates when summer ends. It may be noted that summer tends to end between approximately day 70 and day 90 in all of the figures above. The standard deviation of the unofficial end of summer is equal to approximately 7.4 days. The unofficial summer end dates for each year may be observed more clearly in the following figure:



As it may be noted by observing the above figure, it does not appear as though there is either an increasing or decreasing trend in the unofficial end of summer. Instead, it appears as though the end of summer seems to randomly occur earlier or later from year to year. To test this definitely though, I applied the *cusum* algorithm to the above end of summer data in order to attempt to detect an increase in the yearly end of summer result data, which would indicate that summer is ending later each year on average. The *cusum* formula for detecting an increase is defined as follows:

$$S_t = max\{0, S_{t-1} + (x_t - \mu - c)\}$$

In my *cusum* analysis of the days defined as the unofficial end of summer for each year, I let μ = the mean of the unofficial last day of summer for each year, since there was no other obvious choice of value for μ, and no indication that the summer was ending later based upon the above figure. I let $c = 0.25$ * the standard deviation of the mean summer temperatures. For the critical value c I used a coefficient of *0.25* multiplied with the standard devation, rather than *0.5*, so that the algorithm would be more sensitive to detecting change, since it doesn't appear as though an obvious trend exists. In order to determine, based upon the result of the *cusum* function, whether the summer was ending later, I again included a threshold parameter T. I let $T = 2.5$ * the standard deviation the mean summer temperatures. For the threshold value T I used a coefficient of *2.5* multiplied with the standard deviation, rather *10* as done above, again so that the algorithm would be more sensitive to detecting change. The results of this analysis may be observed in the figure below:



Conclusion

The above plot was designed to mark points which exceed the threshold value T in blue, as done in the other *cusum* plots above. As may be inferred by the lack of blue points in the plot, the parameters I used in my *cusum* analysis did not lead to the detection of an increase. That is, the cusum model did not determine that summer was ending later each year. Given the high sensitivity of the parameters that I used to detect changes in the above *cusum* analysis, I feel confident in saying, based upon my *cumulative sum* analysis of the exponential smoothing model that I created with the *HoltWinters* function, there is no definitive evidence to indicate that summer is ending later each year on average. Rather, the summer simply appears to end earlier or later each year in a seemingly random manner.