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 natural gas OEM and we collect a lot of data on the machinery we sell. This data is used for many purposes, such as analytics or troubleshooting. Since all of the data comes from sensors, there is a lot of variation in the data, which is mostly real, but sometimes not. Another problem is that this data is collected 24 hours a day 365 days a year and takes up a lot of storage space. I would think that exponential smoothing in conjunction with CUSUM could be a very good tool to use for data compression. In order to keep the file sizes low, data compression is used so that the value in the historian storage doesn't change unless the reading from the sensor changes over a certain amount, which is called the dead band. Instead of a simple dead band, exponential smoothing and CUSUM could be used to continually monitor the changes in the sensor readings potentially allowing the historian storage device to react faster during events, so that critical data recorded, and to ensure that when no issues are happening that less important data is not taking up expensive storage space.

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.)
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).

I used the HoltWinters function, with seasonality, to apply the exponential smoothing to the data. The output of the function is shown in red while the original data is shown in black in Figure 1 below.

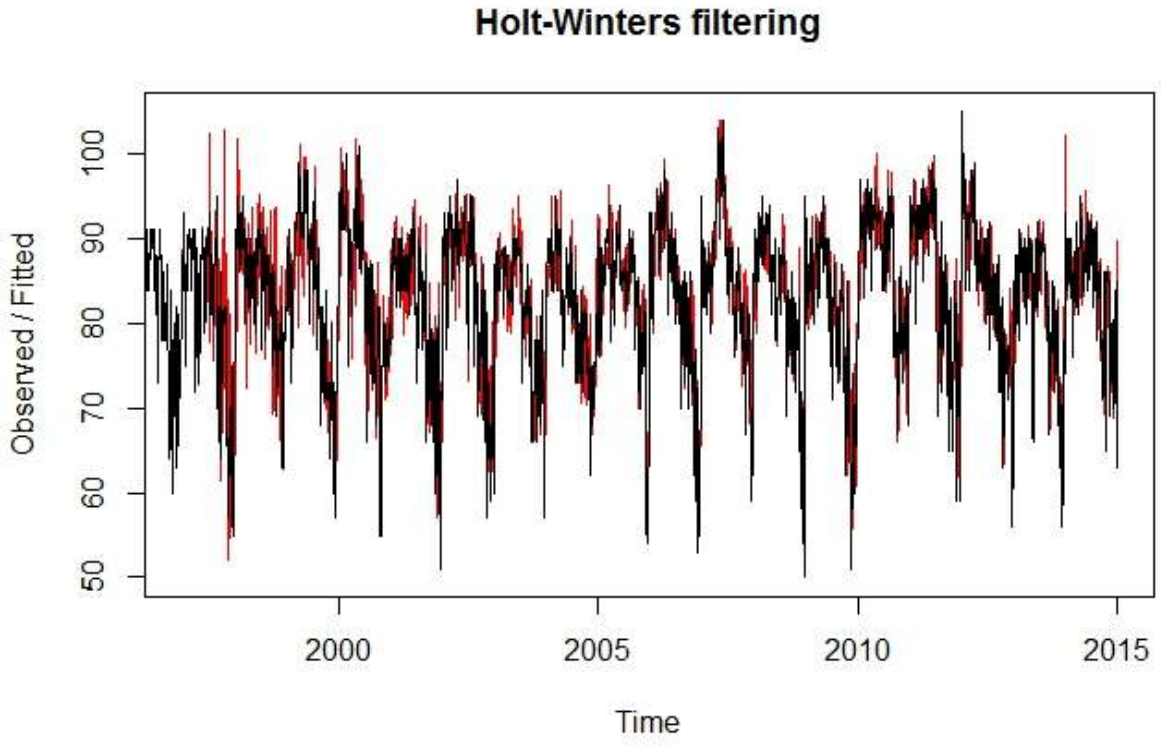

Figure 1. Holt-Winters Plot

Overall, we see a good fit on the data especially on later cycles. This is because as more data points are used to build the trends and seasonality components the model gets better at predicting the future results. It should be noted that the model doesn't start until 1997, after it was able to process two cycles of data.

I then exported the models expected values, or the fitted values, to an Excel file and applied CUSUM to each year of the data. Using a C and T of 6 and 41, this removes all false flags earlier in the

cycles and a very slight trend of summer ending later can be seen. As shown in Figure 2 below, it the summer end shifts from 29 SEP to 5 OCT. However, this is less than a 1 week difference and is probably well within the error of our calculations. This trend could also be due to fact the model gets better calibrated with every cycle, as mentioned above and shown in Figure 1. Therefore, it id plausible that no trend would be seen over this same time period if we were able to incorporate data from years prior to 1995.

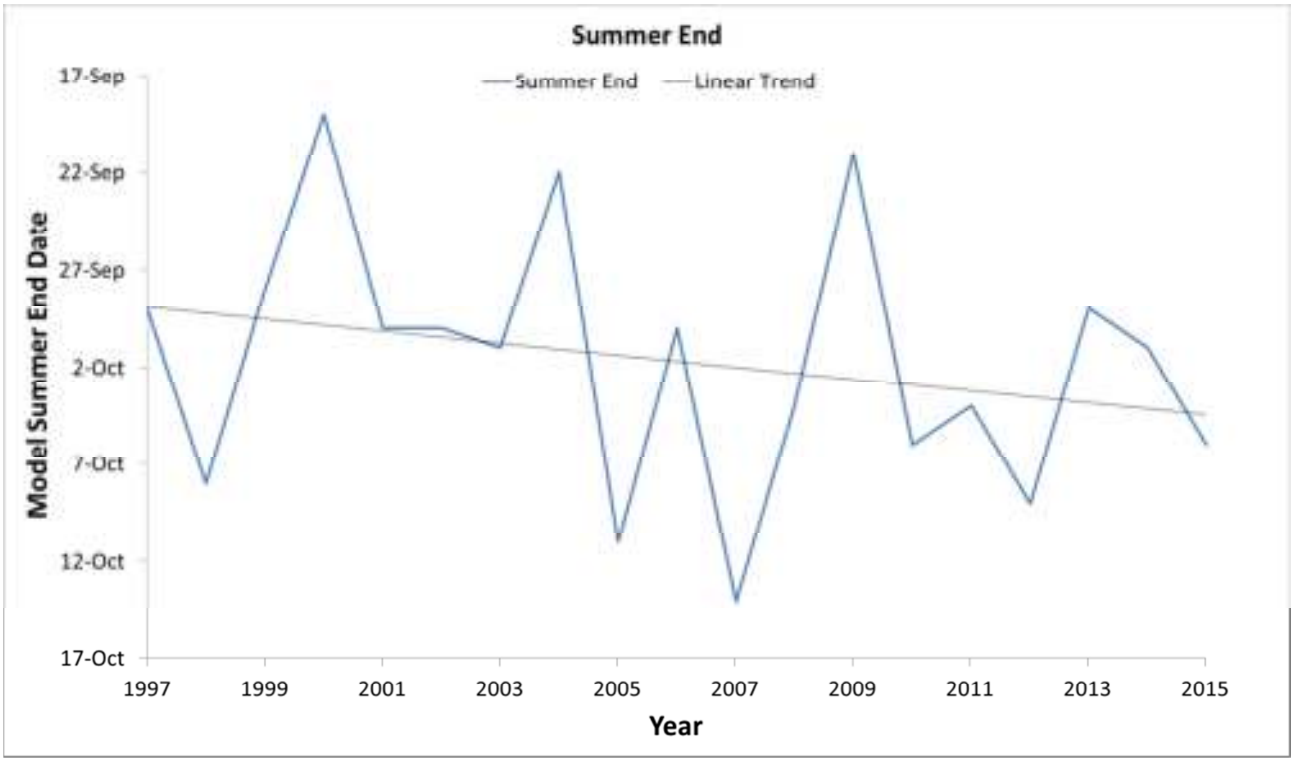

Figure 2. Holt-Winters Plot