**Assignment #7**

**ETS Laboratory**

Due: Wed. Oct 25

(40 pts. Total)

The data set RSGCSN.csv includes monthly retail sales of grocery stores in the US from January 1992 through August 2017 expressed in millions of US dollars.

Source: <https://fred.stlouisfed.org/series/RSGCSN>

Issue the following commands to load the data and convert it into an appropriate time series object:

library(fpp)

library(dplyr)

#  
# Read csv file and make it a time series  
RS <- read.csv("RSGCSN.csv") %>%  
 select(-DATE) %>%  
 ts(start= c(1992,1), frequency=12)

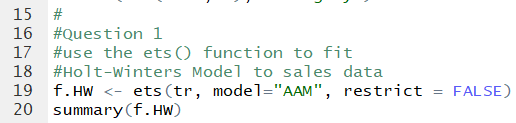
In this assignment we will shift out focus from short-term demand forecasting (on weekly intervals) as it is necessary for the management of distribution logistics, and we focus on longer term forecast as it is appropriate for aggregate planning and/or facilities planning.

In this assignment we are interested in obtaining a 5years+ forecast (68 months to be precise) of the size of the grocery store market in the US, and we want that forecast in monthly (not weekly) intervals. Such a forecast is useful if you are preparing an infrastructure plan or a grocery store chain for example. This forecast is useful to make decisions about number of new stores to open, number of distribution centers and their capacity, personnel and other infrastructure decisions.

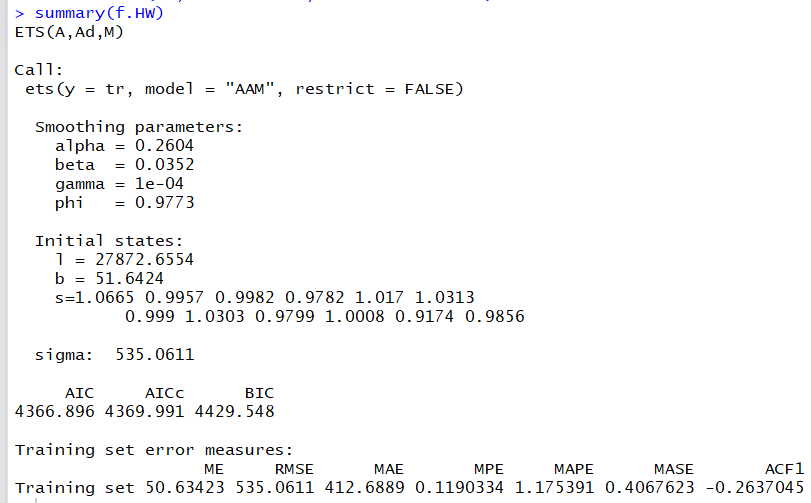
In this assignment we will learn to use the ets(…) function to fit and analyze exponential smoothing models. Before proceeding to fit a model, we examine and divide the data into two sets; a training set tr to fit the models and a testing (or hold-out) data set te to assess the out-of-sample performance of the models. This is accomplished with the following code:

tr <- **window**(RS, end=**c**(2011,12))  
te <- **window**(RS, start=**c**(2012,1))  
  
**plot**(RS)  
**abline**(v=**c**(2011,12), col="grey")

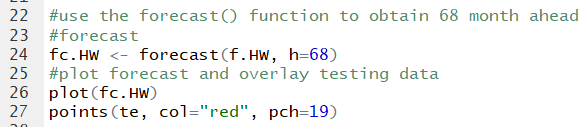
1. (5 pts.) Holt-Winters Model Analysis: part I:
   * Use the ets(…) function to fit a Holt-Winters exponential smoothing model to the sales data. Leave up to the ets(…) function to decide if a damping parameter is necessary (i.e., do not specify the damped directive. Call this model f.HW, and report the model details including the optimized value of each of the constants and smoothing parameters required by the model, the , and values, as well as the in-sample fitting indicators.



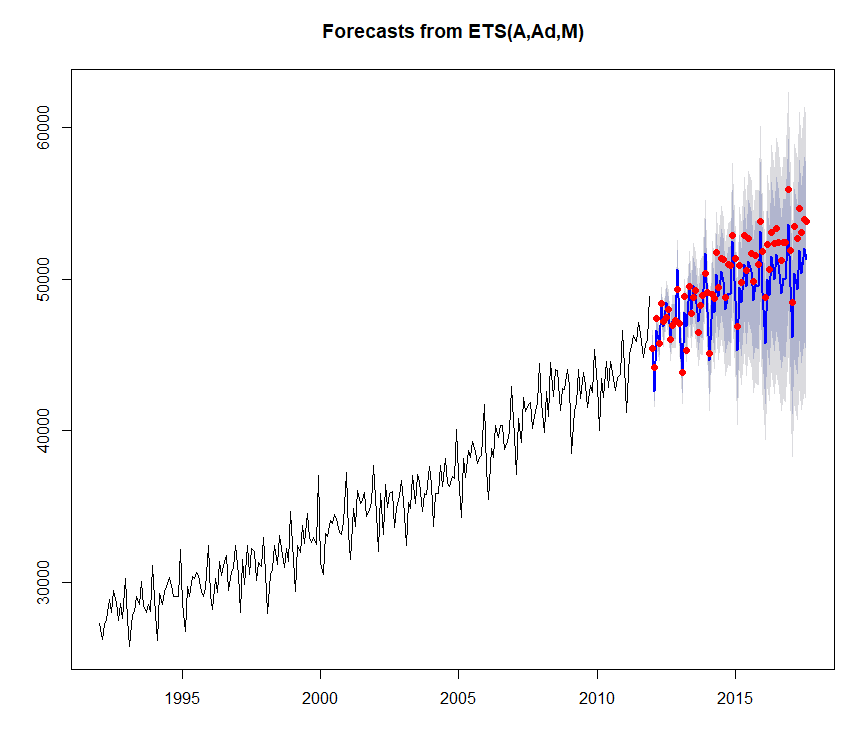
I called the ets() function for Holt-Winters by specifying “tr” for the training data and “AAM” for the model to have additive error, additive trend, and multiplicative seasonality. I did not specify a damping parameter, but as you can see in the summary() call, ets has added a damping factor automatically to the additive trend. The optimized values of each of the constants can be found under “Initial States”, and the Smoothing Parameters under “Smoothing Parameters.” The AIC, AICc, BIC and in-sample fitting indicators are found at the bottom of the summary.



* + Use the forecast(…) function to obtain a 68-month-ahead forecast, name this forecast fc.HW and plot it (i.e. call the plot(fc.HW) function); overlay on this plot the actual sales observed during this testing period (i.e. call the function points(te, col=”red”, pch=19) to overlay the testing data).



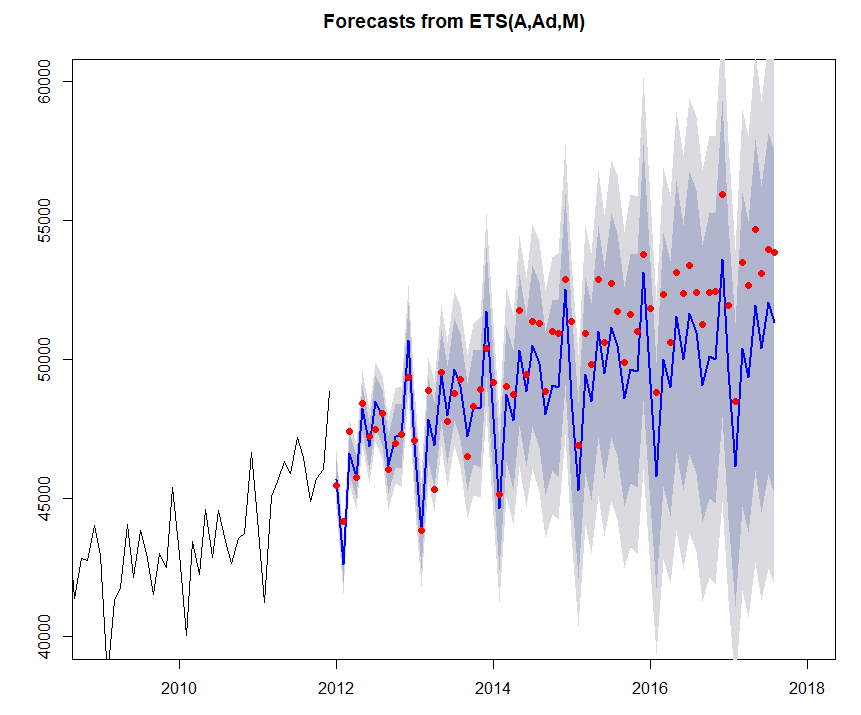
I used the fit model I just created to make a 68 month ahead forecast into the testing range.



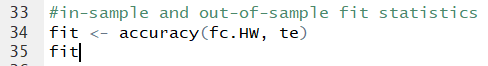
* + Reproduce the plot again, but now zooming on the forecasting period. To do this, include the xlim and ylim parameters in the initial plot call (i.e., use plot(fc.HW, xlim=c(2009,2018), ylim=c(40000,60000) ) to focus the plot on the forecast period). No change is necessary for the points(…) call. For the rest of this assignment **include the above values for the xlim and ylim parameters in every forecast plot in Questions 1 through 7.**



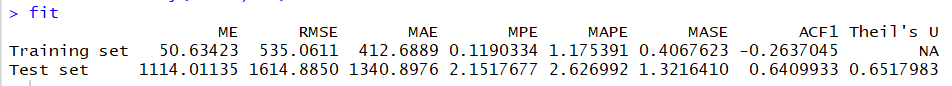
Produces zoomed-in plot.



* + Calculate the in-sample and out-of-sample fit statistics. You can obtain the in-sample and out-of-sample fit metrics comparison by calling the function accuracy(fc.HW, te)



Produce and view the in-sample and out-of-sample fit metrics.

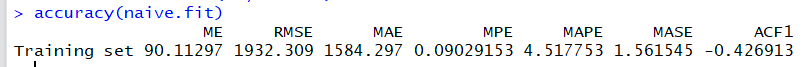


* + Based on your analysis above, discuss the forecast bias and compare the out-of-sample MAE with the MAE that you would obtain if you used the naïve forecasting method. What do you think is driving the poor model performance? Which model/method would you choose for forecasting?

The forecast fits the training set much better than the testing set. This may be a result of overfitting and capturing white noise in the training set, leading to inaccurate forecasts of the testing set. The Mean Absolute Error was 412.6889 with the in-sample data, but was 1340.8976 with the out-of-sample data. Because this measure is absolute, positive and negative errors do not cancel out, so we can say the MAE increased over three times from the in-sample to the out-of-sample data.

I created a naïve fit to compare to my Holt-Winters Model:





Out-of-sample MAE: 1340.8976

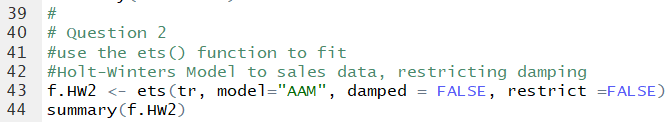
Naïve MAE: 1584.297

The out-of-sample MAE is better than the naïve forecast’s MAE. The Holt Winters model did a slightly better job of predicting future data points than if we had just referred to the previous point as the forecast.

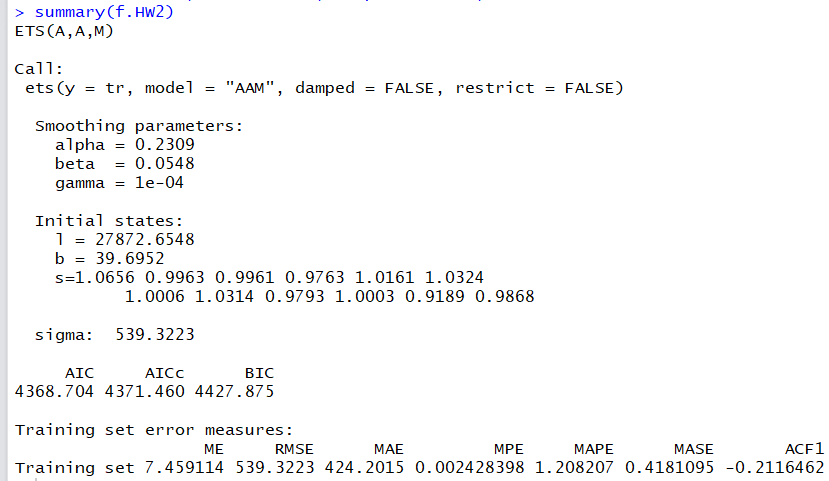
However, the Holt-Winters with Additive damping still has high error numbers. I think that the white noise in the data could be driving poor model performance, as the forecast is picking up the wiggling of the data. I also think that the Additive damping may not be the best way to explain the trend, as the trend looks just additive from the graph to me.

Between Holt-Winters with Damping and naïve, I would pick the Holt-Winters with Additive Damping as it has lower error numbers and thus fits the data better than the naïve forecast.

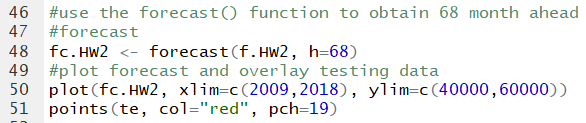
1. (5 pts.) Holt-Winters Model Analysis: part II:
   * Optimize the parameters of a Holt-Winters model disallowing damping of growth (i.e., use the damping=FALSE directive in the call to the ets(…) function). Call the fitted model f.HW2, and report the model details including the optimized value of each of the constants and smoothing parameters required by the model, the , and values, as well as the in-sample fitting indicators.

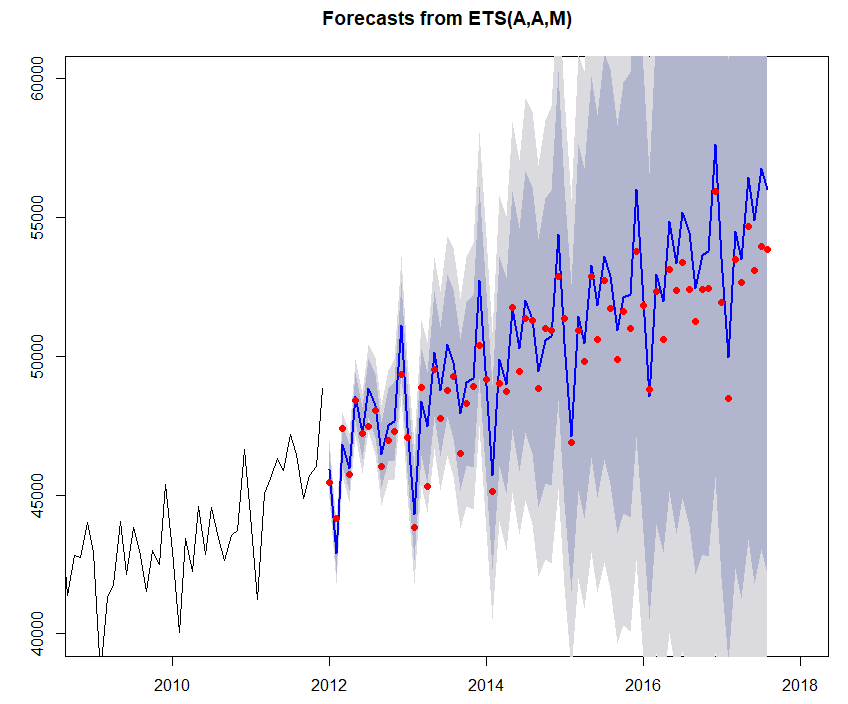


I called the ets() function for Holt-Winters by specifying “tr” for the training data and “AAM” for the model to have additive error, additive trend, and multiplicative seasonality. This time, I disallowed damping in the model. The optimized values of each of the constants can be found under “Initial States”, and the Smoothing Parameters under “Smoothing Parameters.” The AIC, AICc, BIC and in-sample fitting indicators are found at the bottom of the summary.



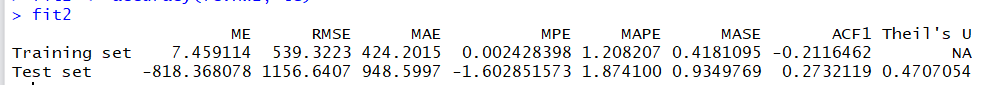
* + Now use the forecast(…) function to obtain a 68-month-ahead forecast, name this forecast fc.HW2 and plot it.





* + Use the function accuracy(…) to calculate the in-sample and out-of-sample fit statistics of the fc.HW2 forecast.





* + As in Question 1, based on your analysis above, discuss the forecast bias and compare the out-of-sample MAE of fc.HW, fc.HW2 and the naïve forecast? Discuss also the confidence interval cone of both models. What do you suspect is making the cone of fc.HW2 much larger? Which model/method would you choose for forecasting?

Like the naïve and fc.HW forecasts, the test set error was much higher than the in-sample error. The mean error for the out-of-sample was very negative, which means the forecast predicted numbers much higher than it should have, and you can see that the cone is higher on the upper half of the graph. However, the out-of-sample MAE for fc.HW2 was much lower than the MAEs for the previous 2 models. The points were predicted closer to the actuals than the previous 2 models.

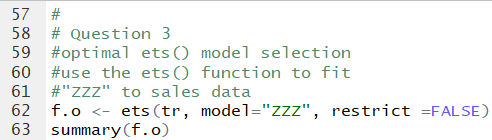
fc.HW2 - HW-not damped Out of Sample MAE: 948.5997

fc.HW -Damped Out-of-sample MAE: 1340.8976

Naïve MAE: 1584.297

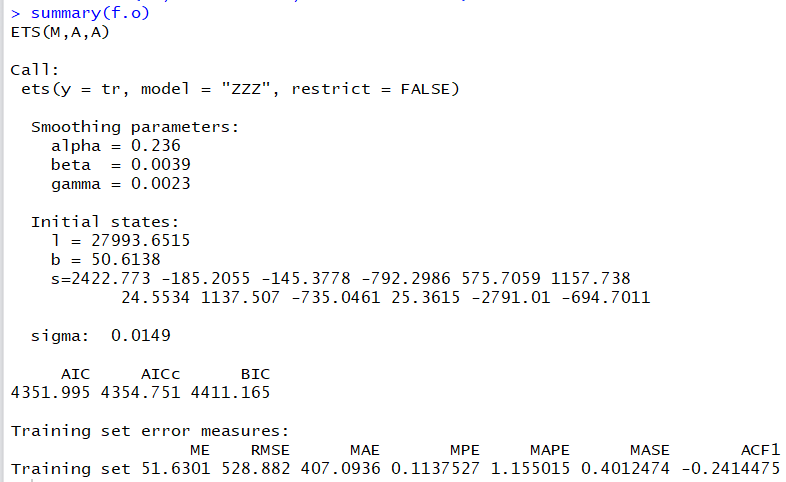
However, the confidence interval cone for fc.HW2 is much larger than the confidence interval cone for fc.HW. I suspect that, by disallowing damping of trend, the model did not fit the data as well, creating a larger critical value and margin of error for mistakes and thus a wider confidence interval. Even though the fc.HW2 model has a larger confidence interval, its forecast resulted in a lower MAE, so I would choose the fc.HW2 model for forecasting.

1. (5 pts) Optimal ETS Model Selection:
   * Now we call the ets(…) function using the model=”ZZZ” directive to optimize the model selection including multiplicative models (i.e., set the restrict=FALSE option). Call the fitted model f.O, and report the model details, the , and values, as well as the in-sample fitting indicators.

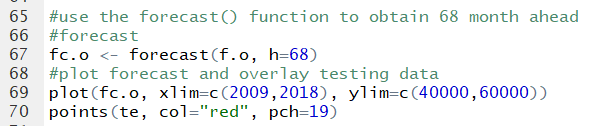


I create the model using the optimal automatic “ZZZ” directive. The summary function will show the model details and fitting indicators.

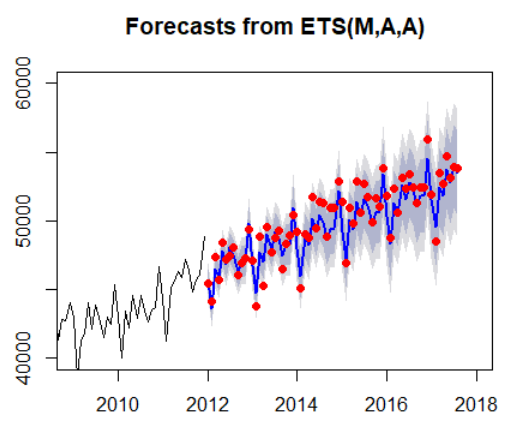
It created a model with Multiplicative Error, Additive Trend, and Additive Seasonality, and no damping.



* + Now use the forecast(…) function to obtain a 68-month-ahead forecast, name this forecast fc.O and plot it.

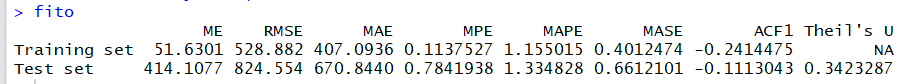


Plotted the forecast using the same parameters and points as the previous models.



* + Use the function accuracy(…) to calculate the in-sample and out-of-sample fit statistics of the fc.O forecast.





* + Compare the out-of-sample MAE of fc.HW, fc.HW2, fc.O and the naïve forecast? Compare the and of models f.HW, f.HW2 and f.O. Which model/method would you choose for forecasting?

fc.o – optimized out-of-sample MAE: 670.8440

fc.HW2 - HW-not damped Out of Sample MAE: 948.5997

fc.HW -Damped Out-of-sample MAE: 1340.8976

Naïve MAE: 1584.297

The out-of-sample MAE decreased from the naïve to fc.HW to fc.HW2 and to fc.o. From looking at the Mean Absolute Error, each model gets better at forecasting the testing data.

AICc

fc.HW 4369.991

fc.HW2 4371.460

fc.o 4354.751

When comparing the AICc of fc.HW and fc.HW2, the AICc goes up slightly when we remove damping. This may be a sign of fc.HW2 overfitting the training data more than the other 2 models. The AICc of fc.o is the smallest, which means that it fits the data better than the other two.

BIC

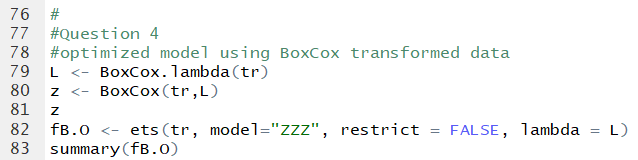
Fc.HW 4429.548

fc.HW2 4427.875

fc.o 4411.165

The fc.o model also has the smallest BIC value, meaning that it is a better fit for the data than the first two models. Based on the out-of-sample MAE, the AICc, and the BIC values, I would choose the fc.o model for forecasting.

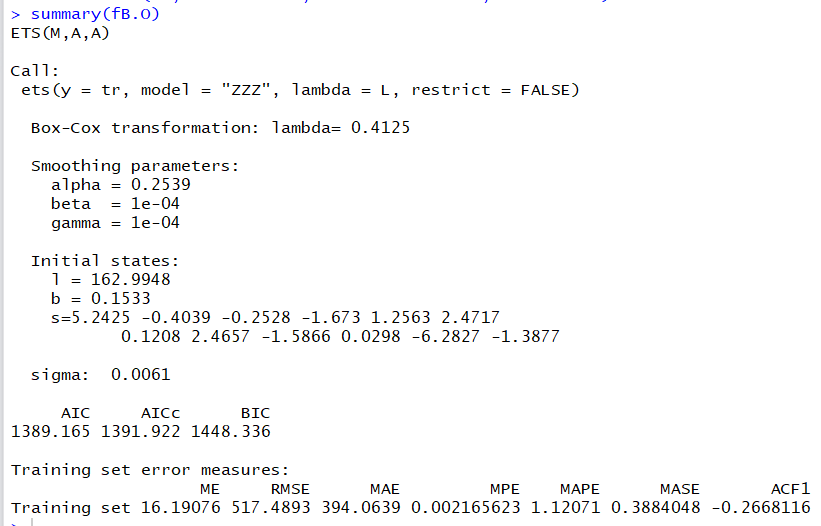
1. (5 pts) Optimized model using BoxCox-Transformed Data:
   * Select the best value of the “lambda” parameter for the BoxCox transformation over the training set tr, and then use the ets(…) function to optimize the model selection as you did in Question 3. Call the fitted model fB.O, and report the model details, the , and values, as well as the in-sample fitting indicators.



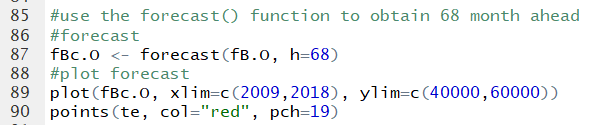
I selected the best value of the lambda parameter by calling the BoxCox.lambda function. I then called ets() over the training data, with the ZZZ model selection as I did in question 3. I added the lambda parameter I found in the first line of code.

The summary function reports the model details and fitting indicators.

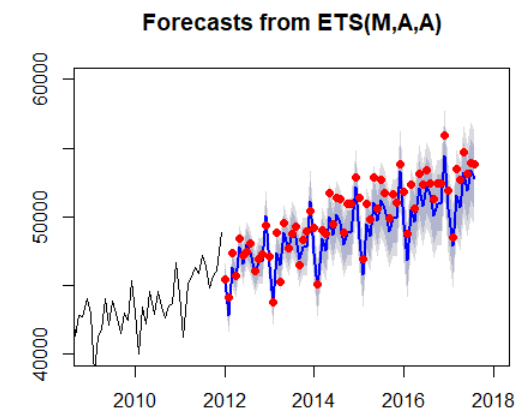
The BoxCox optimization resulted in a model with multiplicative error, additive trend, and additive seasonality, with no damping.



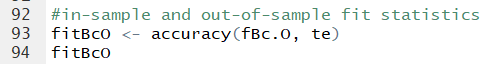
* + Now use the forecast(…) function to obtain a 68-month-ahead forecast, name this forecast fBc.O and plot it.

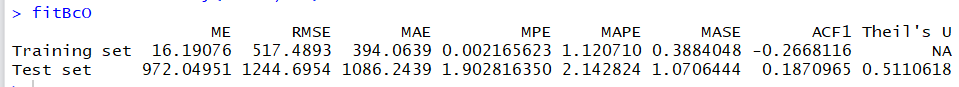


Create and plot the forecast with the same window parameters and points like in the previous 3 questions.



* + Use the function accuracy(…) to calculate the in-sample and out-of-sample fit statistics of the fBc.O forecast.





* + Compare the in-sample and out-of-sample MAE of fBc.O, fc.O and the naïve forecast? Which model/method would you choose for forecasting?

fBc.O – BoxCox optimized out-of-sample MAE: 1086.2439

fc.o – optimized out-of-sample MAE: 670.8440

Naïve MAE: 1584.297

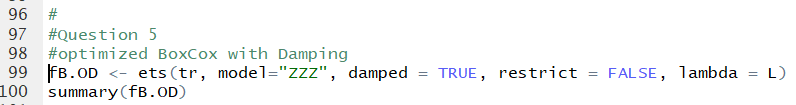
fBc.O – BoxCox optimized in-sample MAE: 394.0639

fc.o – optimized in-sample MAE: 407.0936

Naïve MAE: 1584.297

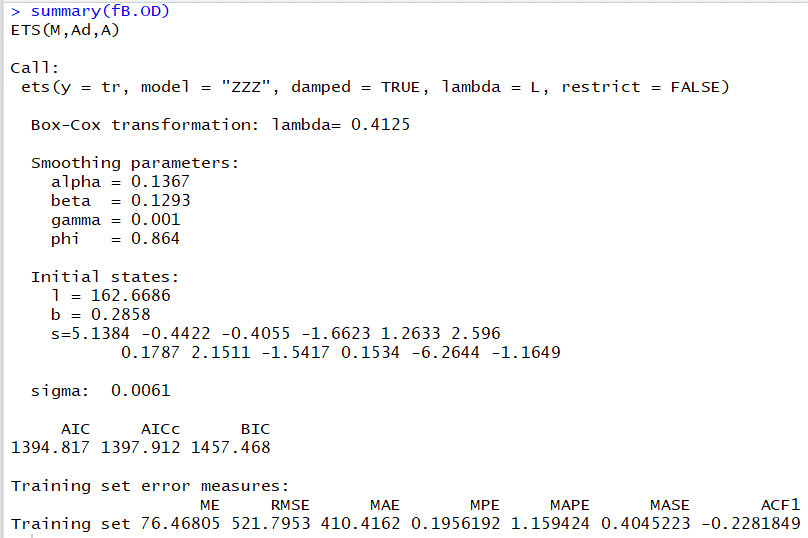
fBC.O has a lower in-sample MAE than fc.O, but a higher out-of-sample MAE. This means that fBC.O has better forecasts for the in-sample data, but this may have been a result of overfitting and picking up white noise. Because fBC.O picked up white noise with the training set, it did not do a good job of fitting the testing data and had a much larger out-of-sample MAE than the fc.O fit. For this reason, I would pick the fc.O forecast because it seems to not overfit the training data.

1. (5 pts) Optimized model with damping using BoxCox-Transformed Data:
   * Using the best value of “lambda” (i.e., the same you used in Question 4), and set damped=TRUE in the ets(…) function. Name the fitted model fB.OD and report the model details and metrics.

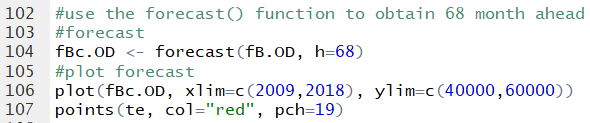


Created the model using the same Lambda = L and the ZZZ optimum model, but now the model includes damping.

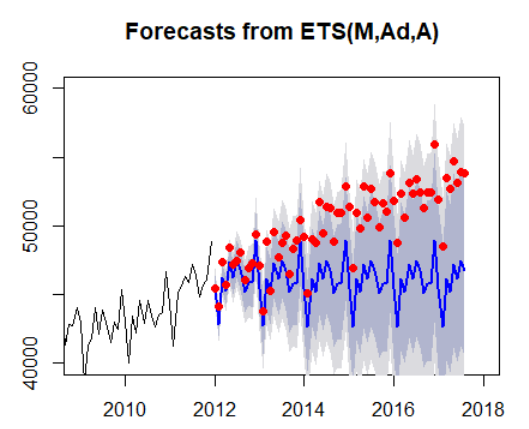
This created a model with Multiplicative Error, Additive damped trend, and Additive Seasonality.



* + Now use the forecast(…) function to obtain a 68-month-ahead forecast, name this forecast fBc.OD and plot it.



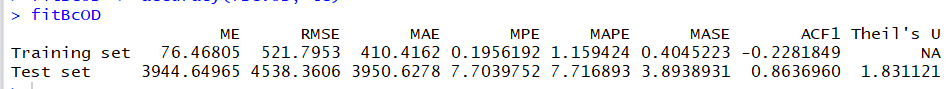
Box-Cox with additive damping model.



The forecasts and the actual data is WAY OFF when we added additive damping.

* + Use the function accuracy(…) to calculate the in-sample and out-of-sample fit statistics of the fBc.OD forecast.





* + Compare the in-sample and out-of-sample MAE of fBc.OD, fBc.O, fc.O and the naïve forecast? Which model/method would you choose for forecasting?

fBC.OD – transformed optimum additive damping out-of-sample MAE: 3950.6278

fBc.O – transformed optimum no damping out-of-sample MAE: 1086.2439

fc.o – optimized out-of-sample MAE: 670.8440

Naïve MAE: 1584.297

fBc.OD – transformed optimum additive damping in-sample MAE: 410.4162

fBc.O – transformed optimum no damping in-sample MAE: 394.0639

fc.o – optimized in-sample MAE: 407.0936

Naïve MAE: 1584.297

When we optimized the fit using BoxCox transformations and “ZZZ,” the forecast gave us the lowest in-sample MAE of 394.0639. When we added in damping during fBC.OD, the damping made the model worse for in-sample MAE. The damping did not fit the training data correctly.

However, the issues with the BoxCox models appear when examining the out-of-sample testing data MAE. When the BoxCox function decided the optimum value of lambda, it was looking at the training data set only. The BoxCox transformation focused only on the trends in the training data, and so it transformed the data to fit a forecast. Because our out-of-sample MAE went up so much with the BoxCox models, it can be assumed that there must be a trend difference between the training and testing data, some kind of shift in the optimum forecast fit.

Because fc.O has the lowest out-of-sample MAE and the BoxCox transformations have much higher out-of-sample MAE values, I would choose the fc.O forecast to predict data. The BoxCox data transformations do not fit the testing data, so maybe some earlier data is not relevant to the newer testing data trends.

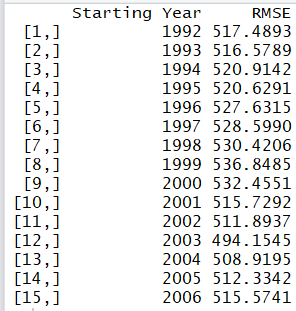
1. (5 pts) In an effort to improve forecasts, this question we want to assess the value of old information and discard the oldest segment of the information that does not have predictive value. To this end we execute the following code:

#  
# Discarding Old Data  
#  
FS <- NULL  
for(sy in (1992:2006)){  
 td <- window(RS,start=c(sy,1), end=c(2011,12))  
 L <- BoxCox.lambda(td)  
 fBC <- ets(td, model="ZZZ", restrict=FALSE, lambda=L)  
 RMSE <- accuracy(fBC)[2]  
 FS <- rbind(FS,c(sy,RMSE))  
}  
colnames(FS) <- c("Starting Year","RMSE")

FS

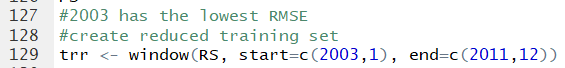
This code loops the evaluation of a moving training set starting from 1992, 1993, etc all the way to starting in 2006. For each starting year, a training data set is formed and called td, then a value of “lambda” is obtained for each training set, an optimized model is obtained, and the in-sample RMSE is extracted. Finally, when the loop terminates its execution a table of the starting year and RMSE values is printed.

* + What starting year results in the lowest RMSE?



2003 – has a RMSE of 494.1545

* + Use the window(…) function to create a “reduced” training set starting the year you identified above, and terminating in December of 2011. Name this reduced training set trr.

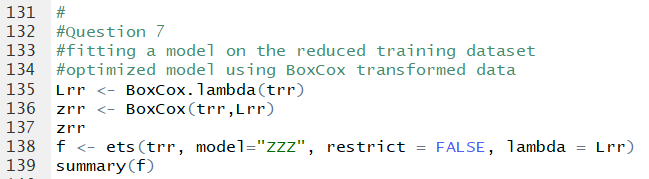


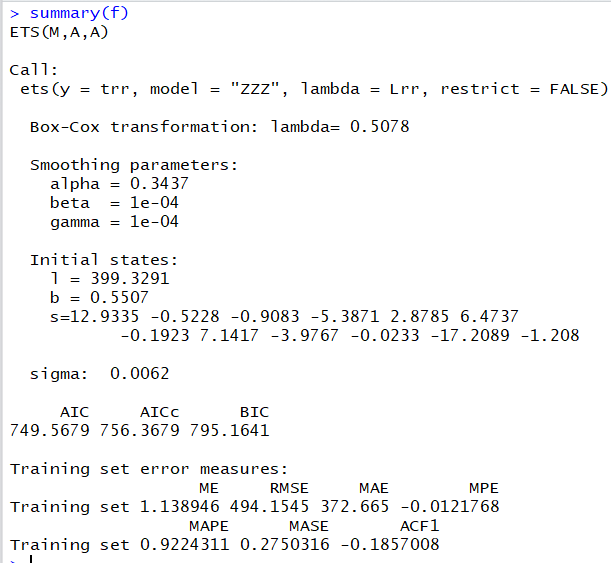
* + Explain why we cannot use the or criteria to select the best starting year for the training data set.

Different data sets – can’t compare AICc values

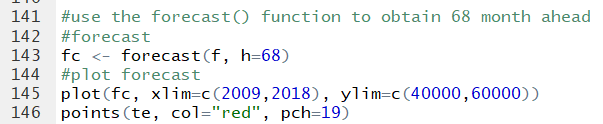
When we iterate through the for loop, a different training data set is created each time. Because a different data set is created each time, and the fit has different lambda values, the ets function is creating a summary of a completely different data set each time. This means that the AICc and BIC values are NOT comparable. As seen in earlier questions, the AICc ranged from around 3400 to around 1900, even when the 3400 AICc had lower in-sample error. We can only compare AICc and BIC values if the training data set was the same. Because of the different training sets, we cannot use the AICc or BIC criteria and instead we used the RMSE to select the best starting year for the training data set.

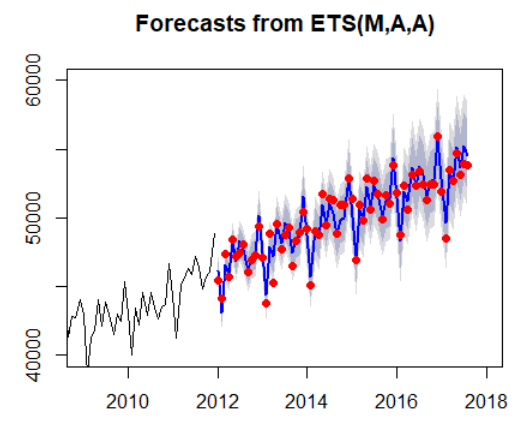
1. (5 pts) Fitting a model on the reduced training dataset:
   * Figure out the best value of the BoxCox lambda value for the reduced training data set trr, and fit the best ETS model to this data. Report the model parameters and metrics. Name this model f.



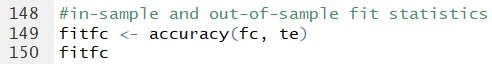


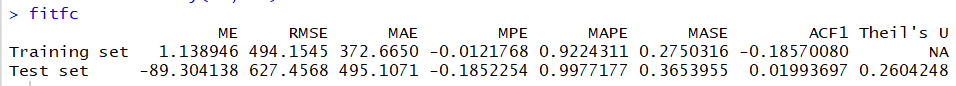
* + Use the forecast(…) function to obtain a 68-month-ahead forecast, name this forecast fc and plot it.





* + Use the function accuracy(…) to calculate the in-sample and out-of-sample fit statistics of the fc forecast.





* + Is the in-sample for model f.O comparable with the in-sample for model f? Explain.

In-sample AICc for model f.o: 4354.751

In-sample AICc for model f: 756.3679

NO – f.o uses all of the training set data, while f uses the reduced training set data. This means that their AICc values are not comparable because the fits are for different data sets.

* + Is the in-sample for model f.O comparable with the in-sample for model f? Explain.

In-sample MASE for model f.o: 0.4012474

In-sample MASE for model f: 0.2750316

YES – you can compare MASE with different data sets because the calculation scales the error. We can see that the model f has 0.27 MASE, while the f.o model has 0.40 MASE, meaning that the model f fits the in-sample data better

* + Is the out-of-sample for model fc.O comparable with the out-of-sample for model fc? Explain. Is the fc forecast truly an out-of-sample forecast? Explain.

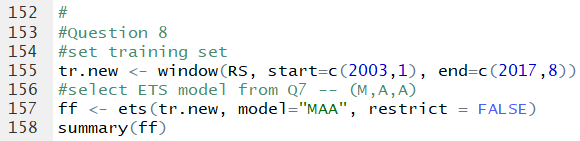
Out-of-sample RMSE for model fc.o: 824.554

Out-of-sample RMSE for model fc: 627.4568

YES – the out-of-sample RMSE for fc.O and fc are comparable. This is because the error is averaged and thus, the number of points for the forecast does not affect the RMSE.

The fc forecast is not truly an out-of-sample forecast because we have the actual data for those points. This is a pseudo out of sample analysis, as we have the actual data points from 2011 – 2017, but we are pretending they are the future points in order to estimate the predictive value of our model. In order to have a true out-of-sample forecast, you would have to not have actual data points yet. In Question 8, we will predict a true out-of-sample forecast.

1. (5 pts.) Aggregate Sales Forecast for 2017—2022:
   * Next we need to prepare a monthly sales forecast through December 2022. To this end we first set the training set to include all the data starting from the year we selected in Question 6 through August 2017. Select the ETS model you analyzed in Question 7, and fit the best parameters to that model. Name the resulting model ff.



Data goes from January 2003 to August 2017.

* + Compare the in-sample fit statistics of ff with those of model f.

In-sample fit of ff:

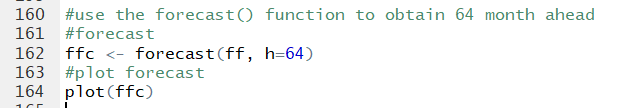


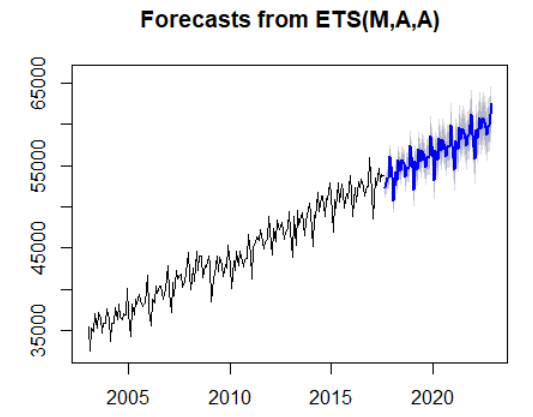
In-sample fit of f:



The in-sample fit statistics of ff are slightly worse than the in-sample fit statistics of f. This makes sense because we are forecasting further into the future with our ff fit, so there will be more in-sample errors. The RMSE of in-sample ff is 541.9784, while the RMSE of f was only 494.1545. As we had more future data, there was more error in ff.

* + Use the forecast(…) function to obtain a 64-month-ahead forecast, name this forecast ffc and plot it (this time do not include the xlim and ylim limits on the forecast plot.





* + Based on your analysis what would you expect the out-of-sample (i.e., the actual) be over the next five years? How about the out-of-sample (actual) ?

I decided to base my expectations by comparing the in-sample and out-of-sample errors of f to the in-sample errors of ff. A proportion of the in-sample to out-of-sample of f should estimate the proportion of the in-sample to out-of-sample of ff.

In-sample MAPE of f 0.9224311

Out-of-sample MAPE of f 0.9977177

In-sample MAPE of ff 0.9580372

Calculation of proportions: 0.9977177/0.9224311\*0.9580372

**Predicted out-of-sample MAPE of ff 1.03623**

In-sample RMSE of f 494.1545

Out-of-sample RMSE of f 627.4568

In-sample RMSE of ff 541.9784

Calculation of proportions: 627.4568/494.1545\*541.9784

**Predicted out-of-sample RMSE of ff 688.1816**