**Data Analytics - Individual Assignment 4**

1. Data Preparation

The goal is to develop a model that can forecast four quarters ahead and account for both trend and seasonality. In R, take the following steps:

* Read in the data.
* Create a time series object from the data.
* Designate the last four observations as the validation data, the rest as the training data.

Feel free to take advantage of the “getting started script” to execute these steps.

1. Data Visualization

Visualize the time series. What are the key components of the series? (recall the four key components from our lecture).

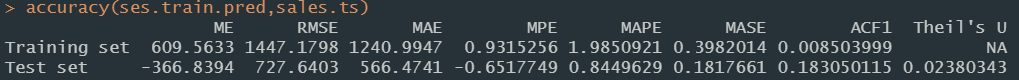
1. Data Driven Models – First Model

Since the data has strong seasonality and trend, we will run a Holt-Winter’s exponential smoothing model. Set α=0.2, β=0.15 and γ=0.05 and set the model specification to run multiplicative Holt-Winter’s method.

Run this method on the data.

* 1. Report the RMSE and MAPE on both the training and validation data

model="MAM", alpha = 0.2, beta = 0.15, gamma = 0.05



* 1. Include a graph that shows the original series, the prediction on the training data and the prediction on the validation data.

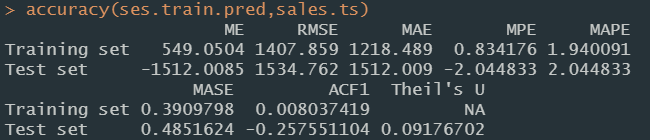
Chart, line chart, histogram

Description automatically generated

1. Data Driven Models – Model Improvement

Now update your model settings (both the type of seasonality and the parameters), and try to improve the prediction, as measured by the error measures on the validation data.

For your selected model:

* 1. Report your *“model”* and other parameter settings.
     1. model="MMM", alpha = 0.16, beta = 0.16, gamma = 0.01
  2. Report the RMSE and MAPE on both the training and validation data
     1. 
  3. Include a graph that shows the original series, the prediction on the training data and the prediction on the validation data.
     1. Chart, line chart, histogram

        Description automatically generated
  4. Analyze the results – how would you characterize the performance?

The changes added more slower learning from seasonality and value, slightly more frequent learning from trend. The performance sees improvement in better prediction overall.

1. Regression Based Models

Fit a regression model with a trend and seasonality. Then create a plot that shows the fit in the training period and validation period.

* 1. Include the plot in your assignment.

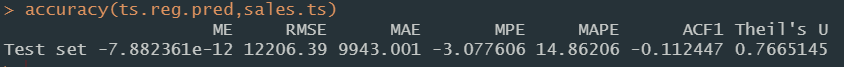
Chart, line chart

Description automatically generated

* 1. Comment on the fit.

The regression is a linear prediction of trend while the original data seems to project an exponential trend. The regression does not catch seasonality at all.

* 1. Report the RMSE and MAPE and compare it to the data driven approach.



The regression line’s prediction seems to have higher error across the board compared to data driven model since it fails to incorporate seasonality and value predictions.

* 1. What would be your next step if you were to try to improve the model (hint: plot the residuals, there should not be a pattern if you have fully captured ).

For trend and seasonality prediction, Improve the fitness of the regression by replacing linear model with exponential model, and introduce more variables in the model to improve accuracy.

1. Model Selection

Review your work.

* 1. Which model would you choose and why? (remember to use quantitative reasoning).

I would choose data driven model because of lower errors, the ability to customize weight on value, trend, and seasonality learning and ability to choose additive or multiplicative approach on time series.