# Airline Forecasting Method

Yovinda Pricila

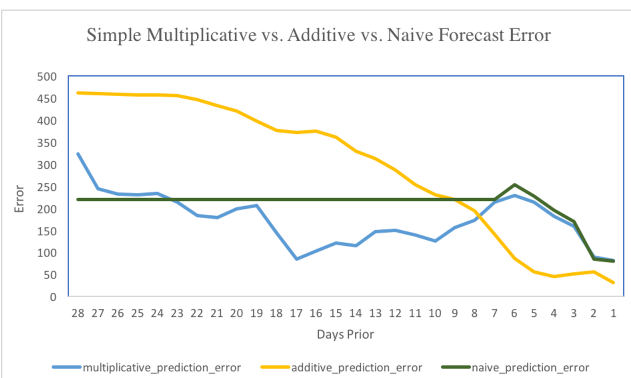
Before starting this project, our research shows:

> "The additive method is more appropriate when the seasonal factors tend to be steady from one year to the next. By contrast, multiplicative method is more widely used since much economic time series have a seasonal factor that grows proportionately with the level of the time series." (Andersen, Semmerloth).

Given this statement, we initially hypothesized that the additive forecast would be much suitable for airline forecasting, assuming that airplane price tickets are mostly driven by seasonality (e.g. airlines price hiked at peak season in mid-June to August, and goes down during off-season).

Therefore, we began our analysis by conducting each empirical approach individually and comparing the MASE and MAD of each model. Our additive approach is to sum the cumulative booking and average of remaining seats based on days-prior, whereas the multiplicative approach is to divide the cumulative booking by booking rate based on days-prior.

The result was far from our initial hypothesis: the Additive forecast method yielded 140.20% MASE, which is much higher than the multiplicative MASE of 83.89%. Proceeding with this result, we tried visualizing the error of each method on the airline Validation Data to see its error pattern and tendency. (refer to the graph below)




Analyzing the error trend, we could see that the additive forecast error trend decreases significantly going towards the departure date; it means that the prediction accuracy of additive forecast improves as days prior approaches zero. On the other hand, the multiplicative forecast has more volatile error. Still, as shown in the figure below, the additive model eventually will have a lower error as the days to departure approaches zero. Furthermore, the naive forecast has a more steady error level than either of the other forecasts, but also gets lowered as the days to departure approaches zero. By this result, we know that solely applying one method would not minimize the error; combining both models would help to further optimize the prediction accuracy.

From this point, we started to think of other factors that might affect the airline demands. We first applied and tested whether or not the MASE decrease if we apply the average remaining seats based on 1.) days-prior, 2.) departure DoW (day of week), and 3.) season (peak vs. shoulder vs. off-season). Logically, we would expect the average remaining seats on a Friday flight during *peak-season* to be much lower than a Tuesday flight during *shoulder-seaso*n. However, slicing the average remaining seats based on these three factors produced a lower MASE than merely applying average remaining seats based on days prior and departure DoW. For this reason, we decided to stick with the model with the lowest MASE: the modified additive dropped the MASE from 140.2% to 78.9%. The same approach, however, did not yield lower MASE value on the multiplicative method.

Therefore, we came up with an idea of combining both modified additive and simple multiplicative methods to find the lowest error. By means, we compare the final forecast of the modified additive and simple multiplicative method on the training data, and picked the final forecast which yielded the lowest error. Given this set of forecast, we utilized this number to forecast the final demand on the validation data; it successfully lowered MASE to 49%.

Although we were able to further reduce MASE, we also found a room of improvement that is yet to be applied on our model. Recognizing anomalies or demand outliers, such as national Holidays or Sports events, we would assume that the flight demand on this particular period to significantly higher. For example, from our given Airline training dataset (2012/05 - 2012/07), the potential error could have been resulted from the increasing travel demand during Fourth of July weekend. This could have been filtered out to further improve the accuracy of prediction, but were not implemented in our model.

This idea also begs the question of what other factors, if added, would reduce the accuracy of our models, such as seasonality, natural disasters, or other incidents that might affect the airline demand, yet were not accounted on our approach.

Citation:
Anderson, Alan, and David Semmerloth . "Time Series Analysis: Forecasting with Decomposition Methods." *Dummies*, Wiley , www.dummies.com/programming/big-data/data-science/time-series-analysis-forecasting-with-decomposition-methods/.