THE UNIVERSITY OF
CHICAGO

MSCA 31006 Time Series Analysis and Forecasting

Assignment #6 - Cross-Validation
Due Date – Beginning of Session #7
Total Points: 6%

Instructions:

- Total number of points is 60. The assignment's final grade will be multiplied by $1/10$ to calculate its weight on the final grade.
- Mark the question number and your final answer clearly (use a textbox.)
- Remember to show and explain your work (*If you can't explain it, you don't understand it.*)
- Please submit your solution through Canvas.

For this assignment, you will use the Monthly Australian short-term overseas visitors May 1985-April 2005 dataset (Dataset: visitors_monthly.rda)

 (5 points) Question 1:

Load and plot the visitors dataset and plot the dataset with and without the Box Cox transformation. Describe the main dataset characteristics.

(5 points) Question 2:

Build two models using the entire visitors dataset

a. Model 1: Let the auto.arima() function determine the best order $ARIMA(p,d,q)(P,Q,D)s$ model.
b. Model 2: Let the ets() function determine the best model for exponential smoothing.

(45 points) Question 3:

In this section you will apply the time-series cross validation method to train and test various models. Use the following values when training and testing the models:

- Set the minimum number of samples required to train the model to 160 (i.e., this is the minimum number of samples in the sliding window and the initial number of samples in the expanding window method.)
- Set the number the forecast horizon, h, to 1 year (i.e., 12 months.)
- Recall that the period, p, is equal to 12 months
- Use a single observation incrementation in each iteration (i.e., shift the training set forward by 1 observation.)

- Note: You are expected to have 80 iterations of cross validation

For each iteration, apply the following 4 forecasts:

1) Use the Arima() function to estimate a sARIMA([1,0,1][0,1,2]12 with drift model for:
   a. Expanding training window and
   b. Sliding training window
2) Use the Exponential Smoothing function ets() to estimate a MAM (Multiplicative Error, Additive trend, multiplicative Season) model for:
   a. Expanding training window
   b. Sliding training window

For each test window record the:

1) One-year forecast horizon error
2) Estimated model AICc value.

For each of the four models above, calculate and plot the

1) Mean Absolute Forecast Error (MAE) vs forecast horizon.
2) Root-square Forecast Error (RMSE) vs forecast horizon.
3) AICc vs iteration number

Discuss your results.

(5 points) Question 4:

What are the disadvantages of the above methods? What would be a better approach to estimate the models? Hint: How were the SARIMA and exponential time series models determined in question 3?