BA 875
Operations and Supply Chain Analytics
Homework Assignment 2 (Due 04/03)

Deliverable: Submit your homework assignment as a single Python Notebook only and clearly state your answer to each component of a question. Ensure to save Notebook with output shown. The delivery requirement has been stated here and in the HW submission portal. *You shall receive a 10 point reduction in grade if you do not submit in the format specified*.

Part I: Stationary Series
Dataset: jeans_data

Think back to the regression **in-class exercise** (NOT the HW assignment 1) using the jeans data for a medium-sized retailer in the southeast region. In that context, we estimated the sales of fashion jeans based on two sales promotions (represented as two dummy variables) and controlling for special events. Once we ran the OLS regression, we also obtained residuals for week-to-week sales.

For this problem on stationary series, you will setup and execute two types of forecast methods, using all 26 residuals (not the sales variable, which was the original demand signal) as the training sample. Note that there is no need for a test or holdout dataset to forecast for this problem. The two methods and their details are as follows:

a. Moving Average (MA) estimations

- Try N = 1 to 5 and forecast estimates for weeks 6 to 26
- Start forecast on week 6 as need 5 weeks of historical data for N=5
- Report resultant MSE values
- MA* = [MA(N) with optimal value of N]

b. Single Exponential Smoothing (ES) estimations

- Try alpha = 0.1 to 1 using 0.1 increments and forecast estimates for weeks 6 to 26
- Assume a burn-in period of 5 weeks
- Report resultant MSE values
- Single ES* = [Single ES(alpha) with optimal value of alpha]

c. Assess MA* versus Single ES*

- Generate plot with residuals, MA* forecast, and Single ES* forecast
- Compare MA* and Single ES* outcomes on training sample and provide your recommendation of method and justification. Simply showing output of running code does not suffice as a justification, provide an interpretation in your own words!

Part II: Non-Stationary Series with Trend
Dataset: bitcoin_data

You are hired as a consultant to apply your forecasting skills for predicting future outcomes of Bitcoin prices. You are given historical Bitcoin data for a range of interest. The data consists of Bitcoin price for 180 days.

a. Test for non-stationarity

- Using the Augmented Dickey-Fuller test, provide your conclusion on the nature of the data based on patterns of the Bitcoin prices over time

b. Double Exponential Smoothing (DES) estimations

- Provided that your test for non-stationarity was supported, forecast estimates by fitting a DES model
- Try alpha and beta from 0 to 1 with 0.05 increments
- Use the first 150 days as training sample
- Assume a burn-in period of 10 days
- Calculate training sample MSE for days 11 to 150
- Forecast for the last 30 days
- DES* = [optimal DES(alpha,beta) with optimal combination of alpha and beta]
- Report resultant MSE values, both for process of forecasting training and test data. Provide a conclusion on whether the DES* performs better on the train or test data.
- Generate plot with Bitcoin prices and DES* forecasts. Explain what is represented in the plot. Simply showing output of running code does not suffice as an explanation, clearly state what is shown in the plot in your own words!

Part III: De-Seasonalizing Data and Forecasting
Dataset: airline_data

Virgin Airlines is seeking your help on forecasting future customer booking behavior based on airline load factor data from the Bureau of Transportation Statistics Monthly. Airline load factor data is a measure of the use of aircraft capacity that compares Revenue Passenger-Miles as a proportion of Available Seat-Miles. This sort of passenger load factor is commonly used to measure the capacity utilization of public transport services like airlines, passenger railways, and intercity bus services. It is typically used to assess how efficiently a transport provider fills seats and generates fare revenue.

In order to assess your forecasting capabilities, the company asks you to forecast the de-seasonalized load factor for the last two months in December 2013, which they already know the outcome for. Note that there is no need for a test or holdout dataset to forecast for this problem.

a. De-seasonalize the historical data

   - The company provided you with the raw data, but have reported that it needs to be adjusted for seasonality before being forecasted. Thus, as a first step, perform the steps to de-seasonalize the data and report the resultant dataset

b. Test for non-stationarity

   - Using the Augmented Dickey-Fuller test (ADF test), provide your conclusion on the nature of the data based on patterns of the (de-seasonalized) airline load factor data over time
   - Based on your conclusion from the ADF test, evaluate the appropriate time series forecasting method(s) from those discussed in class. DO NOT arbitrarily select one or more of the forecasting methods. Be clear on the reason for selecting any appropriate forecasting methods AND the basis for determining the most accurate forecast method to move forward with in the proceeding steps.
   - Use the most accurate forecast method to forecast the last two months in December 2013
   - Assume no burn-in period required
   - Report resultant MSE values from any forecasting method that was used
   - Report forecast of the last two months in December 2013 and make sure to scale back by the seasonal factor for final forecast. Explain the meaning of the reported forecast. Simply showing output of running code does not suffice as an explanation, clearly state what the forecast tells us in your own words!