**DSC 275/475: Time Series Analysis and Forecasting (Fall 2022)**

**HW #1**

**(Total points: 40)**

1. *(8 pts)* The Bureau of Transportation Statistics (BTS) conducted a study to evaluate the impact of Sept 11 attacks (9/11) on U.S. air transportation. The purpose of this study is to provide a greater understanding of the passenger travel behavior patterns of persons travelling by air before and after the event. In order to assess the impact of September 11, BTS took the following approach: Using data before September 11, it forecasted future data (under the assumption of no terrorist attack). Then, BTS compared the forecasted series with the actual data to assess the impact of the event.

The data is available in the file: *BTS\_Air\_Rail\_Vehicle\_Miles.csv*.

1. Is the goal of this study descriptive or predictive ? (2 pts)
2. Create a time series plot of the *Air* data, i.e. a plot yt versus *t, where t=1,2,3 …* What would t=1, 2, 3 refer to in the time series? Which time period does t=1 refer to ? (3 pts)
3. What are the values for y1, y2 and y3 in the time series? (3 pts)
4. *(10 pts)* In addition to air travel data, two additional time series are also provided in the same data file – Rail and Vehicle travel.
5. Which of these components appear in the *Air* and *Vehicle* time series: i) Level; ii) Seasonality; iii) Trend; iv) Noise. List for each data set. (8 pts)
6. For the Rail data set, describe the trend, i.e. how does the trend vary across the time series? (2 pts)
7. *(6 pts)* Forecasting Shampoo Sales: The file *ShampooSales.csv* contains data on the monthly

sales of a certain shampoo over a 3 year period.

1. Create a time series plot of the data. Label the axes. (2 pts)
2. Which of the four components (level, trend, seasonality, noise) are present in this

series? (4 pts)

1. *(6 pts)* The file, *Beverages\_Shipment\_2020.csv*, contains the US beverage product shipments data.
2. Is there seasonality in this time series? (2 pts)
3. Find the sample autocorrelation function for this data set. (2 pts)

(For Python, you can use the “plot\_acf” function in “statmodels” module. Plot at least 25-30 lags)

1. From the autocorrelation plot in (b), what is the seasonal period? (2 pts)
2. *(10 pts)* Data on US coal production is given in *Coal\_Production\_US\_2020.csv*.
3. Plot the coal production data and the sample autocorrelation function. (2 pts)
4. Is the time series stationary or non-stationary ? (2 pts)
5. Plot the first difference of the time series and the sample autocorrelation function of the first difference. (4 pts)
6. What impact has differencing had on the time series? Comment with respect to presence or absence of stationarity (2 pts)