**DSC 275/475: Time Series Analysis and Forecasting (Fall 2021)**

**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)

The goal of the study is descriptive.

1. 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)

T refers to months. t=1 is January 1990

1. What are the values for y1, y­2 and y3 in the time series? (3 pts)

Y1= 35153577

Y2= 32965187

Y3= 39993913

1. *(10 pts)* In addition to air travel data, two additional time series are also provided in the same data file – Rail and Vehicle travel.

Air:

A picture containing text, antenna, screenshot

Description automatically generated

Vehicle:

Chart

Description automatically generated with medium confidence

Rail:

A picture containing chart

Description automatically generated

1. 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)

Air: Trend, Seasonality, Level, Noise

Vehicle: Trend, Seasonality, Level, Noise

1. For the Rail data set, describe the trend, i.e. how does the trend vary across the time series? (2 pts)

The trend was decreasing from 1992 to 1996 and then starts to increase from 1998 to 2004.

1. ­*(6 pts)* Forecasting Shampoo Sales: The file *ShampooSales.csv* contains data on the monthly sales of a certain shampoo over a 3 year period.
2. Create a time series plot of the data. Label the axes.

Chart, line chart

Description automatically generated (2 pts)

1. Which of the four components (level, trend, seasonality, noise) are present in this

series? (4 pts)

It has level, trend, noise but no seasonality.

1. *(6 pts)* The file, *Beverages\_Shipment\_2020.csv*, contains the US beverage product shipments data.

A picture containing text, antenna

Description automatically generated

1. Is there seasonality in this time series? (2 pts)

Yes, it is for about 12 months.

1. 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)

Chart, histogram

Description automatically generated

1. From the autocorrelation plot in (b), what is the seasonal period? (2 pts)

The seasonal period is for 12 months.

1. *(10 pts)* Data on US coal production is given in *Coal\_Production\_US\_2020.csv*.
2. Plot the coal production data and the sample autocorrelation function. (2 pts)

Chart, line chart

Description automatically generated

Chart

Description automatically generated

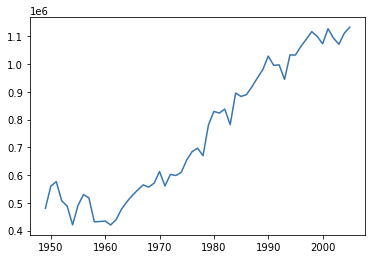
1. Is the time series stationary or non-stationary? (2 pts)

It’s a non-stationary time series

1. Plot the first difference of the time series and the sample autocorrelation function of the first difference. (4 pts)

Chart, line chart

Description automatically generated

Shape, rectangle

Description automatically generated

Chart, scatter chart

Description automatically generated

1. What impact has differencing had on the time series? Comment with respect to presence or absence of stationarity (2 pts)

Due to differencing, the trend in the time series is removed turning it from non-stationary to stationary.