THE UNIVERSITY OF CHICAGO
Booth School of Business
Business 41912-01, Spring Quarter 2020, Mr. Ruey S. Tsay

Homework Assignment #1

Due Date: April 14 (before class). You may use any software to solve the problems. Don't hand in all outputs; use cut-and-paste to select the relevant part of the output.

Data files: Datasets are available from the Internet (including R packages) or course web (UC canvas).

1. Consider the COVID-19 cases in the U.S. New York Times has compiled the data via states and via counties. The data are available on github.com/nytimes/covid-19-data. You can also find the locations (latitude and longitude) of each state (or its capital) on the web. (a) Use the cumulative data of March 31, 2020 to obtain a bubble plot for the number of cases of the US. (b) Use the cumulative data of April 3, 2020 to obtain a bubble plot for the number of cases of the US. (c) Compare and comment of the two bubble plots.

2. Use the bivariate boxplot on the scatterplot of each pair of variables in the US air pollution data to identify any outliers. Compute the Pearson correlation coefficients between the variables using all data, then use data with outlier removed.

3. Compute the Kendall's tau and Spearman's rho between the variables in the air pollution data set. First, use all the data, then use data with outliers removed. Comment of the effect of oultiers on the correlation coefficients.

4. Are the air pollution data jointly normally distributed? Why? You can perform test or use QQ plot.

5. Problem 3.17 of the textbook. Specifically, show that, if $\boldsymbol{X}$ $(p \times 1)$ and $\boldsymbol{Z}$ $(q \times 1)$ are two independent random vectors, then each component of $\boldsymbol{X}$ is independent of every component of $\boldsymbol{Z}$, where $p, q \geq 1$.

Reading assignments: Chapters 1 to 4 of the textbook.