Statistical Sciences
UNIVERSITY OF TORONTO

STA457H1/STA2202HF L5101 – Time Series Analysis

Problem Set #2
Due by Sunday November 3rd, 2019

Problem 1 (35).

Consider the dataset about bitcoin (BTC) price, the time series is from January 1st, 2017 to October 18th, 2019.

(1: Points 6) First read the BTC data, then plot the QQ plot, boxplot and kernel density estimates. Discuss any features you see in the QQ plot, boxplot and kernel density plot. Specifically, address the following questions: Do the data appear to be normally distributed? If not, in what ways do they appear non-normal? Are the data symmetrically distributed? If not, how are they skewed? Do the data seem heavy tailed compared to a normal distribution? How do the left and right tails compare; is one tail heavier than the other?

(2: Points 6) Next, conduct data transformation by taking the square-root and log-arithm to get transformed data, denoted as sqrt.BTC and log.BTC, respec-tively. Further plot QQ plots, boxplots and kernel density estimates, discuss the features you see in the plot similar to question (1).

(3: Points 6) Conduct Box–Cox transformation to the original data and estimate its parameter λ by maximum likelihood.
(Hint: Zoom in on the high-likelihood region with the following function: boxcox (BTC$\sim$ 1, lambda=seq(0, 1, 1/100)); Also use package "MASS").

(4: Points 6) Find a 99% confidence interval (CI) for λ.

(5: Points 6) Try to fit a skewed t-distribution including R code (use package "fGarch" and function "sstdFit").

(6: Points 5) What are the estimates using the skewed t-distribution?

Problem 2 (35).

Consider the Gold price in USD data and this time series is observed from January 3rd, 2017 to October 18th, 2019 (notice that you might need to clean some invalid data in Excel).

(1: Points 6) First read the Gold data calculate the corresponding log returns (say Y). Plot this time series of log returns and write a brief description. Do the series look stationary? Do the fluctuations in the series seem to be of constant size? If not, describe how the volatility fluctuates. (use function "diff").

(2: Points 6) Plot the QQ plot, boxplot and kernel density estimates of the log re-turns and give some explanations.

(3: Points 6) Fit the standardized t-distribution (std) to the log returns. Find MLEs of the mean, standard deviation, and the degrees-of-freedom parameter by using Maximum Likelihood Estimation. (use package "fGarch" and function "Optim").

(4: Points 6) Calculate the AIC and BIC values based on the above optimazition.

(5: Points 6) Modify the code so that the MLEs for the skewed t-distribution are found. Include your modified code with your work. What are the MLEs?(also use package "fGarch" and function "Optim").

(6: Points 5) Which distribution is selected by AIC, the t or the skewed t-distribution?Which distribution is selected by BIC, the t or the skewed t-distribution?

Problem 3. (30)

Consider the dataset of IBM price shares in the time series observed from January 3rd, 2017 to October 18th, 2019.

(1: Points 4) Read the IBM dataset and calculate its sample mean, standard devi-ation, skewness and kurtosis. (need to use package "fGarch").

(2: Points 4) Fit a t-distribution to the data and show the estimates (use function "stdFit").

(3: Points 4) Try to bootstrap the sample mean 1000 times in two cases: model-free and model-based on t-distribution. (use functions "sample" and "rstd").

(4: Points 4) Plot QQ plot and KDEs of ModelFree mean and ModelBased mean. Also, plot side-by-side boxplots of the two samples. Describe any major dif-ferences between the model-based and model-free results. Include the plots with your work.

(5: Points 5) Find 95% bootstrap confidence intervals for the sample mean using the model-based and model-free bootstraps with digits=5.

(6: Points 4) Estimate the bias of the sample mean of IBM based on model-free and model-based bootstraps.

(7: Points 5) Estimate the mean squared error (MSE) of the sample mean of data IBM. (Notice that MSE=variance+bias^2.)