STA4003 Project

Due date: May 14, 2021

- Outstanding projects will be invited to give a presentation in class or on Zoom on April 29, 2021. Students who have given a presentation can receive maximum 10 bonus points in their final exam.
- Students who want to present their work need to submit the project by April 23, 2021. Submissions after April 23 will not be invited for presentation. All students can revise their work before May 14, 2021.
- The submitted `R` codes must be clearly written in the given starter file "Project.R".

This project is a full analysis of a dataset "StockData.Rdata", which contains weekly log returns of SP500 index and 10 stocks for $n_1 + n_2 = 260$ weeks. The data for last $n_2 = 52$ weeks is the test data. In the project, you are required to finish two main tasks.

1. **Forecasting.** Starting from Week $t = n_1 = 208$, you need to build models forecast the log returns of the given 10 stocks, namely Stock i for $i = 1, \ldots, 10$, based on the past observations in Week $t - n_1 + 1$ to Week t for $t = n_1$ to $n_1 + n_2 - h$, when h is the number of weeks to forecast. Let $x_{s+h,i}$ be the true log return of Stock i at Week $s + h$ and $\hat{x}_{s,i}(h)$ be your h-week ahead forecast in Week s, where $h = 1, 2, 3$. The goal of this part is to minimize the mean squared errors,
$$\text{MSE}(h) = \frac{1}{10(52 - h + 1)} \sum_{i=1}^{10} \sum_{s=n_1}^{n_1+52-h} (\hat{x}_{s,i}(h) - x_{s+h,i})^2.$$
2. **Portfolio construction.** Starting from Week $t = n_1$, you are required to construct a portfolio of the given 10 stocks for Week $t + 1$ with the corresponding weights $w_{t,i}$, $i = 1, \ldots, 10$. Note that $w_{t,i}$ must depend on the observations up to Week t only. For example, $w_{t,i} = 1/10$ in the starter codes. Hence, the log return of the portfolio for Week $t + 1$ is
$$r_{t+1} = w_{t,1}x_{t+1,1} + \cdots + w_{t,10}x_{t+1,10}.$$
In this project, the only constraint for $w_{t,i}$ is $\sum_{i=1}^{10} w_{t,i} = 1$. In particular, negative $w_{t,i}$ (short-selling) is allowed. We will compare this log return with that of S&P500 and consider the excessive return
$$e_s = r_s - r_{0,s},$$

where $r_{0,s}$ is the log return of S&P500 in Week s. The goal of this part is to maximize the ratio,

$$\hat{\mu}_e/\hat{\sigma}_e,$$

where $\hat{\mu}_e$ is the sample mean of $\{e_{n_1+1}, \ldots, e_{n_1+n_2}\}$ and $\hat{\sigma}_e^2$ is the sample variance.