**2.3 Analysis of Question 3**

Since the timing of the transfer remains the same, and the number of stocks invested and the percentage of each stock invested remain unchanged from question 1, the question does not indicate that the ten strongest stocks of the previous week should be selected. Therefore, the change in our investment strategy in this question compared to question 1 is to change the reference from last week's data to the previous 30 days' data (or all previous data if less than 30 days) to select the strongest ten stocks. After analyzing a sample of several stocks’ movements, we found that in most cases, the movement of a stock after a certain point in time is most influenced by the previous 30 days or so. A too-long time span results in some factors, which have less impact on the current trend, being taken into account equally. And similarly, a too-short time span results in some stock price fluctuations caused by consecutive time periods not being fully taken into account. By choosing 30 days as the maximum time horizon to be considered, the accuracy of the forecast can be effectively improved.

**2.4 Analysis of Question 4**

Referring to the previous 30 days (or all previous data if less than 30 days), select the ten strongest stocks using the way in question one and invest in the ten stocks. The score of each stock is calculated in the way of question two, and the amount of investment is determined by the percentage of the score, and the time of purchase is referred to the 37% theory. The actual opening time of a business day is 4 hours, and the total opening time of the first three days is 12 hours, so the time point corresponding to the 37% theory is around 10:20 a.m. on Tuesday. Therefore, the lowest price of a stock before 10:20 on Tuesday is used as the reference price, and once the price of the stock is lower than the reference price after that, it will be bought quickly. If the score of the stock is higher than 10%, the stock will be purchased at 10% of the total assets, or if there are unpurchased shares until the last minute, directly at the price at which it will last close. If the position is still less than 50%, the position should be exactly equal to 50% after the purchase is completed, and the weighting of the purchased stocks should be consistent with the score weighting. The total market opening time for a week is 20 hours, so the time point corresponding to the 37% theory is Tuesday at 2:40 p.m. Therefore, after completing the position, the reference price at this time is the lowest price as of Tuesday at 2:40 p.m. If a price lower than the reference price appears again afterwards and the position is not full, buy according to the ratio, and if there is still a position left at the end, buy the stock with the highest score directly.

**III. Explanation of symbols**

Symbol Meaning

𝒎 the principal amount of the stock acquired

𝒐𝒊 the opening price of a stock

𝒄𝒊 the closing price of a stock

i the daily return of a specified stock

𝒊 the average of the specified stock's returns

𝑻𝒊 the trading volume of a specified stock on a given day

𝒕𝒊 trading volume of the specified stock on a given day

𝒙J average price of the specified stock for a week

𝝈𝒊 the standard deviation of the stock price of a given stock

𝝎𝟏, 𝝎𝟐 weight

𝒏 amount of stock purchased after fees

𝒔𝒊 the score of a stock

**IV. Data Pre-processing**

For Appendix I with 30 stock quotes data provided by the topic, we use pandas to read and perform preliminary processing. First, the data are grouped by stock code and indexed by time. Then, the row labels are resampled by week. Since the data given in the topic includes multiple stocks of SZSE and SSE, and the effective trading days of each stock are different, the data are rebuilt using a uniform time series, i.e., the actual trading day series of the two stock exchanges from 2011 to 2020 for indexing the data. The missing data are replaced by the mean of the known data before and after. From the investment strategy, it is known that the closing price on Monday and the opening price on Tuesday of each trading week are needed to determine whether a single week in the resampled series is a complete trading week and to extract data for the dates that meet the requirements.

The Appendix II, which contains Greater Bay Area Index ticker data, provided in the topic is considered as a portfolio in the same format as the model results. An index reconstruction operation similar to the one above is performed on it to ensure that the data are all in the same dimension in the same format.

After considering the computing power of the equipment, the time cost and the need of the model, we decided not to use the minute quotes data provided in the title, as we believe that this data set has limited effect on the overall model.

**V. Development and Solution to the Model**

**5.1 Answer to Question 1**

**5.1.1 Development of Model 1**

Let the actual amount of the stock purchased after deducting the commission be 𝑛, the principal be 𝑚, the closing price on a given day be 𝑐! and the opening price on that day be 𝑜!

Number of shares of this stock purchased on a given day:

Return for that stock on a given day is:

The average rate of return for a given stock is:

The actual amount of shares purchased for a stock is:

The average amount divided to each stock is:

The sum of the proceeds and the principal amount generated by the purchased stock is.

The rate of return on a stock is:

The rate of return on the ten stocks is:

**5.1.2 Solution of Model 1**

Based on the data given in the question, substitute the above model and use python to calculate and process it, the following is obtained:

Graphical analysis: The green line is the yield curve of the investment portfolio generated by the model, the red line is the yield curve based on the Greater Bay Area Index, and the x-axis of the graph is a time series with the week in which January 3, 2011 is located as the 0 point, on a weekly basis, to the week in which October 30, 2020 is located.

It is clear from the graph that the model is generally able to outperform the reference index.

Moreover, the model is able to obtain higher returns in times of financial crisis and poor macroeconomic conditions such as trade frictions between the US and China.

**5.2 Answer to Question 2**

**5.2.1 Development of Model 2**

Assume that the score of a stock is Si

And the amount invested in a particular stock is

The optimal rate of return of this investment strategy is :

**5.2.2 Solution of Model 2**

Based on the data given in the question, substitute the above model and use python to calculate and process it, the following is obtained:

Graphical analysis: The green line is the yield curve of the investment portfolio generated by the model, the red line is the yield curve based on the Greater Bay Area Index, and the x-axis of the graph is a time series with the week in which January 3, 2011 is located as the 0 point, on a weekly basis, to the week in which October 30, 2020 is located.

The investment strategy remains highly effective during periods of stock market instability, such as the financial crisis and the U.S.-China trade friction. The investment return is significantly higher than that of Model 1.

**5.3 Solution of Question 3**

**Development of Model 3**

Set a total score of a stock is equal to Si

The amount invested in the stock is:

The optimal rate of return of this investment strategy is :

**VI. Model evaluation and promotion**

**Advantages of the model:**

1. rich parameter consideration and high fault tolerance.

2. The model generates portfolios that generally outperform the reference index.

3. In several tests with historical data, the model was able to obtain higher returns in poor macroeconomic times.

4. The model requires less computational resources to perform its calculations

**Disadvantages of the model:**

1. The depth of data processing is not sufficient.

2. Standard deviation is not introduced, which makes the model more risky.

**Extensions of the model:**

The model is highly extensible, and more parameters can be considered on this basis in devices with higher computational power.