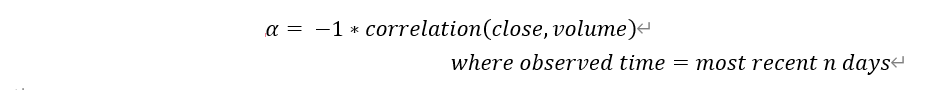
**Multi-factor strategy**

1. **Strategy Intro**

A multi-factor strategy is a stock-picking strategy by combining several factors for analysis. Factors are based on certain yield-related indicators (e.g. High, Close, Volumn...). It calculates a value and compares it to a specified threshold to go long or short a certain number of stocks. For example, we calculate the value of the indicator for each day and compare it to a threshold value endowed. We go long when the indicator value is greater than the threshold value and go short when it is less. The key to the multi-factor model is finding the relationship between factors and returns to make favorable trading decisions.

The key factor in our strategy is modified from the original Alpha-006. The basic logic of this factor is to go long when the stock price and volume are negatively correlated. We first calculate the correlation between the close price and the volume in the most recent *n* days, where *n* is the observed time. The higher the correlation, the closer the opening price and trading volume are to "rise-rise". As we need to find the circumstance of “rise-low” or “low-rise”, we multiply the factor by -1, resulting in the following equation:

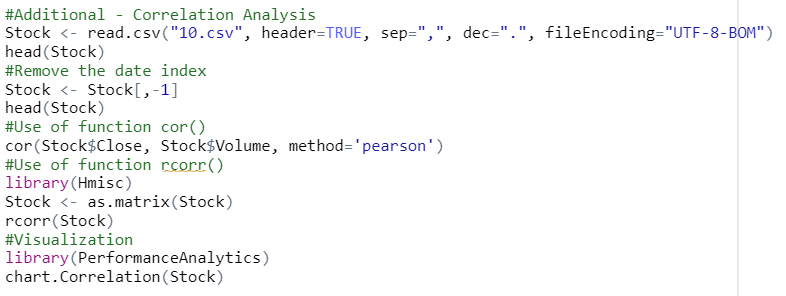


**2.1 Data analysis**

Alpha-006 calculates the correlation between the close price and the corresponding volume. Thus, before applying this factor, we must first make sure that there exist some levels of correlation between the close price and volume on a particular day in the given data. In addition, as we introduce the parameter of observed time *n* in factor Alpha-006, we need to analyze how to choose the value of n based on the volatility of the given data.

***Part A – Prove of correlation between close price and volume***

We use the following R code to analyze the correlation between the given indicators in stock. Set stock 01 as an example.



We get the display of correlation between indicators:

**文本

描述已自动生成**

And a graph that clearly visualized the correlation:

**图表

描述已自动生成**

From the graph displayed, there exist high levels of correlation between indicators of stock 1. By conducting the same operations from stock 01 to stock 10, the correlation index of close price and volume are 0.4284853, 0.4394195, 0.0691163, 0.04081823, 0.1069442, -0.07284416, -0.0242775, 0.07110634, 0.4357325, -0.15527.

***Part B – The stocks’ volatilities and the observed period setting***

In the previous data analysis section, we used Python to calculate the volatilities of the given ten stocks and made a ranking. The result shows that stock no.1, 2 and 8 have the greatest fluctuation while stock no.6, 7 and 10 have the smallest.

For stocks with relatively stable prices, we tend to set the observed period longer so that the decision on a particular day can reference more previous day’s data, resulting in a more reliable trading decision. However, for the stocks with high volatility, the decision using a long-observed period will be greatly affected by fluctuations and become error-prone. Thus, our strategy with a long-observed period input is supposed to work well on stock 6,7,10 and badly on stock 1,2,8, and vice versa.

**3.1 Multi-factor strategy specify**

Our strategy is composed of three factors, namely Alpha006 .

**1. Factor formula analysis**

Original Alpha006: (-1 \* correlation(open, volume, 10))

It is clear that the parameters needed for this factor are: the daily opening price and volume of the stock

**Factor function description：**

correlation(x,y,d)

**Meaning：**

Correlation coefficient of two random variables x,y in the past d days, its value range is [-1,1]. For this factor, it can be interpreted as: the daily opening price of the stock and the volume of the correlation coefficient in the past 10 days.

**Explanation of the formula：**

First, the correlation between the opening price and the volume of the last 10 trading days is calculated. The higher the correlation, the more the opening price and volume tend to be the same. But the final result is added on the negative sign, it means: the higher the value of the factor, the lower the correlation between the stock's opening price and volume over the past 10 trading days.

**Formula logic.**

This is a phenomenon of securities trading: when a new peak in the price of a security occurs, the volume does not increase, but begins to fall. This means that the price of the security is not proportional to the volume of the relationship. In other words, the stock's price has fallen to a minimum and is about to end its downtrend and start an uptrend. This is a buy signal and investors may consider going to buy a certain amount of shares at this time. Conversely, when the stock price continues to rise in phase, but the volume of the stock continues to fall. This means that the stock price has reached its peak and is about to end its uptrend and start a downtrend. This is a sell signal and investors may consider going to sell a certain volume of the stock. And when the above two phenomena are not obvious, it means that the stock does not yet have a strong buy or sell signal and investors can temporarily hold off on trading waiting to trade the stock from the next node where the change is obvious.

**Strategy Setting.**

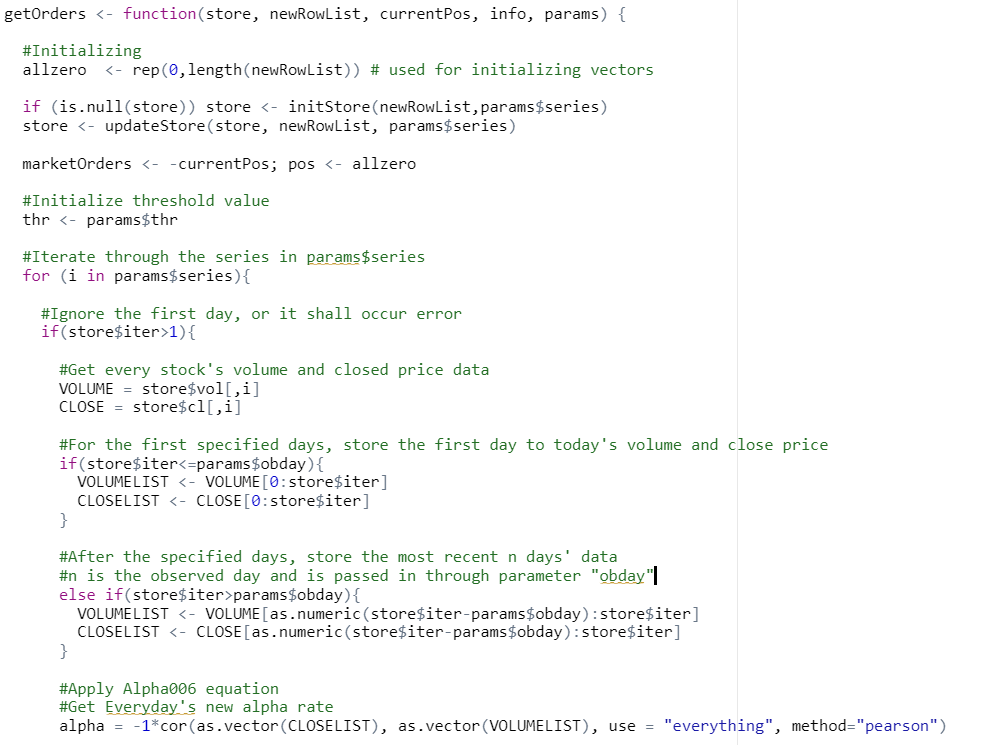
There will be two thresholds in the strategy, a positive number and a negative number.

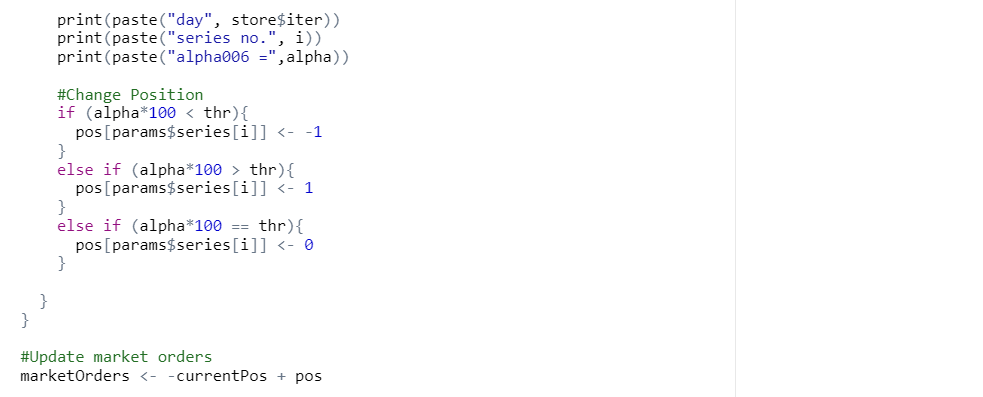
1. When the factor value of the day is greater than the positive threshold, it means that the stock price has fallen to the lowest and is about to end the downtrend and open an uptrend. At this point you should buy, so set the position to 1.

2. When the factor value of the day is less than the positive threshold, it means that the stock price has reached its peak and is about to end its uptrend and start a downtrend. At this point, you should sell, so set position to -1.

3. When the factor value of the day appears in the middle of the two thresholds, the stock does not have a clear signal to buy or sell. You should not trade at this point, so set position to 0.

**Code:**

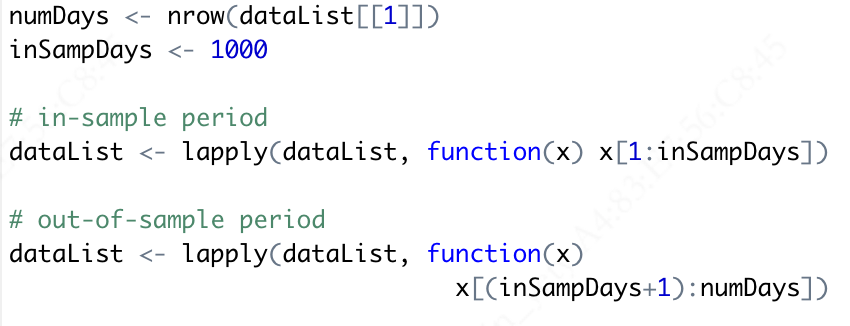




**4.1 Preliminary testing**

**Strategy Implementation：**

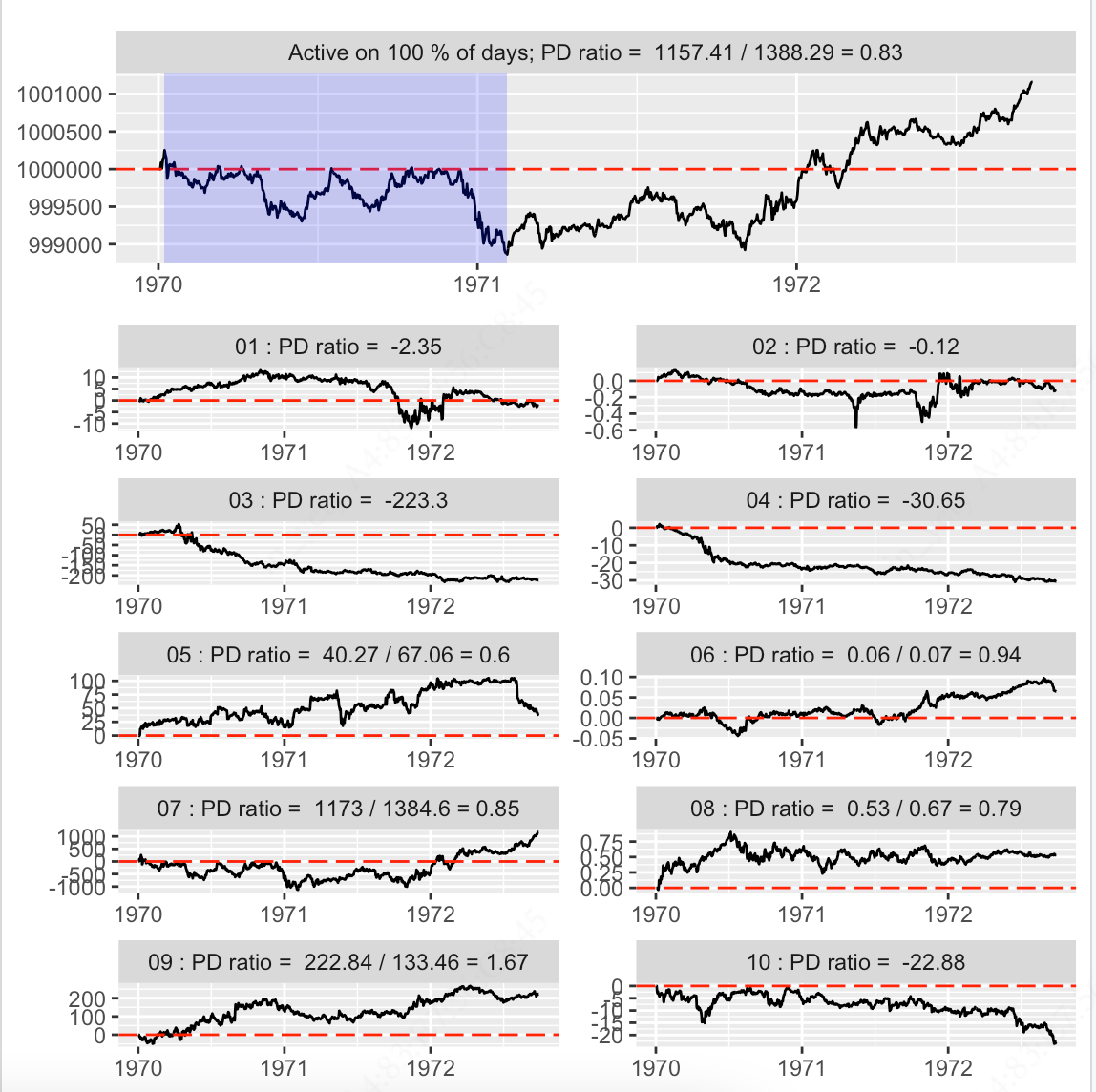
We choose first 1000 days as in sample test

****

We calculate all the stock price return by using this strategy and we set parameters like this:

"alpha101"=list(series=c(1,2,3,4,5,6,7,8,9,10), thr=0, obday=35)

And we get the plot of all ten stock

****

**Strategy Optimisation：**

By the definition of the formula, we need to base the calculation on the previous 10 days of stock trading data. This means that in order to run the strategy, the stock would have to start trading after 10 days. So we made a slight change to the strategy and ran the logic as follows: the strategy starts on the second day and uses all previous trading data as parameters for the calculation of the factors each day. After the tenth day the calculation is then changed to use the previous ten days of trading data. This effectively avoids the disadvantage that stocks can only be traded from 10 days onwards.

**5.1 Future plan**

We have tried to develop a better strategy. But due to some technical or time issues, we are unable to achieve that in this version of the coding.

1. Inability to determine the optimal threshold

Our first thought of this strategy is to find the optimal threshold. A previous version of the code is to loop the threshold from -100 to 100 and simulate the trading, record the highest PD-ratio and the corresponding threshold, then use the best threshold to conduct the strategy again. However, our code meets infinite recursion issue. We need to call backtest() function inside the getOrders() function to simulate the trading and get the PD-ratio, while inside backtest() function it calls the getOrders() function again to conduct the strategy. Thus, we cannot use a functional approach to optimally select the threshold and have to manually set the threshold as a parameter in the current version. We will try to find a solution to solve the infinite recursion issue and try to functionally find the optimal threshold in the future.

2. The incorrectly choose of observed day on some occasions.

Another parameter of this strategy, beside the threshold, is the observed day. In this version of strategy design, we calculate the overall volatility of a stock and determine whether to set a long *n* or a short *n*. However, if there is one stock whose price is stable for a period, but suddenly turns to fluctuate in another period, The *n* we set based on the above data analysis will be inaccurate. This remains an unsolved weakness in our design. For the future improvement plan, this inaccuracy is assumed to be avoided if we could intelligently and intermittently analyze the volatility of each stock over time and set different *n* to different period.

3. Maybe more factors?

Although this strategy is named as multi-factor strategy, only one factor, which is the modified version of Alpha006, has been used. We have tried to modify and add other factor in the Alpha101 series, but have met problems. Some factors require additional stock indicators that we couldn’t get, some require functions that we cannot express correctly in R code, and some are performing an even worse PD-ratio outcome when we directly add to the strategy. In the future plan, we will try to examine and modify more factors in the strategy to make it a literally “real” multi-factor strategy.