Proceedings of the 3rd International Conference on Business and Policy Studies
DOI: 10.54254/2754-1169/106/20241598

Trading Strategy Based on the Leading Effect of Leading Stocks: LLR Model and Correlation Coefficient Effect

Guoquan Zhao[1,a,*,†], Hao Zhang[2,b,†], Xingyi Chen[3,c,†]

[1]*Department of math, Shenzhen University, Shenzhen, 518060, China*
[2]*Aquinas International Academy, Garden Grove, CA, 92840, USA*
[3]*International Business School Suzhou at XJTLU, Xi'an Jiaotong-Liverpool University, Suzhou, 215028, China*
a. 916544047@qq.com, b. 825522871@qq.com, c. 2372445685@qq.com
**corresponding author*
[†]These authors contributed equally.

Abstract: Stocks is one of the most popular financial tools around the world, However, it is essential to determine which stock to buy and the moment we buy in and sell out, only catch up with the proper opportunity, then we can make fortune from the stock market. So how can we find out those stocks which has great potential to grow in the future? Our study focuses on the leading stocks effect, which refers to in a specific industry, when some stocks in that industry have reached their price limit, which we call leading stocks, and in the following trading days, the stocks from the same industry will be affected by leading stocks and have a potential growth. In that case, if we can find out the leading stock in a specific industry and monitor it growth every day, then base on correlation coefficient, we can determine which stock to buy in and make profit via this strategy.

Keywords: LLR model, leading stocks effect, correlation coefficient, stock trading strategy.

1. Introduction

1.1. Idea

We determine industry leading stocks based on LLR, and then use correlation analysis to take the lagging stocks with strong correlation with the leading stocks as our stock pool. Widespread intercorporate ownership might also cause firm fundamentals to move together [1], as this implicitly causes some firms' performance to depend on that of other firms control for size effects in examining the cross-autocorrelation patterns between high volume and low volume stocks. When the leading stock rises by the limit, buy the corresponding lagging stock, and according to the variance and closing price of the leading stock's capital flow, the large fluctuation of the leading stock price is used as a short signal. If the market is efficient, mispricing will be quickly corrected, stocks with abnormal performance will not maintain the performance subsequently [2]. We monitor every day and find that when the leading stocks no longer rise by the daily limit, we believe that leading stocks are ineffective, so we sell the stocks we were long on at this time. When the volatility of leading stocks returns to normal, we will return the stocks we shorted before.

DOI: 10.54254/2754-1169/106/20241598

1.2. Highlight

The top 10% of stocks were selected based on the ranking by the LLR index of stocks in several industries as the leading stocks, and the lagging stocks were determined by computing the correlation coefficient between other stocks in the same industry and leading stocks to build a portfolio. For data accuracy, the study monitors the daily LLR records. We buy the corresponding trend of lagging stocks when the leading stocks appear limit signals. In addition, by calculating the variance of the leading stock, we believe that if the standard variance mean ratio is greater than or equal to 0.02, the stock will have a large fluctuation, and we will use this as a signal to short the corresponding lagging stock. The law is that once the daily purchase monitoring, the leading stocks in the industry will no longer rise by the limit, which means the failure of the leading stocks, once the leading stocks fail, we will sell the corresponding industry to follow the trend of lagging stocks; Once the standard variance mean ratio of the leading stock shorted is less than 0.02, the corresponding lagging stock that was shorted before will be closed.

2. Specification

2.1. Analysis

2.1.1. Qualitative analysis

China's stock market is not mature compared to developed countries; there is a huge information asymmetry. For instance, the market information disclosure system is incomplete due to improper data records and collection, which leads to a serious information gap in the stock market. Therefore, it is difficult for investors to get a completely true operating situation of the listed company's financial statements and their performance since falsification of the financial statement is common. Also, there are many individual investors in China's stock market, and the investment expertise of individual investors has contributed mainly to falsifying financial records due to the phenomenon of following the herd, and the "herding effect" is significant.

2.1.2. Quantitative analysis

First we use lead-lag ratio (LLR) to study the relationship between the two assets after the lead, based on the LLR model we can determined the leading stocks in the market. then we use R^2 and accuracy as dependent variables to detect the real existence of the leading driving effect of leading stocks, which has a predictive effect on other lagging stocks in the sector, if the value of R^2 is big enough (over 0.6), then we reckon the two stocks have a strong correlation.

We use the basic mathematical statistics theory to give a quantitative index to measure the stock price volatility -- standard variance mean ratio, and use this index to divide the volatility into normal volatility and abnormal volatility. And for the standard deviation mean ratio, we obtained A reference value of 0.02 based on the Shenzhen Stock Exchange A-share index.

2.2. Data--Universe & Datasets & Data Sources & Data Range

The study will use secondary sources and adopts a quantitative research method. The research design used is explanatory research using descriptive and inferential statistics to analyze the data. The data is selected randomly for the opening and closing prices from January 2, 2018, to July 17, 2023. The data is inserted in Excel, cleaned through format unification, data standardization, consistency check, and coded before analysis in the Stata software. To get the return of each stock, we subtract the closing price of the current day from the previous day's closing price and divide the previous day's closing price to get the return of each stock. In addition, since the market yield is difficult to obtain, we

DOI: 10.54254/2754-1169/106/20241598

believe that dividends have been reflected in the fluctuations of the stock index. The market index is used as a customary substitute for the market yield. The research will use the CSI 300 index as our daily market yield for accuracy. Also, to obtain the variance of the return rate of each stock and the variance of the market return rate, we prepare the data for the subsequent calculation of the correlation coefficient between the leading stock and other stocks in the trading industry. In addition, we also use data on the flow of funds. Moreover, for the standard variance mean ratio, we obtained A reference value of 0.02 based on the Shenzhen Stock Exchange A-share index.

For the back test data, we selected the data from January 2, 2018, to January 2, 2021, as in-sample data and from January 2, 2021, to July 17, 2023, as out-sample data [3].

2.3. Methodology

2.3.1. Signal Generation

A signal to long:

Different trading sectors in China have different daily stock limits. Therefore, once the leading stocks in a given industry appear to have reached their daily limit phenomenon, it will be good news for the industry since they will play a leading role in the lagging stocks [4]. Therefore, the lagging follower stocks will have a trend to approach the leading stocks in the future. Hence, the daily limits of leading stocks will be used as the signal.

For the main board, the equation can be expressed as follows;

$$\frac{P_{c(n)}}{P_{o(n)}} \geq 0.1$$

For the science and technology board, the equation can be expressed as follows;

$$\frac{P_{c(n)}}{P_{o(n)}} \geq 0.2$$

When the growth rate of the leading stock reaches the limit of its module, we will use it as a buy signal and buy the portfolio of the corresponding industry of the leading stock.

A signal to short:

By calculating the standard variance mean ratio, the price fluctuation of leading stocks can be obtained.

$$\overline{X_n} = \frac{1}{5}\sum_{i=n-2}^{n+2} X_i$$

$$S_n = \sqrt{\frac{1}{4}\sum_{i=n-2}^{n+2}(X_i - \overline{X_n})^2}$$

$$W_n = \frac{S_n}{\overline{X_n}}$$

When the standard variance mean ratio is greater than or equal to the reference value of W_0=0.02, we believe that the price of leading stocks is relatively volatile, and we use this as a signal to short lagging stocks with strong correlation with leading stocks.

2.3.2. Portfolio Construction-Sizing & Timing of trading & Input-output Ratio

2.3.2.1. Sizing

(1) The first step is to find the leading stocks in the industry.

DOI: 10.54254/2754-1169/106/20241598

For two stocks, X and Y (assuming X is the leader and Y is the laggard),

$$\frac{\sum_{i=1}^{p}\left(\rho(i)\right)^2}{\sum_{i=1}^{p}\left(\rho(-i)\right)^2} > 1$$

Huth and Abergel defined the left side of the inequality of the formula (3-1) as the Lead lag ratio (LLR) [5].

$$\rho(i) = Corr(r_t^Y, r_{t-i}^X)$$

$$LLR = \frac{\sum_{i=1}^{p}\left(\rho(i)\right)^2}{\sum_{i=1}^{p}\left(\rho(-i)\right)^2}$$

($\{r_t^X\}$: stock X's return series, $\{r_t^Y\}$: stock Y's return series)

When **LLR**>1, I can judge stock X as the leading stock and stock Y as the lagging follower stock.

Then extend this property to multiple stocks, expand to an entire industry, win_num and $\overline{LLR}$ indicators to sort the stocks in the sector, select the top 10% of the stocks in the sector as leading stocks;

win_num:

If there are N stocks in A sector and **LLR** is calculated by combining this stock with other stocks in pairs to obtain N-1 **LLR**, then win_num is defined as the number of **LLR**>1, the number of times a stock has performed better than another stock.

$\overline{LLR}$:

So far we have learned that according to LLR & gt; 1 can determine that X is more leading than Y, but it is worth noting that the greater the LLR is greater than 1, the stronger X is leading. Therefore, set $\overline{LLR}$ to calculate the average LLR of each stock to make a more reasonable ranking of stocks in the sector.

$$\overline{LLR} = MEAN(LLR)$$

After completing the LLR index ranking of stocks in various industries, we select the top 10% stocks as leading stocks.

(2) In the second step, the study will determine the lagging stocks in our portfolio. Once the leading stocks are identified through the relationship, we select which stocks in the same industry have a leading role, that is, to select lagging stocks. According to statistician Carl Pearson, the correlation coefficient can be used to measure the degree of linear correlation between variables. Therefore, we evaluate the correlation coefficient between other stocks in the same industry and leading stocks as follows [6];

$$R^2 = \frac{\beta^2\alpha_m^2}{\beta^2\sigma_m^2 + \sigma_\varepsilon^2}$$

(σ_m^2: the variance of market returns, σ_ε^2: the variance of individual stock returns, β: The leading stocks are the explanatory variables and individual stocks are the regression coefficients of the explained variables)

Next, we choose lagged stocks greater than 0.6 to construct our investment portfolio. According to the Shanghai Stock Exchange data, it divides the stocks of listed companies into 91 sub-sectors (appendix.). For this purpose, we will find the leading and lagging stocks in each industry and construct the portfolio of each module to form a large portfolio pool.

2.3.2.2. Timing of trading

After completing the portfolio purchase, we monitor LLR daily to check whether there is still a leading effect of leading stocks on lagging stocks on that day.

DOI: 10.54254/2754-1169/106/20241598

The trading time for going long is when the leading stocks rise by the limit of the trailing stocks corresponding to our going long; If the leading stock still reach its price limit, which means there is still a leading effect, we continue to hold positions; If the leading stock fails, we sell the portfolio of the industry corresponding to the failed leading stock [7].

Short trading time is when the price of the leading stock is very volatile, that is, its standard variance mean ratio is greater than or equal to 0.02. When the standard mean variance of the leading stock is less than 0.02, we will pay back all the stocks that were previously shorted.

2.3.3. Allocation of stock investment funds

Assuming that our total investment amount is $1 million.

First, we calculate the variance of the capital flow of leading stocks in each industry, and set the variance of the industry with long position as positive, and the variance of the industry with short position as negative.

The resulting variance is then multiplied by a fixed percentage to get the amount of investment we allocate to different industries.

$$G_i = Var(fund\ \ flow) * porpotion$$

$$porpotion = \left\{0, \frac{1}{n}, \frac{2}{n} \dots\dots \frac{n-1}{n}, 1\right\}$$

$$\overline{G} = \frac{G_i}{\sum_{i=1}^{5} G_i}$$

$$normalize\ \ G_i = G_i - \overline{G}$$

Finally, in various industries, we allocate funds for each stock according to the correlation coefficient between the trailing stocks and the leading stocks selected by us. The stronger the correlation is, the more funds we will invest.

$$\omega_i = C \frac{R_i}{\sum_{i=1}^{p} R_i}$$

(ω_i: The amount of money invested in each stock, C: Total investment amount, R_i: The regression coefficient of each stock to the leading stock)

2.3.4. Trade Execution

2.3.4.1. Fee cost

In our strategy, we based on the average fee cost in China market and set the fee cost equal 0.06% of the total funds, which are taken as a commission according to the exchange regulations.

2.3.4.2. Opportunity Cost

Because we are monitoring the leading stocks to determine whether to buy or sell lagging stocks, we lose the opportunity to buy leading stocks.

DOI: 10.54254/2754-1169/106/20241598

3. Results-- (In-sample performance)

3.1. P&L Graph (Figure 1) & Summary Statistic

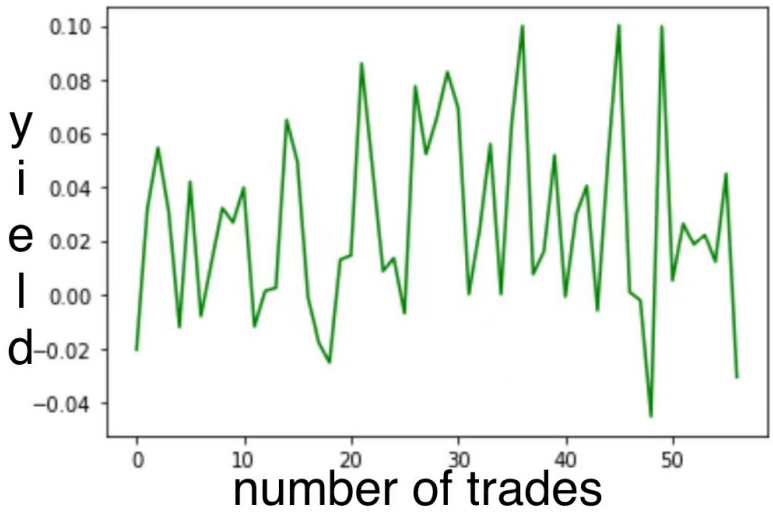

Figure 1: P&L Graph

3.2. Abnormal Analysis

In this trade, we saw an instant drop in the yield on days 45 to 48. And before that, good returns were guaranteed for some time. This phenomenon shows that the leading stock driving effect needs to be detected in time. Otherwise, in the short term, because of the purchase of many stocks, the decline of the sector will lead to a significant reduction in earnings.

4. Refinements

(1) We recommend that investors establish appropriate stop-loss points when using the strategy. This strategy has not yet considered the stop-loss point. Still, it can not be ignored that unexpected situations in the market, such as false signals or excessive confidence, make the leading stock's driving effect on other stocks in the same industry ineffective and may also bring unreasonable plummeting. Therefore, to further avoid risks, it is recommended that investors use this strategy based on setting a reasonable stop-loss point.

(2) We recommend that investors consider the LLR distribution where possible. When sorting LLR indicators in this strategy, $\overline{LLR}$ is also used as one of the indicators. Although it can make up for some defects of the indicators and reduce the possibility of stocks with the same ranking to a certain extent, since $\overline{LLR}$ is the average value is calculated? If the LLR distribution is relatively discrete and there are more extreme values,$\overline{LLR}$ may not be a good ranking indicator. Because it is largely affected by extreme values, a more detailed judgment can be made by considering the LLR distribution to make the strategy more reasonable.

5. Conclusion

5.1. Out-of-sample performance

The average daily yield is 0.0499%.

DOI: 10.54254/2754-1169/106/20241598

5.1.1. P&L Graph & Summary Statistic

Going long (Figure 2):

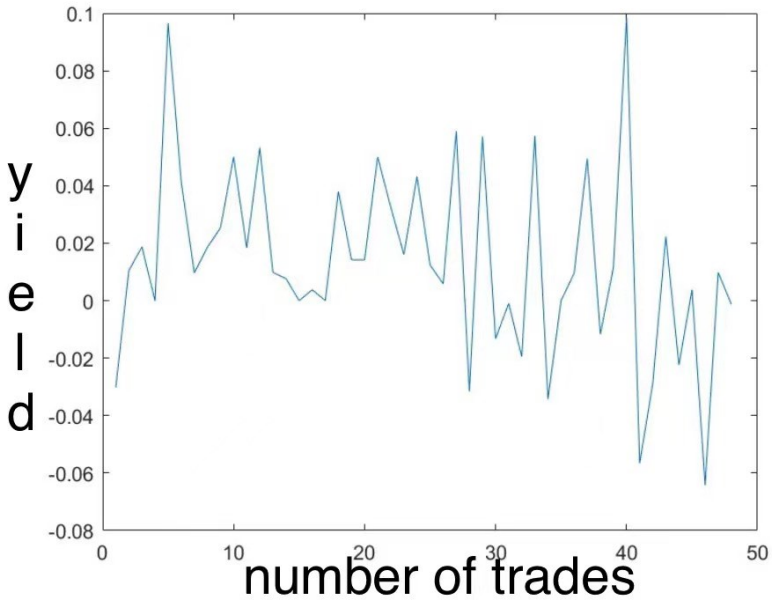

Figure 2: P&L Graph (Going long)

Short-selling (Figure 3):

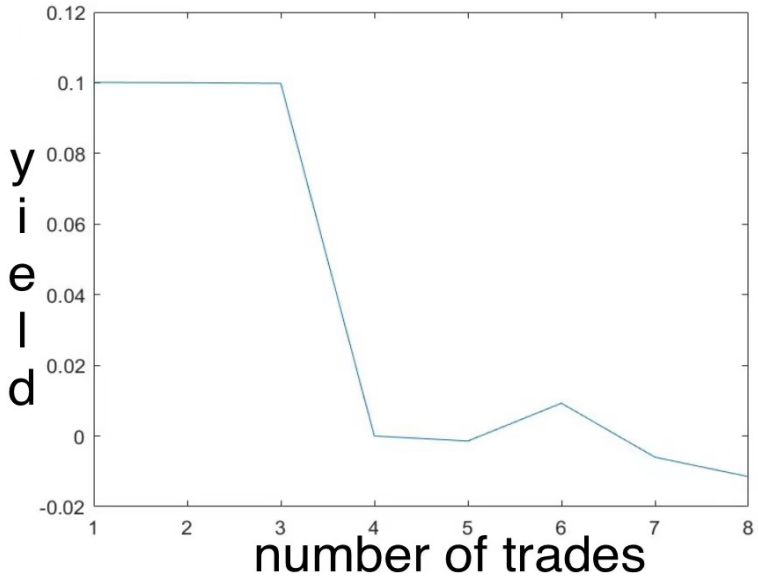

Figure 3: P&L Graph (Short-selling)

5.1.2. Abnormal Analysis

The test was taken from data starting in 2021, which was later affected by the epidemic, resulting in the Chinese stock market being heavily affected and volatile. Therefore, this strategy is greatly affected by macroeconomic conditions, and all in risk is greater.

DOI: 10.54254/2754-1169/106/20241598

5.2. Trading Recommendation

Based on our investment strategy, we recommend that investors who adopt this scheme consider as much as possible the LLR of leading stocks, and do as much testing as possible. For China's A-share market and B-share market, the daily limit is a rare trend. And for our strategy, once the daily limit signal disappears, we must sell all of them. Therefore, real-time detection of LLR data is a very important point. And based on the reason of poor information, the detection of LLR has a delay. Therefore, frequent detection of LLR can help us reduce risks and avoid excessive losses.

The second is to ensure the correctness of real-time data. For the Pearson correlation function, once it is lower than 0.6, it belongs to medium correlation and below. So once it is below 0.6, our strategy will lose its meaning and fail to work successfully.

References

[1] *The Leading Lag Relationship in the Stock Market and the Leading Stock Effect - Based on Empirical Evidence from A-shares*

[2] *ZHANG, J., Zhang, W., Li, Y., & Xu, F. (2021). The role of hedge funds in asset pricing: Evidence from China. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3767866*

[3] *Empirical Analysis of Price Index Fluctuation in China's Stock Market. Hong Tian, Bing Han*

[4] *Analysis of the market opening trend of leading stocks reaching their peak. Fu Weng*

[5] *Huth, N., & Abergel, F. (2014). High frequency lead/lag relationships — empirical facts. Journal of Empirical Finance, 26, 41–58. https://doi.org/10.1016/j.jempfin.2014.01.003*

[6] *Research on the Effect of Leading Stocks and the Effectiveness of Following the Trend Strategy*

[7] *Comparison of performance between leading and following stocks. Fu Weng*

Appendix

Matlab

Calculate the LLR and find out the leading stocks

```
% Specify folder path
folder = " Stock data folder address for a certain industry ";

% Find files and obtain a list of file names
filePattern = fullfile(folder, '*. Desired specific file format ');
fileList = dir(filePattern);

% Create a state marker vector and initialize it as an unreachable state
visitedOuter = false(size(fileList));

LLR_num = [];

% Outer loop until all files are taken out and not repeated
while any(~visitedOuter)
    % Loop through file list
    for i = 1:numel(fileList)
        if visitedOuter(i)
            continue;  % If accessed, skip the current file and continue to the next
loop
        end

        % Mark win_num as Initial value and stored in an empty vector
        win_num = 0;

        % Take out the current file and interact with it
```

DOI: 10.54254/2754-1169/106/20241598

```
        a = fullfile(folder, fileList(i).name);

        % Save the index of the selected file in the outer loop
        outerIndex = i;

        % Create a new state marker vector
        visitedInner = false(size(fileList));  % Initialize to unreachable state

        % Inner loop, interacting with all other remaining files
        for j = 1:numel(fileList)
            if visitedInner(j) || j == outerIndex
                continue; % If a file has been accessed or selected for an external loop,
skip the current file and continue to the next loop
            end

            % Interact with other files (write code logic as needed)
            b = fullfile(folder, fileList(j).name);

            % Calculate the daily return sequence of the first stock
            Stock_Name1=xlsread(a,' The range of data you want to read ');
            open_price1=Stock_Name1(:,1);
            close_price1=Stock_Name1(:,4);
            Re_Stock_Name1 = (close_price1-open_price1) ./ close_price1;

            % Calculate the daily return sequence of the second stock
            Stock_Name2=xlsread(b,' The range of data you want to read ');
            open_price2=Stock_Name2(:,1);
            close_price2=Stock_Name2(:,4);
            Re_Stock_Name2= (close_price2-open_price2) ./ close_price2;

            index = 250;
            P = 100;
            seq_Stock_X = Re_Stock_Name1(index:index+P);
            seq_Stock_X1 = transpose(seq_Stock_X);

            % Define the range of lag length
            lag_range = 1:10;  % N is the number of days behind

            sum_leading_corr = 0;
            sum_lagging_corr = 0;

            % Range of loop traversal lag length
            for lag = lag_range
                % Advance sequence seq1
                seq_Stock_Y_leading = Re_Stock_Name2(index+lag:index+lag+P);

                % Lag sequence seq2
                seq_Stock_Y_lagging = Re_Stock_Name2(index-lag-P:index-lag);

                % Conduct further analysis, such as calculating correlation
                correlation1 = corr(seq_Stock_Y_leading, seq_Stock_X);
                correlation2 = corr(seq_Stock_Y_lagging, seq_Stock_X);
                sum_leading_corr = sum_leading_corr + correlation1;
                sum_lagging_corr = sum_lagging_corr + correlation2;

            end
```

DOI: 10.54254/2754-1169/106/20241598

```
            % Calculate the number of LLR>1
            alpha = sum_leading_corr ./ sum_lagging_corr;
            if alpha > 1
                win_num = win_num + 1;

            end

            % Set inner loop marker points
            visitedInner(j) = true;
        end

        disp(win_num)
        LLR_num = [LLR_num,win_num];

        % Mark the current file as accessed
        visitedOuter(outerIndex) = true;
    end
end

% Identify the top 10% of LLR
n = ceil(0.1 * length(LLR_num));
sortedData = sort(LLR_num, 'descend');
largestValues = sortedData(1:n);

fileNames_leading = cell(size(LLR_num));

for i = 1:n
    idx = find(LLR_num == largestValues(i));
    for j = 1:length(idx)
        fileNames_leading{idx(j)} = sprintf('Leading Stock', idx(j));
    end
end

textToAdd = 'Not Leading Stock'

for p = 1:numel(fileNames_leading)
    if isempty(fileNames_leading{p})
        % If the element is empty, add text to that position
        fileNames_leading{p} = textToAdd;
    end
end

disp(fileNames_leading);
```

Find out the following stocks

```
% Enter data for the first stock
a=" Stock data address of leading stocks ";

% Calculate the daily return sequence of the first stock
Stock_Name1=xlsread(a,' The range of data you want to read ');
open_price1=Stock_Name1(:,1);
close_price1=Stock_Name1(:,4);
Re_Stock_Name1= (close_price1-open_price1) ./ close_price1;
```

DOI: 10.54254/2754-1169/106/20241598

```
% Set folder path and conditions
folderPath = " Folder addresses for other stock data in the same industry ";


% Get all Excel files in the folder
files = dir(fullfile(folderPath, '*. Desired file format '));

% Create an empty matrix to store file names that meet the criteria
fileNames_following = [];
r_sqr_total = []
invest_amount = ' The investment amount calculated in the short strategy '

% Traverse each file
for i = 1:numel(files)
    filePath = fullfile(folderPath, files(i).name);

    % Reading data from an Excel file
    Stock_Name2=xlsread(filePath,' The range of data you want to read ');
    open_price2=Stock_Name2(:,1);
    close_price2=Stock_Name2(:,4);
    Re_Stock_Name2= (close_price2-open_price2) ./ close_price2;
    % Building a linear regression model
    model = fitlm(Re_Stock_Name1, Re_Stock_Name2);
    % View regression results
    disp(model);
    % Extracting regression coefficients
    coefficients = model.Coefficients;
    % Extract R-squared value
    rsquared = model.Rsquared.Ordinary;
    % Calculate the variance of two returns
    variance1=var(Re_Stock_Name1);
    variance2=var(Re_Stock_Name2);

    % Assuming there is a table named T, you want to convert the entire table into a
matrix
    T = coefficients; % Enter your data
    matrixData = table2array(T); % Convert a table to a matrix using the table2array
function
    correlance=matrixData(1,2);
    R_square=(correlance^2)*variance1 ./ ((correlance^2)+variance2);

    % Enter your criteria
    condition = (R_square>0.6);

    % Determine whether the data meets the conditions
    if (condition)
        % Add file names that meet the criteria to the matrix
        fileNames_following = [fileNames_following; files(i).name];
        r_sqr_total = [r_sqr_total,R_square]
    end
end

for l = 1:length(r_sqr_total)
    r_sqr_total(l) = r_sqr_total(l) / sum(r_sqr_total)
end

invest_amount_each = invest_amount .* r_sqr_total
```

DOI: 10.54254/2754-1169/106/20241598

```
output = horzcat(fileNames_following,invest_amount_each)
```

Find out the amount of money we need to do short or do long on one specific stock

```
% Specify folder path
folder_fund_flow = " Address of the data folder for leading stocks ";

% Find files and obtain a list of file names
filePattern_fund_flow = fullfile(folder_fund_flow, '*. Form of document ');
fileList_fund_flow = dir(filePattern_fund_flow);

all_slope = [];
all_var = [];
short_long_name = [];

%  Calculate the r power of all stocks
for i = 1:numel(fileList_fund_flow)
    filePath_fundflow = fullfile(folder_fund_flow, fileList_fund_flow(i).name);
    fund_flow = xlread(filePath_fundflow,' Hourly flow of funds in the first three days
');
    closing_price_W1 = xlread(filePath_fundflow,' Closing price of the first five days
');
    closing_price_today = xlread(filePath_fundflow,' Today's closing price ');
    opening_price_today = xlread(filePath_fundflow,' Today's opening price ');

    W0 = 0.2
    W1 = var(closing_price_W1) / mean(closing_price_W1);

    if W1 > W0
        short_long_name = [short_long_name,fileList_fund_flow(i).name];

        var_fundflow = -var(fund_flow);
        all_var = [all_var,var_fundflow];
    end
    if (closing_price_today-opening_price_today) / closing_price_today > 'different
price limit in different plate of stocks'
        short_long_name = [short_long_name,fileList_fund_flow(i).name];
        var_fundflow = var(fund_flow);
        all_var = [all_var,var_fundflow];
    end
end

for k = 1:length(all_var)
    all_var(k) = all_var(k)-(all_var(k) / sum(all_var));
end

normalized_var = all_var;

% Sort var
sorted_all_var = sort(all_var);

n = length(all_var)-1;
porpotion = [];

for q = 0:n
```

DOI: 10.54254/2754-1169/106/20241598

```
    porpotion_ = q/n;
    porpotion = [porpotion,porpotion_];
end

G = propotion .* sorted_all_var;

normalized_G = [];
for j = 1:length(G)
    beta = G(j)-(sum(G) / length(G));
    normalized_G = [normalized_G,beta];
end

fund_amount = 1000000;
invest_amount = fund_amount * normalized_G;

[~, original_order] = sort(normalized_var);
sorted_G = normalized_G(original_order);

result = horzcat(short_long_name,sorted_G);
```

Back test

Python

Code 1:going long

```
import os
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from matplotlib.font_manager import FontProperties
import tkinter as tk
from tkinter import filedialog, Text, ttk

def load_data(folder_path):
all_data = pd.DataFrame()
for filename in os.listdir(folder_path):
if filename.endswith('.csv'):
data = pd.read_csv(os.path.join(folder_path, filename), encoding='GBK', header=1)
all_data = pd.concat([all_data, data])
all_data.reset_index(drop=True, inplace=True)
all_data[' Daily return '] = all_data.groupby(' Stock code ')[' closing price '].pct_change()
return all_data
```

```
def find_limit_up_stocks(data):
```

DOI: 10.54254/2754-1169/106/20241598

```
data[' Increase amount'] = (data[' closing price '] - data[' Previous closing price ']) / data[' Previous closing price ']
limit_up_stocks = data[data[' Increase amount '] >= 0.1][[' Stock code ', ' Transaction Date ']]
return limit_up_stocks
```

```
def find_correlated_stocks(data, lead_stock_code):
correlations = data.groupby(' Stock code ')[' Daily return '].corr(data[data[' Stock code '] == lead_stock_code][' Daily return '])
correlated_stocks = correlations[correlations > 0.6].index.tolist()
return correlated_stocks
```

```
def backtest(data, lead_stock_code, correlated_stocks, folder_path):
limit_up_dates = data[(data[' Stock code '] == lead_stock_code) & (data[' Increase amount '] >= 0.1)][' Transaction Date ']
returns = []
for date in limit_up_dates:
next_date = data[data[' Transaction Date '] > date][' Transaction Date '].min()
for stock_code in correlated_stocks:
limit_up_close_price = data[(data[' Transaction Date '] == date) & (data[' Stock code '] == stock_code)][' closing price '].values
next_open_price = data[(data[' Transaction Date '] == next_date) & (data[' Stock code '] == stock_code)][' opening price '].values
if len(limit_up_close_price) > 0 and len(next_open_price) > 0:
returns.append((next_open_price[0] - limit_up_close_price[0]) / limit_up_close_price[0])
```

```
pd.DataFrame(returns, columns=['return rate']).to_csv(os.path.join(folder_path, 'returns.csv'), index=False, encoding='utf-8-sig')
plt.plot(returns, color='g')
plt.title(' PAL', fontproperties=FontProperties(fname=r"c:\windows\fonts\simsun.ttc", size=14))
plt.xlabel('number of trade', fontproperties=FontProperties(fname=r"c:\windows\fonts\simsun.ttc", size=14))
plt.ylabel(' Yield ', fontproperties=FontProperties(fname=r"c:\windows\fonts\simsun.ttc", size=14))
plt.show()
daily_return = np.mean(returns)
print(f' daily return: {daily_return * 100-0.0006:.2f}%')
```

DOI: 10.54254/2754-1169/106/20241598

```
root = tk.Tk()
root.title("Backtesting model ")
root.geometry("300x200")

folder_path_label = tk.Label(root, text="")
folder_path_label.pack()

result_text = Text(root)
result_text.pack()

progress = ttk.Progressbar(root, length=100, mode='determinate')
progress.pack()

def select_folder():
folder_path = filedialog.askdirectory()
folder_path_label.config(text=folder_path)

def run_program():
folder_path = folder_path_label.cget("text")
data = load_data(folder_path)
limit_up_stocks = find_limit_up_stocks(data)
progress['maximum'] = len(limit_up_stocks)
for i, lead_stock_code in enumerate(limit_up_stocks[' Stock code ']):
correlated_stocks = find_correlated_stocks(data, lead_stock_code)
backtest(data, lead_stock_code, correlated_stocks, folder_path)
progress['value'] = i + 1
root.update_idletasks()
result_text.insert(tk.END, " Program runs ended!")

select_folder_button = tk.Button(root, text=" Select folder ", command=select_folder)
select_folder_button.pack()

run_program_button = tk.Button(root, text="run the program ", command=run_program)
run_program_button.pack()

root.mainloop()
```

Code 2:short-selling

DOI: 10.54254/2754-1169/106/20241598

```
import pandas as pd

def find_next_day_data_change_ratio(excel_file):
# Read the Excel file, assuming the data is in the first column (column A)
df = pd.read_excel(excel_file, usecols=[0], header=None)

# Calculate the rate of change per day compared to the previous day
df['Percentage Change'] = df.diff() / df.shift()

# Find an index for dates that have fallen by more than 5%
drop_threshold = -0.05
drop_indices = df[df['Percentage Change'] < drop_threshold].index

# Calculate the rate of change between the data for the next day and the data for that day
when the decline exceeds 5%
next_day_change_ratio = []
for idx in drop_indices:
if idx + 1 < len(df):
today_data = df.iloc[idx, 0]
next_day_data = df.iloc[idx + 1, 0]
change_ratio = (next_day_data - today_data) / today_data
next_day_change_ratio.append(change_ratio)

return next_day_change_ratio

# Replace with your Excel file path
excel_file_path = "Your fill adress"
result = find_next_day_data_change_ratio(excel_file_path)
print(result)
```