The novel statistical speculation strategy for stock indexes：Empirical evidence of S&P500, HSI and CSI 500

HUANG, Jian   Student ID: 20665844

HKUST MAFS 5210 project

May,18th, 2020

Abstract

This research proposed a really profitable novel statistical speculation strategy called Differential Trend Motion strategy. This paper investigated this strategy performance of three stock indexes: S&P500, HSI and CSI500. By using the training data to select the optimal strategy and moving average order based on maximum profit, excess profit and information ratio, the return and sharp ratio of strategy outperformed benchmark (buy and hold strategy) in testing set. This research not only showed the robust of strategy but also refuted the weak efficient market hypothesis by proving the long-term data may be valuable in construction of target or reference price (called standard line in this paper) of future transactions.

1. Introduction

In the world of statistic, two things are important, one of which is model, it can refer to machine learning model or mathematical model, another one is features, the valuable indicator generated and processed from raw data. Sometimes, model represent a kind of logic, which bring better results. For example, when the dependent variable has non-linear correlation with independent variables, deep learning models are better than any linear regression model. When the statistical model cannot improve the outcomes, research should consider the inputting features and the intuition behind the data preprocessing approaches. In this research, I focus on the data processing, mathematical logic construction and the basic application of statistical tools rather than any fancy machine learning models.

With respects to trading, there are two types of strategies. One is called passive investment, which choose the valuable share portfolio and update the weight each term. For instance, Fama-French model, Fama-MacBeth regression, Markowitz model, tracking error theory (Barra model), and risk parity theory are classical models. Alternative approach for trading is active investment, it includes technical analysis like Dow theory, technical index analysis like MACD, Bollinger Band, and statistical arbitrage. The main strategy of this research is close to statistical arbitrage or speculation. I name the novel strategy as **Differential Trend Motion.**

2. Literature review

2.1 Efficient market hypothesis and Fundamental analysis

When Fama(1993) did the research of factor model. They proposed a hypothesis stating that the market is efficient and all the information have been reflected in the current price. The weak form of efficient market hypothesis stated that the historical information has been reflected in the price. Thereafter, the history data cannot predict the future price and are useless in trading. Many people focus on fundamental analysis including the analysis of company, industry and macroeconomic environment. However, this approach depends on the believe of investor and different people may deduce different conclusion from same kind of information.

2.2 Technical analysis

In 1970s, technical analysis (Appel, 2005) is popular for investor. People draw the trend line and conduct the technical indexes like MACD (Chong& Ng, 2008), BOLL, KDJ and they can gain a lot of money. However, since many individuals know and apply these skills in the market, many technical indexes cannot work anymore. Besides, another limitation for technical analysis is that investor may know the time to LONG or SHORT the underlying assets, yet they do not know whether it is the optimal time to do the transaction.

2.3 Quantitative analysis and statistical arbitrage

With the development of computer science and trend of big data analysis, more and more researchers begin to applied multiple machine learning methods including regression, SVM (Wang, Liu& Wang, 2013), KNN(Liu &Zhou, 2010), K-mean, LSTM (Baek& Kim, 2018) in forecasting the level, direction or volatility of future price.

This approach called P-Quant (P is prediction). However, the success of prediction does not represent the success of making profit. There is a long way from prediction to a successful strategy. Another quantitative approach is named Q-Quant (Q is quantitative). This approach introduces the random walk into the pricing of assets. Random does not mean that the price can not be forecast at all. The volatility and trend also can be described by Geometric Brownian motion (Mikosch, 2012) and other process. By applying Ito integral, Black-Scholes model (Ursone, 2015) is used in pricing the derivatives. However, this approach assumes that there is no arbitrary space in the market, thus the price will regress to value sharply thus no one can make sure the constant return in the long term. Another issue is that in reality, the market is dynamically changing and investors do not know how long the price will back to the level of value. Therefore, this research proposed a strategy called **Differential Trend Motion** combining the statistical analysis of P-Quant and logical framework of Q-quant.

3. Methodology

3.1 Data selection

There are three stock indexes which are S&P500(Standard& Poor's 500 index), HSI (Hang Seng index) and CSI500(China stock 500 index) in this research. These stock index represent different stock market, which are the market of American, Hong Kong and Republic of China. S&P500 is a stock market index that measures the stock performance of 500 large companies listed on stock exchanges in the United States. HSI is a free-float-adjusted market-capitalization-weighted stock-market index in Hong Kong. CSI500 also called CSI small cap 500 index, which involves 500 small companies listed in China market.

The daily transaction data including open price, high price, low price and close price of S&P 500 and HSI are collected from Yahoo Finance. The daily data of CSI500 is directly gained from Tushare database (A package in Python which provide the China stock market data).

The timeline of S&P 500 and HSI is from January 3rd, 2000 to April 17th, 2020 while the data of CSI 500 is from January 4th, 2005 to May 15th, 2020 since CSI is innovated in 2005. The first 70% of stock index data are training data and the rest of

30% are testing data. For example, the training data of S&P 500 is from January 3rd, 2000 to March 18th, 2014.

3.2 The intuition behind the strategy

The success of the strategy in this research depends on judgement, successful transaction rate, and the profit gain whenever transaction.

3.2.1 Judgement

Judgement in strategy mean that we should judge the close price is higher than a price or lower than a price. Based on the statistic, if we judge the close price will over than a specific price, we should LONG at that price and SHORT at close price. In opposite, SHORT at specific price and LONG at close price.

This research will focus the study of change in moving average of differential price, called **Differential Trend Motion**, which presented by:

$$MA_{t+1}(n) - MA_t(n)$$

If we assume the change(difference) of price follow Brownian motion (Yor, 2001) which follow $N(0, \sigma)$, it is easy to prove that:

(The proof please refer to **Appendix A)**

if $MA_t(n) > 0$

$$P[MA_{t+1}(n) - MA_t(n) < 0] = 0.75$$

which means

$$P[MA_{t+1}(n) - MA_t(n) < 0 | \, MA_t(n) > 0] = 0.75$$

It is equivalent to:

$$P\big((P_{t+1} - P_{t+1-n}) - (P_t - P_{t-n}) < 0 | (P_t - P_{t-n}) > 0\big) = 0.75$$

if $MA_t(n) < 0$

$$P[MA_2(n) - MA_1(n) > 0] = 0.75$$

$$P[MA_{t+1}(n) - MA_t(n) > 0 | \, MA_t(n) < 0] = 0.75$$

$$P\big((P_{t+1} - P_{t+1-n}) - (P_t - P_{t-n}) < 0 | (P_t - P_{t-n}) > 0\big) = 0.75$$

In real situation, the change of stock price may not follow normal distribution like fat tail effect, yet with the other distributions, the conclusion still normally can gain a large probability over 0.6.

Here I want to change P_t into C_t which represent the close price of stock.

when $C_t - C_{t-n} > 0$, it is a high probability that

$$(C_{t+1} - C_{t+1-n}) - (C_t - C_{t-n}) < 0$$

$$C_{t+1} < C_{t+1-n} + C_t - C_{t-n}$$

Let us define the standard line at time $t = C_{t-n+1} + C_t - C_{t-n}$, which denotes as l_t

It means the next close price should lower than l_t with a great probability, here we should SHORT at standard line l_t and LONG at close price.

In opposite, when $C_t - C_{t-n} < 0$,

$C_{t+1} > C_{t+1-n} + C_t - C_{t-n}$ in great probability

The strategy should LONG at standard line l_t and SHORT at close price

However,

whenever $C_t - C_{t-n} > 0$ and we know $C_{t+1} < l_t$ in great probability, and if open price of next day O_{t+1} is over than l_t , namely $O_{t+1} > l_t$, we should SHORT at open price rather than wait for the price cross the standard line. After that, LONG at close price.

Similarly, if $C_t - C_{t-n} < 0$, and we know $C_{t+1} > l_t$ in great probability.

If $O_{t+1} < l_t$, we should LONG at open price and SHORT at close price.

3.2.2 Successful transaction rate

In part of 3.2.1, we know that there are two types of transaction. One is LONG or SHORT at standard line, l_t. The other is doing the transaction at open price.

However, different stock index may have different successful transaction rate. For example, in a specific transaction date for one stock index, it may do the judgement right, yet it cannot do the transaction, because the open price cannot satisfy the condition and the market price also cannot cross the standard line l_t. In opposite, if the we make a wrong judgement, unfortunately, the conditions always can be satisfied and transaction always can be done. It means we frequently loose the money.

Therefore, the success rate of judgement does not represent we finally can do the transaction and earn the profit. The higher profit also depends on whether the market price can satisfy our setting conditions.

3.2.3 Profit gain whenever transaction

Furthermore, if we have the high success judgement rate and high success transaction rate, the strategy may still not earn the profit, since it depends on the conditional profit.

For example, for one strategy original should LONG at standard line and SHORT at close price. Whenever judgement is right, the strategy always earns little money. Nevertheless, whenever judgement is wrong, the strategy always losses huge amount of money. Then, we should totally change our strategy, we should SHORT at standard line and LONG at close price in order to maximum our profit.

The standard form is presented like this:

	Judgement	
	Right	Wrong
Have transaction	Profit	Loss
No transaction	No profit	No profit

Table. 1 The relationship between judgement, successful transaction, and the profit

If $|\text{Profit}| > |\text{Loss}|$, keep the original trading strategy.

If $|\text{Profit}| < |\text{Loss}|$, change the LONG&SHORT direction of trading strategy.

In summary, the total profit gained from strategy depends on judgement, successful transaction rate, and the profit gain whenever transaction. For next part, a mathematical process will achieve and balance these three factors and try to maximum the profit.

3.3 Strategy decision process

3.3.1 Assumptions

(1). the market is not weak form of efficient market and model still can use historical data to predict the future price interval.

(2). the model is robust only when the market style is stable, the long-term trend of market is stable.

(3). the return of market indexes is same as the return of the future of indexes, which means the stock indexes can be traded via future or stock.

(4). the market allow long and short with T+0 mode, which mean we can long and short the market index in one day.

(5). there is no transaction cost considered in this model.

3.3.2 Strategy selection process

$$C_t: close\ price\ at\ time\ t$$

$$O_t: open\ price\ at\ time\ t$$

$$H_t: high\ price\ at\ time\ t$$

$$L_t: low\ price\ at\ time\ t$$

$$l_t: standard\ line\ at\ time\ t = C_{t-n+1} + C_t - C_{t-n}$$

$$\boldsymbol{I_s}: indicator\ function,$$

$$\boldsymbol{I_s} = \begin{cases} 1 & if\ situation\ \boldsymbol{s}\ is\ satistfied \\ 0 & otherwise \end{cases}$$

Suppose:

$$t = i\ which\ is\ the\ time\ i$$

$$S_1 = O_{t+1} - l_t = O_{i+1} - l_i$$

$$S_2 = C_t - C_{t-n} = C_i - C_{i-n}$$

$$S_3 = C_{t+1} - O_{t+1} = C_{i+1} - O_{i+1}$$

$$S_4 = C_{t+1} - l_t = C_{i+1} - l_i$$

$$S_5: is\ a\ situation\ when\ the\ price\ cross\ the\ standard\ line\ \ L_{i+1} < l_i < H_{i+1}$$

Firstly, we define the earning function as follow:

$$E_1 = \sum_{i=1}^{T} \left[(S_3) * \boldsymbol{I}_{\boldsymbol{S_1<0},\, \boldsymbol{S_2<0},\, \boldsymbol{S_3>0}} \right] \qquad (1)$$

$$E_2 = -\sum_{i=1}^{T} \left[(S_3) * \boldsymbol{I}_{\boldsymbol{S_1<0},\, \boldsymbol{S_2<0},\, \boldsymbol{S_3<0}} \right] \qquad (2)$$

$$G_1 = \max(E_1, E_2) - \min(E_1, E_2) \qquad (3)$$

$\forall\, n,$ if $\max(E_1, E_2) = E_1$, which means that the strategy should LONG at open price and SHORT at the close price when $\boldsymbol{S_1} < \boldsymbol{0},\ \boldsymbol{S_2} < \boldsymbol{0}$ otherwise when $\max(E_1, E_2) = E_2$ SHORT at open price and LONG at the close price.

$$E_3 = \sum_{i=1}^{T} \left[(S_4) * \boldsymbol{I}_{\ \boldsymbol{S_1>0},\, \boldsymbol{S_2<0},\, \boldsymbol{S_4>0},\, \boldsymbol{S_5}} \ \right] \qquad (4)$$

$$E_4 = -\sum_{i=1}^{T} \left[(S_4) * \boldsymbol{I}_{\boldsymbol{S_1>0},\, \boldsymbol{S_2<0},\, \boldsymbol{S_4<0},\, \boldsymbol{S_5}} \right] \qquad (5)$$

$$G_2 = \max(E_3, E_4) - \min(E_3, E_4) \qquad (6)$$

$\forall\, n,$ if $\max(E_3, E_4) = E_3$, which means that the strategy should LONG at standard line and SHORT at the close price when $\boldsymbol{S_1} > \boldsymbol{0},\ \boldsymbol{S_2} < \boldsymbol{0}$ and $\boldsymbol{S_5}$ are satisfied otherwise when $\max(E_3, E_4) = E_4$ SHORT at standard line and LONG at the close price.

$$E_5 = \sum_{i=1}^{T} \left[(-S_3) * \boldsymbol{I}_{\boldsymbol{S_1>0},\, \boldsymbol{S_2>0},\, \boldsymbol{S_3<0}} \right] \qquad (7)$$

$$E_6 = -\sum_{i=1}^{T} \left[(-S_3) * \boldsymbol{I}_{\boldsymbol{S_1>0},\, \boldsymbol{S_2>0},\, \boldsymbol{S_3>0}} \right] \qquad (8)$$

$$G_3 = \max(E_5, E_6) - \min(E_5, E_6) \qquad (9)$$

$\forall\, n,$ if $\max(E_5, E_6) = E_5$, which means that the strategy should SHORT at open price and LONG at the close price when $\boldsymbol{S_1} > \boldsymbol{0},\ \boldsymbol{S_2} > \boldsymbol{0}$ otherwise when $\max(E_5, E_6) = E_6$LONG at open price and SHORT at the close price.

$$E_7 = \sum_{i=1}^{T} \left[(-S_4) * \boldsymbol{I}_{\boldsymbol{S_1<0},\, \boldsymbol{S_2>0},\, \boldsymbol{S_4<0},\, \boldsymbol{S_5}} \right] \tag{10}$$

$$E_8 = -\sum_{i=1}^{T} \left[(-S_4) * \boldsymbol{I}_{\boldsymbol{S_1<0},\, \boldsymbol{S_2>0},\, \boldsymbol{S_4>0},\, \boldsymbol{S_5}} \right] \tag{11}$$

$$G_4 = \max(E_7, E_8) - \min(E_7, E_8) \tag{12}$$

$\forall\, n$, if $\max(E_7, E_8) = E_7$, which means that the strategy should SHORT at standard line and LONG at the close price when $\boldsymbol{S_1 < 0},\ \boldsymbol{S_2 > 0}$ and $\boldsymbol{S_5}$ are satisfied otherwise when $\max(E_7, E_8) = E_8$ LONG at standard line and SHORT at the close price.

Then, we define the earning function with all the signals of S_1 are opposite, which mean when the $S_1 > 0$ from E_1 to E_8, change it to $S_1 < 0$ and also change $S_1 < 0$ into $S_1 > 0$. Then we gain:

$$E_9 = \sum_{i=1}^{T} \left[(S_3) * \boldsymbol{I}_{\boldsymbol{S_1>0},\, \boldsymbol{S_2<0},\, \boldsymbol{S_3>0}} \right] \tag{13}$$

$$E_{10} = -\sum_{i=1}^{T} \left[(S_3) * \boldsymbol{I}_{\boldsymbol{S_1>0},\, \boldsymbol{S_2<0},\, \boldsymbol{S_3<0}} \right] \tag{14}$$

$$G_5 = \max(E_9, E_{10}) - \min(E_9, E_{10}) \tag{15}$$

$\forall\, n$, if $\max(E_9, E_{10}) = E_9$, which means that the strategy should LONG at open price and SHORT at the close price when $\boldsymbol{S_1 > 0},\ \boldsymbol{S_2 < 0}$ otherwise when $\max(E_9, E_{10}) = E_{10}$ SHORT at open price and LONG at the close price.

$$E_{11} = \sum_{i=1}^{T} \left[(S_4) * \boldsymbol{I}_{\boldsymbol{S_1<0},\, \boldsymbol{S_2<0},\, \boldsymbol{S_4>0},\, \boldsymbol{S_5}} \right] \tag{16}$$

$$E_{12} = -\sum_{i=1}^{T} \left[(S_4) * \boldsymbol{I}_{\boldsymbol{S_1<0},\, \boldsymbol{S_2<0},\, \boldsymbol{S_4<0},\, \boldsymbol{S_5}} \right] \tag{17}$$

$$G_6 = \max(E_{11}, E_{12}) - \min(E_{11}, E_{12}) \tag{18}$$

$\forall\, n$, if $\max(E_{11}, E_{12}) = E_{11}$, which means that the strategy should LONG at standard line and SHORT at the close price when $\boldsymbol{S_1 < 0},\ \boldsymbol{S_2 < 0}$ and $\boldsymbol{S_5}$ are satisfied otherwise when $\max(E_{11}, E_{12}) = E_{12}$ SHORT at standard line and LONG at the close price.

$$E_{13} = \sum_{i=1}^{T} \left[(-S_3) * \boldsymbol{I}_{\boldsymbol{S_1<0},\, \boldsymbol{S_2>0},\, \boldsymbol{S_3<0}} \right] \qquad (19)$$

$$E_{14} = -\sum_{i=1}^{T} \left[(-S_3) * \boldsymbol{I}_{\boldsymbol{S_1<0},\, \boldsymbol{S_2>0},\, \boldsymbol{S_3>0}} \right] \qquad (20)$$

$$G_7 = \max(E_{13}, E_{14}) - \min(E_{13}, E_{14}) \qquad (21)$$

$\forall\, n,$ if $\max(E_{13}, E_{14}) = E_{13}$, which means that the strategy should SHORT at open price and LONG at the close price when $\boldsymbol{S_1 < 0},\ \boldsymbol{S_2 > 0}$ otherwise when $\max(E_{13}, E_{14}) = E_{14}$ LONG at open price and SHORT at the close price.

$$E_{15} = \sum_{i=1}^{T} \left[(-S_4) * \boldsymbol{I}_{\boldsymbol{S_1>0},\, \boldsymbol{S_2>0},\, \boldsymbol{S_4<0},\, \boldsymbol{S_5}} \right] \qquad (22)$$

$$E_{16} = -\sum_{i=1}^{T} \left[(-S_4) * \boldsymbol{I}_{\boldsymbol{S_1>0},\, \boldsymbol{S_2>0},\, \boldsymbol{S_4>0},\, \boldsymbol{S_5}} \right] \qquad (23)$$

$$G_8 = \max(E_{15}, E_{16}) - \min(E_{15}, E_{16}) \qquad (24)$$

$\forall\, n,$ if $\max(E_{15}, E_{16}) = E_{15}$, which means that the strategy should SHORT at standard line and LONG at the close price when $\boldsymbol{S_1 > 0}$, $\boldsymbol{S_2 > 0}$ and $\boldsymbol{S_5}$ are satisfied otherwise when $\max(E_{15}, E_{16}) = E_{16}$ LONG at standard line and SHORT at the close price.

After that, we can construct the objective function:

$$Profit(n) = \max(G_1 + G_2, G_5 + G_6) + \max(G_3 + G_4, G_7 + G_8) \qquad (25)$$

3.4 Moving average order selection

3.4.1 Order selection based on maximum profit

With respect to different moving average order, the optimal strategy selects the highest total profit among all the moving average orders.

$$\max Profit(n)$$

in which n is the moving average order

3.4.2 Order selection based on maximum excess profit

Excess profit is defined by:

$$\begin{aligned}&\text{Excess profit}\\&= \text{Expected Daily Profit} - \text{Expected daily profit gained by stock index}\end{aligned}$$

Expected daily profit gained by stock index is calculated by $C_t - C_{t-1}$, which is daily profit of buying and holding the benchmark (stock index).

The dataset of calculating total profit of each moving average order is unbalanced, because with the growth of order, the samples for calculating expected daily profit also decrease and may focus on the performance of recent data (short-term performance). If we assume the excess profit performance is constant between long-term performance and short-term performance. Excess profit is a better selection criterion, which can be expressed by:

$$\max \ \text{Excess profit}(n)$$

(3) Order selection based on information ratio

For the third selection, it considers both excess profit and risk, thus we use the information ratio, which defined by:

$$\text{information ratio} = \frac{\text{E(profit return)} - \text{E(benchmark return)}}{\text{Std(profit return} - \text{benchmark return)}} \tag{26}$$

with the condition that

$$\text{Excess profit}(n) \geq R$$

In which R is a constant represent the minimum daily excess profit that investor can accept.

$$\max \text{ information ratio}$$

3.5 Measurement

The first measurement is the total profit in testing period by transaction one unit of stock index. This total profit includes the profit from LONG and SHORT with transaction cost and leverage. Also, total profit is the simple sum of daily profit without any compound interest and reinvestment. The second measurement is the compound interest rate which also have not considered the effect of transaction cost and leverage. The third measurement is the annual return rate which can be calculated by total profit and compound interest rate. Moreover, the risk measurement for strategy is the Sharp ratio and daily 5% VAR.

4. Results

4.1 Optimal strategy for each stock index

S&P500
$\text{line} = C_{t-n+1} + C_t - C_{t-n}$
(1) If $C_t - C_{t-n} < 0$ and $\text{line} - \text{open} < 0$:
LONG at open price and SHORT at close price
(2) If $C_t - C_{t-n} < 0, \text{line} - \text{open} > 0$ and $\text{low price} < \text{line} < \text{high price}$ (when price cross the standard line):
LONG at standard line and SHORT at close price
(3) If $C_t - C_{t-n} > 0$ and $\text{open} - \text{line} > 0$:
LONG at open price and SHORT at close price
(4) If $C_t - C_{t-n} > 0, \text{open} - \text{line} < 0$ and $\text{low price} < \text{line} < \text{high price}$
LONG at standard line and SHORT at close price

HSI
line $= C_{t-n+1} + C_t - C_{t-n}$
(1) If $C_t - C_{t-n} < 0$ and open $-$ line < 0:
SHORT at open price and LONG at close price
(2) If $C_t - C_{t-n} < 0$, open $-$ line > 0 and low price $<$ line $<$ high price (when price cross the standard line):
SHORT at standard line and LONG at close price
(3) If $C_t - C_{t-n} > 0$ and line $-$ open > 0:
SHORT at open price and LONG at close price
(4) If $C_t - C_{t-n} > 0$, line $-$ open < 0 and low price $<$ line $<$ high price
SHORT at standard line and LONG at close price
CSI 500
line $= C_{t-n+1} + C_t - C_{t-n}$
(1) If $C_t - C_{t-n} < 0$ and line $-$ open < 0:
LONG at open price and SHORT at close price
(2) If $C_t - C_{t-n} < 0$, line $-$ open > 0 and low price $<$ line $<$ high price (when price cross the standard line):
LONG at standard line and SHORT at close price
(3) If $C_t - C_{t-n} > 0$ and open $-$ line > 0:
LONG at open price and SHORT at close price
(4) If $C_t - C_{t-n} > 0$, open $-$ line < 0 and low price $<$ line $<$ high price
LONG at standard line and SHORT at close price

Table 2. Optimal strategy for each stock index

4.2 The relationships between moving average order, profit and excess profit.

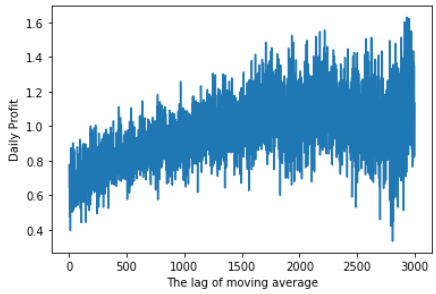
<table>
<tr><th></th><th>moving average order& Profit</th><th>moving average order& Excess Profit</th></tr>
<tr><td>S&P 500</td><td>

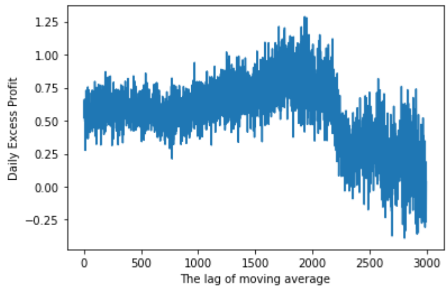
Figure 1</td><td>

Figure 2</td></tr>
</table>

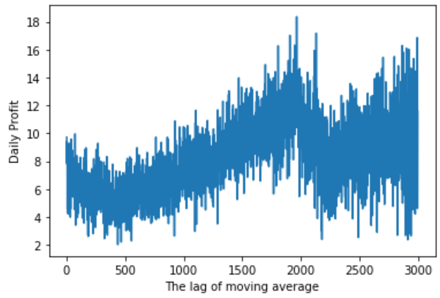
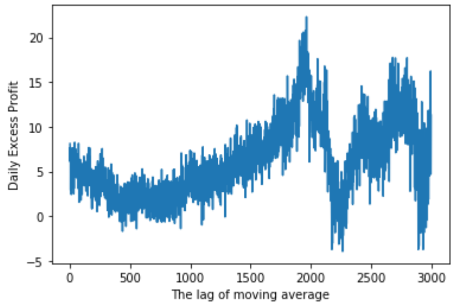
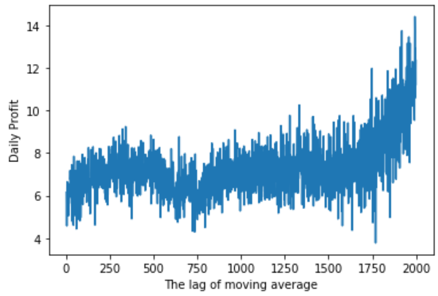
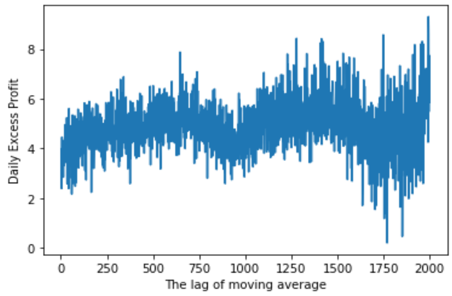
HSI	 Figure 3	 Figure 4
CSI 500	 Figure 5	 Figure 6

Table 3. The change of daily profit, daily excess profit and Moving average order

From Figure 1, 3, and 5, the trend of profit grows with moving average order.

From Figure 2, 4, and 6, the trend of excess profit is fluctuated with moving average order.

4.3 Optimal moving average order

	Optimal moving average order	
	based on maximum profit	based on maximum excess profit
S&P 500	$n = 2931$	$n = 1934$
HSI	$n = 1965$	$n = 1965$
CSI 500	$n = 1993$	$n = 1993$

Table 4. Order selection based on maximum profit and excess profit

	Optimal moving average order	
	condition	based on maximum information ratio
S&P 500	$\text{Excess profit}(n) \geq 1.1$	$n = 1708$
HSI	$\text{Excess profit}(n) \geq 16$	$n = 2996$
CSI 500	$\text{Excess profit}(n) \geq 7$	$n = 1436$

Table 5. Order selection based on maximum information ratio with condition

4.4 Optimal Back-testing for each stock index and each order selection criteria

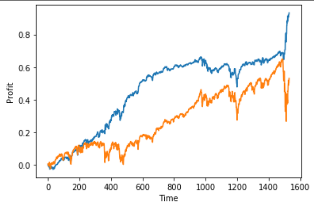
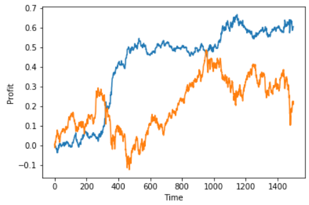
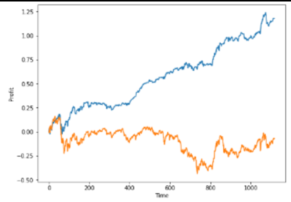
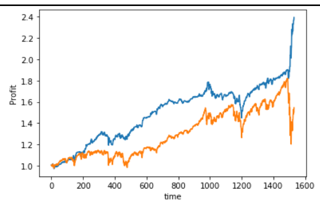
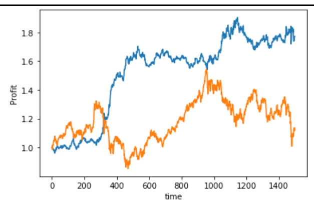
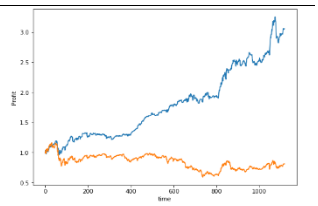
	index	**S&P 500**	**HSI**	**CSI 500**
Without compounding	The optimal growth of accumulated return	 **Figure 7**	 **Figure 8**	 **Figure 9**
	Optimal order	2931	1965	1993
	Annual return of strategy	0.15232932	0.10093018	0.26339797
	Annual return of benchmark	0.08664008	0.03721619	-0.01595927
	Sharp ratio of strategy	1.03561685	0.58389562	1.35135096
	Sharp ratio of benchmark	0.26463479	-0.01555174	-0.22531899
	Daily maximum drawdown	-0.054671	-0.06721295	-0.08300287
	Daily 5% VAR of strategy	-0.01189301	-0.01125934	-0.01825121

Table 6. Optimal back-testing results without compounding

If we regard the highest sharp ratio of strategy as the standard to select the optimal order criteria. The moving average order selection based on maximum total daily profit is the best criteria among these stock indexes.

According to Figure 7, 8 and 9, the growth of strategy (blue line) in testing period outperforms the benchmark that buying and holding the stock indexes (orange line). Either the annual return or sharp ratio of strategy is higher than the index of corresponding benchmark.

	index	**S&P 500**	**HSI**	**CSI 500**
With compounding	The optimal growth of accumulated return	 **Figure 10**	 **Figure 11**	 **Figure 12**
	Optimal order	1708	1965	1993

	Annual return of strategy	0.22735077	0.12884843	0.45926706
	Annual return of benchmark	0.0889652	0.0225699	-0.04250605
	Sharp ratio of strategy	1.05934751	0.5401418	1.40649695
	Sharp ratio of benchmark	0.20468773	-0.08318463	-0.38726207
	Daily maximum drawdown	-0.11083265	-0.12386262	-0.1714184
	Daily 5% VAR of strategy	-0.01996073	-0.01829444	-0.03374125

Table 7. Optimal back-testing results without compounding

In Figure 10, 11 and 12, the blue lines are higher than the orange line in most of time, which means in the condition of compounding, strategy performance is better than benchmark in testing period.

For S&P 500, with compounding, the optimal order criteria change into maximum the information ratio rather than maximum the total profit in training data.

The compounding annual return and sharp ratio overall are better than those index without compounding. CSI 500 can achieve 45% of annual rate which is the best. S&P 500 can gain 22% and HSI can gain 12.8% of annual return. From daily maximum drawdown, S&P 500 is the optimal one, yet with respect to daily 5% VAR, HSI is the best.

For more results of training period, please refer to **Appendix B.**

Discussion and Conclusion

This research proposed a novel strategy called **Differential Trend Motion**. It is really promising and potential in trading. It challenges the weak form of Efficient market hypothesis and prove that we can only use the history price to do the trading. The moving average order selection shows that normally the higher order can gain higher total profit which means the long-term data may include more valuable information to gain the arbitrary opportunity.

S&P 500, HSI and CSI500, are three stock index and represent different stock market. The back-testing show that **Differential Trend Motion strategy** is robust in these stock index. It means that more arbitrary opportunity can be explored by this novel strategy.

For further study, the strategy can be used in trading of stock, future and option markets which allow LONG and SHORT and T+0 transaction (buy and sell within one day). Moreover, the strategy used in this research is merely split the time into training and testing period. The moving average model selection is only one time. Actually, the strategy can be change into dynamic model and the optimal order selection can be updated every month or every year.

Reference

Appel, G. (2005). *Technical Analysis: Power Tools for Active Investors*. Upper Saddle River: FT Press.

Baek, Y. & Kim, H. Y (2018). ModAugNet: A new forecasting framework for stock market index value with an overfitting prevention LSTM module and a prediction LSTM module. *Expert Systems with Applications*, pp. 457-480.

Chong, T. T., and Ng W. (2008). Technical analysis and the London stock exchange: testing the MACD and RSI rules using the FT30. *Applied Economics Letters* 15: 1111–14.

Fama, E. F., and Kenneth R. F. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33: 3–56.

Liu, S.H. and Zhou, F. (2010) On stock prediction based on KNN-ANN algorithm, *2010 IEEE Fifth International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA),* Changsha, 2010, pp. 310-312, doi: 10.1109/BICTA.2010.5645310.

Mikosch, T. (2012) *Elementary stochastic calculus with Finance in view*. Beijing, World Scientific Publishing Company Press.

Ursone, P. (2015) *How to Calculate Options Prices and Their Greeks: Exploring the Black Scholes Model from Delta to Vega.* TJ International Ltd, Padstow, Cornwall, UK

Wang, D. Liu X. and Wang, M. (2013). A DT-SVM Strategy for Stock Futures Prediction with Big Data, *2013 IEEE 16th International Conference on Computational Science and Engineering*, Sydney, NSW, pp. 1005-1012, doi: 10.1109/CSE.2013.147.

Yor, M. (2001) *Exponential Functionals of Brownian Motion and Related Processes* Paris, Springer Press.

Appendix A

If we assume the change(difference) of price follow Brownian motion which follow $N(0,\sigma)$, we are going to prove that

If $MA_t(n) > 0$

$$P[MA_{t+1}(n) - MA_t(n) < 0] = 0.75$$

When $MA_t(n) > 0$, which means

$$MA_t(n) = \frac{(P_t - P_{t-1}) + (P_{t-1} - P_{t-2}) + \cdots + (P_{t-n+1} - P_{t-n})}{n}$$

$$= \frac{(P_t - P_{t-n})}{n} > 0$$

$$P_t - P_{t-n} > 0$$

When $MA_{t+1}(n) - MA_t(n) < 0$, which means:

$$\frac{(P_{t+1} - P_{t+1-n})}{n} - \frac{(P_t - P_{t-n})}{n} < 0$$

$$(P_{t+1} - P_{t+1-n}) - (P_t - P_{t-n}) < 0$$

$$\because \Delta_{t-1}P \sim N(0,\sigma)$$

$$\therefore P_{t+1} - P_{t+1-n} \text{ and } P_t - P_{t-n} \sim N(0, n\sigma)$$

Suppose:

$$P\big((P_{t+1} - P_{t+1-n}) - (P_t - P_{t-n}) < 0 | (P_t - P_{t-n}) > 0\big) * P\big((P_t - P_{t-n}) > 0\big)$$

$$= P\big((P_{t+1} - P_{t+1-n}) - (P_t - P_{t-n}) < 0, (P_t - P_{t-n}) > 0\big)$$

$$= \int_{x=0}^{\infty} P(\,(P_{t+1} - P_{t+1-n}) < x | (P_t - P_{t-n}) = x) d\Phi(x)$$

$$= \int_{x=0}^{\infty} \Phi(x) d\Phi(x)$$

$$= \int_{t=1/2}^{1} \mathrm{t}\, d\mathrm{t}$$

$$= \frac{1}{2}(1)^2 - \frac{1}{2}\left(\frac{1}{2}\right)^2 = \frac{3}{8}$$

$$\because P_t - P_{t-n} \sim N(0, n\sigma) \qquad \therefore P\big((P_t - P_{t-n}) > 0\big) = \frac{1}{2}$$

$$\therefore P\big((P_{t+1} - P_{t+1-n}) - (P_t - P_{t-n}) < 0 | (P_t - P_{t-n}) > 0\big) = \frac{3}{8} \div \frac{1}{2} = \frac{3}{4} = 0.75$$

$$\therefore P[MA_{t+1}(n) - MA_t(n) < 0|\ MA_t(n) > 0] = 0.75$$

Similarly, we can prove

$$P[MA_{t+1}(n) - MA_t(n) > 0|\ MA_t(n) < 0] = 0.75$$

Appendix B

The growth of accumulated return (with compounding) in training and testing period with the order selection by maximum total daily profit.

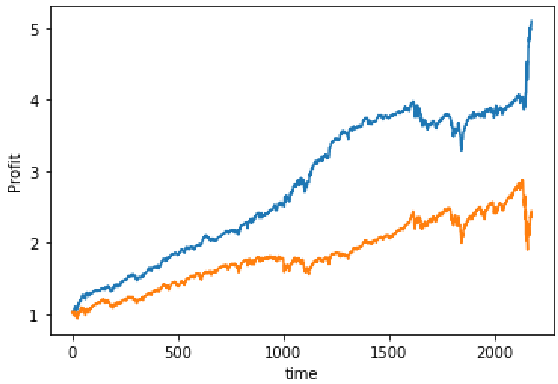


Figure B1. The growth of accumulated return (with compounding) in training and testing period for S&P 500

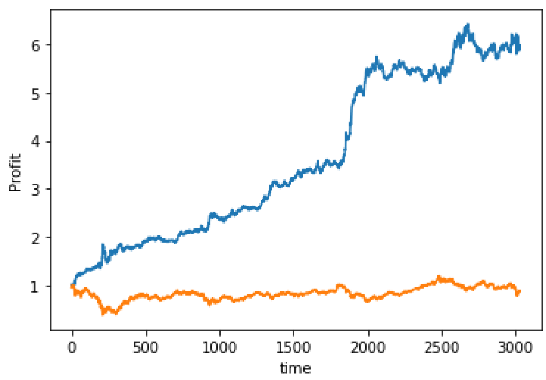


Figure B2. The growth of accumulated return (with compounding) in training and testing period for HSI

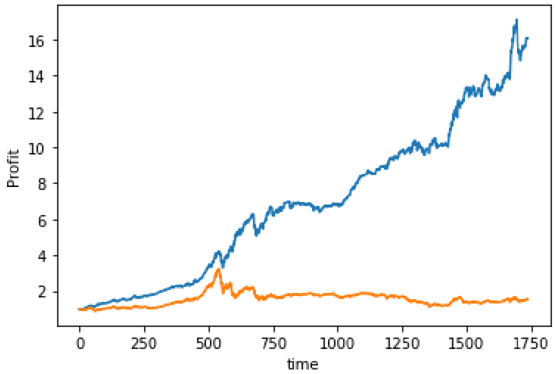


Figure B3. The growth of accumulated return (with compounding) in training and testing period for CSI500

According to Figure B1, B2, B3, both in training period and testing period (about the last 1500 data), the novel strategy outperforms buy and hold strategy. Even if for S&P 500 and HSI, the growth of accumulated return in testing is not as stable as in training period. However, the trend of increasing is constant, namely the trend keeps going up in the testing period and the strategy still can work.

Appendix C

https://github.com/king-yellow/HUANG_Jian_MAFS5210_project3