Statistical Approach to Evaluating Profitability of Stock Markets

Yoshihisa Udagawa
Faculty of Informatics, Tokyo University of Information Sciences
Chiba-city, Chiba, Japan
e-mail: yu207233@rsch.tuis.ac.jp

Abstract - **Candlestick charting is one of the most popular techniques used to predict short-term stock price trends. Despite popularity, there is still no consistent conclusion for the predictability of the technique mainly due to qualitative description of candlestick patterns. This paper proposes a six parameters model that allows us to define both candlestick patterns and price zones where the patterns occur. It is important to grasp buy and sell opportunities for a successful stock trade. Uptrend reversal candlestick patterns are used to find a buy opportunity to enter a trade in a long position. Three exit criteria are proposed to find a sell opportunity to exit a trade for fixing profits or losses. Simulations to estimate profits of markets are performed using historical daily stock data of the US and Asian stock markets with approximately the same parameter values for the six parameters model and the exit criteria in terms of the standard deviation in statistics. Profitability of the proposed stock trade method is statistically examined by linear regression analysis showing that timing to sell stock is significantly related to profits for the three exit criteria. The results of simulations indicate that the US markets are more profitable than Asian markets under the proposed model.**

Keywords - stock price prediction; technical analysis; candlestick patterns; market exit criteria; trailing stop; profit simulation; global market comparison; regression analysis.

I. INTRODUCTION

This paper is an extension of our previous paper on a performance analysis of international stock markets [1]. In the previous paper, we propose a model for finding great opportunities for investors to buy a stock using candlestick chart patterns, and criteria for selling the stock hopefully to keep profits.

In this paper, the following aspects are added to the original work:

(1) Candlestick chart patterns for finding buy opportunities are defined by formulas in terms of successive candlesticks to generalize well-known uptrend reversal candlestick patterns;
(2) A widely-used criterion for finding sell opportunities, named a *trailing stop* [2], is compared with our original criteria in terms of profit;
(3) The period of stock price data used in our experiment is determined based on quarterly and monthly stock price fluctuation analyses.

Forecasting a direction of future stock prices attracts attention of not only financial investors but also researchers on statistics and computer science. Motivation involves to predict the direction of future prices for successful trading and to develop computer system to support it. While many researches on stock price prediction are limited to specific markets, only a few studies are dealing with multiple stock markets.

Dimson, Marsh, and Staunton [3] discuss performances of global markets including emerging markets and developed markets. Though emerging markets have grown to a significant size up to 2007, developed markets, notably the US markets, have outpaced the growth in emerging markets in the 21st century. They conclude that investors should be modest to invest in emerging markets since exchange rate movements are largely affected by inflation that is prevalent in emerging countries.

Ahmad, Ahmed, Vveinhardt and Streimikiene [4] examine Asian stock markets including KSE100 (Pakistan), Nikkei 225 (Japan), KOSPI (South Korea), and BSE (India) in terms of stock return and volatility. The results of statistical analyses lead to a conclusion that volatility is significantly related to return in each market.

To the best of our knowledge, there are few studies that examine profitability of global markets based on simulation using candlestick patterns for estimating buy opportunity and loss stop criteria for finding sell opportunity on daily stock price data.

The contributions of this paper are as follows:

(I) Proposal of a six parameters model to retrieve candlestick patterns that are both similar in price patterns and price zones, i.e., high- or low-price zone in which they occur.
(II) Proposal of three loss stop criteria to exit trade for fixing profits or losses in case of a long position.
(III) Evaluation of profitability of five major markets in the US and Asia using the proposed model for retrieving similar candlestick patterns and the three loss stop criteria through simulations.

The remainder of the paper is organized as follows. Section II reviews related work. Section III gives backgrounds of candlestick patterns. Section IV proposes a model for stock trade using candlestick patterns and exiting criteria from a stock market. Section V presents empirical results on bullish (uptrend) reversal candlestick patterns

using five markets' data in the US and Asia. Section VI concludes the paper with our plans for future work.

II. Related Work

There have been a growing number of studies on predicting future price movements of stock markets. In this section, we review previous studies on performances of global markets and predictabilities of candlestick patterns.

A. Studies on Performances of Global Markets

International investing is believed to bring an advantage of better profits from global markets while managing risks better. Dimson, Marsh, and Staunton [3] discuss that emerging markets achieved a higher profit of 11.7% per year than a developed markets' profit of 10.5% from 1950 to 2019. However, because of the global financial crisis in the 21st century, the average profit on US equities has been an annualized 10.6%, while the world average profit excluding the US has been 5.3% in the 21st century. They conclude that investors should be modest to invest in emerging markets because exchange rate movements are largely affected by inflation in emerging countries in addition to questionable capabilities to maintain a fair market.

Ahmad, Ahmed, Vveinhardt, and Streimikiene [4] study Asian stock markets containing KSE100 (Pakistan), Nikkei 225 (Japan), KOSPI (South Korea), Hang Seng (Hong Kong), Shanghai Stock Exchange (China), and BSE (India) in terms of stock returns and volatility. The results show that KOSPI has the highest average annual return of 12.67%, followed by BSE with 11.61%, while KSE 100 has the least return of 9.31%.

B. Studies disapproving of candlestick patterns

As for candlestick patterns in technical analysis [5], several studies [6]-[8] conclude that they are useless based on the experiments using the stock exchange markets' data in the US, Japan, and Thailand.

Horton [6] studies the profitability of 4 pairs of three-day candlestick patterns on 349 stocks that are selected randomly representing all major industry groups. The main conclusion of his study is that these candlestick patterns create no value for trading individual stocks. Marshall, Young, and Cahan [7] find that for a period of 10 days, candlestick charting strategies are not profitable for Dow Jones Industrial's components from 1992 to 2002 and Japanese equity markets from 1975 to 2004. Based on experiments using stock data in the Stock Exchange of Thailand, Tharavanij, Siraprapasiri, and Rajchamaha [8] conclude that any candlestick patterns cannot reliably predict market directions even with filtering by well-known stochastic oscillators [5].

C. Studies approving of candlestick patterns

Other studies conclude that applying a certain candlestick patterns is profitable at least for short-term trade in the US and Asian stock markets [9]-[15].

Caginalp and Laurent [9] study and favorably evaluate the predictive power of eight three-day reversal candlestick patterns on the S&P 500 index during the period of 1992–1996. They propose to define candlestick patterns as a set of inequalities using opening, high, low, and closing prices. These inequalities are taken over by later studies. Goo, Chen, and Chang [10] define 26 candlestick patterns using modified version of inequalities that are proposed by Caginalp and Laurent. They examine these patterns using stock data of Taiwan markets, and conclude that the candlestick trading strategies are valuable for investors.

Chootong and Sornil [11] propose a trading strategy combining price movement patterns, candlestick chart patterns, and trading indicators. A neural network is employed to determine buy and sell signals. Experimental results using stock data of the Stock Exchange of Thailand show that the proposed strategy generally outperforms the traditional trading methods based on technical indicators [5].

One of the obstacles of candlestick charting is the highly subjective nature of candlestick patterns that are defined using words of natural language and illustrations [5]. Tsai and Quan [12] propose an image processing technique to analyze similarities of candlestick charts for stock prediction instead of using numerical inequality formulas. Their experimental results using Dow Jones Industrial Average index show that visual content extraction and similarity matching of candlestick charts are useful for predicting short-term and medium-term stock movements.

Zhu, Atri, and Yegen [13] examine the effectiveness of five different candlestick reversal patterns for predicting short-term stock movements using data of two Chinese stock markets. The results of statistical analysis suggest that the patterns perform well in predicting price trend reversals.

Jamaloodeen, Heinz, and Pollacia [14] statistically examine whether two of the most popular Japanese candlestick patterns, i.e., *Shooting Star* and *Hammer* patterns [5], have predictive significance to forecast a temporary top and bottom using historical data of the S&P 500 index. They define original formula for each pattern using four parameters, i.e., open, high, low, and closing prices. Their findings include the two patterns are highly reliable when using high price for *Shooting Star* and low price for *Hammer* patterns.

Udagawa [15] proposes a dynamic programing method to skip small and noisy candlesticks to improve predictability of candlestick charting. Experimental results show that the proposed method is effective in predicting both uptrend and downtrend.

III. Candlestick Chart Patterns

This section introduces formation of a candlestick chart. Samples of well-known bullish reversal patterns are described. Criticism of candlestick patterns as a method for predicting stock price movements are also mentioned.

A. Formation of Candlestick

A daily candlestick line is formed with the market's opening, high, low, and closing prices of a specific trading day [5]. Figure 1 represents images of typical candlesticks. The candlestick has a wide part, which is called a *body*, showing the range between the open and close prices of that day's trading. If the closing price is above the opening price,

then a hollow candlestick is drawn indicating a bullish (rising) candlestick. If the opening price is above the closing price then a filled candlestick is drawn showing a bearish (falling) candlestick.

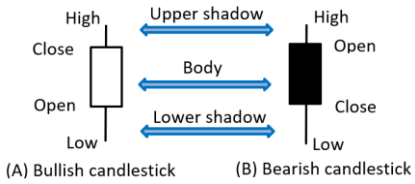

Figure 1. Candlestick formation

The thin lines above and below the body, which are called *shadows*, indicate the high/low ranges. The high price is marked by the top of the upper shadow, and the low price is by the bottom of the lower shadow.

B. Bullish Reversal Candlestick Patterns

Dozens of candlestick patterns are identified and become popular among stock traders [5]. There are three classes of candlestick patterns, i.e., bullish reversal, bearish reversal, and continuation patterns. The reversal patterns are more meaningful because it helps a trader buy at the bottom and sell at the peak of price. This study focuses solely on bullish reversal patterns under the assumption that a trader takes a long position. Triple candlestick patterns are examined because they extend double candlestick patterns with an extra one candlestick for confirmation.

There are four well-known triple candlestick patterns signaling bullish reversal. They are named *morning star*, *three white soldiers*, *three inside up*, and *three outside up*.

Figure 2 shows the *morning star* pattern which is considered as a major reversal signal when it appears in a low-price zone or at a bottom. It consists of three candles, i.e., one short-bodied candle (hollow or filled) between a preceding long filled candlestick and a succeeding long hollow one. The pattern shows that the selling pressure that was there the day before is now subsiding. The third hollow candle overlaps with the body of the first filled candlestick suggests a start of a bullish reversal. The larger the hollow and filled candlesticks are, and the higher the hollow candlestick moves, the stronger the potential reversal.

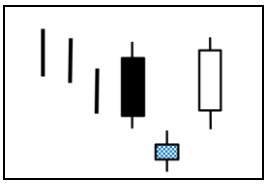

Figure 2. *Morning star* pattern

Figure 3 shows the *three white soldiers* pattern which is interpreted as a strong indication of a bullish market reversal when it appears in a low-rice zone. It consists of three long hollow candlesticks that close progressively higher on each subsequent trading day. Each candlestick opens higher than the previous opening price and closes near the high price of the day, showing a steady advance of buying sentiment.

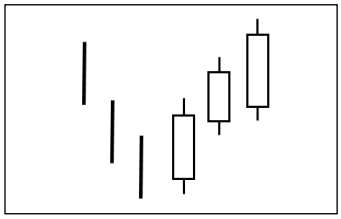

Figure 3. *Three white soldiers* pattern

Figure 4 illustrates the *three inside up* pattern. In this pattern, the first candlestick is a large filled one. The second candlestick is a smaller hollow candlestick contained within the first one. The third candlestick breaks the high price of the second candlestick.

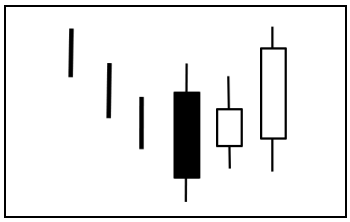

Figure 4. *Three inside up* pattern

Figure 5 illustrates the *three outside up* pattern. It is composed of a small filled candlestick, followed by a longer hollow candlestick that engulfs completely the first one. The third candlestick is a hollow candlestick that closes above the close price of the second one.

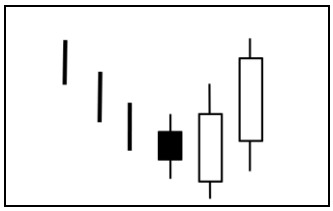

Figure 5. *Three outside up* pattern

In candlestick charting, bullish reversal patterns are deemed to be capable of forecasting price reversal when it appears at bottom after a preceded downtrend. In this study, a price zone where a candlestick occurred is defined by a proposed six parameter model that is described in Section IV.

C. Criticism of Candlestick Patterns

Major criticism of the candlestick chart patterns is that the patterns are qualitatively described with words, such as "long/short candlesticks," "higher/lower prices," "strong/weak signal," supported by some illustrations [5]. Without modeling candlestick patterns in a way that a computer can analyze existence of patterns and perform experiments for measuring a prediction accuracy of future

price trends, arguments on the effectiveness of chart patterns would not come to an end.

In addition, some candlestick chart patterns yield different even oppose forecast depending on whether they appear at a high-price or low-price zone. Formulating a suitable definition of price zones is still an open issue.

It deems that because of the lack of the mathematical definition of the candlestick chart patterns, mixed results are obtained in the studies on candlestick charting. Negative conclusions to the predictability of candlesticks are reported [6]-[8], while positive evidences are provided for several candlestick patterns in experiments including the U.S. and the Asian stock markets [9]-[15].

IV. Proposed Model for Stock Trade

This section describes a model to retrieve a candlestick that is similar in both a price change and a price zone where it occurs. Formulas that abstract well-known bullish reversal patterns are defined. Since market exit criteria are vital to keep profits, three criteria including the popular *trailing stop* [2] are proposed.

A. Six-Parameter Model of Candlestick Retrieval

After trial and error, we propose a six-parameter model that formalizes a zone where a candlestick occurs in addition to a magnitude of price change and a length of candlestick body. Figure 6 illustrates the proposed model with six parameters defined below:

(1) Change of prices w.r.t previous closing price,
(2) Length of candlestick body,
(3) Difference from 5-day moving average,
(4) Difference from 25-day moving average,
(5) Slope of 5-day moving average,
(6) Slope of 25-day moving average.

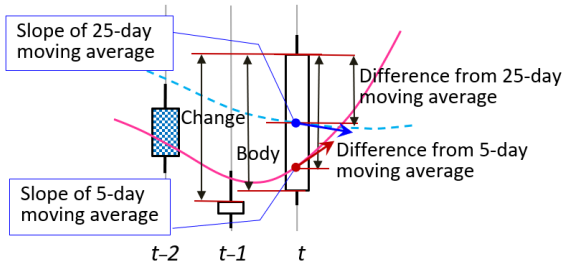

Figure 6. Six parameters to define candlestick and price zone

While most of the previous studies use a series of inequalities or technical indicators to identify a stock price trend, i.e., an uptrend, downtrend or sideway (a stable range), the proposed model is unique in a sense that it uses two moving averages and their slopes. 5-day and 25-day moving averages are used since they are widely used in Japan. They are significant to identify a zone where a candlestick happens. The slopes of the moving averages are also important to identify the price trend.

Tow candlesticks are defined as similar both in a price and zone if all conditions C_1 to C_6 are satisfied.

C_1: if the difference between a closing price change of a given candlestick and that of a candidate candlestick is within the change tolerance (*change_tol*), then C_1 is true.
C_2: if the difference between a body length of a given candlestick and that of a candidate candlestick is within the body tolerance (*body_tol*), then C_2 is true.
C_3: if the difference between a closing price and a 5-day moving average of a given candlestick and that of a candidate candlestick is within the tolerance (*av5diff_tol*), then C_3 is true.
C_4: if the difference between a closing price and a 25-day moving average of a given candlestick and that of a candidate candlestick is within the tolerance (*av25diff_tol*), then C_4 is true.
C_5: if the slope of a 5-day moving average of a given candlestick and that of a candidate candlestick is within the given tolerance (*slope5_tol*), then C_5 is true.
C_6: if the slope of a 25-day moving average of a given candlestick and that of a candidate candlestick is within the given tolerance (*slope25_tol*), then C_6 is true.

B. Finding Buy Oppotunities of Stock Trade

Profit in stock trade in a long position comes from the difference between a buy price and a sell price of a stock. So, buying a stock at a low price and selling it at a higher price is essential for a successful stock trade.

We define formulas that intend to be a generalization of three-day bullish reversal patterns including the *morning star* pattern [5], etc. The formulas in combination with the six parameters model of similar candlesticks are used to find buy opportunity in a low-price zone.

Let CP(*t*) and OP(*t*) denote close and open prices of a given market day *t*. A bullish reversal candlestick pattern is defined as follows:

$$CP(t) > OP(t) \quad (1)$$
$$(CP(t) + OP(t)) / 2 > CP(t–1) \quad (2)$$
$$CP(t) > CP(t–2) \quad (3)$$

Figure 7 depicts a pattern defined by (1)-(3). Inequality (1) means the body of the candlestick is hollow with signaling a rise in stock prices. Inequality (2) specifies that the close price of day *t–1* is below the average of the open and close prices of day *t*. Inequality (3) describes that the close price of day *t–2* is below the close price of day *t*. Inequalities (2) and (3) are satisfied even when the close price of day *t–2* is far below the close price of day *t–1*.

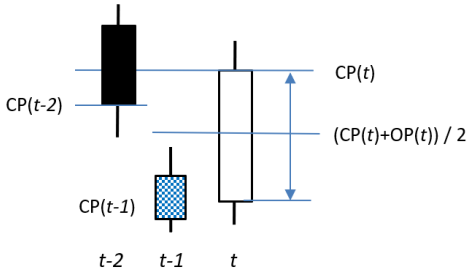

Figure 7. Pattern defined by (1)-(3)

Inequalities (1)-(3) exclude a condition on the length of a candlestick body. In effect, the length of the candlestick body of day *t* is specified by the condition C_2 because (1)-(3) are used with the proposed six parameters model. Since there is no specification on other candlesticks bodies of day *t–1* and *t–2*, (1)-(3) generalize four bullish reversal patterns.

C. Finding Selling Oppotunities of Stock Trade

To successfully complete stock trade, we need to find preferable opportunities to sell a stock in a long position. Candlestick patterns claim that they can be applied to find sell opportunities. However, because there are tens of candlestick downtrend patterns known so far, it is difficult to implement all the patterns.

A capable method named a *trailing stop* [2] is proposed to decide when to sell a stock. The *trailing stop* criteria is designed to lock in profits and suppress losses. Figure 8 illustrates a concept of the criteria. A trader typically specifies a stop price by means of setting a percentage of a loss that can be tolerable on a trade. If a stock price rises in trader's favor, the stop price is continuously reset to a higher value. In case a stock price falls against trader's expectation and exceeds the tolerable percentage of a loss, then the *trailing stop* criterion signals selling a stock.

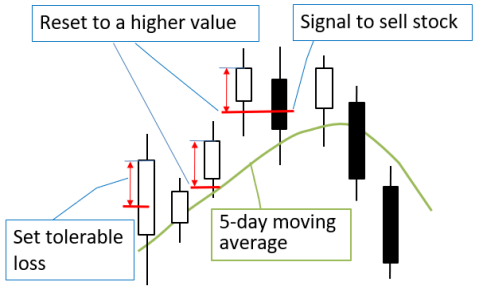

Figure 8. Concept of *trailing stop* criterion

To compare performance of the *trailing stop* criterion with others, we implement two original criteria named a sum of negative price change criterion *(SumNC)*, and a sum of negative price changes below a 5-day moving average (*SumNC5av).* Their concepts are depicted in Figure 9.

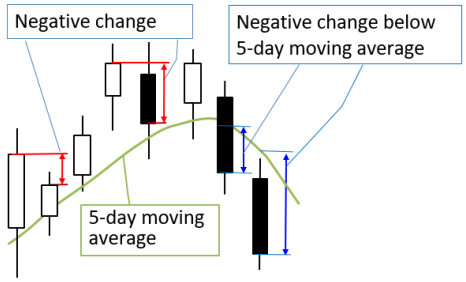

Figure 9. Concept of *SumNC* and *SumNC5av* criteria

The *SumNC* criterion signals selling a stock when the sum of negative price changes exceeds a pre-defined tolerable value. This criterion works the same way as a stop-loss when a price of a stock continues to decline contrary to trader's expectation.

Moving averages are often used in a trading strategy, especially over 5, 25, and 75-day periods in Japan. The *SumNC5av* criterion is devised as a criterion of stock trading with respect to a moving average. The criterion keeps holding a stock until the sum of the negative differences between a stock price and a 5-day average reaches below a pre-defined value. Because falls of a stock price often keep above a 5-day price average in an uptrend, e.g., the fourth candlestick from the right in Figure 9, the *SumNC5av* criterion tends to hold a stock longer than the *SumNC* criterion.

V. Empirical Results

After outlining processes of the performed experiments, statistical analyses of price fluctuations on Dow Jones Industrial Average, NASDAQ Composite index, Nikkei 225 Stock Average, Hang Seng index, and Shanghai Composite index are presented. Results of profit simulations using historical daily stock data of five stock markets are discussed.

A. Data Conversion

Stock prices are converted to the ratio of closing prices to reduce the effects of highness or lowness of the stock prices. The formula below is used for calculating the ratio of prices as a percentage.

$$R_i = (CP_i - CP_{i-1})*100 / CP_i \ (1 \leqq i \leqq n) \qquad (4)$$

CP_i indicates the closing price of the i-th market date. CP_n means the closing price of the current date. R_n is the ratio of the difference between the closing price CP_n of the current date and the closing price CP_{n-1} of one day before to the CP_n.

The daily stock data from Mar. 1, 2007 to June 30, 2020 are used in the experiment. The number of data is approximately 3,358 for each market. Daily stock data are downloaded from a website that provides historical data of major world markets [16].

B. Statistics of Candlestick Parameters

As the first step of experiments, quarterly statistics about six parameters concerning the proposed six-parameter model of a candlestick pattern are calculated for the period between Apr. 1, 2007 and Jun. 30, 2020. Table I summarizes statistics of the six parameters for the five markets, i.e., Dow Jones, NASDAQ, Nikkei 225, Hang Seng, and Shanghai.

TABLE I. SUMMARY OF STATISTICS OF SIX PARAMETERS DURING PERIOD BETWEEN APR. 1, 2007 AND JUN. 30, 2020

	DowJones		NASDAQ		Nikkei 225		Hang Seng		Shanghai	
	Average	Deviation	Average	Deviation	Average	Deviation	Average	Deviation	Average	Deviation
Body length	0.0124	1.1479	0.0141	1.1259	−0.0271	1.1309	−0.0569	1.0728	0.0985	1.4620
Change	0.0302	1.2691	0.0525	1.4107	0.0191	1.5401	0.0184	1.5391	0.0143	1.6197
Difference of price and 5-day average	0.0296	1.2811	0.0663	1.4376	−0.0100	1.6750	−0.0111	1.6688	−0.0253	1.8162
Difference of price and 25-day average	0.1819	3.2450	0.3976	3.6178	−0.0464	4.2567	−0.0623	4.1579	−0.1692	4.9980
Slope of 5-day average	0.0200	0.5084	0.0397	0.5708	0.0042	0.6682	0.0035	0.6664	−0.0024	0.7478
Slope of 25-day average	0.0213	0.2164	0.0407	0.2470	0.0068	0.2889	0.00473	0.2881	0.0002	0.3571

Averages of all six parameters are positive for Dow Jones and NASDAQ indicating that the two markets are generally on an uptrend. Nikkei 225 and Han Seng markets mark negative values for three parameters, i.e., the candlestick body length, the difference between price and 5-day average, and the difference between price and 25-day average. Shanghai market has negative values of three parameters, i.e., the difference between price and 5-day average, the difference between price and 25-day average, and the slope of 5-day average. These negative values suggest that the Asian markets are less profitable than the US ones.

Figure 10 shows price fluctuations as a percentage of the five markets for each quarter. "200706" in the x-axis of Figure 10 indicates the second or April-June quarter of calendar year 2007, for example. We see that the prices of Shanghai and Hang Seng markets fluctuate larger than those of the other markets, notably during the period from the second quarter of 2007 to the fourth quarter of 2015.

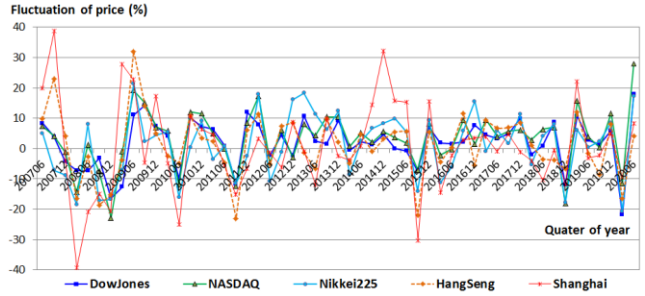

Figure 10. Price fluctuations of five markets for each quarter

We take the period between Apr. 1, 2015 and Jun. 30, 2020 for further examination, because the price fluctuations of the five markets are somehow linked during this period as observed in Figure 10. Table II summarizes monthly statistical results of the six parameters. The average values of all six parameters are positive for Dow Jones and NASDAQ markets. All average values are barely positive for Nikkei 225 market. The five average values out of six parameters are negative for Han Seng and Shanghai markets.

TABLE II. SUMMARY OF MONTHLY STATISTICS OF SIX PARAMETERS DURING PERIOD BETWEEN APR. 1, 2015 AND JUN. 30, 2020

	DowJones		NASDAQ		Nikkei 225		Hang Seng		Shanghai	
	Average	Deviation	Average	Deviation	Average	Deviation	Average	Deviation	Average	Deviation
Body length	0.0046	0.9296	0.0236	0.9684	−0.0195	0.9551	−0.0499	0.8753	0.1086	1.2899
Change	0.0361	1.2520	0.0630	1.3101	0.0201	1.3242	0.0057	1.2012	−0.0069	1.4673
Difference of price and 5-day average	0.0412	1.2807	0.0922	1.3172	0.0039	1.4657	−0.0166	1.3327	−0.0588	1.6885
Difference of price and 25-day average	0.2537	3.4524	0.5442	3.4125	0.0370	3.7052	−0.0839	3.4422	−0.3279	4.5487
Slope of 5-day average	0.0261	0.5084	0.0521	0.5252	0.0088	0.5962	−0.0017	0.5474	−0.0198	0.6864
Slope of 25-day average	0.0277	0.2261	0.0514	0.2321	0.0117	0.2505	−0.00060	0.2394	−0.0137	0.3159

Figure 11 shows price fluctuations of the five markets for each month. For example, "202006" in the x-axis of Figure 11 indicates Jun. 2020.

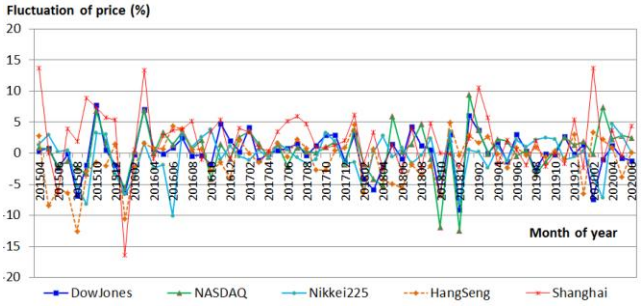

Figure 11. Price fluctuations of five markets for each month

In Figure 11, it is observed that stock price fluctuations of Dow Jones and NASDAQ overlap in many months. Stock prices of Asian markets roughly move in the same direction with some degrees of time delays.

C. GUI for Experiments

Figure 12 shows a GUI that is used in the experiments. It provides parameter values for the six-parameter model in Figure 6 and the three market-exit criteria. The *File* button allows users to choose a CSV file containing a set of stock price data. The full path of the CSV file is displayed. In Figure 12, a file named *DowJones.csv* is chosen. The right two text boxes in the first-row are used to specify periods of market days, i.e., 20150401 to 20200630, used in the experiments.

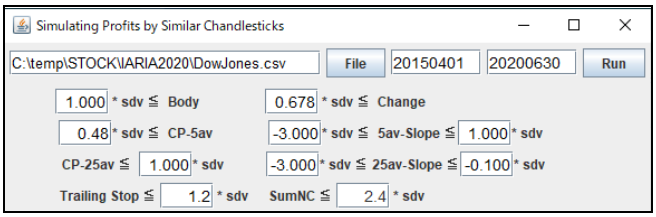

Figure 12. GUI used in experiments

A developed simulator calculates the averages and the standard deviations of the six parameters for the specified period. The text box in the second-row labeled by *"* std ≦ Body"* specifies the magnitude of standard deviation that the length of a candlestick body need to satisfy. Because the experiments are performed for a long position, the length of a candlestick body is required to be longer than the specified magnitude of the standard deviation. The text box labeled by *"* std ≦ Change"* specifies the magnitude of standard deviation that price changes need to satisfy.

Three text boxes in the third-row are used to define ranges of the difference between a close price and a 5-day moving average (labeled by *"CP-5av"*), and a slope of a 5-day moving average (labeled by *"5av-Slope"*). In order to spot candlesticks in uptrend reversal, the developed simulator is designed to retrieve candlesticks whose *CP-5av* are greater than the specified magnitude. As for *5av-Slope*, we need to specify both lower and upper limits.

Three text boxes in the fourth-row are used to define ranges of the parameters concerning 25-day average. The

three text boxes in the fourth-row play the similar role as those in the third-row.

Tow text boxes labeled by "*Trailing Stop"* and "*SumNC"* at the bottom of the GUI specify parameter values for the *Trailing Stop* and *SumNC* criteria. The value labeled by "*SumNC"* is also applied to the *SumNC5av* criterion.

Trade price is calculated by the average of the high and low prices on a trading day. This calculation is feasible because the high and low prices can be observed during stock trading time. Traders can decide whether to keep or sell a stock based on the prices. Simulated profits are calculated using the trade price, i.e., the average of the high and low prices.

A typical commission fee of online brokers is between 0.05% and 0.15% depending on the order size of trade. Because we assume swing trading [17] that attempts to capture profit over a period of a few days to several weeks, a commission fee is treated as negligible costs in this study.

D. Experiments on Profit Estimation by Simulation

Table III shows an experimental result performed on Dow Jones daily data using parameters shown in Figure 12. The column named "*Trade day"* in Table III lists market days that satisfy all conditions defined by (1) to (3), and C_1 to C_6. Strictly, conditions defined by (2) and (3) are embedded in source code and cannot be seen on the GUI.

The *trailing stop*, *SumNC*, and *SumNC5av* criteria are used to make a decision to sell a stock. The parameter value for the *trailing stop* is set to 1.2 times the standard deviation of price changes. Because the standard deviation of price changes is 1.2520% as shown in Table II, the tolerable loss for the *trailing stop* is 1.5024% (=1.2*1.2520%).

The parameter values for the *SumNC* and *SumNC5av* are set to 2.4 times the standard deviation of price changes. The tolerable loss is analogously calculated to be 3.0048% (=2.4*1.2520%). The other parameters are carefully adjusted for each market to retrieve 26 days for buy opportunities so that the number of the days is suitable for being analyzed based on the theory of the normal distribution [18].

The columns named "*Exit day"*, *"Days to hold stock"*, "*Total profit"*, and "*Profit per day"* indicate the day to sell stocks, the number of holding days of a stock, a simulated total profit, and a rate of a profit to the number of holding days of a stock, respectively. Averages of profits are 2.53%, 2.247%, and 2.322% for the *trailing stop*, *SumNC*, and *SumNC5av* criteria, respectively.

Figure 13 shows a candlestick chart of the trade that buy a stock on Feb. 16, 2016 as listed in the 9th line of Table III. The *trailing stop*, *SumNC*, and *SumNC5av* criteria generate a signal to close the trade after 52, 15, and 38 days, respectively.

The *trailing stop* criterion signals selling a stock on Apr. 29, 2016, i.e., 52 days after the stock trade. Because the nature of the *trailing stop* criterion, there is always a predefined amount of loss from the maximum profit before closing a trade, i.e., 1.5024% in this experiment.

As for the *SumNC* criterion, the sum of negative price changes exceeds the predefined limit value of 3.0048% on Mar. 8, 2016, i.e., 15 days after the stock trade.

TABLE III. ESTIMATED PROFITS ON DOW JONES DAILY DATA

No.	Trade Day	Trailing Stop				SumNC				SumNC5av			
		Exit Day	Days to Hold Stock	Total Profit	Profit per Day	Exit Day	Days to Hold Stock	Total Profit	Profit per Day	Exit Day	Days to Hold Stock	Total Profit	Profit per Day
1	20150810	20150811	1	−0.166	−0.166	20150811	1	−0.166	−0.166	20150820	8	−1.909	−0.239
2	20150827	20150831	2	0.376	0.188	20150901	3	−1.348	−0.449	20150901	3	−1.348	−0.449
3	20150908	20150909	1	0.832	0.832	20150909	1	0.832	0.832	20150922	10	0.263	0.026
4	20150915	20150918	3	−0.025	−0.008	20150918	3	−0.025	−0.008	20150923	6	−1.391	−0.232
5	20151002	20151112	29	8.154	0.281	20151109	26	9.487	0.365	20151112	29	8.154	0.281
6	20151216	20151217	1	0.063	0.063	20151217	1	0.063	0.063	20151218	2	−1.835	−0.918
7	20160122	20160125	1	−0.285	−0.285	20160125	1	−0.285	−0.285	20160209	12	−0.125	−0.010
8	20160129	20160202	2	−0.086	−0.043	20160202	2	−0.086	−0.043	20160209	7	−1.655	−0.236
9	20160216	20160429	52	10.115	0.195	20160308	15	5.544	0.370	20160411	38	9.559	0.252
10	20160524	20160613	13	1.012	0.078	20160617	17	0.194	0.011	20160615	15	0.352	0.024
11	20160629	20160802	23	4.240	0.184	20160812	31	5.635	0.182	20160822	37	5.339	0.144
12	20161107	20170321	91	14.798	0.163	20170109	42	9.855	0.235	20170119	49	8.947	0.183
13	20180214	20180220	3	1.309	0.436	20180228	9	2.393	0.266	20180301	10	0.428	0.043
14	20180223	20180227	2	1.719	0.860	20180228	3	0.508	0.170	20180301	4	−1.420	−0.355
15	20180329	20180402	1	−1.605	−1.605	20180402	1	−1.605	−1.605	20180402	1	−1.605	−1.605
16	20180410	20180411	1	−0.396	−0.396	20180411	1	−0.396	−0.396	20180424	10	−0.619	−0.062
17	20181016	20181017	1	0.235	0.235	20181018	2	−0.472	−0.236	20181024	6	−2.599	−0.433
18	20181031	20181112	8	1.911	0.239	20181112	8	1.911	0.239	20181113	9	0.714	0.079
19	20181226	20190103	5	2.744	0.549	20190103	5	2.744	0.549	20190208	30	12.105	0.404
20	20190104	20190306	41	10.897	0.266	20190207	23	8.405	0.365	20190305	40	11.180	0.280
21	20190604	20190731	40	7.344	0.184	20190725	36	8.059	0.224	20190729	38	8.244	0.217
22	20190808	20190809	1	0.170	0.170	20190812	2	−0.797	−0.399	20190814	4	−1.746	−0.436
23	20190819	20190820	1	−0.249	−0.249	20190820	1	−0.249	−0.249	20190919	22	4.008	0.182
24	20190829	20190903	2	−0.795	−0.397	20190926	19	2.328	0.123	20191001	22	1.928	0.088
25	20191011	20191203	36	2.126	0.059	20191129	34	4.569	0.134	20191202	35	4.066	0.116
26	20200302	20200303	1	1.330	1.330	20200303	1	1.330	1.330	20200303	1	1.330	1.330
		Average	13.923	2.530	0.122	Average	11.077	2.247	0.062	Average	17.231	2.322	−0.051

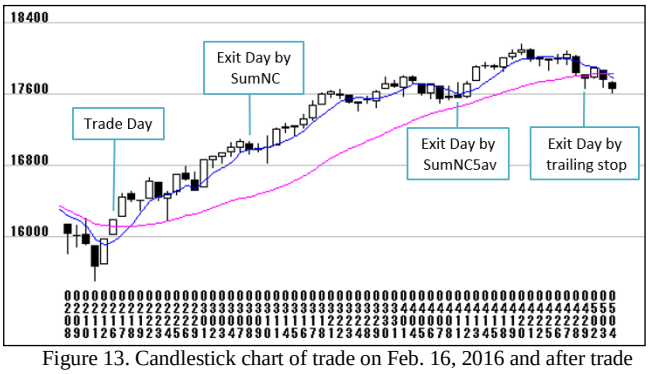

Figure 13. Candlestick chart of trade on Feb. 16, 2016 and after trade

The *SumNC5av* criterion, accumulating negative changes below 5-day moving average, reaches the limit value of 3.0048% on Apr. 11, 2016, i.e., 38 days after the stock trade. Because the stock prices generally keep a steady uptrend after the trade day, the period to hold a stock based on the *SumNC5av* criterion is long more than twice of that of the *SumNC* criterion.

E. Regression Analysis

Regression analysis is a reliable mathematical method to estimate the relationship between two or more variables of interest. It is widely used to examine the influence of one or more independent variables on a dependent variable. In this section, we evaluate profitability and diversity of the proposed stock trade method using regression analysis.

Table IV summarizes the result of regression analysis that is applied to the results for the *trailing stop* criterion by specifying *Profit* as an independent variable and *Days to hold stock* as a dependent variable. *R Square* is 0.8672 suggests that 86.72% of *Profit* can be explained by *Days to hold stock*. The table *ANOVA* (analysis of variance) shows the results of the F-test for measuring the probability that *Profit* is related to *Days to hold stock* by chance. As the value of *Significance F* and *P-value* are 5.16817E-12 (<0.05), which means that *Profit* is significantly related to *Days to hold stock*.

TABLE IV. SUMMARY OF REGRESSION ANALYSIS

Summary Output

Regression Statistics	
Multiple R	0.9312
R Square	0.8672
Adjusted R Square	0.8617
Standard Error	1.5597
Observations	26

Analysis of Variance (ANOVA)

	df	SS	MS	F	Significance F
Regression	1	381.3205	381.3205	156.7494	5.16817E-12
Residual	24	58.3842	2.4327		
Total	25	439.7047			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	0.0674	0.3636	0.1855	0.8544	-0.6831	0.8180
Days to Hold Stock	0.1768	0.0141	12.5200	5.17E-12	0.1477	0.2060

Figure 14 presents a scatter plot for the *trailing stop* criterion with *Profit* as the x-axis and the *number of days to hold stock* as the y-axis. Figure 14 depicts a significant correlation between the two variables. 16 out of 26 trades are terminated within less than four days. Because of early decision, losses are limited within 1.605% that is shown in the 15th line of Table III. On the other hand, when price moves in a favorable direction, the *trailing stop* criterion leads to hold stock for a rather long period resulting in high profits up to 14.798% that is shown in the 12th line of Table III.

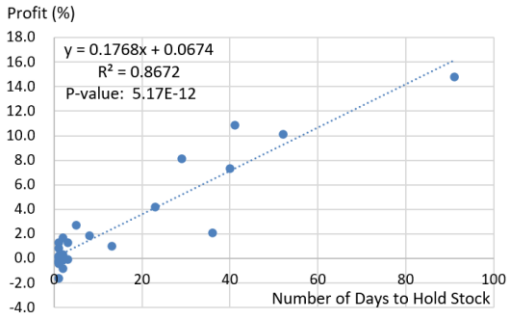

Figure 14. Scatter chart for the *trailing stop* criterion

Figure 15 shows a scatter plot for the *SumNC* criterion. *R Square* is 0.7677. *P-value* is 4.4816E-09 (<0.05). 14 out of 26 trades are stopped within less than four days. Because selling stock is performed based on the sum of the negative prices, the *SumNC* criterion is apt to stop trade earlier than the *trailing stop* criterion.

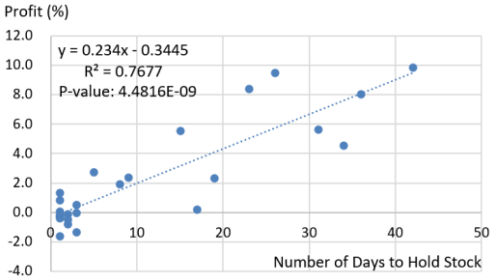

Figure 15. Scatter chart for the *SumNC* criterion

Figure 16 is a scatter plot for the *SumNC5av* criterion. *R Square* is 0.8068. *P-value* is 4.82E-10 (<0.05). In the *SumNC5av* criterion, a sell decision is made using the sum of negative differences between a stock price and a 5-day average. Since a negative price change is ignored while the price keeps above the 5-day average, the *SumNC5av* criterion is insensitive to price fluctuations. Accordingly, 4 out of 26 trades are stopped within less than four days.

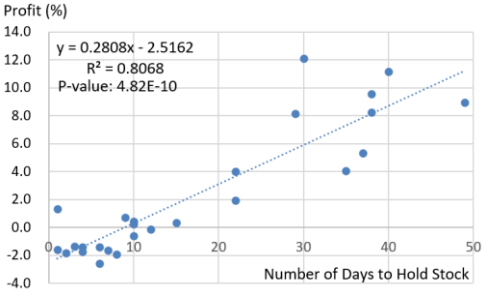

Figure 16. Scatter chart for the *SumNC5av* criterion

All three criteria show significant correlations between the two variables, i.e., *Profit* and *Days to hold stock*, with *R Squares* are greater than 0.7677.

F. Profits for Each Market

Figure 17 shows a graph on the *trailing Stop* exit criterion with profits of each market in the y-axis and multiples of the standard deviation of price changes in the x-axis. In NASDAQ market, the profits increase as the multiples of the standard deviation increase. In Dow Jones market, by contrast, profit reaches a peak of profits at 1.2 multiples of the standard deviation of price changes. Hang Seng and Shanghai markets reach peaks at 1.0 multiples of the standard deviation. Hang Seng and Nikkei 225 markets are notably less profitable than the others.

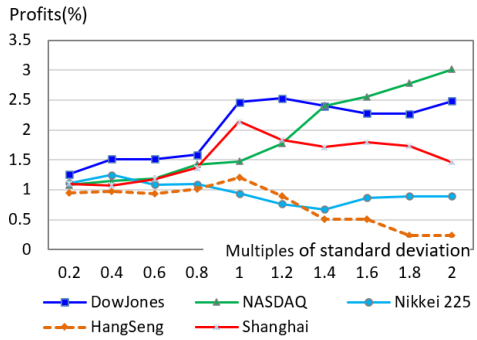

Figure 17. Profits and times of standard deviation using *Trailing Stop*

Figure 18 shows a graph concerning the *SumNC* exit criterion with profits in the y-axis and multiples of standard deviation of price changes in the x-axis. In NASDAQ market, the profits increase as the multiples of the standard deviation increase. In the other four markets, profits reach the highest points at a certain multiple of the standard deviation of price changes. In Dow Jones market, for example, profit reaches a peak at 2.8 multiples of the standard deviation of price changes.

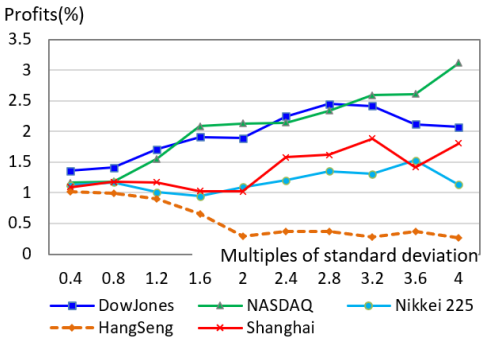

Figure 18. Profits and times of standard deviation using *SumNC*

Figure 19 shows a graph on the *SumNC5av* exit criterion. Profits of Hang Seng market show less than those of the other markets with losses at three multiples, i.e., 1.6, 2.0, and 4.0.

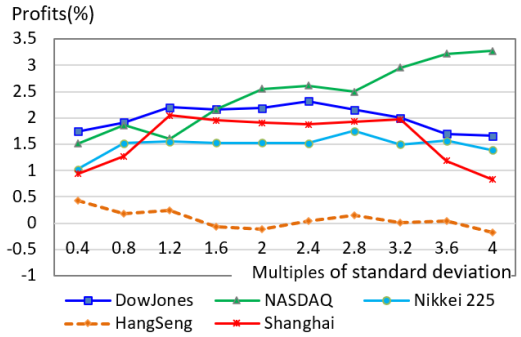

Figure 19. Profits and times of standard deviation using *SumNC5av*

For all three exit criteria, profits of the NASDAC market continue increasing in the range of multiples examined in the experiments. The fact suggests that it is a better decision to hold a stock even if prices fall. Because NASDAQ's stock prices generally keep rising, they tend to recover in a short period. Meanwhile, profits of Dow Jones market always show higher than those of Shanghai, Nikkei 225, and Hang Seng markets, implying that Dow is a leading index of Asian markets.

G. Holding Days for Each Market

Figure 20 shows a graph concerning the *Trailing stop* exit criterion with the number of days to hold a stock in the y-axis and multiples of standard deviation of price changes in the x-axis. The criterion generates a signal to sell a stock when a stock price falls more than a predefined percentage from the highest price. So, the longer the days to hold a stock become, the fewer chances of the stock price plummets are expected. Figure 20 suggests that Dow Jones market has fewer plunges than the other markets.

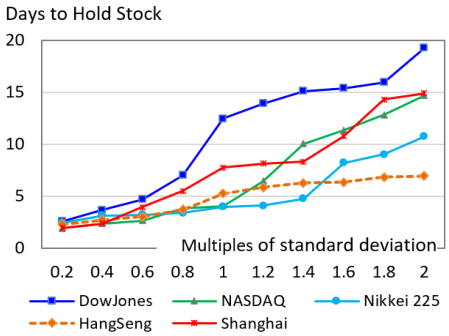

Figure 20. Days to hold stock and times of standard deviation using *Trailing Stop*

Figure 21 is a graph on the *SumNC* exit criterion showing a relationship between the number of days to hold a stock and multiples of the standard deviation of price changes. The number of days to hold a stock apparently linearly depends on multiples of the standard deviation. Since the *SumNC* criterion is based on the sum of the negative price changes, the longer days to hold stock mean the smaller chances of negative price changes during the periods of

trade. Dow Jones and Shanghai markets are more likely to continue price uptrends with less falls in stock price than Hang Seng and Nikkei 225 markets.

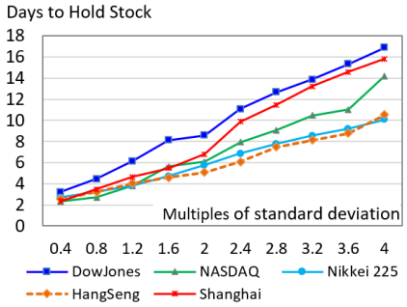

Figure 21. Days to hold stock and times of standard deviation using *SumNC*

Figure 22 shows a graph on the *SumNC5av* exit criterion. Due to the definition of the *SumNC5av* criterion, days to hold stock are tend to be longer than those of the *SumNC* criterion. For example, the maximum number of days to hold in Dow Jones market is 24 for the *SumNC5av* while it is 17 for the *SumNC5* as shown in Figure 21. The graph apparently shows linear dependency between the two variables.

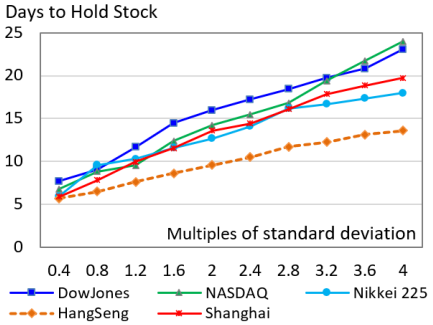

Figure 22. Days to hold stock and times of standard deviation using *SumNC5av*

Days to hold a stock in Dow Jones market tend to be longer than those in the other markets on the three exit criteria, likely leading to high profitability compared with the other markets.

H. Visualizing Profit and Loss Pattern

Figure 23 shows simulated profit-and-loss patterns in a bar graph using the *trailing stop*, *SumNC*, and *SumNC5av* criteria for Dow Jones market. Parameters for the *trailing stop*, *SumNC*, and *SumNC5av* are set to 1.2, 2.4, and 2.4 multiples of the standard deviation, respectively. The profits are sorted in ascending order. Roughly, the three exit criteria result in comparable profits and/or losses. Approximately ten out of 26 trade days result in small amounts of losses with large amounts of profits for the rest of days.

Since the *SumNC5av* criterion tends to hold a stock longer than the *SumNC* criterion, profits and losses obtained by the *SumNC5av* criterion are greater than those by the *SumNC*, suggesting that return and risk are always correlated. Profits gaind by the *trailing stop* criterion seems to be better than those of the other two criteria in a sense that the criterion yields larger profits with less losses.

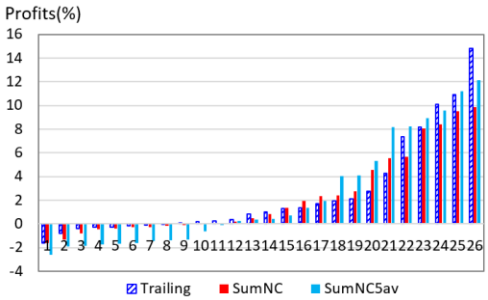

Figure 23. Bar graph of profit and loss for Dow Jones market

Figure 24 shows profit-and-loss patterns in a bar graph for NASDAQ market. Like Dow Jones market, approximately ten out of 26 trade days result in failure. The *SumNC5av* criterion outperforms the other criteria in profits and losses. The maximum profit is about 12%, and the maximum loss is about –3%. Figures 23 and 24 indicate that NASDAQ market is roughly comparable in profit-and-loss patterns to Dow Jones market.

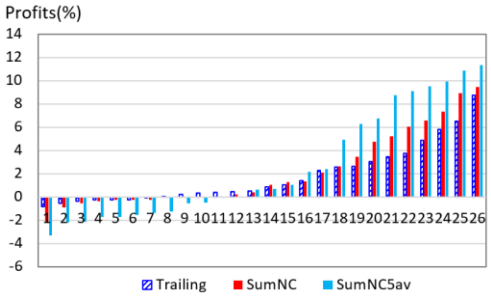

Figure 24. Bar graph of profit and loss for NASDAQ market

Figure 25 shows a profit-and-loss bar graph for Nikkei 225 market. While ten out of 26 trade days are failure like Dow Jones and NASDAQ markets, the maximum profit is about 8%, and the maximum loss is about –4%. The bar graph suggests that Nikkei 225 market seems to have similar price movement patterns, but it is less profitable than Dow Jones and NASDAQ markets.

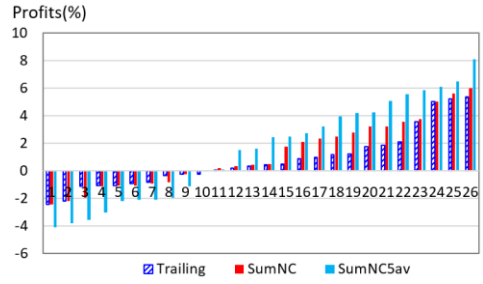

Figure 25. Bar graph of profit and loss for Nikkei 225 market

Figure 26 shows profit-and-loss patterns for Hang Seng market. Roughly, there are 12 out of 26 chances of failure. The maximum profit is estimated about 9%, which is roughly the same as the maximum profit of Nikkei 225 market.

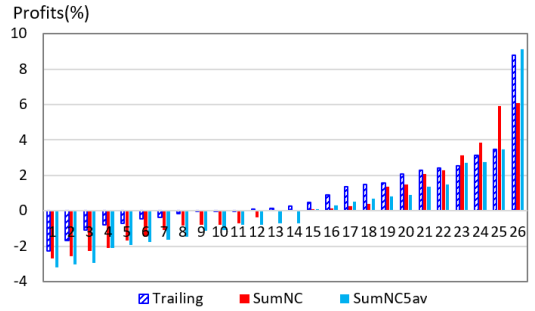

Figure 26. Bar graph of profit and loss for Hang Seng market

Figure 27 shows profit-and-loss patterns for Shanghai market. The results of simulation include a trade on Jan. 4, 2019 with 21.96% profit for the *trailing stop* and *SumND5av* criteria. The trade is deemed to be treated as a special case.

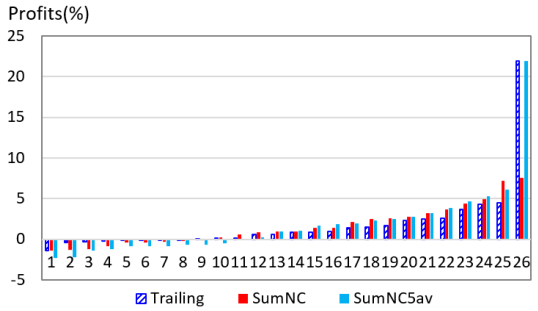

Figure 27. Bar graph of profit and loss for Shanghai market

A feature that is unique in Shanghai market is that profits simulated by the *SumNC* and *SumNC5av* criteria are the same in many cases. The difference between stock price and 5-day average is negative in the Shanghai index as shown in Table II. It is considered that the negative difference between a stock price and a 5-day average leads to the same values of *SumNC* and *SumNC5av* criteria.

I. Summarizing Profit for Each Manket

Table V summarizes "*average profit*", "*success ratio"*, and "*potential profit"* for each criterion. The *trailing stop*, *SumNC*, and *SumNC5av* criteria yield 2.53%, 2.25%, and 2.32% of average profits, respectively, for Dow Jones market as an example. A *success ratio* is obtained by dividing the number of profitable days by the total number of days, i.e., 26. A *potential profit* is calculated by multiplying the *average profit* by the *success ratio*.

TABLE V. SUMMARY OF PROFIT AND SUCCESS RATIO FOR EACH MARKET

	Average Profit(%)			Success Ratio(%)			Potential Profit(%)		
	Trailing	SumNC	SumNC5av	Trailing	SumNC	SumNC5av	Trailing	SumNC	SumNC5av
Dow Jones	2.53	2.25	2.32	69.2	61.5	57.7	1.75	1.38	1.34
NASDAQ	1.78	2.15	2.61	73.1	65.4	57.7	1.30	1.40	1.51
Nikkei 225	0.77	1.21	1.52	61.5	65.4	61.5	0.47	0.79	0.93
Hang Seng	0.90	0.37	0.03	57.7	46.2	46.2	0.52	0.17	0.02
Shanghai	1.84	1.58	1.88	69.2	65.4	61.5	1.27	1.03	1.16

The cell with the largest value among the three criteria is highlighted. The *trailing stop* criterion marks the notable *potential profit* in Dow Jones, Hang Seng, and Shanghai markets. The *SumNC5av* criterion achieves the preferable *potential profit* for NASDAQ and Nikkei 225 markets.

VI. CONCLUSION AND FUTURE WORK

This paper proposes a six-parameter model for retrieving similar candlesticks. The model also deals with the 5-day and 25-day moving averages to identify price trends in addition to a price zone where a stock price occurs. The proposed model is devised to find a buy opportunity. Since a successful stock trade is significantly depends on a sell opportunity, three criteria are proposed and profits that each criterion generates are simulated.

Bullish (uptrend) reversal candlestick patterns consisting of three candlesticks are focused to find a buy opportunity in the empirical study. Three inequality formulas are defined to abstract the bullish reversal patterns. The parameter values used in experiments are determined statistically in terms of the standard deviations of the proposed six parameters to minimize the differences of characteristics among stock markets.

The empirical results show that the US markets, i.e., Dow Jones and NASDAC, are more profitable than Asian markets, i.e., Nikkei 225, Hang Seng, and Shanghai. As for profitability of international markets, the results generally support what is stated in the paper of Dimson, Marsh, and Staunton [3]. The popular *trailing stop* criterion [2] gives the best result among the three exit criteria.

This study only focuses on the bullish reversal patterns in a downtrend, which leads to limitations of the study. Future work may include experiments on the proposed method to measure profitability of other candlestick patterns including bearish (downtrend) reversal patterns that are profitable in short position, and continuation patterns that predict a price trend is likely to remain. Additional studies need to be carried out to measure profitability of global markets to meet demands of finding the most profitable market in the world.

REFERENCES

[1] Y. Udagawa, "Statistical Analysis of Stock Profits to Evaluate Performance of Markets," The Sixth International Conference on Big Data, Small Data, Linked Data and Open Data (ALLDATA 2020) IARIA, Feb. 2020, pp. 14–21, ISSN: 2308-4138, ISBN: 978-1-61208-250-9.

[2] C. Mitchell, "Trailing Stop Definition and Uses," Available from: https://www.investopedia.com/terms/t/trailingstop.asp/ May 22, 2020.

[3] E. Dimson, P. Marsh, and M. Staunton, "Should you invest in emerging markets?" London Business School, Available from: https://www.london.edu/think/emerging-markets/ Apr. 2019.

[4] N. Ahmad, R. R. Ahmed, J. Vveinhardt, and D. Streimikiene, "Empirical Analysis of Stock Returns and Volatility: Evidence from Asian Stock Markets," Technological and Economic Development of Economy, vol. 22, Nov. 2016, pp. 808–829.

[5] "Technical Analysis," Cambridge Univ., pp. 1–179, Available from: http://www.mrao.cam.ac.uk/~mph/Technical_Analysis.pdf Feb. 2011.

[6] J. M. Horton, "Stars, crows, and doji: The use of candlesticks in stock selection," Quarterly Review of Economics and Finance, vol. 49, Nov. 2007, pp. 283–294.

[7] R. B. Marshall, R. M. Young, and R. Cahan, "Are candlestick technical trading strategies profitable in the Japanese equity market?" Review of Quantitative Finance and Accounting, vol. 31, Aug. 2008, pp. 191–207.

[8] P. Tharavanij, V. Siraprapasiri, and K. Rajchamaha, "Profitability of Candlestick Charting Patterns in the Stock Exchange of Thailand," SAGE journals, Oct. 2017, pp. 1–18.

[9] G. Caginalp and H. Laurent, "The predictive power of price patterns," Applied Mathematical Finance, vol. 5, Jun. 1998, pp. 181–206.

[10] Y.-J. Goo, D.-H. Chen, and Y.-W. Chang, "The application of Japanese candlestick trading strategies in Taiwan," Investment Management and Financial Innovations, vol. 4, Jan. 2007, pp. 49–79.

[11] C. Chootong and O. Sornil, "Trading Signal Generation Using a Combination of Chart Patterns and Indicators," International Journal of Computer Science Issues, vol. 9, Nov. 2012, pp. 202–209.

[12] C.-F. Tsai and Z.-Y. Quan, "Stock Prediction by Searching for Similarities in Candlestick Charts," Journal ACM Transactions on Management Information Systems (TMIS), vol. 5, Jul. 2014, pp. 1–21.

[13] M. Zhu, S. Atri, and E. Yegen, "Are candlestick trading strategies effective in certain stocks with distinct features?" Pacific Basin Finance Journal, vol. 37, Apr. 2016, pp. 116–127.

[14] M. Jamaloodeen, A. Heinz, and L. Pollacia, "A Statistical Analysis of the Predictive Power of Japanese Candlesticks," Journal of International & Interdisciplinary Business Research, vol. 5, pp. 62–94, Available from: https://scholars.fhsu.edu/jiibr/vol5/iss1/5/ Jun. 2018,

[15] Y. Udagawa, "Dynamic Programming Approach to Retrieving Similar Candlestick Charts for Short-Term Stock Price Prediction," International Journal on Advances in Software, IARIA, vol. 11, Dec. 2018, pp. 440-451.

[16] "Major World Market Indices," Available from: https://www.investing.com/indices/major-indices/ Dec. 2020.

[17] M. Hall, "Introduction to Swing Trading," Available from: https://www.investopedia.com/trading/introduction-to-swing-trading/ Jun. 10, 2020.

[18] S. Glen, "Statistics How To: T-Distribution," StatisticsHowTo.com, Elementary Statistics for the rest of us! Available from: https://www.statisticshowto.com/probability-and-statistics/t-distribution/ 2020.