A Stock Trading Strategy with Daily-Low-Price Predictions

Korivi Ganesh[1], Akito Sakurai[2]
Graduate School of Science and Technology, Keio University, Yokohama 223-8522, Japan

Abstract: Conventional stock price predictions or stock return predictions focus mainly on predicting closing price of stocks in a period of a day, a week, a month, or an hour. The predicted prices of a stock at time *t*+1 made at time *t* are considered to determine orders to trade stocks: to buy it, to sell it, or nothing to order. This is a natural approach but it does not utilize any data other than the closing price. In almost all records of stock price movements, e.g. daily stock price movements, high and low prices are given, which are the highest and lowest prices, respectively, in the period. We naturally believe that the more the information is used, the higher the prediction accuracy is achieved. We, therefore, focused on the high and low price and sought for a new trading strategy and found a possibly profitable rule that uses the predicted low price where the prediction is done based on historical prices. We showed the profitability of the trading strategy by obtaining profits in simulated trading on historical prices of the components in Nikkei225 index of Tokyo Stock Exchange in the period between 1986 and 2015.

Keywords: Daily-Low-Price Predictions; Stock Trading Strategy; linear regression; machine learning algorithms.

I. INTRODUCTION

Asset price prediction in developed markets has attracted much attention of many chartists, practitioners, and researchers far from the time when we have had concise records of transactions. Even now the question “To what extent can the history of a common stock's price be used to make meaningful predictions concerning the future price of the stock?” [16] remains open.

If the direction of the price movement is correctly predicted we could earn profits by buying or selling the stocks, but the prediction of the price movements as well as the price itself is very difficult. The financial markets such as stock markets in the developed countries are thought to be efficient or very close to efficient in the sense of the efficient market hypothesis [16]. In these situations, at least for research for predictions of movement, the first and the least thing to is to find if it is possible to predict the price movement in accuracy sufficient to yield profits. The methods to be used are in general some of machine learning methods to predict the price movement or the price itself to trade and counter-trade the stocks.

In earlier researches, the prediction methods adopted are as simple as a description expressing being in upward trend, downward trend, or stationary using technical indicators including utilization of moving averages. In recent researches, the methods become more complex, such as neural network [1-4, 9, 11], fuzzy systems [2, 12, 14, 15], SVM [23-25], genetic alogrithms [25-28] and others. These techniques are used to predict mainly: (1) the direction of the movement of the closing price of the stock, or (2) closing price value of the stock.

The researchers seem to believe that after the recent development of information technology and surge of involvements of many non-professional traders, the markets become more efficient, and therefore we need more sophisticated method to find any regularity or anomaly in the markets.

We thought differently in our current research. We supposed that still some simple method would work to predict the best time to buy or sell stocks by utilizing price information we could get from records of price movements, if the time is not restricted to the opening or closing of the market of the day, week, or month, but the time allowed in between by utilizing limit orders. The limit order is an order placed with a specification of the price to buy or sell. If we place a limit order to buy, it may not be executed if no one wants to sell a stock at the specified price or if the orders of the same price other than ours are executed but these counter-orders are not large enough in volumes to sell them to us. But if our price set in the limit order is appropriate, the order is executed sometime in the day. This means that if we could correctly predict the price, which would give us a profit and the order of which could be executed with high probability, we could get profits with higher probability than by just predicting the opening/closing price or its direction of movements.

In many current researches, stock prices are tracked by their closing prices of the months, days, and/or hours, and the consecutive or successive difference of a stock price sequence is the target of prediction [1-5,9,11-15]. The trading strategy behind the prediction scheme is that when the price difference is predicted to be positive (or negative) on a day, we are going to buy (or sell) a stock at the closing price of the day just before its closing and sell (or buy, respectively) it at the closing price of the next day just before its closing. Note that we assume implicitly that we are able to buy or sell at the closing price that was recorded as historical data by placing a market order. A market order is an order without a specification of price and is executed at the best available current price.

Unlike these frameworks, ours is to predict the lowest price of the day, called low price of the day, of a stock and to place a limit buy order with the predicted price and close the position by a market order at just before the closing time of the day supposing that we can sell the stock at the closing price of the day.

Our proposal is (1) to predict the lowest price of the day of a stock by using historical price data, (2) to buy the stock at the predicted price by placing a limit order and sell it at the closing time of the day, (3) the assessment of the prediction is not by its accuracy but by profits that the method yield. The points that exist in the choice (i.e., the reason why we focus on the lowest price of the day) are as follows: (1) orders based on the erroneous prediction which is lower than the actual low price of the day do not contribute to loss since the limit orders are not executed, and (2) we do not need to trade on credit, because if we are to start trading by selling stocks without having them, we need to sell on credit or to sell short. Trading on credit is done in a different scheme from spot stock trading. For example, not only we can sell stocks that we do not have, but we can buy stocks with money less than their prices. We need to assume conditions to calculate returns and risks varying on markets and exchanges.

The prediction in our proposal could be assessed by, e.g., mean squared error since it is a regressor for prices. But the error has little meaning in the setup because an outcome of the strategy is profit the entire method might yield. The profit depends on fluctuation of not only low prices but also closing prices which could be dependent. Therefore, we adopt the profit as a measure of performance.

High and low prices have been rarely paid attention to except for volatility estimation [29] and [30]. The former is interested in the fractional cointegration nature of high and low prices and the latter focuses not only its nature but also trading strategy which is a contrarian strategy for minute trading based on predicted daily H-L bands and not on daily trading.

For the prediction we used linear regression models. We tested nonlinear regression by SVR with nonlinear kernels in preliminary experiments, but the results were worse than linear regressions.

The remainder of the paper is organized as follows: Section 2 describes the Method, Section 3 describes the experiments, Section 4 presents the results and analysis, and finally Section 5 reports discussion followed by a brief conclusion in Section 6.

II. METHOD

Our proposed method consists of a linear regressor to predict low price of the day, and the trading strategy that accepts the predicted value as an estimator of the low price and places a limit order to buy a stock at the estimated low price at opening time of a market, and sell it at the closing of the market of the day. As an evaluation measure, we adopt the amount of profit obtained in simulated tradings using historical data.

As we mentioned in Introduction, there are many researches to employ more complex expressions and learning methods than just linear models to express formula for predictions. One of the merit of using these complex expressions is that they could give us and in fact gave us a good prediction by fitting more complex expressions or nonlinear models to the historical data better than fitting the linear models. In machine learning, it is accepted that more complex model could be over-fitted to the data with higher probability. Because financial time series are close to random and therefore include many noises that could be fitted to, we believe that we had better try simpler model first in any situations, i.e., we believe that a linear model has good generalization capability under very random condition as in the stock market. In fact, in this research, we find that linear model is suited in the simulated trading in experiments. One another reason why we used the linear model is that we can easily infer the most and least effective features, if exist, in the features

It is commonly accepted that the predictability of returns of financial asset should be estimated by out-of-sample tests rather than by in-sample fitting performance. Out-of-sample tests of prediction of financial time series lead to test of profitability of simulated trading based on the prediction rather than statistical measures such as mean squared error of prediction which is not directly related to profitability of the trading strategy.

In this paper, we examined and compared the goodness of prediction mainly by logarithmic return obtained by simulated trading based on the predictions.

Let us suppose that $l(d,s)$ be the low price of day d and stock s, $o(d,s)$ be the opening price of day d and stock s, and $c(d,s)$ be the closing price of day d and stock s.
The set of these values, i.e., some number of the historical values in low-price sequence $l(d-1,s)$, $l(d-2,s)$, ..., $l(d-n,s)$ for some n and the opening price of the day $o(d,s)$ and the historical sequence $o(d-1,s)$, $o(d-2,s)$, ..., $o(d-n,s)$ consist the features of a sample for the day, i.e., a vector representation of a sample for the day $x(d,s)$ in the following. The low price $l(d,s)$ of the day is the target value $y(d,s)$ of prediction to be done. The number of the historical low price values n is not predefined and therefore we conducted experiments by varying n before the targeted simulated traingn. The model to be used for prediction is a linear combination of the feature values in this research.

To simulate trading, we need to define a trading strategy based on prediction, which specifies how and how many shares of stocks we place orders to buy and sell. We also need to specify how to allocate capitals to stocks. We will not consider different stocks separately, but we do not consider management of portfolio of the stocks, because there are other problems to be solved in the management. We will adopt the following way of allocating capitals.

A. Overview of the trading strategy

There are two phases in our trading strategy:

- Phase1: At the time of opening of the stock exchange, instantly after the opening price is fixed and announced, we predict low price of the day, and we place a limit buying order at the $pred\_l(d,s)$, i.e., predicted low price of the day.
- Phase2: When the limit order is executed, we will sell them by a market order at just the closing time of the stock exchange.

Prediction is done by linear regression. For day d and stock s, the target is $l(d,s)$ and the features are:

1. $l(d-i,s)$ for $i = 1,\ldots, n-1$ for a fixed n, i.e., historical data of the target values. n is called the number of features in this paper.
2. $o(d-i,s)$ for $i = 0,\ldots, n-1$, i.e., historical data of the opening prices and the opening price of the current day.

When predicted $l(d,s)$ is higher than or equal to the actual $l(d,s)$ ($pred\_l(d,s) \geq act\_l$ (d,s)) i.e., if the actual low price is lower than or equal to the predicted low price, then the limit order is supposed to be executed or otherwise not. We also suppose that the time when we place the limit order is just the time when a market is opened, i.e., we suppose that the order is filled even when the predicted $l(d,s)$ is equal to $o(d,s)$, i.e., the low price of the day is equal to the opening price $o(d,s)$ of the day d and stock s. We further suppose that the market order placed at just before the closing time is also executed at the closing price of the stock. These conditions are customarily assumed.

For buying a stock we suppose that we buy as many shares as possible but we cannot buy fractional shares. In Tokyo Stock Exchange, although depending on stocks, we could not by a share in general. Regulation is now in transition and in a few years, all the stocks are traded in multiples of 100 shares. We supposed in the simulated trading that only multiples of 100 shares can be bought.

For the allocation of capitals, we adopt the following allocation method. We suppose that orders are placed daily, once for a stock and continue trading for years from 1990 to 2015 and that we have a fixed amount of capital which is equally allocated for each stock at the start of simulated trading. For each stock the allocated capital is solely invested in the purchase of the stock until rebalancing.

Yearly Rebalancing: At the end of every year, the capitals we have are gathered from stocks and reallocated equally to each stock.

B. Overview of the performance metrics

We consider performance metrics as logarithmic return & return on investment (ROI), as follows.

1. Logarithmic return and return on investment (ROI):

Logarithmic return and return on investment (ROI) are used as performance metrics in this paper. Because logarithmic return (log-return hereafter) and ROI is related to each other in a simple expression: log-return = log (1 + ROI), sometimes we will refer to only one of them. Log-return and ROI are defined for a period and an investment object, i.e., a stock in this paper. Because we do not keep stocks in hand overnight, i.e., we keep only cash at night from the closing till opening of market, return is clearly defined.

Suppose that we had a capital m_1 being ready for buying a stock at the start of a period, i.e., at the morning of the first day of the period, and we have a capital m_2 after the last day of the period and at the morning before the next period after all the shares that we had were sold, ROI for the stock and period is $(m_2 - m_1)/ m_1 = -1+m_2 / m_1$ and log-return is $\log (m_2 / m_1) = \log (1+$ ROI) and ROI = $-1+ \exp$ (log-return) where log is the natural logarithm.

We will define specific log-return and ROI to be used as metrics of performance in the following:

- *$ROI^{1,1day}$ and log-return^{1,1day}*:

Suppose that we had m_1 before the opening of a market, bought a stock at price p_1, and sold it at price p_2, then we have $m_2 = (p_2 - p_1) * k *100 + m_1$, where $k = m_1 \div (p_1 *100)$ and $\div$ is the integer division. Log-return and ROI for the stock and the period are defined by this m_2 and m_1 as above.

- *$ROI^{1,1year}$ and log-return^{1,1year}*:

Suppose we had capital m_1 before year y_1 and m_2 after year y_2. Log-return and ROI are simply $\log (m_2 / m_1)$ and $-1+ (m_2 / m_1)$.

- *$ROI^{1,n\ years}$ and log-return$^{1,n\ years}$*:

Suppose we had capital m_1 before year y_1, m_2 before year $y_2,\ldots,$ and m_n before year y_n, and m_{n+1} after year y_n. Then, log-return for this period is: $\log (m_{n+1} / m_1) = \log (m_2 / m_1) + \ldots + \log (m_{n+1} / m_n)$, i.e., sum of log-returns of each year. Therefore annualized log-return for the period is naturally defined as $(1/n)*\log (m_{n+1} / m_1)$. ROI for this period is $-1+ m_{n+1} / m_1 = -1+ \exp (\log (m_{n+1} / m_1))$ and corresponding annualized ROI is naturally defined by $-1+ \exp ((1/n)*\log (m_{n+1} / m_1))$.

Suppose we consider a set of stocks S, with s stocks in it, and the same amount of capital m_0 is assigned to the stocks in S.

- *$ROI^{s,1year}$ and log-return^{s,1year}*:

Because yearly rebalancing is adopted and capital m_i for stock i in S is obtained after a year, ROI of the investment for S for the year is: $(m_1 + m_2 + \ldots + m_s)/(s* m_0) -1 = (1/s)((m_1 / m_0 -1) + \ldots + (m_s / m_0 -1)) = (1/s)(\mathrm{ROI}_1^{1,1\mathrm{year}} + \ldots + \mathrm{ROI}_s^{1,1\mathrm{year}})$, where $\mathrm{ROI}_i^{1,1\mathrm{year}}$ is ROI of stock i in S for this period. Log-return for S and this period is naturally defined as:

log-return$^{s,1\mathrm{year}}$ = $\log ((1/s)(\mathrm{ROI}_1^{1,1\mathrm{year}} + \ldots + \mathrm{ROI}_s^{1,1\mathrm{year}}) +1) = \log ((1/s) (\exp (\text{log-return}_1^{1,1\mathrm{year}}) + \ldots + \exp (\text{log-return}_s^{1,1\mathrm{year}}))$, where log-return$_i^{1,1\mathrm{year}}$ is log-return of stock i in S for this period.

When we are to consider returns of a set of stocks S for years $y_1, y_2,\ldots, y_n$, log-return and ROI depends on the existence of rebalance.

- *$ROI^{s,n\ years;\ reb}$ and log-return$^{s,n\ years;\ reb}$*:

When yearly rebalancing is adopted, supposing that capital m_1 is allocated equally to the stocks in S before year y_1, m_2 before year $y_2= y_1+1,\ldots,$ and m_n before year $y_n=y_1+(n-1)$, and m_{n+1} after year y_n by virtually allocating equal capital to each stock. Then log-return$^{s,n\ \mathrm{years;\ reb}}$ is: $\log (m_{n+1} / m_1) = \log (m_2 / m_1) + \ldots + \log (m_{n+1} / m_n)$ = log-return$_1^{s,1\mathrm{year}} + \ldots +$ log-return$_n^{s,1\mathrm{year}}$, where log-return$_y^{s,1\mathrm{year}}$ is log-return for a set of stocks S and year y.

Therefore, the annualized log-return is naturally defined as $1/n$ of it. ROI for this period and corresponding annualized

- ROI is defined from the log-return:

1) $\mathrm{ROI}^{s,n\ \text{years; reb}} = \exp(\text{log-return}^{s,n\ \text{years; reb}}) - 1$, and
2) Annualized $\mathrm{ROI}^{s,n\ \text{years; reb}} = \exp((1/n)\ \text{log-return}^{s,n\ \text{years; reb}}) - 1$

III. EXPERIMENTS

We first conducted experiments to determine hyper parameters such as the number of features, the number of training samples, and the ranking of stocks according to logarithmic returns.

In general, the larger the number of training samples is, the more accurate the learned model is. But in the financial time series, such as the stock price sequence, this may not be true, because the economic conditions for the newest sample may be different from the economic conditions in the older samples in the training samples. Therefore, in our preliminary experiments, we have conducted many experiments with varying the number of training samples. We give in this paper the results for 300, 600, 900, and 1000. Also, the number of features was varied. We give results for 3, 6, 9, and 30. Since the period of 1986 to 1989 including set aside for training has only 1076 days of trading (see the next paragraph) and therefore 1000 is almost maximum.

The stock market we consider in this paper is Tokyo Stock Exchange. Because we want to show that our trading strategy is valid for many companies and long periods, we obtained daily data, i.e., open, high, low, and close prices of Nikkei 225 components from 1986 to 2015 [31]. Although Nikkei 225 contains 225 stocks at any time, the number of the companies that are listed as components throughout the period 1986-2015 is 174. In this paper, the survivorship bias is not considered, although these 174 companies may be the winners in their highly competitive economic markets. We used 1990-2015 daily data for training and prediction. 1986-1989 data were not used for prediction, but were used for training for days in 1990. This was done to make the results comparable between experiments with different length of training data.

For assessment of performance, we compared our model with four investment strategies in terms of log-returns and ROI, i.e., $\text{log-return}^{s,n\ \text{years; reb}}$ and $\mathrm{ROI}^{s,n\ \text{years; reb}}$:

Table 1. Logarithmic returns obtained by simulated trading over 1990 to 2000 with yearly rebalancing varying the number of features and the number of learning samples. The left table is of our proposed strategy and the right is of Comparison 4 (conventional strategy). These tables are used to determine the number of features and the number of learning samples to be used in main experiments.

# samples / #features	300	600	900	1000
3	0.2215	0.2310	0.2370	0.2380
6	0.2151	0.2252	0.2321	0.2321
9	0.2065	0.2222	0.2299	0.2315
15	0.1925	0.2143	0.2212	0.2229
30	0.1517	0.1852	0.2024	0.2045

Table 1(a). Proposed method

# samples / # features	300	600	900	1000
3	0.2215	0.2310	0.2370	0.2380
6	0.2151	0.2252	0.2321	0.2321
9	0.2065	0.2222	0.2299	0.2315
15	0.1925	0.2143	0.2212	0.2229
30	0.1517	0.1852	0.2024	0.2045

Table 1(b). Comparison 4 (conventional method)

Comparison 1: Interests of Japanese government bond.
Comparison 2: Buy and hold Nikkei225 index.
Comparison 3: Buy and hold components of Nikkei225.
Comparison 4: Predict the direction of closing price movement of each stock and buy it today and sell it tomorrow if the predicted direction is up. . More precisely, predict at just before the closing of market if the price will go up or down. If the price is predicted to go up, buy the stock, if otherwise, do nothing for the stock (because we are not dealing with short selling or selling on credit in this paper, we do not sell stock if the predicted direction is down). On the next day at just before the closing of market, sell the stock. We call this conventional method in the following.

IV. RESULTS AND ANALYSIS

Before the main experiments, we conducted experiments to decide the number of features and the number of training examples to be used in the following experiments. Because we are dealing with time series, we have chosen days in earlier time. For historical data between 1990 to 2000, we have conducted simulated trading varying the number of feature and training samples. The results are shown in Table 1(a). The annualized log-returns in Table 1 is of $\text{log-return}^{s,n\ \text{years; reb}}$ where S is a single-element set of each stock.

As is seen in Table 1(a), the combination of the number of features and training samples that gives the largest increase of asset, i.e. the largest logarithmic return from 1990 to 2000 among stocks is the one with 3 features and 1000 training examples. We, therefore, used the condition in the following experiments.

Then we conducted simulated trading with proposed trading strategy with prediction. Limit buy order is placed every day. If the order is executed, market sell order is placed in the day. We adopted the rolling window method. Therefore, the trading is done day by day as calendar goes.

The detail of trading process is as follows. Suppose that we are now just before opening of a market on a day d. First a prediction model for low price of day d is built by a machine learning algorithm by providing training data which are constructed solely by data prior to the day d. Just at the time of opening of the market and the opening price of the stock is determined, prediction for low price of the day d is calculated by supplying feature values based on previous data and the opening price just given. The calculation of prediction takes less than a day *d*. We place a market order to sell the stock that we bought today. The price is supposed to be the closing price. We suppose that at the very start of simulated trading, we have a fixed amount of capital which is equally divided and invested into each stock trading. In the simulation, to make the simulation closer to practice, we suppose that we invest 1,000,000 yen per few milliseconds and a limit order to buy the stock is placed. Then suppose that we are just before closing of the market on stock and we allow us to buy them by multiple of 100 stocks. In the current practice in TSE (Tokyo Stock Exchange), many stocks are bought/sold with 100 shares as unit. Although in 1990's and 2000's the situation was different, we keep this style in the simulation.

Table 2: Annualized log-return and ROI of the proposed strategy with yearly rebalancing. Each row repesents a set of stocks that are ranked in decreasing order of log-returns for 1990 to 2000. The two columns with name 2001-2015 show preformance of stocks ranked according to 1990-2000 performance. The rightmost two columns are for reference. They show performance for the whole period of simulated rading including learning and testing.

Rank for 1990-2000	Annualized log return for 1990-2000	Annualized ROI for 1990-2000	Annualized log return for 2001-2015	Annualized ROI for 2001-2015	Annualized log return for 1990-2015	Annualized ROI for 1990-2015
top1-20	0.3324	39.4%	0.0825	8.6%	0.1882	20.7%
top1-40	0.2745	31.6%	0.0712	7.4%	0.1572	17.0%
top1-60	0.2385	26.9%	0.0688	7.1%	0.1406	15.1%
top1-80	0.2088	23.2%	0.0724	7.5%	0.1301	13.9%
top1-100	0.1839	20.2%	0.0671	6.9%	0.1165	12.4%
top1-120	0.1630	17.7%	0.0650	6.7%	0.1065	11.2%
top1-140	0.1437	15.5%	0.0663	6.8%	0.0990	10.4%
top1-160	0.1244	13.2%	0.0673	7.0%	0.0915	9.6%
top1-174	0.1097	11.6%	0.0669	6.9%	0.0850	8.9%

The results are summarized in Table 2. By the simulated trading in 1990 to 2000, we calculate annualized log-return log-return$^{s,n \text{ years; reb}}$ for each stock and sort them in decreasing order to obtain ranking of stocks in terms of performance of our proposed algorithm. We gather top 20, top 40, to top 160, and all the stocks to forma sets and calculate annualized logarithmic returns and ROIs for each set, i.e., we obtained log-return$^{s,n \text{ years; reb}}$ where *S* is one of top1-20 and others. For the sets, annualized ROI for 2001 to 2015 is from 11% to 15%, which is greater than Comparison 1 to 3 shown below. Even when we invest to all the stocks, i.e., the stocks including ones that the proposed strategy may not be good at earning profit, we get about 13.8% annualized ROI which is still about twice of the largest in Comparison 1 to 4 shown below.

Table 4 is an example of results of simulated trading when trading commission is introduced. Commission proportional to the execution price of an order has been common. It is around 0.5% to 1% and cannot be covered by profits. But in these days, commission per transaction with ceiling becomes available. Some of brokerage firmas claim just around 800 yen per transaction in Japan. Table 4 is log-returns of cases when the commission is 800 yen per transaction, i.e., we need to pay each for buying and selling. The results in Table 4 show us: (1) top 20 stock group decreases ROI by around 1/7, and (2) other groups decrease ROI much. The transaction commission has been decreasing rapidly in Japan in the last few years, but from a viewpoint of automatic trading, the commission is still high.

Table 3: Annualized log-return and ROI of the conventional strategy (Comparison 4) with yearly rebalancing. Directional change of daily closing price is predicted. If the prediction is up, market buy order is placed. Each row repesents a set of stocks that are ranked in decreasing order of log-returns for 1990 to 2000. The two columns with name 2001-2015 show preformance of stocks ranked according to 1990-2000 performance is proportional

Rank for 1990-2000	Annualized log return for 1990-2000	Annualized ROI for 1990-2000	Annualized log return for 2001-2015	Annualized ROI for 2001-2015	Annualized log return for 1990-2015	Annualized ROI for 1990-2015
top1-20	0.3702	44.8%	0.1065	11.2%	0.2181	24.4%
top1-40	0.3234	38.2%	0.1123	11.9%	0.2016	22.3%
top1-60	0.2868	33.2%	0.1355	14.5%	0.1995	22.1%
top1-80	0.2600	29.7%	0.1315	14.0%	0.1859	20.4%
top1-100	0.2360	26.6%	0.1257	13.4%	0.1724	18.8%
top1-120	0.2169	24.2%	0.1229	13.1%	0.1627	17.7%
top1-140	0.1967	21.7%	0.1397	15.0%	0.1638	17.8%
top1-160	0.1760	19.2%	0.1345	14.4%	0.1521	16.4%
top1-174	0.1587	17.2%	0.1294	13.8%	0.1418	15.2%

For Comparison 1, coupon interest rate of 10-year Japanese government bond is used. To calculate returns of, for example, 15 years, buy and holding JGB of 10-year redemption period is not appropriate. We simply calculate average of log-return obtained from coupon interest from 2001 to 2015. It is 0.01162. Note that we collected coupon interest rate data from web-site of Ministry of Finance, Japan. Nikkei225 index was 13,898.09 at the opening of Jan 4, 2001 and 19,033.71 at closing of Dec. 30, 2015. Therefore buy-and-hold strategy gives us annualized log-return 0.02096 and annualized ROI 2.12%.

Supposing that equal amount of investment was done on each stock in the stock set we consider in Nikkei225 and that simply any fractional shares could be bought, we get 0.0736 for annualized log-return and 0.0764 for annualized log-return.

Comparison 4 is a trading strategy with prediction, where the prediction is done for the direction of change of daily close prices, which is a common practice in these days. Predicting model is built by linear regression. The best number of parameters and the number of learning samples are determined as our proposal. The searching results is in Table 1(b) which shows that the number of features is 3 and the number learning samples is 1000 is to be used.

Table 4: Annualized log-return and ROI of the proposed strategy with yearly rebalancing. Different from the results in Table 2 is that in this table transaction commission is charged for each transaction. Some brokerage firm request 800 JPY + tax as maximum for a transaction, which corresponds to 1600 JPY + tax per day when traded in our strategy.

Rank for 1990-2000	Annualized log return for 1990-2000	Annualized ROI for 1990-2000	Annualized log return for 2001-2015	Annualized ROI for 2001-2015	Annualized log return for 1990-2015	Annualized ROI for 1990-2015
top1-20	0.3082	36.1%	0.0820	8.5%	0.1777	19.4%
top1-40	0.2504	28.5%	0.0926	9.7%	0.1594	17.3%
top1-60	0.1992	22.0%	0.0526	5.4%	0.1146	12.1%
top1-80	0.1508	16.3%	0.0200	2.0%	0.0753	7.8%
top1-100	0.1029	10.8%	0.0039	0.4%	0.0458	4.7%
top1-120	0.0557	5.7%	-0.0011	-0.1%	0.0229	2.3%
top1-140	0.0166	1.7%	-0.0038	-0.4%	0.0048	0.5%
top1-160	-0.0176	-1.7%	-0.0063	-0.6%	-0.0111	-1.1%
top1-174	-0.0404	-4.0%	-0.0074	-0.7%	-0.0214	-2.1%

Simulated trading results of Comparison 4 based on conventional strategy is in Table 3. In all the cases, the proposed method outperforms Comparison 4 by 50% to 100%. When commission is charged, the difference of returns is larger. If we invest in top 20, 40, or 60 stocks where they are ranked to 1990 to 2000 performance the propsed method yield positive returns whereas Comparison 4 (conventional method) yield negative returns in top 20 stocks.stocks.

To evaluate the performance of our proposed strategy we use logarithmic return, or simply speaking, profits obtained by the strategy at the simulated trading of components of Nikkei 225 in Tokyo Stock Exchange for 1990 to 2000 and for 2001 to 2015.

Table 5 Annualized log-return and ROI of the Comparison 4 (conventional) with yearly rebalancing. Different from the results in Table 3 is that in this table transaction commission is charged for each transaction. Some brokerage firm request 800 JPY + tax as maximum for a transaction, which corresponds to 1600 JPY + tax in our strategy.

Rank for 1990-2000	Annualized log return for 1990-2000	Annualized ROI for 1990-2000	Annualized log return for 2001-2015	Annualized ROI for 2001-2015	Annualized log return for 1990-2015	Annualized ROI for 1990-2015
top1-20	0.1212	12.9%	-0.0239	-2.4%	0.0375	3.8%
top1-40	0.0078	0.8%	-0.0358	-3.5%	-0.0173	-1.7%
top1-60	-0.0503	-4.9%	-0.0309	-3.0%	-0.0391	-3.8%
top1-80	-0.0893	-8.5%	-0.0288	-2.8%	-0.0544	-5.3%
top1-100	-0.1162	-11.0%	-0.0276	-2.7%	-0.0651	-6.3%
top1-120	-0.1379	-12.9%	-0.0263	-2.6%	-0.0736	-7.1%
top1-140	-0.1584	-14.6%	-0.0250	-2.5%	-0.0814	-7.8%
top1-160	-0.1743	-16.0%	-0.0249	-2.5%	-0.0881	-8.4%
top1-174	-0.1891	-17.2%	NA	NA	NA	NA

While the evaluation usually includes statistical tests, we do not do them in this paper. Because standard statistical tests for stock price predictions now accepted widely are the Directional Accuracy (DAC) test proposed by Pesaran and Timmermann [17] and Excess Profitability test proposed by Anatolyev and Gerko [18]. These tests, however, suppose that the prediction to be tested is done in the case that the price may go up or down, i.e., the true direction of the price movement may be positive or negative and the prediction is, naturally, done for the direction, positive or negative. In our proposed strategy, true movement of price is always going up, i.e., the price at which we are going to buy a stock is the lowest price of the day, that is, there is no other price higher than it in the day and therefore the closing price at which we are going to sell the stock is always higher than the price at which we bought the stock. The first and firm restriction that the DAC test requires us to follow is the price change (the price difference between the first action for the stock and the second action for it) must be positive and negative. It does not suppose that true price change is always positive, simply because in that case it is always the case that we should predict that the direction is positive. We surely predict so and always try to buy a stock and will never try to sell a stock at the start, although there may be chances that we cannot actually buy a stock because our proposed price in the limit order is lower than the actual low price of the day. In this case, the true sign of price change and the predicted sign are the same all the time, but it does not mean that the predictions are correct. True evaluation should not be done on the true low price but should be done on predicted low price.

From the above argument, it may be considered that the sign of true closing price minus predicted low price could be tested based on binary distribution. The problem is that we cannot know true parameter of its distribution, if we know the distribution. There may be the case that the true positive rate is different from 50%. Because we do not know the positive true rate nor a way to estimate it, we cannot statistically test the prediction.

The profitability test also requires that the price changes should not be persistently in the same direction, which hinders application.

V. DISCUSSION

If the proposed method is applied to all the components in Nikkei225, the annualized log return is 0.1294 or ROI 0.138 for 2001 to 2015, which is higher than the annualized log - return of buy-and-hold strategy applied to the components of Nikkei225, buy-and-hold of Nikkei225 index, and 10-year Japanese Government Bond.

Nikkei225 index is closed on the first trading day of 2001 at 13,691.49 which is fallen from 19,002.86 the first closing price of 2000. Nikkei225 has recovered in 2015 to close at 19,033.71 at the end of the year. The total gain in the period 2001 to 2015 is just 39% which is 2.22% annually. The index has been repeatedly going up and down, so that simple buy-and-hold strategy does not give steady earnings. Our proposed method overcomes this staggering situation to give annualized ROI 13.8%. Japanese Government Bond too is low in earnings. Its average rate during the period is 1.16% per year. The results show that our proposed method clearly outperformed this, too.

VI. CONCLUSION

In this paper, we proposed a new trading strategy of stocks and examined its effectiveness by applying it to simulated trading of stocks of the companies listed as components of Nikkei225 of Tokyo Stock Exchange from 1986 to 2015. The annualized logarithmic return of our proposed trading strategy investing in all the stocks in Nikkei225 for 2001 to 2015, where the invested capital is rebalanced at the end of each year by allocating equally to the stocks for the next year, is 0.1294 which is greater than 0.02096 that Nikkei225 index (13,898.09, opening price of Jan 4, 2001 to 19,033.71, closing price of Dec. 30, 2015) gives as annualized logarithmic return. It is also greater than 0.01162, average of logarithmic annual return of 10-year Japanese Government Bond from 2001 to 2015. Our proposed trading strategy is more profitable, by about twice depending on the groups of stocks, than a trading strategy based on the prediction of directional changes of daily closing price, which is a commonly adopted scheme.

VII. REFERENCES

[1]. H. B. Kekre, Hitesh Makhija, and Pallavi N. Halarnkar. “Prediction of Closing Price of Stock Using Artificial Neural Network”. International Journal of Scientific & Engineering Research, Volume 5, Issue 3, 2014.

[2]. Mustain Billah, Sajjad Waheed, and Abu Hanifa. “Predicting Closing Stock Price using Artificial Neural Network and Adaptive Neuro Fuzzy Inference System (ANFIS): The Case of the Dhaka Stock Exchange”. International Journal of Computer Applications, Volume 129, No.11, November 2015.

[3]. Adam Fadlalla and Farzaneh Amani. “Predicting Next Trading Day Closing Price of Qatar Exchange Index Using Technical Indicators and Artificial Neural Networks”. Intelligent Systems in Accounting, Finance and Management, Volume 21, Issue 4, October/December 2014, Pages 209–223.

[4]. Jun Zheng. “Forecast of Opening Stock Price Based on Elman Neural Network”. Chemical Engineering Transactions, Vol. 46, 2015, Pages 565 – 570.

[5]. Chenggang Li, Bing Yang, and Min Li. “Forecasting Analysis of Shanghai Stock Index Based on ARIMA Model”. MATEC Web of Conferences, Article Number: 02029, Volume 100, 2017, DOI: 10.1051/matecconf/201710002029.

[6]. Mustafa Göçken, Mehmet Özçalıcı, AslıBoru, and Ayşe TuğbaDosdoğru. “Integrating metaheuristics and Artificial Neural Networks for improved stock price prediction”. Expert Systems with Applications, Volume 44, February 2016, Pages 320-331.

[7]. Rahim Ghasemiyeh, Reza Moghdani and Shib Sankar Sana. “A Hybrid Artificial Neural Network with Metaheuristic Algorithms for Predicting Stock Price”. Journal Cybernetics and Systems, Volume 48, Issu 4, 365-392, 2017

[8]. Esmaeil Hadavandi, Hassan Shavandi, and Arash Ghanbari. “Integration of genetic fuzzy systems and artificial neural networks for stock price forecasting”. Knowledge-Based Systems, Volume 23, Issue 8, December 2010, Pages 800-808.

[9]. Jigar Patel, Sahil Shah, Priyank Thakkar, and K Kotecha. “Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques”. Expert Systems with Applications, Volume 42, Issue 1, January 2015, Pages 259-268.

[10]. Mohammad Sultan Mahmud, and Meesad. “An innovative recurrent error-based neuro-fuzzy system with momentum for stock price prediction”. Soft Computing, Volume 20, Issue 10, pp 4173–419, October 2016.

[11]. Rajashree Dash, Pradipta Kishore Dash. “Efficient stock price prediction using a Self-Evolving Recurrent Neuro-Fuzzy Inference System optimized through a Modified Differential Harmony Search Technique”. Expert Systems with Applications, Volume 52, 15 June 2016, Pages 75-90.

[12]. Eugene F. Fama. The Behavior of Stock-Market Prices. Journal of Business, Volume 38, Issue 1 (Jan., 1965), 34-105

[13]. M. Hashem Pesaran and Allan Timmermann. A Simple Nonparametric Test of Predictive Performance. Journal of Business & Economic Statistics, Vol. 10, No. 4 (Oct., 1992), pp. 461-465

[14]. Stanislav Anatolyev and Alexander Gerko. A Trading Approach to Testing for Predictability. Journal of Business & Economic Statistics, 2005, vol. 23, 455-461

[15]. J. T.-Y. Kwok, “The evidence framework applied to support vector machines,” IEEE Transactions on Neural Networks, vol.11, no. 5, pp. 1162–1173, 2000.

[16]. L. Cao, “Support vector machines experts for time series forecasting,” Neurocomputing, vol. 51, pp. 321–339, 2003.

[17]. S Deng, K Yoshiyama, T Mitsubuchi, A Sakurai. Hybrid method of multiple kernel learning and genetic algorithm for forecasting short-term foreign exchange rates. Computational Economics 45 (1), 49-89

[18]. D Shangkun, S Yizhou, S Akito. Robustness test of genetic algorithm on generating rules for currency trading. Procedia Computer Science 13, 86-98, (2012)

[19]. D. de la Fuente, A. Garrido, J. Laviada, andA.G´omez, “Genetic algorithms to optimise the time to make stock market investment,” in Proceedings of the 8thAnnualGenetic and Evolutionary Computation Conference (GECCO ’06), pp. 1857–1858, ACM, Seattle,Wash, USA, July 2006.

[20]. F. Allen and R. Karjalainen, “Using genetic algorithms to find technical trading rules,” Journal of Financial Economics, vol. 51, no. 2, pp. 245–271, 1999.

[21]. [29] Jozef Baruník and Sylvie Dvořáková. An empirical model of fractionally cointegrated daily high and low stock market prices. Economic Modelling 45 (2015) 193–206

[22]. Massimiliano Caporin, Angelo Ranaldo, and Paolo Santucci de Magistris. On the predictability of stock prices: A case for high and low prices. Journal of Banking & Finance 37 (2013) 5132–5146

[23]. Market CD-ROM, Japanese Stocks. Pan Rolling, Inc., Tokyo, Japan