2015 IEEE First International Conference on Big Data Computing Service and Applications

Fast Quantitative Analysis of Stock Trading Points in Dual Period of DMAC

Yi Wei
Department of Computer Science and Engineering,
University at Buffalo, SUNY
yiwei@buffalo.edu

Vipin Chaudhary
Department of Computer Science and Engineering,
University at Buffalo, SUNY
vipin@buffalo.edu

Abstract**—**We propose a novel volatility segmentation approach to detect effective trading points from 2679 stocks of NASDAQ. The buy and sell points are derived from dual periods of DMAC based on daily and weekly periods by estimating the amplitude and interval. The proposed approach is very accurate in that only 373 stocks (out of 2679) in NASDAQ have the average rate of profit of overall buy points higher than 3% and only 193 stocks (out of 2679) in NASDAQ have the average rate of stop-loss of overall sell points higher than 3%. The volatility segmentation approach reduces the uncertainty of stock estimation in single period DMAC. This approach, however, is very computationally intensive requiring 2382.82s for evaluating buy points and 2688.53s for evaluating sell points. A parallel implementation using 240 cores reduced the time to 16.01s and 13.50s, respectively.

Keywords**—***Stock Trading Points; Dual Period DMAC; Parallel Computing; Quantitative Analysis*

1. INTRODUCTION

It is important for investors to have the ability to correctly determine the effective trading points for stocks. In order to achieve this goal, many researchers have made efforts such as the US expert investor J.E.Granville who had proposed the concept of moving average (MA) in the mid 20th century. Adopting the principle in statistics, the stock price averages over one period of time can be drawn as a curve, which indicates the trend of historical stock price fluctuations so that it reflects future changing trend in stock price [9]. MA is widely used to describe stock trend and extract stock trading points as one of the important technical indicators.

SGM Fifield et al. have examined the performance of moving averages rules for 15 emerging stock markets and 3 developed stock markets. The results indicate that the return behavior of the emerging markets studied differed markedly from that of their developed market counter parts. It also indicates that the emerging markets have more persistent profitability comparing with developed stock markets [8].In addition, M.Metghalchi, et al. have studied the data between 1990 and 2006 in 16 European markets, the result shows that MA rules indeed have predictive power being able to discern recurring price patterns for profitable trading [12].

In a sense, high-performance computing (HPC) technology has deepened people's understanding of certain areas, the special advantages of this technology allows us to reveal the internal law of things from some aspects [11]. In the stock market, in many cases, it is difficult, in a very short period of time to analyze the overall stock market data, and establish appropriate trading strategy. However, high-performance computing technology solves this problem; it meets the needs of fast processing, analysis and integration on massive stock market data. For instance, the technique of stream data analysis has been applied in the stock market, the clustering of stock rates can be generated by calculating stock stream data, so as to discover the evolution similarity of stock data streams over time in a specific sense [7]. Related research results have contributed to provide guidance for investors to develop reasonable trading strategies.

In this study, we utilize the dual period DMAC (Dual Moving Average Crossover) to quantitatively analyze 2679 NASDAQ stocks using parallel computing via Matlab clusters. This enables quantitative analysis of stocks to eventually extract the required effective trading points. We propose using an HPC based dual period DMAC approach to determine the effective buy points that generate more profit and the effective sell points that prevent further loss, as well as to shorten the running time on massive stock data analysis.

2. METHODOLOGY

2.1 Stock Data

In this study, 2679 selected NASDAQ stocks data from October 31, 2011 to October 31, 2014 including 756 trading days are analyzed, they are collected from Yahoo Finance website http://finance.yahoo.com/. The selected in-market trading data are based on the opening price and the closing price items for each of stocks. Some factors in terms of ex-right, ex-dividend and etc. are not taken into consideration for data testing.

- Data sequence for stock price

  For real-time simulation, each transaction split from the historical data downloaded from Yahoo Finance website will be integrated into the data sequence, in which its size is supposed to be gradually increased by adding incoming data.

978-1-4799-8128-1/15 $31.00 © 2015 IEEE
DOI 10.1109/BigDataService.2015.63

When the selected NASDAQ stocks data is analyzed per second, each trading day may generate data exceeding 893.49GBper day.

- Data sequence of DMAC

For the dual moving average crossover analysis, all occurrences of the golden cross and dead cross are found from the collected real-time data sequence, and their extracted information are stored as a new independent sequence for further analysis.

2.2 Computing Platform

G.Sharma and J. Martin have illustrated the effective use of Matlab parallel computing in terms of design goal, framework, infrastructure and script language and its broad impact on big data analysis [14]. The cluster facilities for data analysis in our research are provided by the Center for Computational Research (CCR) of University at buffalo. Its clusters have more than 8000 cores CPU, with 48G memory per node and 40GB/s InfiniBand network connections. The resource management system for task scheduling we used in clusters is Simple Linux Utility for Resource Management(SLURM), the data and scripts are submitted to clusters from the front-end server, which will be running on the nodes that deployed the Matlab Distributed Computing Server(MDCS).

2.3 Technical Indicators

- Moving Average

Moving average is one of today's most commonly used technical indicators, it helps traders to identify current trend, predict the future trend and discover turning point that is about to appear. Moving average is a calculation to analyze price data points in specific period by establishing a series of averages of different subsets of the full price data set. It can be classified as simple moving average (SMA) , exponential moving average (EMA) and etc.

The commonly used MA indicator is used to be displayed in 5 days (MA5), 10 days (MA10), 30 days (MA30), 60 day (MA60), 120 days (MA120) and 250 days (MA250). Especially, MA5 and MA10 are used for short-term trading, known as the day MA indicator; MA30 and MA60 are used for mid-term trading, and known as the seasonal MA indicator; MA120 and MA250 are used for long-term trading, known as the annual MA indicator.

- Dual Moving Average Crossover(DMAC)

K.Miwaand Ueda's research shows that the daily DMAC have remarkable influence on stock prediction, it can be used as an effective indicator for technical analysis in stock trading [13]. We have adopted the DMAC technical indicators, because when daily MA and weekly MA are in the same interval, the golden cross and dead cross simultaneously generated by them will exclude those non-resonant MA crossovers . Upon that, we will describe this occurring phenomenon as the resonance of dual period DMAC. In our study, the selected trading points(TP) combined with volatility segmentation approach will eventually produce the so called effective resonance trading points(ERTP). These points are an important basis for quantitative analysis in stock amplitude and interval in the next section of this paper.

2.4 Research Workflow

In our research, we have established the entire process including stock data collection, stock trading point extraction, and quantitative analysis. Its purpose is to compare the difference between single-period DMAC and dual period DMAC. After receiving the stock data at a certain time, the data will be firstly split into several data chunks which will be distributed to each node in clusters for regression analysis, through the parameter estimation and verification of polynomial model, the data chunk on each node will be fitted and produce trend curve. Then by applying the differential to each stock data, the corresponding local extrema points of each stock can be discovered, so that the stock price volatility can be segmented. On the other hand, we need to extract the daily and weekly trading points respectively by using MA turning points to find resonance trading points of dual period DMAC. Finally all effective resonance trading points of the analyzed stocks are stored for quantitative analysis. By determining the volatility distribution of amplitude and interval for all stock trading points, we can establish the proper stock dataset which has trading points that can help to make profits and implement stop-loss (Fig. 1).

3. IMPLEMENTATION

3.1 Regression Analysis

In order to find out the stock data trend, we need to re-construct the obtained data using regression. In this process, the system will eliminate the large volatile data point as noise. Thus when analyzing data using polynomial regression, we assume the mean of random error is 0, define the variance as σ^2 which is independent with the defined data point x_i, the noise is defined as ϵ_i, and the model is defined as:

$$y_i = \sum_{i=1}^{p} a_i x^i + \epsilon_i \tag{1}$$

Where i = 1,2, ... n, p is the highest order, whether the change of data trend can be reflected correctly depends on how the parameter p is determined. Parameter estimation mainly has direct and indirect methods, the prior includes least-square method, Yule-Walker equations method, Ulrych-Clayton method and etc., the latter includes LUD, BSMF, Burg method etc. In this study, we prefer to use least-square method which is

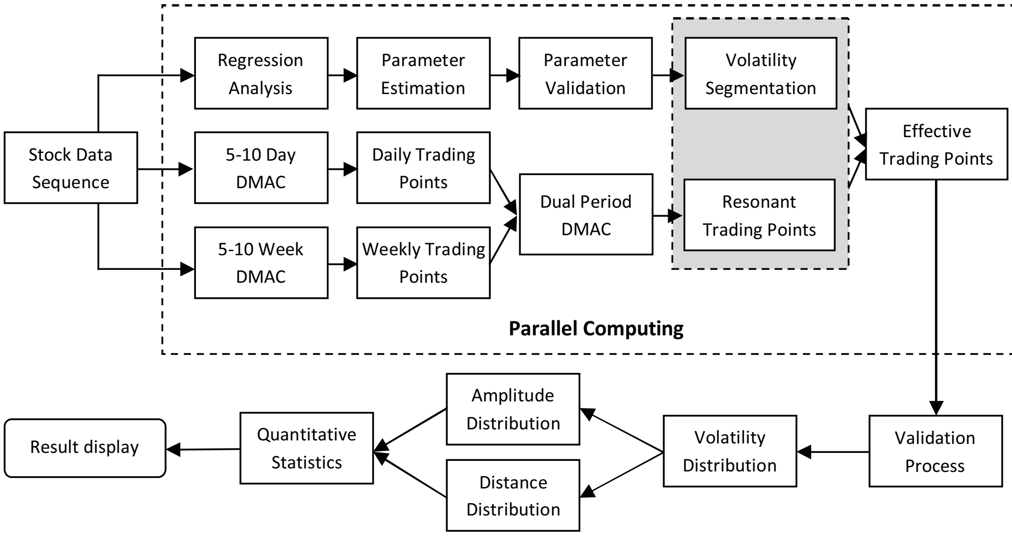


Fig. 1. Research workflow chart

Note: Shaded area generates the interval of effective resonance trading point.

a simple parameter estimation, with an unbiased estimation, the equation is as follows:

$$Y = X\Phi + a \tag{2}$$

Where, $X = \begin{bmatrix} x_n & x_{n-1} & \cdots & x_1 \\ x_{n+1} & x_{n+2} & \cdots & x_2 \\ \vdots & \vdots & \ddots & \vdots \\ x_{N-1} & x_{N-2} & \cdots & x_{N-n} \end{bmatrix}$,

$Y = [x_{n+1} x_{n+2} \ldots x_N]^T$, $\Phi = [\varphi_1 \varphi_2 \ \ldots \ \varphi_N]^T$,
$a = [a_{n+1} a_{n+2} \ldots a_N]^T$

So, the least-square estimation of parameter Φ is:

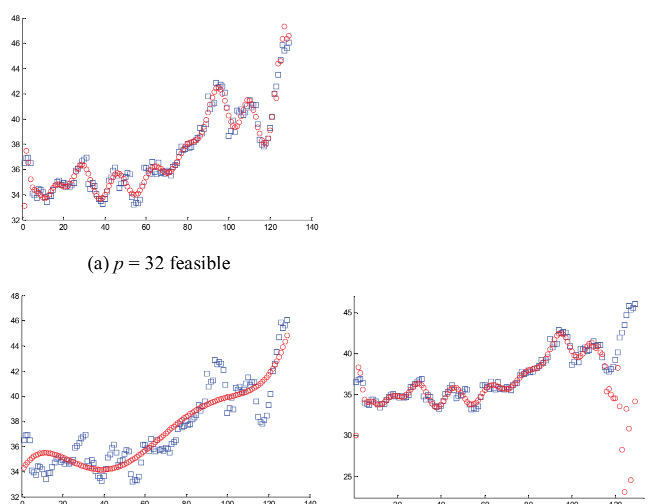
$$\Phi = (X^T X)^{-1} X^T Y \tag{3}$$

3.2 Noise Elimination

Some stock price data have higher volatility, which has strong interference to the changing trends, these data points are treated as noise. The existence of noise will increase the divergence of the fitted data and real price data, making it impossible to find the relatively accurate extrema, which leads to improper segmentation on price fluctuations. Sometimes one single volatile point may destroy the fitting seriously, so that we cannot properly extract the effective trading points. About stock noise elimination, Band and Russell have done research to separate microstructure noise from stock price volatility [5].

3.3 Parameter Estimation and Validation

In order to guarantee that the fitting results reflect the change trend of stocks correctly, the criterion functions are applied to validate the estimated parameters. In our research, we leverage three criterion functions including FPE [2], AIC [3] and BIC [4] to validate the estimated parameter, its result shows they have consistency on validation (Fig. 2).



(a) $p = 32$ feasible

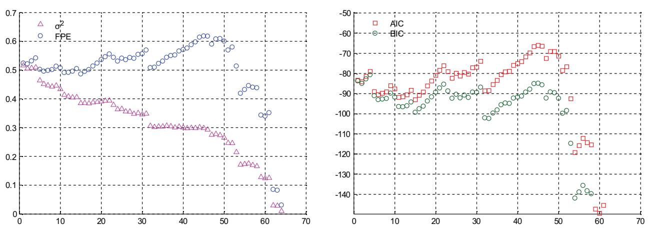
(b) $p = 5$ deviation

(c) $p = 60$ overfitting

Fig. 2. Comparison chart of different parameter estimation

Note: ○represents fitting data, □represents real data.

The noise data should be eliminated appropriately before estimating the order p of polynomial parameter to reduce the interference error from the data fitting with excessive volatility, and so that to improve the accuracy of the model fitting. However, when the parameter is set too small [Fig. 2 (b)], the data curve becomes too smooth and the deviation becomes large, which cannot reflect the real changes in the stock price; when the parameter is set too large [Fig. 2 (c)], the data curve has over-fitting phenomenon, because fitting on the price changes is sensitive to the response of fluctuations, and the deviation will increase either. Therefore, it is necessary to select appropriate parameter to simulate real fluctuations of stock data [Fig. 2 (a)], so that the deviation can be reduced. After the noise elimination process, the data analysis can take the advantage from the weakened noise influence, so as to avoid such interference due to excessive deviation caused by higher volatility.



(a) σ^2 and FPE (b) AIC and BIC

Fig. 3. Validation curve of criterion function

Note: x-axis is parameter p, y-axis is error.

From the validation results of criterion function, we can conclude that:

- With the increase of polynomial order, the overall trend of σ^2 goes downward (Fig. 3), when p is small, it descends rapidly. With the increase of p value, it will decrease in exponentially rate until n = N / 2 (N is even number), σ^2 is reduced to 0.
- The changing trend of criterion function keeps the consistency with that of σ^2. In contrast with the results of the criterion function, we can choose the appropriate order for polynomial regression of stock price data.

The selected parameters has generated the analysis results which can be seen from the validation curve of criterion function that (Fig. 3), although the variance σ^2 can be reduced ideally to 0 with the increase of the polynomial order. But with such a high order of the model, the computational cost becomes inevitably high. Considering the tradeoff between computing speed and fitted effect, we need to choose the parameter with relatively minimal error and moderate value. Therefore, the use of validation criterion function can be used to determine a reasonable order p for polynomial model, and help to generate accurate trend curves which reflect the price fluctuations.

3.4 Volatility Segmentation

In order to find the independent changing interval in the price data, the differential equations are applied to extract extrema, i.e., to determine the beginning and end of each interval, so that we can calculate and verify the trading point within its interval. In the analysis, we will extract all the relatively highest price(maxima) and relatively lowest(minima) in pairs from fitted price fluctuation curves, in which two adjacent maxima and minima will establish one tradable period of time. Therefore, to track the internal changes of stock data, the price fluctuation data need to be differentiated. The results show that the zero crossing point is what we are looking for - local extrema.

E.Keogh et al. have introduced a novel and more accurate online algorithm for efficient and effective segmentation in piecewise linear representation [11]. Because the characteristics of stock data changes distinctly with other time-series data in its representations, several high level representations in segmentation are proposed such as Fourier Transform [1] and Wavelet Transform [6], however their variation of amplitude changes almost randomly and can hardly find matching or duplicated patterns in segmentations. For simplicity, we have adopted differential fitting for time-series stock sequence to segment stocks movements in their volatilities.

Define the stock price equation with equal time interval as:

$$x_k = x_0 + kh \ (k = 0,1, \dots, n) \tag{4}$$

Where x_0 is the initial stock price with time interval h at time k, as the forward differential is commonly equivalent to the differential used in discrete data, the differential equation is defined as:

$$\Delta f(x_k) = f(x_{k+1}) - f(x_k) \tag{5}$$

Given equation (5), the differential stock data set is：

$$S_{\Delta c} = \frac{dS_c}{d\tau_{Close}} = \{ \Delta f(x_1) \cup \Delta f(x_2) \cup \dots \cup \Delta f(x_{k-1})\} \tag{6}$$

$S_{\Delta c}$ is the differential dataset on stock close price, in which the volatility segmentation plays an important role in examining stock trading points, because it determines how to correctly divide continuous stock price period into several movement intervals, and also determines whether the trading points have practical operability. The stock price data are divided into segments by using differentials, it reveals the local information including the associated upward and downward movements in each segment. Taking an example of stock YHOO, the data through 01/01/2014 to 08/10/2014, a total of 152 trading days are implemented segmentation, we can see the overall trend of continuous movement of relatively upward and downward (Fig.4). According to the result of volatility segmentation, we can examine the trading points selected by dual period DMAC.

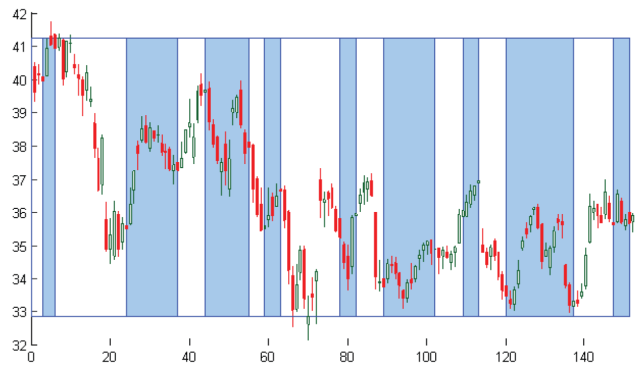


Fig. 4. Volatility segmentation graph of YHOO stock price.

Note: x-axis is date p, y-axis is stock price.

4. Results

4.1 Trading Points Validation of DMAC

As mentioned earlier, the trading points of daily MA and weekly are selected from their daily DMAC and weekly DMAC respectively, while the select of resonant trading points of dual period DMAC depends on whether the daily DMAC and weekly DMAC have overlapped resonance intervals. However, the generated trading points for buy and sell cannot guarantee a higher probability of profit or stop loss, the in-depth analysis and selection on their price amplitude and trading interval are necessary, so as to ensure that they have the potential for profit or stop loss. Such trading points are defined as effective buy point and effective sell points respectively. When the trading points of daily DMAC and weekly DMAC appears simultaneously in same interval, we define such points as the resonance trading points (RTP) of dual period DMAC.

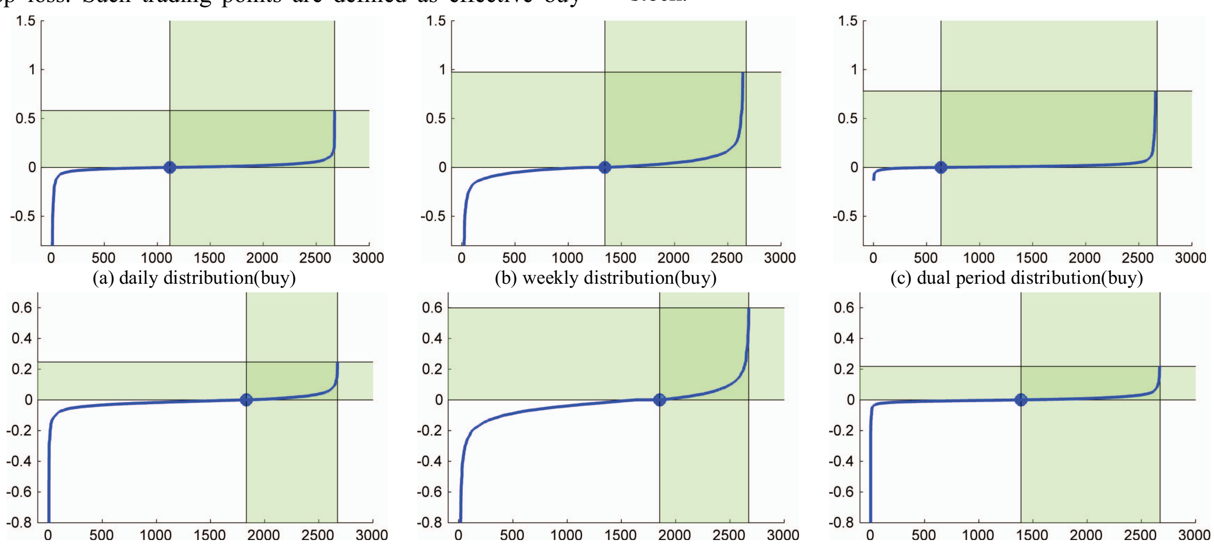
We define the local extreme in differential as the segmentation boundary, the distance between two adjacent local extrema is defined as the segmentation interval. By using regression analysis, we will obtain the local extrema of stock price, it is the segmentation boundary that validates and examines effectiveness of the resonance trading points selected by dual period in associated segmentation interval in each period, thus the resonance trading points is determined by the relative distance of segmentation interval and local extrema.

In our research, stock amplitude and interval are used to estimate the performance of resonance trading points, which are defined as follows:

Amplitude is the stock price difference between the price at resonance trading point and the price at its neighboring segmentation boundary.

Interval is the stock time span between the date at a resonance trading point and the date at its neighboring segmentation boundary.

After the analysis on selected stocks, we can obtain the trend changes of effective trading points in daily DMAC, weekly DMAC and dual period DMAC, which we call it as stock distribution, i.e., the curve of average amplitude in increasing order of the resonance trading points of every single analyzed stock.



(a) daily distribution(buy) (b) weekly distribution(buy) (c) dual period distribution(buy)

(d) daily distribution(sell) (e) weekly distribution(sell) (f) dual period distribution(sell)

Fig. 5. Trading Points Validation of DMAC

Note: x-axis represents each single stock, in Fig.5 (a), (b), (c), y-axis is the average rate profit, in Fig.5 (d), (e), (f), y-axis is the average rate of stop-loss.

The stock distribution with RTP including ETP and non-ETP illustrates that (Fig. 5), the dual period DMAC indicator is able to make up the deficiency of daily DMAC and weekly DMAC in stock selection. From the generated stock distributions in daily and weekly, the number of selected stocks by either buy points or sell points is less than that of dual period DMAC. According to the results, the primary conclusion is that to some extent, the suggested buy signal and sell signal imparted by dual period DMAC appears earlier in time than that of single period DMAC in terms of daily and weekly DMAC. Therefore, we can make the assumption that regardless of the number of selected stocks and the early warning of real-time trading, as an indicator, dual period DMAC indicator has better performance than single period DMAC indicator.

Through the examination of ETP of dual period DMAC), a few number of buy points of daily MA satisfies profitable condition [Fig. 5 (a)], their amplitude are small. Similarly, the number of buy points of weekly MA [Fig. 5 (b)] is small as well, but their amplitude is large, indicating that those stocks with weekly MA have more profitable space in longer intervals.

From the analysis result, the total number of stocks with sell points selected by dual period DMAC is more than that of single period DMAC [Fig. 5 (f)], it is possible in the situation that the frequency of stock price going down is higher than that of going up. In addition, in this case, it also illustrates that it has higher practical value for the dual period DMAC as a trading indicator.

4.2 Data and Statistics

Through the analysis of amplitude and interval, we can further verify the reliability on whether resonance trading points are able to make profit or stop loss. Therefore, the performance of overall profit and the performance of overall stop-loss are determined by the amplitude and interval simultaneously. The average amplitude represents the average of price difference for one stock happened within the time span between all its resonance trading points and its corresponding neighboring segmentation boundaries in a selected period of time. When average amplitude imparts signals, we will know how much the estimated profit obtained or loss prevented. The average interval represents the average of time intervals for one stock happened within the time span between all its resonance trading points and its corresponding neighboring segmentation boundaries in a selected period of time. When average interval imparts signals, we will know how long it takes to generate estimated profit or to reduce the losses.

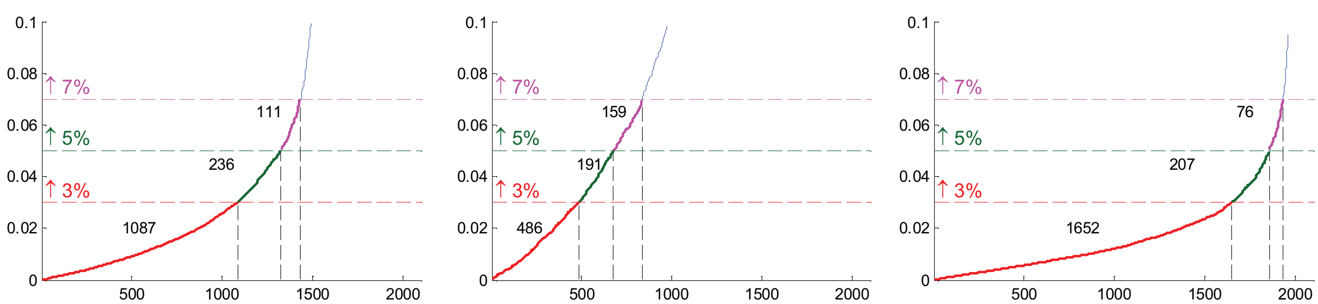
4.2.1 Amplitude Analysis

The effectiveness validation of stock trading points is an important step of quantitative analysis in our research. Its implementation has guaranteed the in-depth information retrieval to the analyzed data results and the understanding of comprehensive price trend of stock market. To achieve the above goal, the local price extrema points should be estimated and extracted. Let N be the total number of trading points of one stock, the opening price is defined as α, and the closing price is defined as β, and let γ be the price at segmentation boundary, and the average rate of profit(ARP) is defined as δ_{profit} :

$$\delta_{profit} = \frac{\sum_{i=1,j=1}^{N,M}[(\beta_i-\alpha_i)/\alpha_i-(\gamma_j^{high}-\beta_i)/\beta_i]}{min(N,M)} \qquad (7)$$

The ARP represents the average price difference of all buy points and its corresponding local neighboring maxima, which represents the consolidated earnings performance of the buy points that one stock may have[equation (7)].

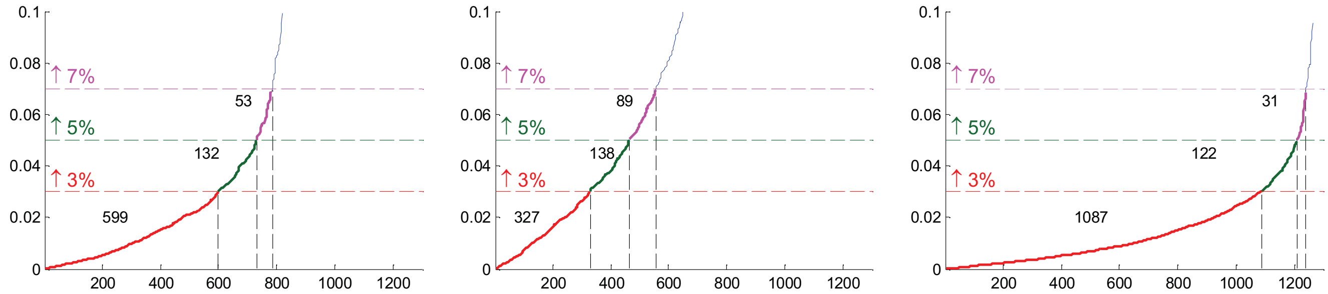
From the stock distribution of profit trading points (Fig. 6), it can be seen that dual period DMAC approach has selected 1652 stocks with ARP between 0%-3% [Fig.6 (c)], taking 61.66% as majority, while daily DMAC approach has selected 1087 stocks with ARP between 0%-3% [Fig.6 (a)], and weekly DMAC approach has selected 486 stocks with ARP between 0%-3% [Fig.6 (b)].The above figures indicate that the dual period DMAC approach has the ability to select stocks that generate higher investment gain.



(a) daily ARP of buy points (b) weekly ARP of buy points (c) dual period ARP of buy points

Fig. 6. ARP distribution of effective buy points

Note: x-axis represents each single stock, y-axis represents all positive ARP, and the dash line represents the different stages of ARP.



(a) daily ARS of sell points (b) weekly ARS of sell points (c) dual period ARS of sell points

Fig. 7. ARS distribution of effective sell points

Note: x-axis represents each single stock, y-axis represents all positive ARS, and the dash line represents the different stages of ARS.

The average rate of stop-loss(ARS) is defined as $\delta_{stop-loss}$:

$$\delta_{stop-loss} = \frac{\sum_{i=1,j=1}^{N,M}[(\beta_i - \alpha_i)/\alpha_i - (\gamma_j^{low} - \beta_i)/\beta_i]}{min(N,M)} \quad (8)$$

The ARS represents the average price difference of all sell points and its corresponding local neighboring minima, which represents the consolidated stop-loss performance of the sell points that one stock may have[equation (8)].

From the stock distribution of stop-loss trading points(Fig.7), it can be seen that dual period DMAC approach has selected 1087 stocks with ARS between 0%-3% [Fig. 7 (c)], while daily DMAC approach has selected 599 stocks with ARS between 0%-3% [Fig. 7 (a)], and weekly DMAC approach has selected 327stocks with ARS between 0%-3% [Fig. 7 (b)], the above figures indicate that the dual period DMAC approach has the ability to select stocks with better performance of stop-loss.

When ARP and ARS of stocks are arranged in increasing order, their trend distribution can be emerged, all the ARP and ARS value of y-axis greater than zero represents the stock which has the effective trading points for profit and stop-loss, respectively (Fig. 5).

Table 1. The number of stocks with effective buy points

ARP	Number of Stocks		
	Daily DMAC	**Weekly DMAC**	**Dual period DMAC**
0 $< \tau \leq$ 3%	1087	486	1652
3 $< \tau \leq$ 5%	236	191	207
5 $< \tau \leq$ 7%	111	159	76
$\boldsymbol{\tau} >$ 7%	121	493	90

Note: the number of stocks at given stages at 3%, 5%, 7% ARP.

As it can be seen, the stock distributions with effective trading points are distinguished from each other in different periods. However, from all given stages of ARP, the stocks selected by dual period DMAC indicator yield higher overall earnings than that of daily DMAC and weekly DMAC(Table 1).

Similarly, from all given stages of ARS, the stocks selected by dual period DMAC indicator have better stop-loss effect than that of daily DMAC and weekly DMAC(Table 2).

Table 2. The number of stocks with effective sell points

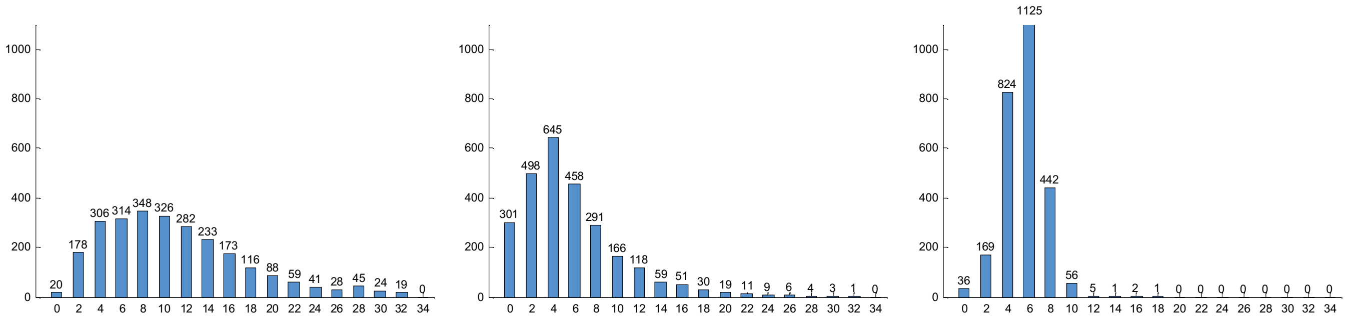
ARP	Number of Stocks		
	Daily DMAC	**Weekly DMAC**	**Dual period DMAC**
0 $< \tau \leq$ 3%	599	327	1087
3 $< \tau \leq$ 5%	132	138	122
5 $< \tau \leq$ 7%	53	89	31
$\boldsymbol{\tau} >$ 7%	59	267	40

Note: the number of stocks at given stages at3%, 5%, 7% ARS.

The result shows that the dual period DMAC indicator will help to select more stocks than that of single period DMAC with respect to profit rate(Fig. 6) and stop-loss rate(Fig. 7), which shows the potential value of this approach in real market trading.

4.2.2 Interval Analysis

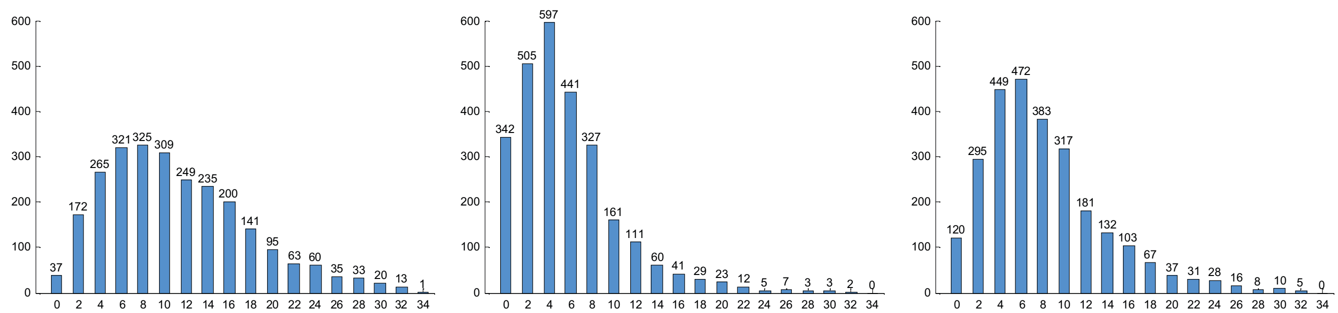
To determine effectiveness of the stock trading points, we need to understand the changes in their amplitude, as well as its associated interval changes, i.e. the amplitude change beginning from the current trading point to the next N days' time interval. To describe the time span N that the amplitude changes may take for one stock, it is necessary to collect statistics of the average time span it takes for all effective trading points of that stock. The time point of the current trading point τ is defined as τ_i^c, its time point at segmentation boundary of the upgoing interval is defined as $\tau_j^{seg} \uparrow$.



(a) daily ARP interval (b) weekly ARP interval (c) dual period ARP interval

Fig. 8. The interval distribution of ARP

Note: x-axis represents ARP interval, y-axis represents the number of stocks.



(a) daily ARS interval (b) weekly ARS interval (c) dual period ARS interval

Fig. 9. The interval distribution of ARS

Note: x-axis represents ARS interval, y-axis represents the number of stocks.

And the profit interval φ_{profit} is defined as:

$$\varphi_{profit} = \frac{\sum_{i=1,j=1}^{N,M}(\tau_i^c - \tau_j^{seg}\uparrow)]}{min(N,M)} \tag{9}$$

Its time point at segmentation boundary of the downgoing interval is defined as $\tau_j^{seg}\downarrow$, and the stop-loss interval $\varphi_{stop-loss}$ is defined as:

$$\varphi_{stop-loss} = \frac{\sum_{i=1,j=1}^{N,M}(\tau_i^c - \tau_j^{seg}\downarrow)]}{min(N,M)} \tag{10}$$

From the stock distribution, we can see that, the average profit intervals of the majority of stocks are located within a small range. 2596 stocks are selected in less than two weeks, taking 96.9% of all stocks [Fig. 8 (c)]. Similarly, the average stop-loss intervals of the majority of stocks are also located within a small range. 1719 stocks are selected in less than two weeks, taking 64.16% of all stocks [Fig. 9 (c)].

In addition, the dual period DMAC is dedicated to impart the signal of turning points known as effective trading points at an earlier time, so that investors can obtain the stock with candidacy ahead of time, and to improve earning rate or stop-loss rate, it also confirms our previous assumption.

4.3 Result Analysis

It is highly possible that the earning expectation has uncertainty to some extent merely based on daily DMAC or weekly DMAC as trading principle, therefore, in most cases, it is unreliable sometimes to extract trading point by using single period DMAC. Thus buying and selling stocks based on that will lead to unpredictable results frequently. By using volatility segmentation approach to dual period DMAC, the estimated trading points can be quantitatively analyzed in two aspects of amplitude and interval, so as to select the stocks that have effective trading points within one period of time. The results show that our proposed approach can generate investment gain on certain level by using the selected effective trading points (Table 3).

Table 3. ARP of partial tested stocks

Stock	DMAC		
	Day	Week	Dual period
GALT	2.38%	18.18%	17.44%
AMBI	4.06%	1.81%	16.97%
AQXP	4.49%	4.60%	16.06%
ARNA	-8.22%	10.07%	16.03%
GNVC	-35.98%	0.00%	15.91%
ARCW	-1.96%	-36.20%	15.72%
SRPT	-45.33%	18.48%	15.27%
MTSN	0.54%	1.22%	12.15%
RGLS	-2.52%	-2.62%	11.05%
BONT	0.53%	-19.49%	10.98%

Note: The ARP of partial tested stocks in all periods of DMAC, they are independent each other.

The performance of the extracted effective sell points in dual period DMAC for stop-loss is relatively good (Table 4), although the result of sell points in daily and weekly DMAC is inaccurate, the trading points in dual period DMAC can still be guaranteed to be effective.

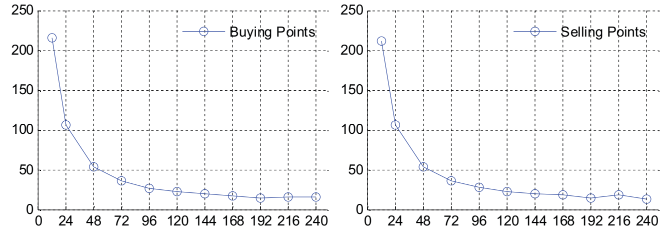
Table 4. ARS of partial tested stocks

Stock	DMAC		
	Day	Week	Dual period
AMDA	-0.29%	-0.27%	14.05%
GNCA	0.20%	0.00%	11.86%
EGLT	6.92%	3.46%	11.66%
FREE	12.57%	23.39%	10.66%
BNFT	-0.54%	12.49%	9.23%
ADHD	1.31%	-4.89%	9.20%
TVIX	-42.10%	3.54%	8.79%
LIQD	5.68%	0.00%	7.95%
DXM	0.38%	-23.99%	7.85%
ENG	-7.68%	0.12%	7.85%

Note: The ARS of partial tested stocks in all periods of DMAC, they are independent each other.

4.4 Computing Performance Comparison

In this study, we have used a total of 240 cores on Matlab Distributed Computing Server to perform the quantitative analysis on 2679 selected NASDAQ stock from 10/31/2011 to 10/31/2014 with the data processing capacity 39.1 MB/s. Considering the correlated trend changes in analysis process, the calculation and analysis are implemented on buy and selling points respectively. It is convenient to observer whether there is a greater difference between them in computational performance of quantitative analysis, but also conducive to the data management. When analyzing the data, the stock data are grouped into chunks and distributed to each node for computing, the intermediate result will be aggregated for statistical purpose. The entire running process takes longer time for sequential computing.

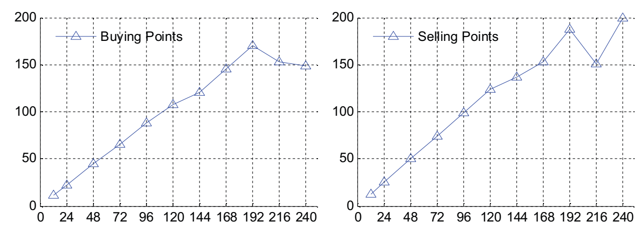


(a) effective buy points (b) effective sell points

Fig. 10. Performance comparison of effective trading points

Note: x-axis is the number of CPU, y-axis is running time.

By analyzing and comparing the performance in different number of CPUs, it can be seen that(Fig. 10), with 240 cores, the calculation on effective buy points in daily DMAC takes 14.21 seconds, while the sequential time of using single core is 2357.13 seconds, which is approximately 165 times faster; the calculation on buy points in weekly DMAC takes 5.77 seconds, its sequential time is 215.01 seconds, which is approximately 37 times faster; and the calculation on buy points in daily DMAC takes 16.01 seconds, its sequential time is 2382.82 seconds, approximately 148 times faster.



(a) effective buy points (b) effective sell points

Fig. 11. Speedup comparison of effective trading points

Note: x-axis is the number of CPU, y-axis is the speedup.

By analyzing and comparing the speedup in different number of CPUs, it can be seen that(Fig. 11), with 240 cores, the calculation on effective sell points in daily DMAC takes 14.9 seconds, while the sequential time of using single core is 2363.56 seconds, which is approximately 158 times faster; the calculation on sell points in weekly DMAC takes 6.39 seconds, its sequential time is 262.98 seconds, which is approximately 41 times faster; and the calculation on sell points in daily DMAC takes 13.5 seconds, its sequential time is 2688.53 seconds, approximately 199 times faster.

5. Conclusions

We applied volatility segmentation approach to effectively extract the buy and sell trading points in HPC clusters from 2679 NASDAQ stock data based on dual period DMAC resonance of daily and weekly moving average, it is the implementation of data-intensive quantitative analysis for stock data. When the data is analyzed per second, the sequential computing time for detecting buy and sell points takes 2382.82s and 2688.53s, respectively, while parallel computing takes only 16.01s and 13.5s with 240 cores, respectively to obtain all selected NASDAQ stock trading points from simulated real-time data analysis. The proposed dual period DMAC resonance trading strategy via parallel computing can greatly improve the estimation accuracy of determining buy and sell points, as well as quickly find out investable stocks with profitable trading points. This approach also provides inspiration for feasibility planning to increase investors' returns and reduce the risk of loss.

This study has achieved the goal to quickly understand the macroscopic change of NASDAQ stock price data. The provided stock data distribution with effective trading points plays an important role in real-time trading. According to the statistical results, the dual period DMAC approach has better outcome than both golden-cross and dead-cross in daily and weekly period. It also has practical significance to determine the trading points with respect to buy and sell, and theoretical significance for the research on expecting stock profits and stop-loss. The market data analysis illustrates that the number of stocks which have the average rate of profit of overall buy points higher than 3% is 373 including GALT, etc.; the number of stocks which have the average rate of profit of overall sell points higher than 3% is 193 including AMDA, etc. In this study, the volatility segmentation approach reduces the uncertainty of stock selection and estimation in single period DMAC, which helps investors to achieve the desired objectives to improve the stock investment income. It provides a novel application of data-intensive computing on quantitative analysis stock trading. We expect to reduce the computation time to significantly below a second by implementing our program using MPI instead of Matlab by making our code more efficient and scalable.

References

[1] R. Agrawal, C. Faloutsos, and A. Swami, "Efficient similarity search in sequence databases," Proceedings of the 4th Conference on Foundations of Data Organization and Algorithms. 1993.

[2] H. Akaike, "Fitting autoregressive models for prediction," Annals of the Institute of Statistical Mathematics, 21(1): 243-247, 1969.

[3] H. Akaike, "A new look at the statistical model identification," Automatic Control, IEEE Transactions on, 19(6): 716-723, 1974.

[4] H. Akaike, "A new look at the bayes procedure," Biometrika, 65(1): 53-59, 1978.

[5] F. M. Bandi, and J. R. Russell, "Separating microstructure noise from volatility," Journal of Financial Economics, 79(3): 655- 692, 2006.

[6] K. Chan, and W. Fu, "Efficient time series matching by wavelets," Proceedings of the 15th IEEE International Conference on Data Engineering, 1999.

[7] J. Beringer, and E. Hllermeier, "Online clustering of parallel data streams," Data & Knowledge Engineering, 58(2): 180- 204, 2006.

[8] S. G. M. Fifield, D. M. Power, and D. G. S. Knipe, "The performance of moving average rules in emerging stock markets," Applied Financial Economics, 18(19): 1515–1532, 2008.

[9] J. Granville, "Strategy of Daily Stock Market Timing," Prentice Hall, Englewood Cliffs, NJ, 1960.

[10] E. Keogh, S. Chu, D. Hart, and M. Pazzani, "Segmenting Time Series: A Survey and Novel Approach," In: Last, M., Kandel, A., Bunke, H. (eds) Data Mining in Time Series Databases. World Scientific Publishing, Singapore, 2004.

[11] V. Mayer-Schönberger, "Big data: A revolution that will transform how we live, work, and think," Eamon Dolan/Houghton Mifflin Harcourt; 1 edition, 2013.

[12] M. Metghalchi, J. Marcucci, and Y.H. Chang, "Are moving average trading rules profitable? evidence from the european stock markets," Applied Economics, 44(12): 1539–1559, 2012.

[13] K. Miwa, and K. Ueda, "The influence of investor's behavioral biases on the usefulness of the Dual Moving Average Crossovers," New Generation Computing, 23(1): 67-75, 2005.

[14] G. Sharma, and J. Martin, "Matlab: A language for parallel computing," International Journal of Parallel Programming, 37(1): 3-36, 2009.