

Multisource Data Analysis for Stock Prediction

Xuebo Jin, Yingbo Wang, Chen Liu, Tianyi Liu and Tingli Su

*School of Computer and Information Engineering, Beijing Technology and
Business University, Beijing, 100048, China
jinxuebo@btbu.edu.cn*

Abstract

Over the last decade, quantitative investment has been adopted by the international investment community as a new method that combines financial data with mathematics and computer technology. This paper proposes a quantitative investment prediction model that combines multifactor and candlestick models to predict stock prices. First, relevant indicators with good future trends were identified to construct a stock pool containing stocks. Then, the candlestick model was used to decide when to buy or sell stocks in the pool. The method developed here can be used for quantitative investment strategies, and the experimental results show it can have good returns in the investment stock market based on the transaction data.

Keywords: *multifactor model, candlestick model, stock prediction, multisource data, quantitative investment*

1. Introduction

In the stock market, fundamental analysis, technical analysis, and evolutionary analysis are the three main classical methods for predicting stock prices [1]. Fundamental analysis involves analyzing a business' financial statements, the macroeconomic situation influenced by stock prices, and the prospects for industry development [2-4]. Technical and evolutionary analyses are usually employed to determine when to buy or sell. There are similarities and differences between these methods. Technical analysis is a security-analysis methodology used to forecast the direction of prices by studying past market data—mainly price and volume. The founders of technical analysis believe the stock market is related to the historical data, and the volatility of stock prices can be quantified and forecasted [5]. Technical analysts typically make decisions using theories such as Dow theory and the Elliott wave principle [6].

Today, with the rapid development of science and technology, data-based prediction methods have generated great interest. As a promising technology, quantitative investment technology has quickly developed in recent years [7]. Quantitative investment combines computer technology with mathematical models to realize investment philosophies and strategies.

The multifactor model is the most widely used quantitative stock-selection model; its basic principle involves using a series of factors as stock-selection criteria [8]. The advantage of factor models is the ability to get more comprehensive information and select stocks as a result; there have been significant achievements in both theoretical research and practical development showing the effectiveness of factors, as is widely known to investors [9]. In addition, given the general and consistent use of factors by investors, the performance of multifactor models is relatively stable. That said, under different market conditions, various factors can play significant roles [10]. Therefore, investors and quantitative researchers have developed several factor models based on different strategies. The key difference between models concerns the way different factors are scored to obtain an optimal, comprehensive judgment. At the same time, factor

models have certain limitations. For example, they can determine the future trend for a specified period but cannot pinpoint a specific operating point in time; thus, there is no guarantee of profit [11].

Meanwhile, the candlestick chart, with its unique form of representation, was introduced into the analysis of stock and futures markets. This type of chart analysis is currently popular throughout Southeast Asia [12-13]. Using a candlestick chart, we can completely record the daily or cyclical performance of a market in the form of special areas on a diagram, with different forms showing different meanings. From these morphological changes, we can infer some regular patterns [14]. Candlestick forms can be divided into the inversion of the form, the arrangement of the form, gap and trend lines, *etc.* From the candlestick, we can completely observe real market changes and see stock-price trends, while also grasping daily market fluctuations [15-16]. In practice, stocks with good future trends should be selected in advance to guarantee high profits [17].

Each multifactor and candlestick method has advantages and disadvantages. Multifactor models can show stocks with good future trends but cannot tell the trader when to buy or sell, which is important for obtaining investment returns. While candlestick charts can signal when to buy or sell, the large number of stocks in the market makes it difficult for the trader to find the target stock using a candlestick chart. In practice, therefore, factor and candlestick methods can only be used for reference as technical analysis tools [18]. To overcome these drawbacks, we developed a model that combines multifactor and candlestick methods with moving average (MA), relative strength index (RSI), and other technical indicators to improve the stock index forecasting performance of the proposed model [19-20]. Importantly, combining multifactor and candlestick models gives the model a high training speed, a highly effective learning algorithm, and structural simplicity [21].

Moreover, the combination model has the advantage of intelligent forecasting and compatible input and output for mapping stock data. When you want to illuminate which stocks to buy or sell, there is no requirement for prior information or characteristics. The combination model can find potentially profitable stocks and has low time complexity. In the existing literature, there is a knowledge gap regarding the identification of combinatory multifactor and candlestick models. The present study aims to fill that gap using the proposed model.

The next section presents the multisource information fusion method. Then, Section 3 introduces the multifactor model, while Section 4 discusses the candlestick model. Finally, the results and conclusion are presented in Section 5 and 6, respectively.

2. Multisource Information Fusion Method

This section illustrates the core technology, which mainly consists of candlestick and multifactor models. The multifactor model is the most widely used quantitative model for stock selection. The basic principle is to use a series of factors as stock-selection criteria, buying when the conditions are met and selling when they are not. A stock with good development momentum is selected from a large number of stocks and placed in the stock pool. Meanwhile, the candlestick model, which has developed advanced characteristics over hundreds of years, involves numerous chart patterns representing gradual change. It is very suitable for long-term stock operation and can be used to predict future trends. This method is most commonly used in technical analysis. Using the multifactor model to select stocks and the candlestick model to grasp future trends is an important combinatory method of quantitative investment technology. The whole system is shown in Figure 1.

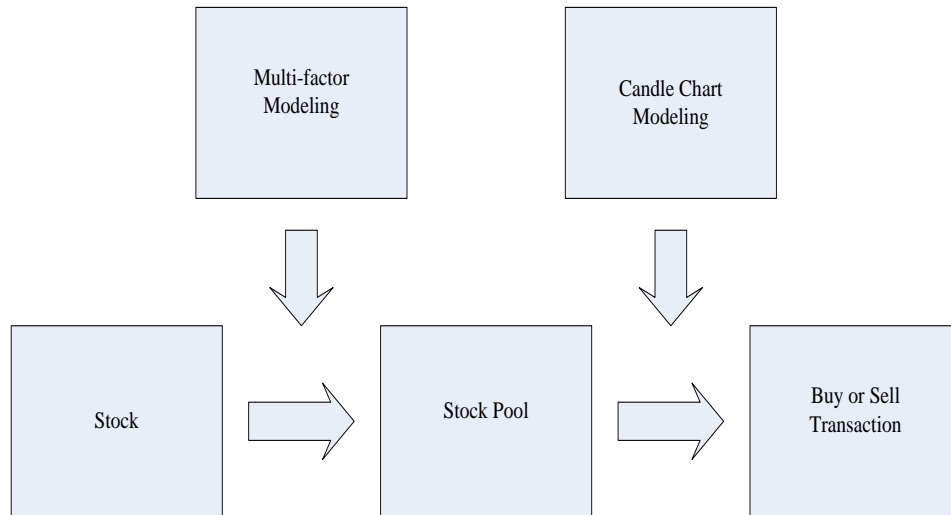


Figure 1. Flowchart of the Entire System

2.1 Multifactor Model

The multifactor model is relatively stable since, even in different market conditions, there are always some factors that can play a role. The multifactor model can synthesize large amounts of information to obtain quantitative stock-selection results. Since the model's performance is relatively stable, investors and researchers have developed many different multifactor structures.

In general, the multifactor quantitative stock-selection model has two kinds of judgment methods. One is the scoring method, and the other is the regression method. The scoring method considers the size of various stock factors, weighs them according to a certain weight ratio to get the total value, and then selects the stock based on that value. The advantage is that the system is stable and not easily affected by extreme values. This type of investment practice builds a quantitative model that is highly favored.

The regression method is more sensitive to stock and more vulnerable to extreme values. For this study, we used the scoring method to establish a quantitative stock-selection model. The multifactor stock-selection modeling process consists of four steps: selecting candidate factors, removing similar factors, establishing a comprehensive scoring model, and continuously improving the model.

Candidate factors were chosen according to the correlation between stock yields and factors. To ensure the multifactor model would be scientific, we tried to choose candidates from multiple types of factors. Each type of factor needed to have a maximum correlation factor, and we removed what remained. This was because the same kinds of internal drive factors can cause an earnings impact with consistent modeling results. Finally, the selected effective factors were combined into a multifactor model, ensuring that each factor would have a correlation effect on the yield.

2.2 Candlestick Model

In its earliest application, the candlestick graph model was used during the shogunate era to analyze rice prices, and it has been widely used ever since. Candlestick charts can be used alone or in combination with other techniques. Such combinations can accurately portray stock trends. The candlestick model is suitable for predicting stock market trends, even over a very short period of time. By using pattern recognition based on the candlestick model, we can accurately observe a period of time in the future.

First, we constructed a matrix of candlestick chart patterns. The characteristics included candles, CCI, MA information, and rising and falling trends. To train the input and output sets, a support vector machine (SVM) machine learning model was used. Then, the SVM training was used to get the stock trading point.

3. Multifactor Quantitative Model for Stock Selection

Some factors were related to a company's stock price and the value of the company, including the basic index of the company (earnings, discount rates, *etc.*) and the technical indicators of the company's stock price (momentum, exchange rate, *etc.*).

The factors were selected for a given stock using the following steps:

(1) Assume we consider N stocks, whose rate of return is denoted as R_i , where $i=1, \dots, N$. Define f_j representing the j th factors $j=1, \dots, M$. The correlation coefficient of R_i and f_j is $AR_{ij} = Corr(R_i, f_j)$, which should satisfy the relationship $-1 \leq AR_{ij} \leq 1$, where

$$AR_{i,j} = \frac{1}{N-1} \sum_{i=1, j=1}^N \left(\frac{R_i - \bar{R}}{\sigma_R} \right) \left(\frac{f_j - \bar{f}}{\sigma_f} \right) \quad (1)$$

The Pearson product-moment correlation coefficient was used to measure the relationship between two variables (linear relationship) — the value between -1 and 1.

Among these parameters, $\frac{R_i - \bar{R}}{\sigma_R}$, \bar{R} , and σ_R represent standard deviation, sample mean, and standard deviation, respectively.

We should mention that the value of AR_{ij} refers to the different relation between R_i and f_j . For example, $AR_{ij} > 0$ means the i th stock has a positive correlation relation with the j th factor, which shows that if the j th factor becomes larger, then the i th stock will go up in the future. On the other hand, $AR_{ij} < 0$ shows that the i th stock has a negative correlation relation with the j th factor, which means that if the j th factor becomes larger, then the i th stock will go down in the future.

(2) Only the factors that could give a strong and obvious signal of a stock's future were considered and chosen. Thus, we defined two reference parameters, $0 < AR_{top} < 1$ and $-1 < AR_{bottom} < 0$. For the i th stocks, if the j th factor satisfied $AR_{ij} > AR_{top}$ or $AR_{ij} < AR_{bottom}$, then the j th factor was selected as the factor feature.

(3) For the last step, we used the selected factors to select the potential advantage stock in the future. The i th stock was scored using the selected factors; the details are as follows:

For a given stock, the stock got a positive score, +1, if the factors with the positive correlation relation went up or those with a negative correlation relation went down. Otherwise, the stock would get a -1 score. When all scores were added together, if it was larger than +2, the stock was selected for the stock pool.

With an increasing number of stock investors using multifactor models, some factors will gradually fail while other new factors may be validated. In our method, we chose factors every six months, meaning we could obtain a new stock pool every half year. After we had the stock pool, we used the candlestick model to determine the buying and selling points.

4. Candlestick Chart Modeling

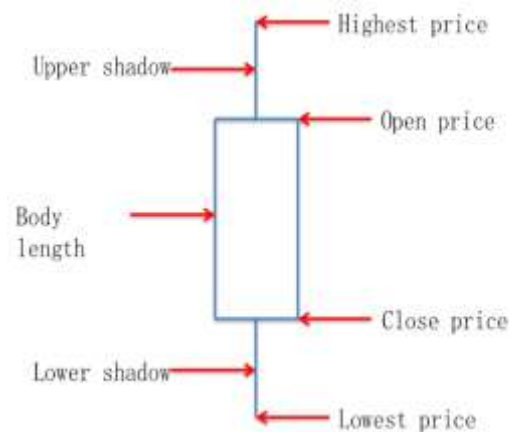


Figure 2. A Candlestick Chart

A candlestick chart is a type of financial chart used to show the price movement of a derivative, security, or currency. Candlesticks can be shown for a given time period (*e.g.*, five minutes, one hour, one day, one month). After a period of recording, the chart shows a particular region or morphology, and different forms express different meanings.

A candlestick represents the open price (*o*), close price (*c*), highest price (*h*), and lowest price (*l*). Different kinds of candlesticks are defined by the body color (*t*), body length (*b*), upper shadow (*u*), and lower shadow (*d*). The body color (*t*) is determined by the difference between the open price and close price.

Candlestick patterns can consist of a single candlestick or multiple candlesticks. In this study, we used ten of the most widely used patterns: hammer and hanging-man lines, the engulfing pattern, the morning star, the evening star, dark-cloud cover, piercing pattern, the morning doji star, the evening doji star, the shooting star, and the inverted hammer. These strongly indicate future uptrends and downtrends in stock prices. The patterns are shown in Figure 3.

4.1 Candlestick Chart Model

Since candlestick figures contain a wealth of information regarding price changes, they are important tools for predicting price movements. Observing changes in the candlestick form, it is easy to identify changes over a series of trading days. For example, the “three ravens model,” often seen in the stock market, arrives at the end of an uptrend, which means there is a possibility of falling stock prices; thus, investors can implement a short operation to prevent losses.

However, since a computer cannot recognize candlestick patterns as we see them, they need to be converted into numerical values. To convert them, we first defined the following parameters:

(1) We defined the peak according to the rising and falling peaks of the stock—that is, whether it was in the uptrend peak or downtrend peak.

(2) We defined max as the larger of the open and close values and min as the smaller value.

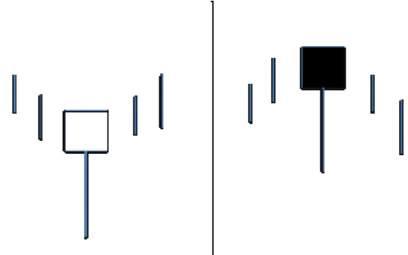
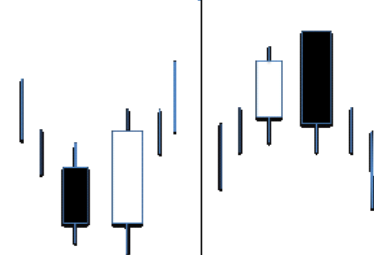
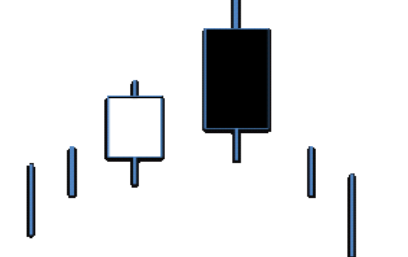
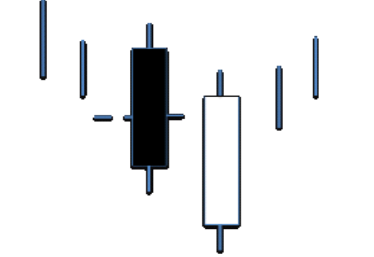
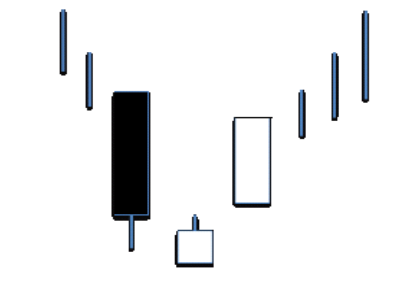
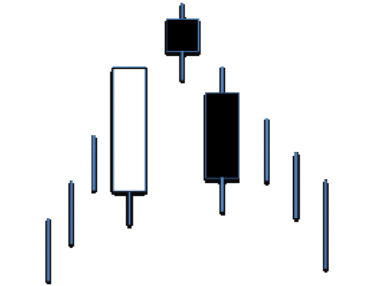
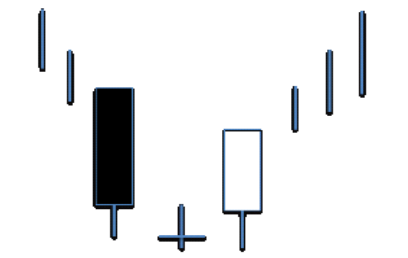
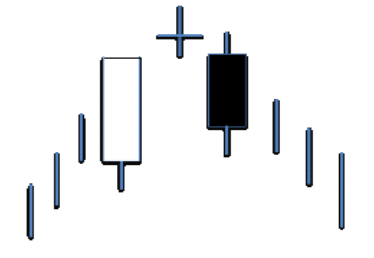
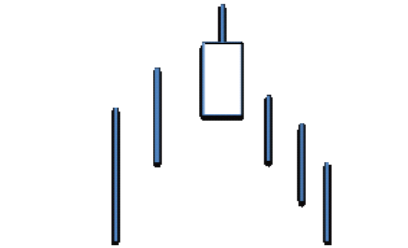
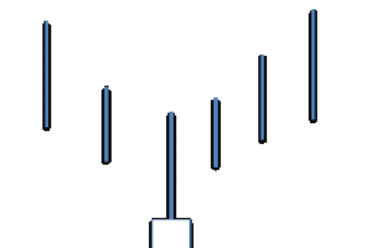
<p>Pattern 1(chuizi) Hammer and hanging-man lines</p> 	<p>Pattern 2(tunmo) Engulfing pattern</p> 
<p>Pattern 3(wuci) Park-cloud cover</p>	<p>Pattern 4(wuci) Piercing pattern</p>
	
<p>Pattern 5(qi Huang) The morning star</p>	<p>Pattern 6(qi Huang) The evening star</p>
	
<p>Pattern 7(szqh) The morning doji star</p>	<p>Pattern 8(szqh) The evening doji star</p>
	
<p>Pattern 9(ld) The shooting star</p>	<p>Pattern 10(ld) The inverted hammer</p>
	

Figure 3. Ten Candlestick Chart Patterns

(3) In Figure 3, we can see the patterns showing a stock's upward and downward trends. We defined 1 as representing the future uptrend and -1 as representing the future downtrend.

(4) We used i to represent the observed day.

(5) We used "crow" to represent the days we used to predict the value of next day.

We adopted 10 names to represent the 10 abovementioned patterns: *chuzi* for pattern 1, *tunmo* for pattern 2, *wuci* for patterns 3 and 4, *qihuang* for patterns 5 and 6, *szqh* for patterns 7 and 8, and *ld* for patterns 9 and 10 (Figure 3). The value of each pattern could only be 1, 0, or -1. Further, the same name could represent more than one pattern (e.g., *tunmo* represents patterns 5 and 6; Figure 3). The names of the process patterns are also shown in Figure 3. Moreover, if we could predict that the value would rise the next day, -1 represented down, and 0 meant the pattern was not shown on the observed day.

After that, we converted all patterns into numerical values. In addition, we also used other indexes to build our model more authentically.

4.2 Moving Average

Although candlestick charts are widely accepted by analysts, there are many features for evaluating uptrends and downtrends in stock prices. We selected moving averages (MA) for stock analysis, including short-term moving average and long-term moving average. This method is described as follows: A short-term moving average crossing a long-term moving average indicates a future uptrend. We chose 13 days as the short average and 30 days as the long average; 1 indicates an uptrend, and -1 indicates a downtrend.

4.3 Relative Strength Index

Relative strength index (RSI) is a technical indicator used in the analysis of financial markets. It charts the current and historical strength or weakness of a stock or market based on the closing prices of a recent trading period. RSI is figured by comparing the gains of up sessions with the losses of down sessions in a given time frame. The calculations depend only on closing prices.

The RSI index is calculated (within N days) as:

$$RSI = \left[A \div (A + B) \right] \times 100\% \quad (2)$$

where A is the sum of the closing price of up sessions within N days, and B is the sum of closing price of down sessions within N days.

As an overbought/oversold indicator, RSI implies that the market is overbought if it approaches the upper end of the band (i.e., above 70 or 80). Conversely, an RSI range below 30 or 20 reflects an oversold condition. In such an environment, there is the potential for short covering moves.

4.4 Commodity Channel Index

The commodity channel index (CCI) is a versatile indicator that can be used to identify a new trend or warn of extreme conditions. CCI is calculated as the difference between a commodity's typical price and its simple moving average, divided by the mean absolute deviation of the typical price. The index is usually scaled by an inverse factor of 0.015 to provide more readable numbers.

CCI is designed for extreme situations, which means it has no effect in a normal market. CCI is calculated (within N days) as:

$$CCI = (TP - MA) \div MD \div 0.015 \quad (3)$$

where $TP = (\text{highest_price} + \text{lowest_price} + \text{closing_price}) \div 3$, $MA =$ (sum

of past N days) $\div N$, MD = sum of past N days abstract of (MA-closing_price) $\div N$, and 0.015 is the calculation coefficient; N is the calculation cycle.

4.5 Combination Features Vector

According to the above features, we defined $A(i, j)$ as the input matrix, where i is the day of the historical data and j is the features vector, which includes the candlestick patterns *chuzi*, *tunmo*, *wuci*, *qihuang*, *szqh*, and *ld*, and the features *RSI*, *CCI*, and *MA*. The matrix $A(i, j)$ has only 1, -1, and 0 (1 for a positive pattern, -1 for a negative pattern, 0 if nothing arises). Each column represents the numerical value we got from the discussed patterns.

Table 1. Combination Features Vector

1	2	3	4	5	6	7	8	9
RSI	CCI	MA	chuzi	tunmo	wuci	qihuang	szqh	ld

4.6 SVM Learning To Generate Buy and Sell Points

The support vector machine (SVM) is considered one of the most important breakthroughs in the field of machine learning and can be applied in both classification and regression [22]. The goal of SVM modeling is to select the optimal hyperplane in a high-dimensional space, ensuring the upper boundary of generalization error is minimal. Given a set of data points (x_i, y_i) , ($i=1, \dots, n$, $x_i \in R$, $y_i \in \{-1, 1\}$), randomly and independently generated from an unknown function, SVM approximates the function in the following form: $g(x) = \omega\phi(x) + b$, where $\phi(x)$ is the feature nonlinearly mapped from the input space; ω , x and b are coefficients and can be estimated by minimizing the regularized risk function.

We define the future uptrend and downtrend as:

$$stcok_change = (now_price - yesterday_price) \div yesterday_price \quad (4)$$

where *yesterday_price* means yesterday's closing price. The output vector for SVM learning is the future uptrend and downtrend of the next day to each reference day. The input vector is the combination vector of each reference day shown in Table 2.

5. Results and Analysis

5.1. Procedures and Results

We applied our method to the stock market. First, a multifactor model was used to select the potential stocks and construct the stock pool. Then, using a candlestick graph model, the input and output vectors for the SVM were calculated for each stock. Without loss of generality, we chose the current day as the reference time and 100 days before the reference time as the learning interval.

Using the Pearson correlation coefficient to calculate the value between the candidate factor and returns, we selected five stocks with the most development potential—600322.XSHG, 600189.XSHG, 600790.XSHG, 600291.XSHG, and 600284.XSHG—and constructed the stock pool (Table 3).

Table 2. Five Selected Stocks

No.	Code	Score	Name
1	600322	4515	Tianjin Reality Development (GROUP) Co., Ltd.
2	600189	4242	China Jilin Forest Industry Group Co., Ltd.
3	600790	4239	Zhejiang China Light & Textile Industrial City Group Co.,Ltd.
4	600291	4184	West Water shares Co., Ltd.
5	600284	4179	Shanghai Pudong Road & Bridge Co., Ltd.

Next, we analyzed each stock in the pool using the candlestick chart method with nine combination feature vectors. If the pattern rose, we assigned a value of 1; for the opposite, 0. The results were the input matrix. According to the actual situation of whether the stock price rose or fell—that is, whether the day’s closing price was higher than that of the previous day—if yes, a value of 1 was assigned; otherwise, -1. The results were the output matrix. Thus the model was built. In subsequent transactions, the patterns above can give us the signal for when to buy or sell stock.

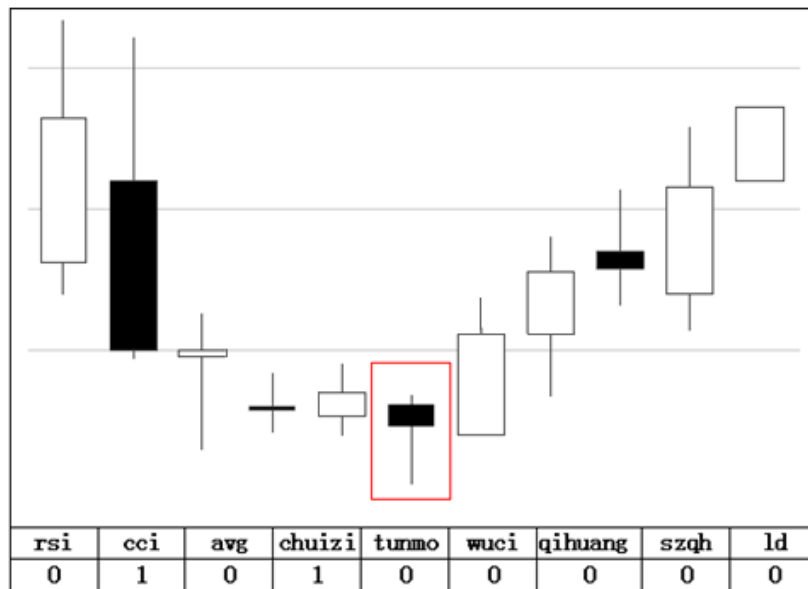


Figure 4. 600790 from Dec. 7, 2014, to Dec. 25, 2014

5.2. Strategic Return Results

We selected five stocks covering several domains (industry, sales, banks, etc.). Samples were collected from January 2015 to July 2016. The strategic return figures are shown below in Figures 4–8 (— Strategic return, — Benchmark return).

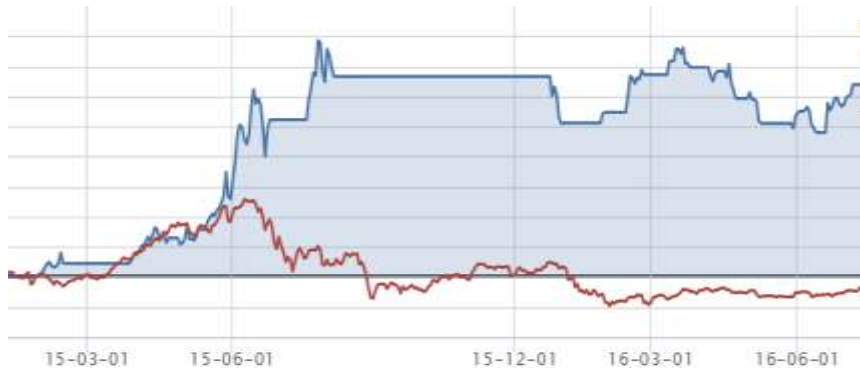


Figure 5. Returns on Stock 600189.XSHE

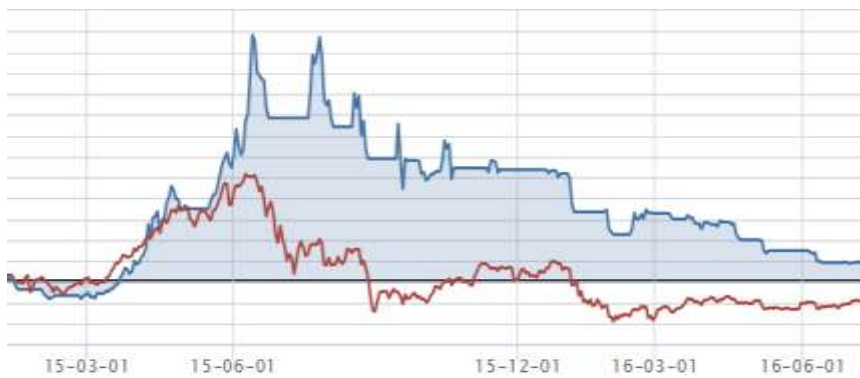


Figure 6. Returns on Stock 600284.XSHE

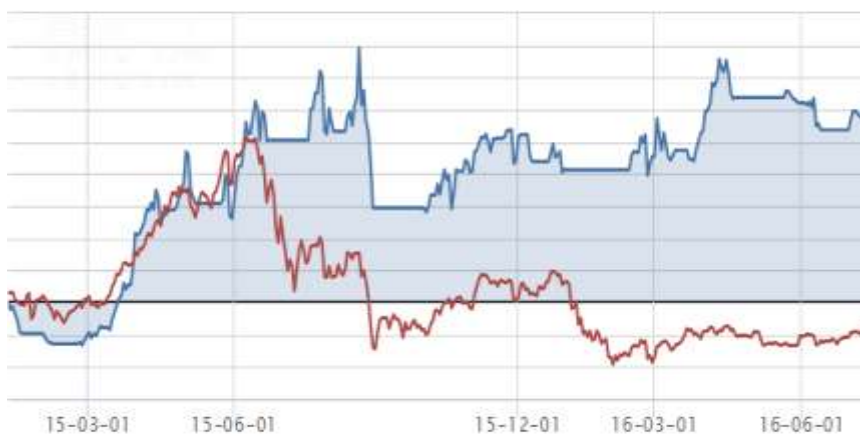


Figure 7. Returns on Stock 600322.XSHE



Figure 8. Returns on Stock 600790.XSHE

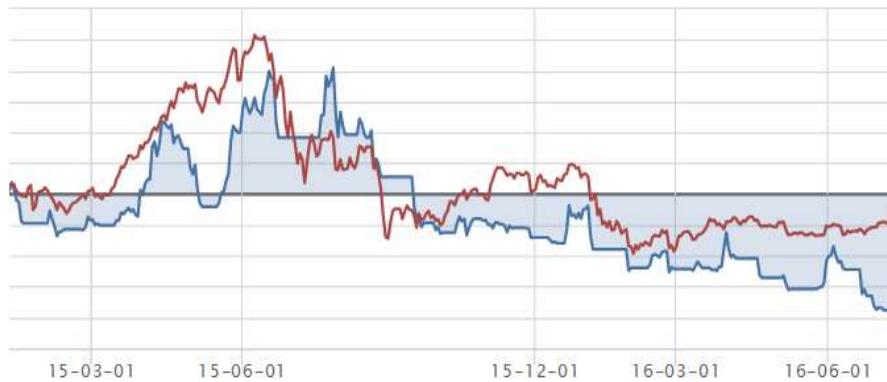


Figure 9. Returns on Stock 600291.XSHE

As shown in Figures 5–9, the developed method yielded significantly higher returns than the benchmark for most stocks in the pool. This means that forecasting accuracy can be greatly improved by using multifactor stock-selection models combined with candlestick models. However, good predictive results were not achieved for 600291.XSHE. This means that to improve accuracy, we need more information for different industries, more understanding of different interference factors, and a fuller range of considerations.

6. Conclusion

Predictions are needed before making plans or decisions. With the development of the futures market, predicting future prices has attracted the attention of researchers and investors. This market is complicated, and the prices present strong inaccuracies.

In our method, we used multifactor and candlestick models to predict stock prices. First, we selected stocks with good development momentum using the multifactor model. Then, the candlestick model was used to match methods to predict stock trends and identify buy and sell points. Combining an extensive back testing strategy with a constantly optimized system structure for predicting stock price trends, this method can obtain ideal predictive performance.

Acknowledgments

This work is partially supported by NSFC under Grant No. 61273002, 61673002 and the Key Science and Technology Project of Beijing Municipal Education Commission of China No. KZ201510011012.

References

- [1] C. D. Kirkpatrick and J. R. Dahlquist, *Technical Analysis: “The Complete Resource for Financial Market Technicians”*, 2st ed.; Publisher: Financial Times Press, U.S.A., (2006), p. 3.
- [2] G. L. Morris and R. Litchfield, *“Candlestick Charting Explained: Timeless Techniques for Trading Stocks and Futures”*, 3rd ed.; Publisher: McGraw-Hill, U.S.A., (2006).
- [3] Y. M. Tanaka and S. Tokuoka, “Adaptive use of technical indicators for the prediction of intra-day stock prices”, *Physica A383*, no. 1, (2007), pp. 125-133.
- [4] P. Anagnostidis and C. J. Emmanouilides, “Nonlinearity in high-frequency stock returns: Evidence from the Athens Stock Exchange”, *Physica A Statistical Mechanics & Its Applications*, (2015), pp. 421, 473–487.
- [5] J. M. Matias and J. C. Reboredo, “Forecasting Performance of Nonlinear Models for Intraday Stock Returns”, *J. Forecast*, (2012), pp. 172-188
- [6] L. E. Baum and N. A. Weiss, “Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains”, *Annals of Mathematical Statistics*, vol. 41, no. 1, (1970), pp. 164-171.
- [7] Y. S. Chen, C. H. Cheng, C. L. Chiu and S. T. Huang, “A study of ANFIS-based multi-factor time series models for forecasting stock index”, vol. 5, no. 2, (2016), pp.1-21.
- [8] F. Xudong, L. Hui and Z. Zhipu, “A quantitative investment model based on multi-fractal theory and support vector machine”, (2014), pp.34-35
- [9] J. Stanković, I. Marković and M. Stojanović, “Investment Strategy Optimization Using Technical Analysis and Predictive Modeling in Emerging Markets”, *Procedia Economics & Finance*, vol. 19, (2015), pp. 51–62.
- [10] J. M. Matias and J. C. Reboredo, “Forecasting Performance of Nonlinear Models for Intraday Stock Returns”, *J. Forecast*, (2012), pp. 172-188
- [11] S. Chenxi, Z. Fengli, X. Zhongsheng, L. Kun, S. Yun and W. Zi, “Quantitative estimation of forest stock volume based on multi-source data”, (2010), pp.75, pp.34-36
- [12] L. Chmielewski, M. Janowicz, J. Kaleta and A. Orłowski, “Pattern Recognition in the Japanese Candlesticks”, *Soft Computing in Computer and Information Science*, (2015), pp. 342, 227-234.
- [13] S. Barak and T. Tichý, “Wrapper anfis-ica method to do stock market timing and feature selection on the basis of japanese candlestick”, *Expert Systems with Applications*, vol. 42, no. 23, (2015), pp. 9221-9235
- [14] T. C. Fong and Q. Z. Yu, “Stock Prediction by Searching for similarities in Candlestick Charts. TMIS”, vol. 5, no. 2, (2014), pp.1-21
- [15] N. Nison, *“Japanese Candlestick Charting Techniques”*, 2nd ed.; Techniques of the Far East Nyif Ed Publisher, (2001).
- [16] H. Xie, K. Fan and S. Wang, “The role of Japanese candlestick in dvar model”, *Journal of Systems Science and Complexity*, vol. 5, (2015), pp.1-17.
- [17] C. Y. Lee and D. Gunawan, “Stock investment decision making based on quantitative association rules”, *Icic Express Letters Part B Applications An International Journal of Research & Surveys*, vol. 6, (2015), pp. 1181-1187.
- [18] J. Bollen, H. N. Mao and X. J. Zeng, “Twitter mood predicts the stock market”, *Arxiv working paper*, Indiana University, (2010).
- [19] S. K. Khatri, H. Singhal and P. Johri, “Sentiment analysis to predict Bombay stock exchange using artificial neural network”, *Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions)*, 3rd International Conference on. IEEE, (2014), pp.1-5.
- [20] X. F. Liu and K. Chi, “Tse. A complex Network Perspective of World Stock Markets: synchronization and volatility”, *International Journal of Bifurcation and Chaos*, (2012), pp. 62-66.
- [21] U. Harun, “A two-stage feature selection method for text categorization by using information gain, principal component analysis and genetic algorithm”, *Knowledge-Based Systems*, vol. 24, no. 7, (2011), pp. 1024-1032.
- [22] SVM-Support Vector Machines, Introduction to Support Vector Machine (SVM) Model <http://www.dtrek.com/svm.htm>

Authors



Jin Xue-Bo, Dr. Jin Xue-bo was born in Liaoning in 1972, China. She received the B.E. degree in industrial electrical and automation and the Master degree in industrial automation from Jilin University, Jilin, China, in 1994 and 1997, and the Ph.D. degree in control theory and control engineering from Zhejiang University, Zhejiang, China, in 2004. From 1997 to 2012 she was with College of Informatics and Electronics, Zhejiang Sci-Tech University. Since 2012 she has been with College of Computer and Information Engineering, Beijing Technology and Business University as a Professor. Her research interests include multi-sensor fusion, statistical signal processing, target tracking and dynamic analysis, distributed parallel computing and image processing. In particular, her present major interest is Bayesian estimation, Graph model and big data tendency analysis. JIN is the Fellow of information fusion branch in China Society of Aeronautics and Astronautics, the Member of China Computer Society and Chinese Society of Artificial Intelligence. JIN has published about 50 papers in the statistical, mathematics, information fusion and image processing journals. She also developed a new estimation technology known as closed-loop estimation and owned multiple patents.

