Physica A 442 (2016) 50–66

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Financial technical indicator based on chaotic bagging predictors for adaptive stock selection in Japanese and American markets

Tomoya Suzuki *, Yuushi Ohkura

Department of Intelligent Systems Engineering, College of Engineering, Ibaraki University, 4-12-1 Nakanarusawa-cho, Hitachi, Ibaraki 316-8511, Japan

HIGHLIGHTS

- We propose new technical indicators based on chaotic bagging predictors.
- Our technical indicators evaluate the consensus of the ensemble given by predicted price movements.
- We adaptively select the most reliable and the least risky stock whose indicator shows the highest consensus.
- We demonstrate the predictability and profitability of financial markets by our technical indicators.

ARTICLE INFO

Article history:
Received 9 December 2014
Received in revised form 1 July 2015
Available online 8 September 2015

Keywords:
Technical analysis
Nonlinear prediction
Ensemble learning
Econophysics

ABSTRACT

In order to examine the predictability and profitability of financial markets, we introduce three ideas to improve the traditional technical analysis to detect investment timings more quickly. Firstly, a nonlinear prediction model is considered as an effective way to enhance this detection power by learning complex behavioral patterns hidden in financial markets. Secondly, the bagging algorithm can be applied to quantify the confidence in predictions and compose new technical indicators. Thirdly, we also introduce how to select more profitable stocks to improve investment performance by the two-step selection: the first step selects more predictable stocks during the learning period, and then the second step adaptively and dynamically selects the most confident stock showing the most significant technical signal in each investment. Finally, some investment simulations based on real financial data show that these ideas are successful in overcoming complex financial markets.

1. Introduction

Financial technical analysis has started from Sakata's five methods with candlestick chart patterns in 18th century in the Japanese rice market. After that, Charles Dow proposed Dow theory in 19th century and built the foundation for the present technical analysis. However, because technical analysis is one of prediction methods, its validity has been a controversial topic in terms of the efficient-market hypothesis (EMH). Moreover, some of technical analyses are subjective and are difficult to be considered as scientific approaches. On the other hand, there are some positive studies that report the validity of technical analysis such as Refs. [1–6]. This discussion is detailed in Ref. [7]. Also, the behavioral economics explains that

* Corresponding author.
E-mail address: tomoya.suzuki.lab@vc.ibaraki.ac.jp (T. Suzuki).

http://dx.doi.org/10.1016/j.physa.2015.08.042
0378-4371/

traders are inefficient and irrational, and some anomalies such as the return reversal and the volatility clustering can be observed in real financial markets. These gaps from the random work enhances the predictability of real markets. Moreover, thanks to the development of computational power, machine learning methods have been used to detect better trading rules. Accordingly, it has been popular that some technical indicators are used together by applying the genetic algorithm, the neural network, etc. However, these complex models often cause the problem of overfitting the learning data, and sacrifice the easy understanding that the original technical analysis has. In addition, big data obtained from online stock boards [8] or Twitter [9] are also used to extract market trends. In a sense, these machine learning methods are widely considered as technical analyses, but the calculation process tends to be a black box.

On the other hand, an advantage of the traditional technical analysis [10] is its simplicity, which can make it easy for traders to understand the given results. Instead, its prediction power might be sacrificed for this easy understanding. As another problem, previous technical analyses are often late in signalizing entry timings because they are based on statistical values such as the moving average and the standard deviation of the recent historical data. Until we have corrected enough historical data, we cannot notice a trend occurrence or a trend reversal.

To solve these problems, we propose new technical indicators based on the nonlinear dynamical theory, as simple as possible to keep the advantage of technical analysis. For this purpose, we apply a chaos prediction model [11,12], which is relatively simple compared to other nonlinear prediction methods. This chaos prediction model can use spatial neighbors for prediction instead of temporal neighbors (i.e., recent historical data), and therefore technical signals can be indicated more quickly. Then, if a financial market has a few dynamics, this dynamics can be reconstructed in a multidimensional state space by the embedding method [13] as a spatial pattern hidden in the original financial data.

Next, to make the most of these spatial neighbors, we also apply the bagging algorithm proposed by Ref. [14]. This is a kind of the ensemble learning technique, which can reproduce many predictors and improve the prediction accuracy by accepting the majority of their predicted values (see Ref. [15]). In the field of the nonlinear dynamical theory, Ref. [16] applied this ensemble learning to a chaos prediction model, and reported its improvement. Moreover, Ref. [17] focused on the standard deviation of the ensemble set and used it to estimate the prediction risk. Then, in terms of financial engineering, Ref. [18] utilized this prediction risk to compose an efficient stock portfolio, but it leaves a problem that estimating a covariance requires so large embedding dimension for the limited learning data that nonlinear predictions do not work well due to the so called "curse of dimensionality". In addition, these Refs. [16–18] simplified the ensemble learning for easy calculation and less computational cost. If the ensemble learning is exactly applied, we have to resample all of the learning data, not only local neighbors, to generate bootstrap ensembles as mentioned in Section 3. For example, Ref. [19] used the exact bootstrap method to estimate the Lyapunov exponents. However, these Refs. [16–18] did not discuss this viewpoint.

As another difference with Ref. [18], the present study focuses on the single-stock investment in order to reduce the problem of large embedding dimension, and adaptively selects the best stock which shows the highest confidence in prediction. Here, to quantize the confidence, we propose new technical indicators based on the ensemble learning, and these indicators are designed in the similar way of the conventional technical indicator: the relative strength index (RSI) [20] for traders familiar with the technical analysis. Then, we also discuss the difference of how to generate bootstrap ensembles as pointed out above. Finally, to confirm the validity of our ideas, we perform investment simulations with real financial data during the four terms in the Japanese market and the American market.

2. Classical technical indicators

As shown by Ref. [10], plenty of technical indicators have been proposed. Basically, classical indicators tend to be sluggish to recognize a trend occurrence or a trend reversal because they have to collect enough evidence from the recent historical data. In this section, we introduce the relative strength index (RSI) as an example, which will be applied to our indicators in Section 3.

Let $x(t)$ denote the stock price at time t, and let $\Delta x(t)$ denote the price movement. Thus, $\Delta x(t) = x(t) - x(t-1)$. The RSI is used to quantify the momentum of market trends through the recent historical price movements $\Delta x(t-n)$, $(n = 0, 1, \ldots, T-1)$, and to predict trend reversals. Here, if $\Delta x(t-n) \geq 0$, we denote its rise width $\Delta x(t-n)$ as $U(t-n)$ and set its fall width $D(t-n)$ to 0. On the other hand, if $\Delta x(t-n) < 0$, we set $U(t-n) = 0$ and $D(t-n) = |\Delta x(t-n)|$. The original definition of RSI is as follows:

$$I_{\mathrm{RSI}}(t) = \frac{\sum_{n=0}^{T-1} U(t-n)}{\sum_{n=0}^{T-1} U(t-n) + \sum_{n=0}^{T-1} D(t-n)} \times 100[\%], \tag{1}$$

where $I_{\mathrm{RSI}}(t) \in [0, 100]$. As $I_{\mathrm{RSI}}(t)$ is larger or smaller than 50[%], we can more strongly consider that the recent trend has been bullish or bearish, and therefore this trend would be reversed soon. Namely, the original RSI is used as an oscillator indicator. For example, if $I_{\mathrm{RSI}}(t)$ is less than a threshold θ, it is better to make a long (buy) position at this time t.

Fig. 1 shows an example of the original RSI. As mentioned above, because conventional indicators are based on temporal statistical values, the movement of $I_{\mathrm{RSI}}(t)$ is so slow that we can notice a trend changing after it has happened. In the next section, to reduce this time delay, we use a chaotic prediction model to learn moving patterns hidden in the historical data.