Winter 12-9-2003

An Behavioral Finance Analysis Using Learning Vector Quantization in the Taiwan Stock Market Index Future

An-Pin Chen

Yi-Chang Chen

Pei-Chen Lee

A Behavioral Finance Analysis Using Learning Vector Quantization in Taiwan Stock Market Index Future

An-Pin Chen
Institute of Information Management, National Chiao Tung University
Hsinchu, Taiwan
apc@iim.nctu.edu.tw

Yi-Chang Chen
Institute of Information Management, National Chiao Tung University
Hsinchu, Taiwan
ycchen@iim.nctu.edu.tw

Pei-Chen Lee
Institute of Information Management, National Chiao Tung University
Hsinchu, Taiwan
ricelee@iim.nctu.edu.tw

Abstract

There are various types of trading behavior in the stock market. And the buying or selling activities in many investment strategies are influenced by numerous factors respectively, such as fundamental analysis, macroeconomic analysis, and news analysis. Consequently, various factors will reflect on market price. Random Walk in financial engineering is not the focus in this paper. Otherwise, the importance of the technique analysis about Taiwan Stock Index Futures will be emphasized in this research.

It is the intention of this paper to investigate the information content of Open, High, Low, Close prices in the previous trading day and relative higher and lower points in the prior period of the current trading day, as well as their prices in analyzing Taiwan Stock Index Future. The predictability of Learning Vector Quantizationl Network can clearly be seen from the empirical result.

Keywords： Neural Network, Learning Vector Quantization, Taiwan Stock Index Future

1. Introduction

From economic viewpoint, the prices of goods are determined by supply and demand. Buyers and sellers made their own trading decisions according to different view about the target goods. The price will rise if demand falls short of supply; On the contrary, if supply falls short of demand, price will decline. In other words, the price of goods is a common consensus for all participants about the value of goods itself. It is the same with financial commodities.

The prices of financial commodities are the trading results of buyers and sellers. Buyers and sellers made their own trading decision according to different trading rules, including fundamental analysis, macroeconomic analysis, and news analysis. Therefore, these factors have influence on determining stock prices as well. Figure 1 describes how these trading rules affect the stock prices continuously, such as fundamental factors, macroeconomic factors and etc. In this paper, we assumed that all factors are functions of time and will be continuous time relative data.

$$Stock\ price = Real\ price + Effect\ price \quad (1)$$

$$Effect\ (fundmental) = f_1(t) \quad (2)$$

$$\vdots$$

$$Effect(macroeconomics) = f_2(t) \quad (3)$$

Because of the restrictions on trading hours, the curve of market prices is discrete in time-axis. Therefore, it is impossible for market prices to reflect immediately the effect of all factors. After the market closing, all elements influencing market prices will be reflected on the Open prices in the next trading day. The non-cash-period represents expectation formation process.

For this reason, this research will focus on whether all factors between 13:45 and 08:45 (period T) from previous day will affect the Open price (point E) of next trading day. Furthermore, it is suspected that the sum of all factors will be the Open price of the next trading day.

$$price_{Open} = \sum_{e=1}^{n} Effect\ (e) = \sum_{e=1}^{n} f_e(T) \quad (4)$$

$Open$ ：open price
e ：element influencing prices
n ：number of elements
T ：time period between two trading times

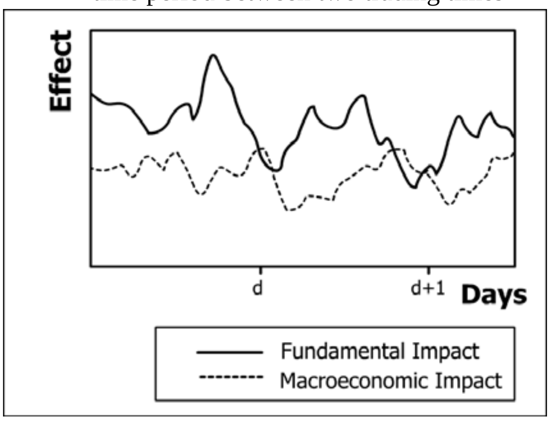

Figure 1. The factors influence the stock prices continuously, such as fundamental factors, macroeconomic factors and etc.

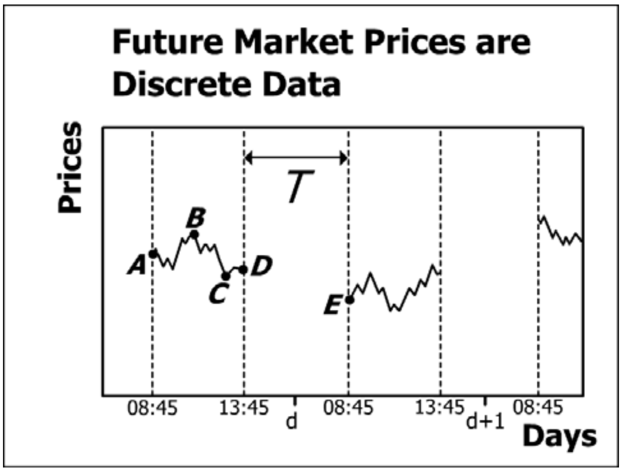

Figure 2. An illustration of the Open, High, Low, Close prices in the previous trading day as point A, B, C and D.

The relevance between the Open, High, Low, Close prices in the previous trading day, and relative higher and lower points in the prior period of the current trading day is discussed by using artificial Intelligence in this paper. The purpose is to make profits and to avoid risks by analyzing Taiwan Stock Index Future.

2. Literature

2.1 Literature Review

2.1.1 Literature in Stock Market Behavior

The overnight effect on the index future is discussed in this paper. Research on forecasting stock returns, such as Campbell and Shiller in 1988, and Bollerslev and Hodrick in 1992, have indicated that stock returns are predictable by using linear models.

Menkhoff (1998) indicated that changing order flows play an important part in determining market prices. Tian Shyue Lee and Nen Jing Chen [8] showed that the lead-lag relationship between the futures market during the non-cash-period and the cash market during its opening period does exist. The empirical results of the above research show that the prediction result of the model is better than random walk model. Therefore, this research is based on non random walk process.

2.1.2 Literature in Application of Artificial Intelligence.

Neural network has been widely applied in several financial areas, such as exchange rate prediction, bankruptcy forecasting and stock market prediction. For exchange rate prediction, Jingtao Yao and Chew Lim Tan [2] stated that a neural network model is applicable to the prediction of foreign exchange rates. For bankruptcy forecasting , Moshe Leshno and Yishey Spector [5] stated that the prediction capabilities of neural network model in this bankruptcy case are clearly more accurate than other classical discriminate analysis model. For stock market prediction, Mary Malliaris and Linda Salchenberger [4] used neural network to forecast the S&P 100 implied volatility.

2.2 Candlestick Analysis

Open, High, Low and Close prices defined in candlestick analysis are used to be input factors in this research. Norbert M. Fiessa and Ronald MacDonald indicated that the difference between High and Low prices marks the intradaily trading range and measures volatility. The difference between the Open and the Close prices thus serves as a measure of the direction and the extent of intradaily trends. The academic survey on support and resistance levels by Curcioand Goodhart (1992), DeGrauwe and Decupere (1992) showed clearly that High and Low prices reveal information about shifts in the demand and supply structure. Therefore, Open, High, Low and Close prices are the inputs to the model in this research.

2.3 Learning Vector Quantization

Learning Vector Quantization of neural network is selected in this paper. Neural network is one part of Artificial Intelligence. It is an information-processing paradigm inspired by the way the densely interconnected, parallel structure of the mammalian brain processes information. The key element of the ANN paradigm to be differentiated form other methods of Artificial Intelligence is self-learning. Neural network is also well known for its ability to process nonlinear problem.

Ramazan Gençay[6] pointed out that strong evidence of non-linear predictability is found in the stock market returns by using the past buy and sell signals of moving average rules. Therefore, artificial neural network is applied in this paper to improve predictability of stock market and to raise the profitability.

Learning Vector Quantization is a competitive learning algorithm said to be a supervised version of the Self-Organising Map (SOM) algorithm by T. Kohonen[7].

Reference vectors are the weights of the connections leading from the input neurons to the output neurons. For each training sample, the reference vector that is closest to it is determined. Only the weights of the connections to this winner neuron are updated. The direction of the adaption depends on whether the class of the training sample and the class assigned to the reference vector identify or not. If the training sample is placed into the correct class, the reference vector is moved closer to the training sample; otherwise it is moved farther away. This movement of the reference vector is controlled by a parameter called the learning rate [Christian Borgelt, 2000].

The algorithm of LVQ is listed as follows：

1. Initialize weights; initialize learning rate.
2. Do step 3-7 until stopping condition is true.
3. For each input vector X ，repeat steps 4-5
4. Find node J so that $\left\| X - W_J \right\|$ is a minimum distance.
5. Update W_J according to follows ：

   If X is classified to the correct class $W_J^{new} = W_J^{old} + \alpha(X - W_J^{old})$

   If X is classified to the incorrect class $W_J^{new} = W_J^{old} - \alpha(X - W_J^{old})$
6. Reduce learning rate.
7. Test stopping condition. It may be a fixed number of iterations or the learning rate reaching a sufficiently small value.

3.Architecture

3.1 The Research Process

First, data of Taiwan Stock Index Future is collected. The data of each trading day is processed as follows.

1. According to candlestick analysis，every Open, High, Low and Close prices are selected using daily minute ticks.
2. All apexes are found by the sign of multiple of two neighboring slopes, where these slopes are determined from each two neighboring points in daily minute ticks.

The input nodes of neural network model will be selected after completing the above steps. Then, the model will be trained with part of collected data. At last, there will be a performance validation with the rest of collected data. The research architecture is showed as Figure 3.

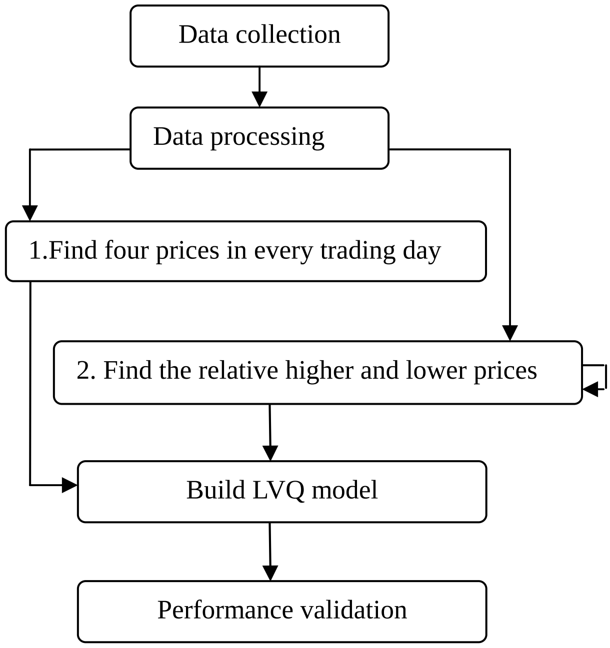

Figure 3. The research process of the study

3.2 The Neural Network Model

According to candlestick analysis mentioned in section 2.2, it is the belief that Open, High, Low and Close prices have a higher information content than other prices to extract featured prices of intradaily trading. These four prices are used as input nodes in neural network model. The relative higher and lower points found after the opening of future market and stock market are also parts of the input nodes in neural network model. Hence, the Open, High, Low and Close prices in the prior period can be represented by these relative higher and lower points after future market and stock market opening. Based on fractal theory[3], pattern feature extraction is irrelevant to the length of period. Consequently, four prices in the previous trading day and four prices in the prior period after opening are considered as the factors to influence future prices. The input nodes and output nodes are listed as follows:

Table 1 Input nodes of neural Network model

1. True
2. False

Table2 Output nodes of neural Network model

1. The 1st relative higher point after future opening
2. The 1st relative lower point after future opening
3. The 2nd relative higher point after future opening
4. The 2nd relative lower point after future opening
5. The 1st relative higher point after stock opening
6. The1st relative lower point after stock opening
7. The 2nd relative higher point after stock opening
8. The 2nd relative lower point after stock opening
9. Open price in the previous trading day
10.High price in the previous trading day
11.Low price in the previous trading day
12.Close price in the previous trading day

4.Experiment Result and Analysis

In this paper, the minute ticks of Taiwan Stock Index Future between 2001/04/13 and 2002/04/12 are used. Every minute price in one trading day is characterized as a record, totaling 204 records. 80% of records are taken as the training sample to build the neural network model. The remaining 20% records are tested for the accuracy of this model.

The first step is to process original data by computing the slope of every two neighboring points. The trend of stock market index could be determined by using the sign of slops as formula (5).

$$m = \frac{p_2 - p_1}{t_2 - t_1} \quad (5)$$

m : slope
t : time point
p_t : future price at time t

By using the sign of the product of the two neighboring slopes, the apex is determined to be the relative higher point and lower point. The formula is listed as follows:

$$s = sign(m_t) * sign(m_{t+1}) \quad (6)$$

s : sign of product of two slopes

According to formula (5) and formula (6), relative higher and lower points can be found as Figure 4.

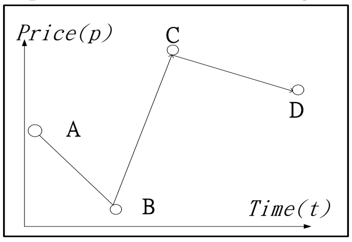

Figure 4. The relative higher and lower points

The slope of AB (m_t) is negative, and the slope of BC (m_{t+1}) is positive. So the product of m_t and m_{t+1} is negative, Because of a negative sign, point B is a relative higher and lower point.

This research takes relative higher and lower points and the four prices as the input nodes in this neural network model. The nodes are listed as Table 1 and Table 2. Nodes will be placed into True group if it is higher than or equal to the *higher price*, which is 0.5% higher than Open price calculated in formula (7), and the total number of True item will increase accordingly; otherwise, it will be placed into False group.

The formula to *higher price* is listed as follows:

$$h = O + 0.5\% * O \quad (7)$$

h : *higher price*
O : Open price

$$\text{If } H > h \text{ then } T$$
$$\text{If } H < h \text{ then } F \quad (8)$$

H : High price

The ratio of success is the percentage of correct identification. The formula is listed as follows:

$$S = \frac{T}{T + F} \quad (9)$$

S : ratio of success
T : the number of correct classification
F : the number of incorrect classification

After data processing, Learning Vector Quantization is used to build a Neural Network model. The sequence of the input samples is randomly selected. This research also uses random weights and SOM (Self- Organizing Maps, SOM) to get the initial weights. After several runs, the representative results are listed in Table 3.

Table 3 Statistics of the experiment (iteration is100)

experiment run	initial weights	learning rate	ratio of success
1	**(W_1, W_2)**	**0.2**	**80%**
2	(W_1, W_2)	0.2	35%
3	(W_1, W_2)	0.1	65%
4	(W_1, W_2)	0.1	37.5%
5	(W_1, W_2)	0.2	35%
6	(W_1, W_2)	0.2	30%
7	(W_1, W_2)	0.1	17.5%
8	(W_1, W_2)	0.1	27.5%

(w_1, w_2) are different initial weights in different runs of experiment.
$W_1 = \{w_{1i}, i = 1,2,3,...,12\}$, $W_2 = \{w_{2i}, i = 1,2,3,...,12\}$, (w_1, w_2) are showed as Figure 5:

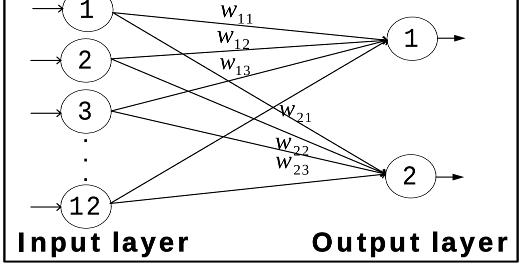

Figure 5. The structure of neural network model

Table 3 shows that the first four runs used random initial weights, and the last four runs used SOM initial weights. When using random initial weights, success rate can achieve 80% with learning rate of 0.2 and iteration number of 100. Hence, it is obvious that selecting random initial weights is better than selecting SOM weights.

This research is the application of classification in Neural Network. In this paper, there are two analytic methods, including BPN (Back-Propagation Network) and LVQ (Learning Vector Quantization). Figure 6 shows the result of BPN training process in this experiment, and also shows the unstableness of BPN. On the contrary, LVQ is more stable instead. Figure 7 describes the increase on success rate and stability of LVQ. Therefore, Learning Vector Quantization is chosen in this paper.

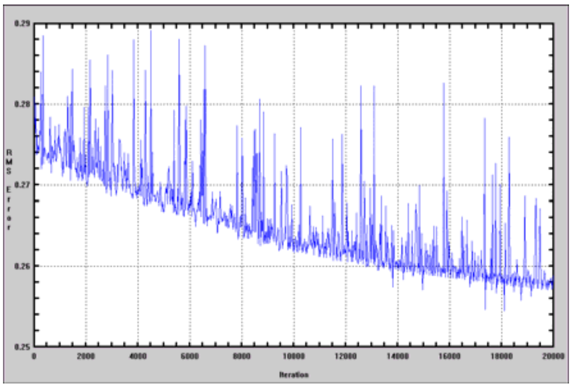

Figure 6. Te result of BPN training process in this experiment, and also shows the unstableness of BPN

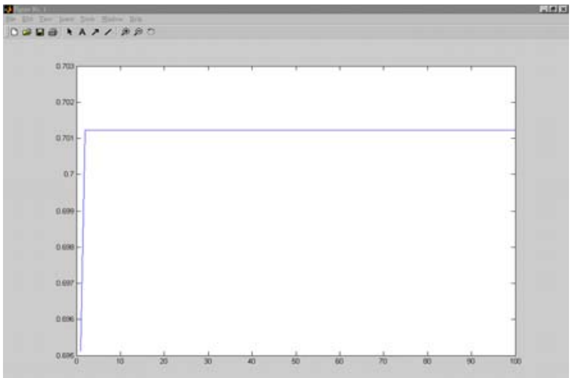

Figure 7. The increase on success rate and stability of LVQ

The relevance between the four prices in the previous trading day and the relative higher points and lower points in the prior period of current trading day is well analyzed using LVQ. The success rate of 80% represents the existence of behavior model, and disproved the unpredictableness of the stock market. In other words, this research emphasized analysis on behavior model. The phenomenon happened in the past can imply the occurrences of investment behavior in the future. This concept is a concept in behavioral science under conditional probability, not Random Walk Process stressed in financial engineering.

5.Conclusion

By using Learning Vector Quantization, this paper discussed about the gap of the market price between the two trading time. Furthermore, it showed that Open, High, Low, Close prices in the previous trading day and the relative higher and lower points are related to Taiwan Stock Index Future. The result can be applied to the strategy of arbitrage and speculation. As arbitrage between TAIMEX and SIMEX, arbitragers should have a short position on TAIMEX and have a long position on SIMEX or have a opposite position on SIMEX and TAIMEX. If the analysis on TAIMEX or SIMEX is correct, it is possible for investors to lower the trading cost of one. For example, Open price of TAIMEX is high, but the analysis shows that there will be a higher price in the trading hours. Arbitragers can offset the short position of TAIMEX, and repurchase the position back before closing. For arbitragers, this trading strategy can increase total profit by lowering the cost at the loss position. It can avoid the fluctuation of prices between the two trading times and make fixed profit.

Further research can focus on the application of Fuzzy system to judge amplitude of prices volatility and the application of trading volume to reflect momentum of buying and selling with the purpose to approve the accuracy of neural network model. In other words, it is intended to minimize the probability of loss and to increase profit.

Reference

[1] Campbell JY, Shiller RJ. 1988. “The dividend-price ratio and the expectations of future dividends and discount factors”. *Review of Financial Studies* **1**: 195–228.
[2] Jingtao Yao, Chew Lim Tan. “A case study on using neural networks to perform technical forecasting of forex,” *Neurocomputing*, 2000, 34, 79-98
[3] Mandelbrot, Benoit B. "The Variation of Certain Speculative Prices," *Journal of Business*, 1963, 36, 394-419.
[4] Mary Malliaris, Linda Salchenberger. “Using Neural Networks to Forecast the S&P 100 Implied Volatility,” *Neurocomputing*, 1996, 10, 183-195
[5] Moshe Leshno, Yishey Spector. “Neural Network Prediction Analysis: The Bankruptcy Case,” *Neurocomputing*, 1996, 10, 125-147
[6] Ramazan Gençay, “Non-linear prediction of security returns with moving average rules,” *Journal of Forecasting*, 1996, 15,165-174
[7] T. Kohonen. “Learning vector quantization,” *Helsinki University of Technology, Laboratory of Computer and Information Science Report* , 1986TKK-F-A-601.
[8] Tian Shyue Lee, Nen Jing Chen, “Investigating the information content of non-cash-trading index futures using neural networks,” *Expert Systems with Applications*, 2002, 22, 225-234