ISSN 1 746-7233, England, UK
World Journal of Modelling and Simulation
Vol. 2 (2006) No. 1, pp 36-44

Fundamental analysis of stock price by artificial neural networks model based on rough set theory *

Wei Wu , Jiuping Xu †

Uncertainty Decision Making Laboratory
School of Business and Administration, Sichuan University, Chengdu 610064, P. R. China

(*Received July 26 2005, Accepted December 1 2005*)

Abstract. It has been widely accepted that predicting stock price is not a simple task since many market factors are involved and their structural relationships are not fully known. In this study, we use both rough set theory and neural networks approach to get an effective model of stock price movement for China's young stock market. The model is modified and tested by the most recent 6 years of data collecting from China's stock market to make sure it is updating and withstanding in a long time. Consequently, a group of most important fundamental indicators are selected by rough set theory and these indicators are successful in detailing and predicting stock price movement in a long-term using neural networks approach. Results indicate that neural networks approach based on rough set theory is efficient in modelling and more accurate in prediction.

Keywords: stock price, rough set, artificial neural networks

1 Introduction

Stock price prediction is a famous problem which has been studied by both academics and practitioners for years. Although many studies have investigated the prediction of price movements in the stock market, stock price movements as a financial time series are too complex and noisy to forecast[27]. Hence, the efficient market hypothesis states that all available information affecting the stock price has already been exerted by the market before the general public can make trades based on it [8]. Nonetheless, this is still a debating issue because there are considerable evidences that markets are not fully efficient, and it is possible to predict the future stock prices or indices with results that are better than random results[14]. In fact, many researchers provide evidence that stock market returns are predictable by means of publicly available information such as time-series data on financial and economic variables and so on [1, 2, 9, 24].

The publicly available information mainly include financial and economic variables that affecting on the firms and stock markets. Meanwhile, most of the factors which were considered by various research to exploit the information changing stock price are basically belong to either technical analysis or fundamental analysis. Technical analysis studies the historical data surrounding price and volume movements of the stock by using charts as a primary tool to forecast future price movements[16]. Investors base their analysis on the premise that some historical patterns of stock prices are assumed to repeat in the future, thus, these patterns can be used for predictive purposes. On the other hand, fundamental analysis concentrates on the economic factors that cause prices to move up and down [23]. The relevant economic factors affecting stock prices are

* This research was supported by the National Science Foundation for Distinguished Young Scholars, P. R. China (Grant No. 70425005) and the Teaching and Research Award Program for Outstanding Young Teachers in Higher Eduction Institutions of MOE of P. R. China.(Grant No. 20023834-3)
† Corresponding author. Tel.: +86-28-85418522; fax: +86-28-85400222.
E-mail address: xujiuping@openmba.com, wuwei@eyou.com.

examined in order to determine the intrinsic value of securities. When attempting to determine the likely price changes, the two techniques approach the objective of predicting stock movement from different directions: fundamental analysis studies the cause of market movement, while technical analysis studies the effect. The technical aspect can include psychological factors, supply-demand relations and random variation, etc.. All these factors make the stock price go up and down, however, the movement of price can not always deviate the intrinsic value of that security. Therefore, in this study, we concentrate mainly on more essential things, that are the fundamental factors causing stock price change.

Recently, a considerable number of studies have been done attempting to address the ability of neural networks to predict future stock movements by using fundamental factors, such as Desai and Bharati [4], Dropsy [5], Motiwalla and Wahab [15], Poddig and Rehkugler [20], Qi [21], Qi and Maddala [22]. Although most of the above-mentioned studies provide valuable findings, they each rely on either a fewpopular fundamental factors or some famous technical indicators.The reason why choosing such factors have not been fully detailed. And the mechanism of various factors affecting on the stock price is not fully exploited in the literature. In addition, pure Neural Networks have some limitations in learning the patterns since stock price data has tremendous noise and complex dimensionality[10]. Recently, promising results were obtained by incorporating a data mining technique used in machine learning to uncover the predictive powers of numerous financial and economic variables[26]. This approach seems particularly attractive in selecting the variables when the usefulness of the data is unknown, especially when non-linearity exists in the financial market and there are a considerable number of noisy and errors.

Besides the improved method, there are many advantages to study the stock price movement in China[13]. Firstly, it is a young market with plenty of asymmetric information status, which cause effective market hypothesis is not fully exerting its affection. Secondly, stock price can be change obviously by the newly coming information from annual report. Thirdly, lacking sufficient number of institute investors makes China's stock market more sensible to the various information. Therefore, the stock price movement in China stock market deviate far from the random walk, which is suitable for finding trading rules and profitable trading model/system.

In this paper, we will concentrate mainly on analyzing the fundamental factors which affect on the stock price in a long-term and a forecasting model will be developed to estimate the exact price changes. The remainder of this paper is organized as follows: a summarized stock price problem will be presented in the next section, where we will focus on the mechanism of stock pricing and theoretical foundation of our analysis. The third section discusses the data pre-processing and some other details of our modelling. The fourth section analyzes the results getting from the modelling process. The last section summarizes our conclusion and do some discussions.

2 Approach

2.1 Paradigm

All the fundamental factors source from both macro-environment and micro-environment of a certain firm. Theoretically, the macro-environment should include all the external factors of a firm, which are able to affect on not only a firm, but also the whole stock market, such as macroeconomic factors, political factors, industrial factors, market factors, and so on. However, most of these factors are not available or too difficult to be quantified. Hence, it is hard to model stock price movement with all such kind of factors. Yet, we can use Stock Market Index (SMI) to represent all the macro-environment factors affecting stock price. This is testified to be highly correlated to stock price in the literature[27].

Micro-environment can be divided into financial factors and non-financial factors. Generally, there are some specific which construct the financial factors of stock price movement: Profitability, Risk, Development, Current and Operation. Profitability specific is the one of the most important aspects for stock price, and we select some of most representative factors to describe such specific, that are Earnings Per Share, Net Assets Per Share, Return Ratio on Assets, Major Business Profit Ratio, Rate of Return on Sale, and Net Assets Income Ratio. Risk specific is another important aspect to stock price, which contains both current debt and non-current debt. Current Ratio and Quick Ratio are main factor of current debt, while Debt to Assets Ratio is representative to non-current debt. Development specific reflecting the continuity of a firm's profitability, can be denoted as Major business Growth and Net Assets Growth. Current specific mainly means Net Cash Flows per Share from operating activities reflecting firm's ability to gain cash. Operation aspect tells public the firm's efficiency, it can be summarized into Accounts Receivable Turnover, Inventory Turnover, and Total Assets Turnover. Table 1 illustrates the selected factors of our model and their computation.

Table 1. Financial factors list

Specific	Factor	Abbreviation
Profitability	Earnings Per Share	EPS
	Net Assets Per Share	NAPS
	Return Ratio on Assets	RRA
	Major Business Profit Ratio	MBPR
	Rate of Return on Sale	RRS
	Net Assets Income Ratio	NAIR
Risk	Current Ratio	CR
	Quick Ratio	QR
	Debt to Assets Ratio	DAR
Development	Major Business Growth	MBG
	Net Assets Growth	NAG
Current	Net Cash Flows per Share from operating activities	NCFS
Operation	Accounts Receivable Turnover	ART
	Inventory Turnover	IT
	Total Assets Turnover	TAT

Non-financial factors also sway stock price, such kind of factors usually can be found in technical analysis. As we focus on fundamental factors, we only select non-financial factors which are able to change the stock price for a long time. The change of Total Shares illuminates the firm has had some action on its stock share structure, such as dividend or allotment of shares, which no doubt affects stock price. Moreover, the Number Of Shareholders (NOS) reflects the degree of shares centralization: when shares centralization is high, a few people may change the stock price as they want. Therefore, all the above mentioned two non-financial factors are considered as important to stock price in this study.

Although some other detailed criteria are also able to affect on stock price, most of the fundamental information are incorporated in our selected criteria.

2.2 Theorem

Rough set theory is a method qualified for discerning relation among factors, and a powerful soft-computing tool to handle vagueness and uncertainty in data sets, which was proposed by Pawlark[18]. The concept of rough sets theory is founded on the assumption that every object of the universe of discourse is associated with some information[11]. Objects characterized by the same information are indiscernible in view of their available information. The indiscernibility relation generated in this way is the mathematical basis of the rough sets theory. Important problems that have been solved by rough sets theory are: checking dependencies between attributes, reducing attributes, analyzing the significance of attributes, and generating decision rules[19]. For a detailed review on the rough sets theory, readers may refer to Pawlak[19], Komorowski et al.[12].

The information regarding the objects is supplied in the form of a data table, whose separate rows refer to distinct objects, and whose columns refer to different attributes considered. Each cell of this table indicates an evaluation of the object placed in that row by means of the attribute in the corresponding column. The information table is composed of 4-tuples, $S =< U, Q, V, f >$, where $U = \{x_1, x_2, \ldots, x_n\}$ is a finite set of objects(Universe), $Q = \{q_1, q_2, \ldots, q_m\}$ is a finite set of attributes, $V = \cup_{q\in Q} V_q$ is a set of attribute values while V_q is the domain of the attribute q, and $f : U \times Q \rightarrow V$ is a total function such that $f(x, q) \in V_q$ for each $q \in Q$, $x \in U$, called information function. Any pair (q, v) from the sets $q \in Q$, $v \in V_q$ is a descriptor in S. The attributes in Q are composed of two disjoint subsets, condition attributes C and decision attributes D, such that $Q = C \cup D$ and $C \cap D = \varnothing$.

To every non-empty subset of attributes P is associated an indiscernibility relation on U, denoted by I_P : $I_P = \{(x, y) \in U \times U : f(x, q) = f(y, q), \forall q \in P\}$. If $(x, y) \in I_P$, it is said that the objects x and y are P-indiscernible. Thus an equivalence relation can be defined by the indiscernibility relation. The family of all the equivalence classes of the relation I_P is denoted by $U|I_P$ and the equivalence class containing an element $x \in U$ by $I_P(x)$. The equivalence classes of the relation I_P are called P-elementary sets because it represents the smallest indiscernible groups of the P subset of attributes. All the elementary sets of a universe form basic granules of knowledge about the universe. Therefore, two equivalence relations I_C and I_D can be defined for an information table S.

A great practical importance of Rough set theorem is that superfluous data can be eliminated, in fact, without deteriorating the information contained in the original data table. Let $P \subseteq Q$ and $p \in P$. It is said that attribute p is superfluous in P if $I_P = I_{P-\{p\}}$; otherwise, p is indispensable in P. The set P is independent if all its attributes are indispensable. The subset P' of P is a reduct of P (denoted by $Red(P)$) if P' is independent and $I_{P'} = I_P$.

The reduct of P may also be defined with respect to an approximation of a partition Y of U. It is then called Y-reduct of P (denoted by $Red_Y(P)$) and specifies a minimal subset P' of P which keeps the quality of classification unchanged, i.e. $\gamma_{P'}(Y) = \gamma_P(Y)$. In other words, the attributes that do not belong to Y-reduct of P are superfluous with respect to the classification Y of objects from U. More than one Y-reduct of P may exist in a data table. The set containing all the indispensable attributes of P is known as the Y-core. Formally

$$Core_Y(P) = \bigcap Red_Y(P).$$

Obviously, since the Y-core is the intersection of all the Y-reducts of P, it is included in every Y-reduct of P. It is the most important subset of attributes of Q, because none of its elements can be removed without deteriorating the quality of classification.

In our application of Rough set theory, objects can be interpreted as samples of data observed in the stock trading, and attribute can be interpreted as factors influencing the stock price. Rough set has the ability to reduce the dimensionality of the data matrix. Then, the reduced data matrix will be used to train a neural network. In general, a neural network is a set of connected input and output units where each connection has a weight associated with it. For each training sample the input variables are fed simultaneously into a layer of processing units making up the input layer. The weighted outputs of these units are, in turn, fed simultaneously to a second layer of processing units known as a hidden layer. The hidden layers weighted outputs can be input to another hidden layer, and so on. The weight outputs of the last hidden layer are input to units making up the output layer which issues the networks prediction for a given set of samples. During the learning phase, the network learns by adjusting the weights so as to be able to correctly predict the output target of a given set of input samples. Feed-forward neural networks as a wildly used type of neural networks, is selected to be implemented in this study to handle the results getting from rough set theory selection.

During neural network modelling, it has been suggested that the proper number of hidden layer nodes requires validation techniques to avoid under-fitting and over-fitting. It has been widely accepted that a three-layer feed forward network with an identify transfer function in the output unit and logistic functions in the middle-layer units can approximate any continuous functions arbitrarily well, given sufficiently many middle-layer units[21]. The generic three-layer network model can be expressed in equation 1.

$$Y_t = f[X, \alpha, \beta] = \sum_{j=1}^{n} \alpha_j \log sig(\sum_{i=1}^{k} \beta_{ij} x_i + \beta_{0j}) \tag{1}$$

where Y_t is the networks output, X is the input vector, x_i is i th input, n is the number of units in the middle layer, k is the number of inputs, α represents a vector of the weights from the middle to output layer units, β indicates a matrix of the coefficients from the input to middle-layer units, α_j is the weight of the output layer that connects the j th hidden layer unit to the output, $\beta_j = \{\beta_{ij}, i = 1, 2, ..., k\}$ is the weight vector of the j th unit of the middle layer, β_{0j} is the bias weight of the j th unit of middle layer unit, and $\log sig$ is the logistic transfer function $\log sig(a) = 1/[(1 + exp(-a)]$.

Backpropagation is by far the most popular neural network training algorithm that has been used to perform learning on feed-forward neural networks. It is a method for assigning responsibility for mismatches to each of the processing units in the network, which is achieved by propagating the gradient of the activation function back through the network to each hidden layer, down to the first hidden layer. The weights are then modified so as to minimize the mean squared error between the networks prediction and the actual target. The output value for a unit j is given by the following function:

$$O_j = G(\sum_{i=1}^{n} w_{ij}x_i - \theta_j), \tag{2}$$

where x_i is the output value of the i th unit in the preceding layer, w_{ij} is the weight on the connection from the i th unit, θ_j is the threshold, n is the number of units in the preceding layer, and G is the activation function. Fig 2.2 shows the configuration of our experiment single-hidden layer neural network which will be used later in this study. Readers who are interested in greater detail can refer to [41] for an explanation of the backpropagation algorithm used to train feed-forward neural networks.

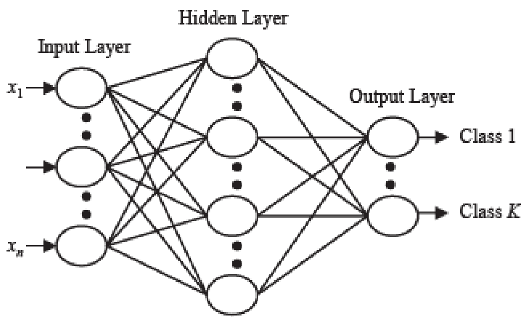

Fig. 1. Neural networks stucture

3 Modelling

3.1 Preparation

The data for this study are obtained from the following sources: (1) CSMAR Databases which is commercially available at Shenzhen GTA Information Technology company Ltd; (2) Firms' annual reports, stock price and stock market index which are available at Shanghai Stock Exchange (http://www.sse.com.cn) and Shenzhen Stock Exchange (http://www.szse.cn). The initial samples are selected from all stock traded in one of the above-mentioned China's two stock exchanges during the period from December 1999 to April 2005, and those data which come from firm year with dividing or allotting shares are deleted to make sure the share structure remain unchanged between two consecutively comparable years. Again, we concentrate our research on firms which are in normal business activities, i. e., we only select samples from stock that EPS is greater than ¥0.05, and NAPS is more than ¥1. Therefore, some abnormal phenomena of price movement result from firms' operation are exclude from this experiment, so as to make the result generally hold.

As far as stock market index is concerned, we use Shenzhen Stock Exchange Sub-Index to reflect the stock market's external environment. This sub index is a market value weighted index which consists of dozens of stocks, and is considered as historical comparative, accurate to market movement, and admissive ones[6]. Meanwhile, to test the fundamental factors affecting on stock price, the study primarily bases on annual reports which usually published from late January to April every year. Following Ou and Penman[17] and Holthausen and Larcker[7], the data for fundamental analysis are used with a 4-month lag to insure that investors actually have access to the data at the time of the investment decision. This means that the accounting ratios are used to forecast 1-year-ahead returns that are cumulated from months 4 to 16 following the publication of annual fiscal year accounting data. We set the average closing Shenzhen Sub-Index of a December as base period index, every stock's average exchange price of a December as its base period price; and the average closing Shenzhen Sub-Index of the next April as current period index every stock's average exchange price of the next April as its current period price respectively. The ratio of both index and price change can be calculated then.

Finally, the number of samples is 1106, which is about 15.32% of all the stock data samples generated in the 6 years. Hence, the results could be considered as statistical significance and representatives. All the samples are divided into two group randomly, the first group composed 900 samples, about 80% of all the samples, are used to rough set selection and neural networks training, while the remaining 206 untouched samples (about 20% of all samples) are used to test the forecast ability of the whole model. That is, the first group is used for training and validating the forecasting models, while the second group is reserved for out-of-sample evaluation and comparison of performance among the forecasting models.

To utilize rough sets theory, we also discretize all the sample data. We use quartiles to separate continuous numbers into discrete numbers [0,1,2,3], which is useful to reduce noisy containing in the data. When the number of a criterion is smaller than the first quartile, change the number to 0; When the number of that criterion is smaller than the second quartile and not smaller than the first quartile, change the number to 1; When the number of that criterion is smaller than the third quartile and not smaller than the second quartile, change the number to 2; for all other cases, change the number to 3. For the change of stock market index, we set 0.045 as the threshold, this means if the index rising ratio above 0.045, then the macroenvironment of this period is consider good, let it be 1; if the index change ratio is less than 0.045, then the macroenvironment of this period is consider normal, let it be 0; if the index falling ratio above 0.045, then the macroenvironment of this period is consider bad, let it be -1.

3.2 Process

During the experiment, we firstly use rough set theory to select the useful factors from all factors. Using rough set reduct arithmetic compiled in C language, one can get the reducts. Secondly, we delete all the redundant factors indicated by the reducts and use the factors in the reducts as input of neural network, and linearly transform those data into interval scale such as [0, 1] or [-1, 1], since the neural network is good at explain interval variables[3]. Thirdly, we train feed-forward neural networks to simulate the stock price movement of all the fundamental factors.

For the feed-forward neural networks, a sigmoid hyperbolic tangent function is selected as the activation function to generate an even distribution over the input values. A single hidden layer is also chosen for the neural network model since it has been successfully used for financial classification and prediction[25]. Accordingly, the feed-forward neural networks are built with three layers, including the input layer, hidden layer, and output layer. In addition, the resilient backpropagation-learning algorithm was employed to train the feed-forward neural networks since this optimization method is generally much faster than the standard steepest descent algorithm and requires only a modest increase in memory requirements.

The connection weights and biases are initially randomized and then set during the backpropagation training process. The appropriate learning rate for maximizing performance is also determined during neural network training. We also test the number of hiden nodes from at least 3 to the number which generates obvious out-fitting. And set the training epochs as 200, train accuracy as 0.01, goal target as 0, learning rate as 0.5, and momentum rate as 0.1. Finally, we use the best network to forecast the second group data.

To test the rough set theory's ability to optimize the neural network approach, direct input of original data matrix to neural network model is also testified in this experiment. With the same method, two consequent training phrase are proceeded, and the training results are listed below:

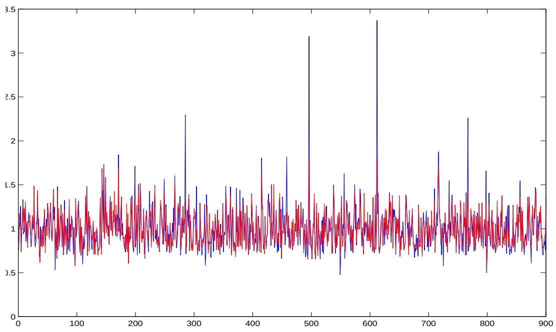

Fig. 2. The best train result of full data

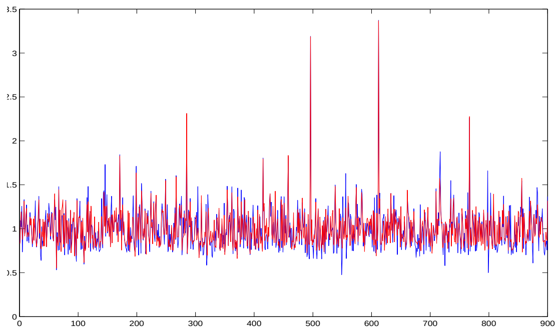

Fig. 3. The best train result of reduced data

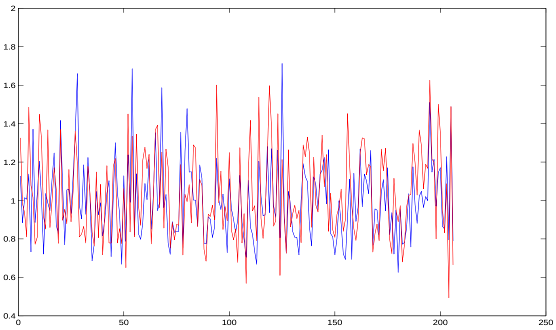

Fig. 4. The best test result of full data

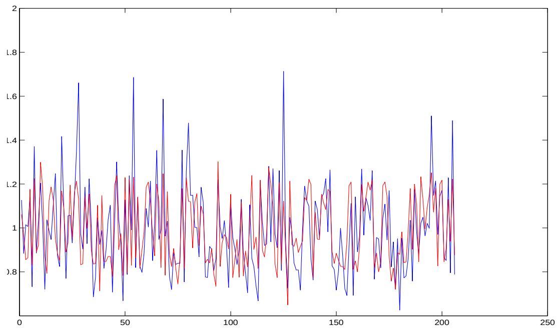

Fig. 5. The best test result of reduced data

Table 2. Modelling results (represented by mean square error)

Hiden Nodes	Original Data		Hiden Nodes	Reduced Data	
	Training	Prediction		Training	Prediction
3	0.0255137	0.0362	3	0.0169575	1.6446
4	0.0515445	0.0468	4	0.0153309	0.1729
5	0.0101872	0.2551	5	0.0186569	0.0302
6	0.0135799	0.0498	6	0.0116944	0.1526
7	0.024835	0.0366	7	0.0239849	0.0268
8	0.028866	0.0399	8	0.017852	0.0299
*9	0.0286887	0.0343	*9	0.0123403	0.0211
10	0.0515209	0.0473	10	0.0109238	0.0660
11	0.00802364	0.0829	-	-	-
12	0.0461758	0.0414	-	-	-
13	0.00322362	0.1713	-	-	-

$*$: The Best Prediction results of modelling.

4 Analysis

From the rough set theory selection, we get all the discernable factors as the following order:

$$\text{SMI} \rightarrow \text{RRS} \rightarrow \text{NAPS} \rightarrow \text{CR} \rightarrow \text{NOS} \rightarrow \text{MBG} \rightarrow \text{IT} \rightarrow \text{NAIR} \rightarrow \text{NCFS}.$$

However, there are two groups of equal important factors: CR& QR, and NCFS & DAR & RRA & ART. Since rough set theory failed to discern such two pairs of factors, we can randomly select one of them without deteriorating the descriptive ability of original data. The first group of equal important factors is identical to the result of theoretical analysis, since CR and QR are two common accounting factors used to describe a firm's debt-paying ability in a short time. The second group of equal important factors infers to four aspects, which are Current, Risk, Profitability and Operation aspects respectively. According to rough set theory, the equal important factors have the same ability of discerning the relation of inputs and outputs. Hence, in our model, we select CR and NCFS as the corresponding two factors at last.

From table 2, one can notice that no matter which set of data is used, the best number of hidden nodes is 9, while the original inputs number varies from 9 to 17. It is reasonable to make the hypothesis that the reduced number of attributes in a reduct have something to do with the number of hiden nodes. Moreover, one can observe that the training results of neural networks using full data matrix are better than those using reduced data matrix. However, it is not true for the prediction results. It can be conclude that using full data matrix, neural networks are easy to get local optimum since there are much more variables in such circumstance. On the other side, the neural networks, using reduced data matrix can get better performance in the prediction phase, which means rough set theory optimize the result of neural network approach indeed.

The neural network is able to detail not only the direction of stock price change, but also the exact value of changing price. The MSE of the best training model is about 0.011282 for the neural network approach incorporating Rough Set theory. Form figure 3, we can easily notice that the stock price forecast has some error comparing to the target value. The reason may be that there are some factors which are not included in the model's input. However, as the total error is within our tolerance, the model can be considered as significant.

To test the application ability of this model, we use the second group of samples which is untouched in previous proceed of modelling, to simulate Out-of-Sampling Valediction and Prediction. Figure 5 details the result of prediction using the best training model. Again, we can get close enough output to the target, which means our model is useful to forecast future stock price with trustable accuracy.

Also, one can notice that there are much more upward abbreviations than downward abbreviations in the prediction results. It can be explained resulting from the short sell constrain in China's stock market. Under such circumstance, investors can only gain profit from the rising of stock price, hence, it is nature that they are unwilling to sell their holding stock when they face to a large number of lose in stock price which cause the stock price stop from falling. It is also concluded that phycological factors and other technological factors significantly affect on stock price, which are valuable to forecast stock price.

5 Conclusions

Similar to other research using data from other stock market, this study shows fundamental analysis can be powerful to predict stock price, especially by the aid of neural networks approach and rough set theory. In this study, we are comparatively successful in designing a model to forecast future stock price movement caused by the fundamental factors. This result owe to the cooperation of rough set theory and neural network approach. Using the proposed model, stock price can be predicted significantly. However, this model depends highly on the environment of the stock market. The Stock Market Index which stands for the macroeconomic environment can affect the price movement of a stock seriously. Fortunately, we use only [-1,0,1] the 3 nominal numbers to indicate the environment, which make the model robust to tiny measureless changes of the environment.

This study also stir up some issues to be solved in future work. Since our experiments use only Market index to presents the environment of finance, more research should be done on how to comprise much more specifical factors of the external environment, such as supply-demand relation, investors' psychological factor, and so on. In addition, this study mainly be done on the base of normally operated firms' stock price, there are lots of abnormally operated firms to be studied in the future. Moreover, the profit distribution policy of stock should be include to explain its stock price movement.

Finally, the technical analysis should also be paid much more attention to forecast stock price in the near future. This study covers only the analysis of fundamental available information, while the technical analysis technique remains nearly intact. Technical analysis nonetheless has been shown to provide another aspect for stock price and stock return forecasting. If both technical and fundamental approaches are thoroughly examined and included during the variable relevance analysis modeling, it would no doubt be a major improvement in predicting stock price.

References

[1] R. J. Balvers, T. F. Cosimano, B. McDonald. Predicting stock returns in an efficient market. *Journal of Finance*, 1990, **55**: 1109–1128.

[2] W. Breen, L. R. Glosten, R. Jagannathan. Predictable variations in stock index returns. *Journal of Finance*, 1990, **44**: 1177–1189.

[3] A. N. Burgess, A. N. Refenes. Modelling non-linear co-integration in international equity index futures. **in:** *Proceedings of the Third International Conference on Neural Networks in the Capital Markets, Singapore*, World Scientific, London, 1995.

[4] V. S. Desai, R. Bharati. A comparison of linear regression and neural network methods for predicting excess returns on large stocks. *Annals of Operations Research*, 1998, **78**.

[5] V. Dropsy. Do macroeconomic factors help in predicting international equity risk premia? testing the out-of-sample accuracy of linear and nonlinear forecasts. *Journal of Applied Business Research*, 1996, **12**: 120–133.

[6] S. S. E. I. Group. Shenzhen sub-index: Witnessing china's stock market for 10 years. http://www.cninfo.com.cn /finalpage /2005-01-24 /14964626.PDF, 2005.

[7] R. W. Holthausen, D. F. Larcker. The prediction of stock returns using financial statement information. *Journal of Accounting and Economics*, 1992, **15**: 373–411.

[8] M. Jensen. Some anomalous evidence regarding market efficiency. *Journal of Financial Economics*, 1978, **6**: 95–101.

[9] D. Keim, R. Stambaugh. Predicting returns in the stock and bond markets. *Journal of Financial Economics*, 1986, **17**: 357–390.

[10] I. Kim, K. J. and. Han. Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index. *Expert Systems with Applications*, 2000, **19**: 125–132.

[11] K. Kim, I. Han. The extraction of trading rules from stock market data using rough sets. *Expert Systems*, 2001, **18**(4): 194–202.

[12] J. Komorowski, Z. Pawlak, L. Polkowski, A. Skowron. Rough sets: A tutorial. **in:** *Rough Fuzzy Hybridization: A New Trend in Decision-making* (S. Pal, A. Skowron, eds.), Springer, Singapore, 1999, 3–98.

[13] L. Liu, J. Yu. Trading volume of annual report and inverstor tracing: An empirical research. *Journal of Financial Research*, 2002, **266**(8): 39–47.

[14] A. W. Lo, A. C. MacKinlay. Stock market prices do not follow random walks: evidence from a simple specification test. *Review of Financial Studies*, 1988, **1**: 41–66.

[15] L. Motiwalla, M. Wahab. Predictable variation and profitable trading of us equities: A trading simulation using neural networks. *Computer & Operations Research*, 2000, **27**: 1111–1129.

[16] J. J. Murphy. *Technical Analysis of the Financial Markets: A Comprehensive Guide to Trading Methods and Applications*. New York Institute of Finance, New York, USA, 1999.

[17] J. A. Ou, S. H. Penman. Financial statement analysis and the prediction of stock returns. *Journal of Accounting and Economics*, 1989, 295–329.

[18] Z. Pawlak. *Rough Sets. Theoretical Aspects of Reasoning about Data*. Kluwer, Dordrecht, 1991.

[19] Z. Pawlak. Rough sets. **in:** *Rough Sets and Data Mining* (T. Y. Lin, N. Cercone, eds.), Kluwer Academic Publisher, Dordrecht, 1997, 3–8.

[20] T. Poddig, H. Rehkugler. A world of integrated financial markets using artificial neural networks. *Neurocomputting*, 1996, **10**: 251–273.

[21] M. Qi. Nonlinear predictability of stock returns using financial and economic variables. *Journal of Business & Economic Statistics*, 1999, **17**: 419–429.

[22] M. Qi, G. S. Maddala. Economic factors and the stock market: A new perspective. *Journal of Forecasting*, 1999, **18**: 151–166.

[23] J. D. Schwager. *Fundamental Analysis*. John Wiley & Sons, New York, USA, 1995.

[24] W. Schwert. Stock returns and real activity: a century of evidence. *Journal of Finance*, 1990, **45**: 1237–1257.

[25] G. S. Swales, Y. Yoon. Applying artificial neural networks to investment analysis. *Financial Analysts Journal*, 1992, **48**: 78–80.

[26] S. Thawornwong, D. Enke. The use ofdata mining, neural network models, and validation techniques for predicting excess stock returns. **in:** *Proceedings of the International ICSC Congress on Computational Intelligence: Methods and Applications*, Bangor, Wales, UK, 2001, 329–335.

[27] S. Thawornwong, D. Enke. The adaptive selection of financial and economic variables for use with artificial neural networks. *Neurocomputing*, 2004, **56**: 205–232.