Stock Market Prediction Based on Fundamentalist Analysis with Fuzzy-Neural Networks

RENATO DE C. T. RAPOSO[1] AND ADRIANO J. DE O. CRUZ[2]
Núcleo de Computação Eletrônica, Instituto de Matemática,
Federal University of Rio de Janeiro
PO BOX 2324, Rio de Janeiro, RJ, CEP 20001-970
BRAZIL

1. raposo@attglobal.net
2. adriano@nce.ufrj.br

Abstract: - In this article, we discuss the application of a combination of Neural Networks and Fuzzy Logic techniques to fundamentalist analysis of stock investment. Several researchers have used neural networks models in a variety of ways to predict short and long-term stock forecasting, but most of these models use technical indicators as inputs. This paper presents the results of the application of Fuzzy Neural Networks to predict the evolution of stock prices of Brazilian companies traded on the São Paulo Stock Exchange. The network indicates if a trader would have to keep, sell or buy a stock using a combination of information extracted from balance sheets (released every three months) and market indicators. In our experiments, we concentrated on a segment of the Brazilian industry, companies from the textile sector. The results show that the network was able to deliver good results depending on the quality of the available data.

Key Words: - Fuzzy Neural Networks, Stock Market Prediction, Fundamentalist Analysis

1 Introduction

Artificial Neural Networks (ANN) are loosely based on the operation of the brain and the functioning of biological neurons. ANNs are composed of many parallel computing devices, richly connected one to another, the function it computes is determined by the pattern of these connections. They have been widely studied and applied in financial analysis [1,2,3,4]. Fuzzy Logic is derived from Fuzzy Set Theory and provides a mechanism for decision systems to deal with imprecise information [5]. Fuzzy Logic is an attempt to allow computers overcome the problems of Boolean logic. Humans reason by manipulating imprecise concepts, categories and use vague terms to define these ideas. Humans use terms like "quite expensive", "extremely risky" and "growing very fast" when discussing business options.

Fuzzy Neural Networks are a connecting link between fuzzy logic and neural computing. The goal is to combine the advantages of each approach in order to process vague information and deal with human based rules. In this research, we applied a model of Fuzzy Neural Network introduced in [6].

Several researchers have used neural networks models in a variety of ways to predict short and long-term stock forecasting and most of these models use as input technical indicators, which are obtained from the stock market behavior. This paper discusses the application of Fuzzy Neural Networks to assess the evolution of stock prices of Brazilian companies traded on the São Paulo Stock Exchange, on a three month basis using a combination of information extracted mainly from balance sheets and some and market indicators. In our experiments, we concentrated on a segment of the Brazilian industry, the textile companies. We plan to test the architecture with other sectors as well. The network will produce three outputs: keep, sell or buy a stock. This information will serve as an indicator for the next three months, until the next balance is released and the network is retrained, including the new data.

2 Input Data and Sector Choice

In our experiments, we utilized a database containing information about all Brazilian companies traded on the São Paulo Stock Market. The database, we used, includes balance sheets from the first trimester of 1986 to the second trimester of 1998 making 51 trimesters.

In this work we choose companies from only one sector. The idea was to find groups of companies that have similar characteristics, therefore resulting less economic indicators to consider. Another important characteristic taken into account was the production cycle. We decided to select a short-term production sector. The cycle of manufacturing/selling on short-term sectors is around six months. Considering the number of semesters available at our database, a short-term sector has more cycles of high-low prices. Some of the short-term sectors are: Textile, Food, Car Parts, Drink-Tobacco, Electric-Electronic and Commerce. In this work we have not trained the fuzzy neural network in all these sectors, nor tested companies from other sectors. However, we tested the resulting fuzzy neural network against companies not included in the training set.

After searching among all the short cycle sectors available on the database, we chose one that has more companies, more data and data that are more consistent. Following these criteria, textile sector was selected. This sector has 28 companies listed on our database. Table 1 shows the sectors analyzed in order of importance.

Sector	*Nº of Companies*	*Data since 1986*	*Performance on the stock market (%)[1]*	*Appreciation rank among all 14 sectors*
Textile	28	Almost Complete	63,99	2
Food	21	Almost Complete	18,01	7
Electric-Electronic	8	Almost Complete	27,07	5
Commerce	13	Incomp.	33,00	4
Car-Parts	13	Incomp.	15,66	8
Drink-tobacco	2	Incomp.	15,45	9

(1)Sirotsky & Associados (http://www.sirotsky.com.br/news.html)

Table 1. Rank of Sectors

3 The Choice of Economic Indicators

After choosing the sector we selected which economic indicators would be used as input to the fuzzy-neural network. The database has 52 economic indicators generated automatically from the balance sheets of the companies traded on the São Paulo stock market. The database also has indicators such as Brazilian inflation rate, reference interest rate, exchange rate, etc. So we decided to reduce this number in order to speed up the training process. Two approaches were used to select the economic indicators. One would be based on the know-how of market analysts and references from the Brazilian economic literature. The idea behind this choice was to create an automatic system based on know-how from specialists. The other selection would result from a statistical analysis of the database, therefore creating an automatic system from scratch.

We gathered information from seven different sources in order to find which were considered the most significant. The sources are listed below and the number shows the total of indicators selected by the source.

A. First senior economic analyst (29);
B. Second senior economic analyst (17);
C. Selection used by Economática (http://www.economatica.com.br), a Brazilian company that commercializes data and software for analysis of the 1,000 major companies in Latin America and the 200 most representative US companies (21);
D. Book "Análise Financeira de Balanços", D. C. Matarazzo. Editora. Atlas (pg 158), “Finance Analysis of Balance Sheets” (11);
E. Selection used by the Brazilian magazine EXAME to rank the 500 biggest Brazilian companies of the year. EXAME is an important source of economic information in Brazil (10);
F. Book "Princípios de Administração Financeira", L.J. Gitman, Editora Harbra, pp 128/129 (10);
G. Selection used by "Lopes Filho & Associados, Consultores de Investimentos”, (http://www.lopesfilho.com.br) (16);

Fundamentalist Indicators
Price/Profit
Firm Value/EBITDA
Liquidity Indicators
Current Liquidity Ratio
Performance Ratios
Asset Turnover
Operating Margin
Net Margin
Return on Assets
Return on Equity Basis
Return on Equity/Net Profit

Table 2. Indicators selected by specialists

The result of this research reduced the set of 52 available indicators on the database to nine, that were either used by almost all sources or were considered very important by the senior economic analysts interviewed. Table 2 lists these indicators and the class they belong to.

The second approach to select the economic indicators used Principal Component Analysis (PCA) that reduced to ten the number of indicators. This method transforms a set of correlated variables to a new set of uncorrelated variables. The PCA [6] finds components that are close to the original variables but arranged in decreasing order of variance. In order to illustrate the method, consider $X^T = [X_1, X_2, \ldots, X_p]$ a p-dimensional random variable with mean μ a covariance matrix Σ. PCA transforms X^T in $Y^T = [Y_1, Y_2, \ldots, Y_p]$ where each Y_j is a linear combination of the X's, so that

$$Y_j = a_{1j}X_1 + a_{2j}X_2 + \ldots + a_{pj}X_p$$

After applying PCA to the data, we identified ten components. These ten components retained almost 65% of the total variance of the sample as is indicated in the Figure 1. The PCA was applied to data from twenty companies from the textile sector, which has twenty-eight companies. These twenty were the companies that have more consistent data.

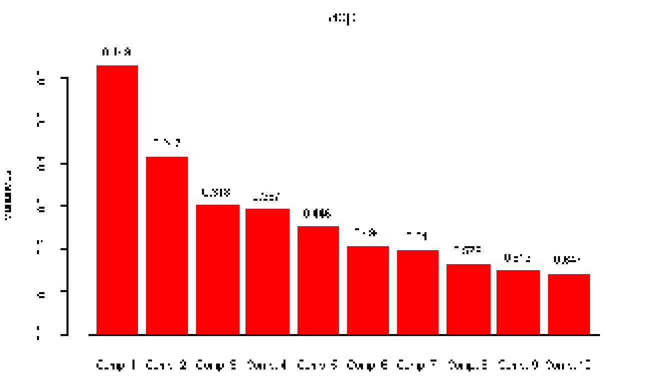

Figure 1. PCA results

4 Fuzzy-Neural Model

The fuzzy-neural model used is a feed-forward architecture with five layers of neurons as indicated in the Introduction section. It maps a fuzzy system to a neural network that will simulate the inference process executed in the fuzzy system. The first layer of the fuzzy neural system receives input values and feeds them to the second level, so the architecture defined has to have either nine (specialists selection) or ten inputs (PCA). The second layer determines the degree of membership of each variable to the fuzzy sets to which it belongs. The third layer represents the fuzzy rules that will combine the input variables using rules of the type **if-then**. In the next layer, each node will represent one fuzzy set from the consequent elements of the rules, the output variables. The last layer executes the process of defuzzification, yielding an exact value for each output variable. Several tests were realized to find the best architecture and the resulting numbers of neurons in each layer are 9 or 10, depending on the number of inputs, 3, 20, 3 and 3. The three outputs are: keep, buy and sell.

5 Training and Testing

In order to train the network, we divided the data into two sets, one training set and one test (recall) set. It is important to notice that the problem has two characteristics. First it behaves like a temporal series, since we have data, collected every trimester since 1986, from balance sheets of 28 companies. Second we want to classify companies into three groups: keep, buy and sell companies. Taking into account these characteristics, at the beginning of the training process, we separated the data into 35 trimesters to training and the last 8 to testing. Some trimesters were left out because not all data were available.

In order to check if the network was producing correct answers the target was defined according to following criteria. Whenever the stock price increased more than 5% from one trimester to another, the network had to indicate a buy option at the beginning of this trimester. If the price went down the same percentage then the answer had to be sell. A stable price indicated that the stock had to be kept.

The goal of the research is to produce a network that will give an indication of the best option until the next balance is released in three months time. After this period the network is retrained including the new released data. Retraining of the network is not a problem due to the tree month interval in between samples. Another test made was to check if the network was able to give meaningful answers two trimesters in advance.

We used a windowing like system. At first a window of 35 trimesters with a step of one trimester was moved through the data. So the network was trained using the first 35 trimesters, after that an evaluation of the results and the window moved to the right one trimester after a retraining. This process went through the next four trimesters.

One interesting outcome of the training process was the possibility of using a window with fewer trimesters. A window of 14 trimesters gave the same performance, with the additional advantage of reducing the training time. This reduction shows that economic information older than three and half years was not relevant to our solution. This may be due to the very unstable situation of the Brazilian economy in the last two decades. It would be interesting to test on data from other countries.

The data set was composed of data from 20 textile companies spawning a period 43 trimesters. Some companies were left out on purpose so that a test on companies not trained could be performed later. The percentage of each target in the data set is shown in Table 3. Each line shows the number of targets at each trimester window. It should be noted that the figures for the three targets are not evenly distributed and this will show up at the network performance as it will be showed later. Remember that we are using real data extracted from balance sheets of companies listed at São Paulo stock market and they show the state of the economy during the period considered. First, it is important to note that buy targets are more frequent, showing that companies were growing most of the time. Another characteristic is the low frequency of keep targets showing few periods of stability.

Window last trimester	Buy	Buy %	Sell	Sell %	Keep	Keep %	Total
35º (06/96)	160	59,3	79	29,3	31	11,5	270
36º (09/96)	148	54,2	91	33,3	34	12,5	273
37º (12/96)	141	51,5	97	35,4	36	13,1	274
38º (03/97)	134	48,7	103	37,5	38	13,8	275
39º (06/97)	120	43,5	113	40,9	43	15,6	276

Table 3. Percentage of targets at each training stage.

Tables 4 summarizes the prediction capabilities of the network obtained from a selection of indications from market specialists and references. This network will be labeled as REF network. Table 5 shows the results obtained by the network generated from the Principal Component Analysis and this will be referenced as PCA network. The last column (Not Class) is the percentage of not classified inputs. These numbers were obtained after testing the 20 companies during 5 trimesters. Note that the PCA network obtained the best results. This may result from the method that we used to aggregate the choices of the specialists. Another contribution to these better results may be due to the fact that the PCA method takes into account all economic indicators when creating the linear combination.. Tables 4 and 5 shows that the results for the buy and sell targets were reasonable. As for the keep target the networks did not perform well due to the low percentage of samples of this kind at the beginning of the considered period. Partial results showed that when only the last two time windows were considered, this percentage is higher and the performance improves. Table 6 and 7 shows the results for the tests of 39th and 40th trimesters respectively. Another conclusion is that the network has not enough information to predict two semesters ahead.

Prediction %	**Total**	**Buy.**	**Sell**	**Keep**	**Not Class**
Last trained trimester	100	100	100	100	0
One trimester ahead	70.1	75.0	75.5	41.6	10.2
Two trimesters ahead	51.1	74.1	45.2	26.6	10.2

Table 4. Average prediction capabilities of the REF network.

Prediction (%)	**Total**	**Buy**	**Sell**	**Keep**	**Not Class**
Last trained trimester	100	100	100	100	0
One trimester ahead	75.0	77.4	77.5	58.3	6.1
Two trimesters ahead	53.6	72.7	59.5	31.2	2.0

Table 5. Average prediction capabilities of the PCA network.

Prediction (%)	**Total**	**Buy**	**Sell**	**Keep**	**Not Class**
39th	72.2	80.0	77.8	50.0	10.0
40th	84.2	100.0	81.3	100.0	5.0

Table 6. REF network results for the 39th and 40th trimesters.

Prediction (%)	Total	Buy	Sell	Keep	Not Class
39th	75.0	75.0	80.0	66.7	15.0
40th	73.7	100.0	68.8	100.0	5.0

Table 7. PCA network results for the 39th and 40th trimesters.

Tables 8 and 9 shows the results of the tests applied to 2 companies not in the set of trained companies. The results were similar to obtained from trained companies indicating that both networks have good generalization capacity.

Prediction (%)	Total	Buy	Sell	Keep	Not Class
One trimester ahead	70.0	80.0	66.6	50.0	0

Table 8. REF network results for 2 not trained companies.

Prediction (%)	Total	Buy	Sell	Keep	Not Class
One trimester ahead	77.8	66.6	100.0	50.0	10.0

Table 9. PCA network results for 2 not trained companies.

6 Conclusions

This study gave rise to important points regarding the suitability of applying fuzzy neural networks for long-term prediction about stock prices using indicators from fundamentalist analysis. First, it indicated that it is possible to expect good performance, especially after the observation that their results were considerably improved by the incorporation of data that have the numbers of targets evenly distributed. The second one is that, it is rather implausible to expect that the networks could provide acceptable predictions of two semesters in advance. Nevertheless, this point is under consideration and further investigations are under way.

Another point is that the experiments showed that the network performance was not reduced when the window of time was reduced drastically, indicating that old information lost significance as the economic factors changes.

The network presented similar results to trained and not trained companies showing capacity of generalization.

References

[1] R. R. Trippi, E. Turban. *Neural Networks in Finance and Investing*, Probus Publishing Company, Chicago, Illinois, USA, 1993.

[2] E. M. Azoff. *Neural Network Time Series Forecasting of Financial Markets*, John Wiley & Sons Chichester, England, 1994

[3] G.M. Papadourakis, G. Spanoudakis and A. Gotsias. Application of Neural Networks in Short-Term Stock Price Forecasting, *Proceedings of 1st International Workshop Neural Networks in the Capital Markets*, London, November 1993.

[4] I. Matsuda. Application of Neural Sequential Association to Long-Term Stock Price Prediction. *Proceedings of IJCNN'91*, Singapore, 1991

[5] L. Zadeh. Fuzzy Sets, *Inf. Control*, vol 8, pp 338-353

[6] N. Kasabov. *Foundations of Neural Networks, Fuzzy Systems and Knowledge Engineering*, MIT Press, Cambridge, MA, USA, 1996

[7] W. R. Dillon, M. Goldstein. *Multivariate Analysis: Methods and Applications*, John Wiley & Sons, New York, USA,