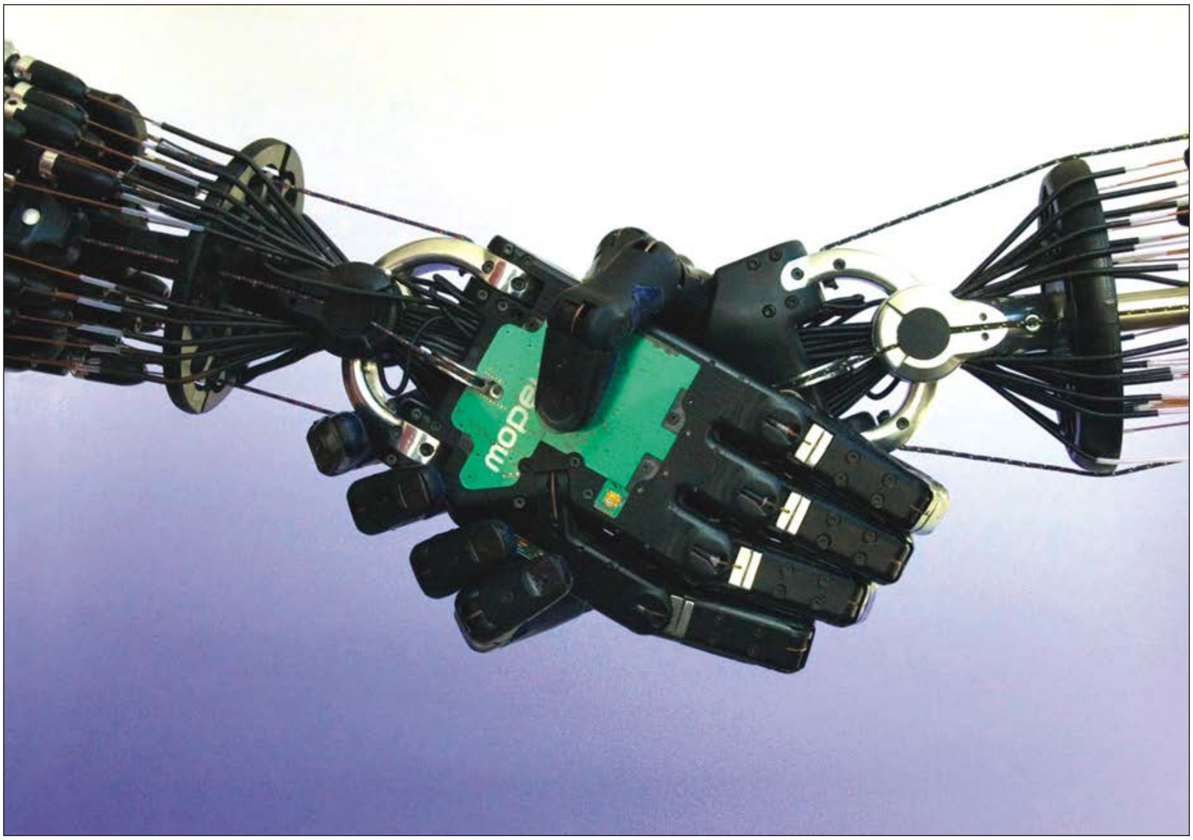

Journal of Artificial Intelligence

ISSN 1994-5450

science alert

ANSInet
an open access publisher
http://ansinet.com

Journal of Artificial Intelligence 1 (2): 70-77, 2008
ISSN 1994-5450

Stock Price Direction Prediction Using Artificial Neural Network Approach: The Case of Turkey

D. Şenol and M. Özturan
Department of Management Information Systems, Boğaziçi University, Turkey

Abstract: In this study, it is aimed to illustrate that Artificial Neural Network (ANN) can be used for predicting the stock price behaviour in terms of its direction. Financial daily statistical data, derived from raw price data obtained from Istanbul Stock Exchange (ISE), which is the only stock market in Turkey, have been defined in terms of five independent variables that are grouped in seven different Prediction System (PS) models to which eight different ANN and Logistic Regression (LR) models have been applied. For this purpose, a software library package is developed using C#.NET to run the ANN models whereas a commercial statistical analysis software package is used to run the LR model. At the end of the study; the best PS and ANN models are determined for ANN methodology by comparing the average mean squared errors of training sets and the best PS model is determined for LR methodology by eliminating the insignificant independent variables; the outputs of the developed software library package and a commercial ANN software are compared on the basis of prediction success rate and the accuracies of prediction by ANN and LR methodologies are compared on the basis of coefficient of determination. The results show that; the best results are obtained for the PS model that has used stochastic indicator for 14 days (K14%), stochastic moving average (D3%) and relative strength index of 14 days (RSI14) simultaneously for both ANN and LR methodologies whereas the best ANN model has consisted of three inputs, 11 hidden neurons in single hidden layer and one output; developed software library package performs statistically same as the commercial software; statistically ANN methodology outperforms LR methodology; and there is relevant empirical evidence that ISE-30 is not weak form efficient.

Key words: Neural network approach, stock market, prediction, logistic regression, Turkey

INTRODUCTION

Stock price behavior has been a widely questioned and not a mutually agreed area of researchers, where the main question is whether stock price behaviors are predictable or not.

Researchers, who believe that stock prices do not follow a trend, act in a random walk and cannot be predicted, are usually followers of a hypothesis called the Efficient Market Hypothesis (EMH). EMH has been a widely accepted theory which claims that the prices are defined in a random walk procedure, making price behavior completely unpredictable. It also suggests that it is not possible for any kind of prediction algorithm to outperform a buy and hold strategy (a long term trading strategy based on the concept that in the long run financial markets give a good rate of return) consistently for a long period of time. This hypothesis has been discussed, expanded and deepened by Reilly and Brown (1997), Dutt and Ghosh (1999) and Dietrich *et al*. (2001).

As oppose to EMH, various studies have been done using different methodologies and different indicators to predict stock price behaviour. According to Hellstrom and Holmstrom (1997), there are

Corresponding Author: D. Şenol, Department of Management Information Systems, Boğaziçi University, Turkey

four main methodologies to predict stock market; fundamental analysis, technical analysis, time series forecasting and machine learning. As indicators of stock price behaviour, different combinations of various indicators such as; closing stock price, stock market index value, foreign exchange rate, interest rate value, vector curve, turnover, moving average, momentum, relative strength index, stochastic and moving average of stochastic have been used in previous researches (Kimoto *et al*., 1990; Tsibouris and Zeidenberg, 1995; Yao and Poh, 1995; Fernandez-Rodriguez *et al*., 2000; Egeli *et al*., 2003).

In recent studies, Artificial Neural Network (ANN), which is the most popular machine learning methodology, with various sets of indicators as inputs and with various topologies, has been utilized for stock price behaviour prediction and contradictory to EMH, has shown that stock price behaviour can be predicted and ANN approach can outperform conventional methods (Van Eyden, 1996; Yao and Poh, 1995; Fernandez-Rodriguez *et al*., 2000; Phua *et al*., 2000; Egeli *et al*., 2003; Versace *et al*., 2004; Yümlü *et al*., 2004).

The main objective of this study is to show that, with a well chosen set of indicators and ANN topology, ANN method has the capability to predict stock price direction and in this context, outperforms the conventional technique, Logistic Regression (LR).

MATERIALS AND METHODS

Stock price direction, as stated before, is mostly predicted by financial indicators and the act of selecting the true indicators, in other words designing a correct Prediction System (PS) model, is not easy and varies from market to market and even stock to stock. Based on the previous studies discussed before and the opinions of the experts, the following financial indicators are chosen to be the indicators of the PS models in this study:

- Moving average of 14 days (MA14)
- Moving average of 37 days (MA37)
- Stochastic indicator for 14 days (%K14)
- Stochastic moving average (%D3)
- Relative strength index of 14 days (RSI14)

Considering these indicators, seven different PS models (PSM1 to PSM7) consisting of different sets of these indicators have been considered for the prediction of stock price direction:

PSM1: $f_O = f_{Model1}(RSI_{14})$
PSM2: $f_O = f_{Model2}(K_{14}, D_3)$
PSM3: $f_O = f_{Model3}(MA_{14}, MA_{37})$
PSM4: $f_O = f_{Model4}(RSI_{14}, K_{14}, D_3, MA_{14}, MA_{37})$
PSM5: $f_O = f_{Model5}(RSI_{14}, K_{14}, D_3)$
PSM6: $f_O = f_{Model6}(RSI_{14}, MA_{14}, MA_{37})$
PSM7: $f_O = f_{Model7}(K_{14}, D_3, MA_{14}, MA_{37})$

Thus, the effectiveness of different combinations of financial data has been investigated for the stock price direction prediction.

Istanbul Stock Exchange (ISE-30) have been chosen for the data set of this study. Daily closing prices of each stock in ISE-30 for each day have been acquired from a private data feeder company and these prices are then used to calculate the indicators of the PS models. Statistical summary of this data is given in Table 1.

Table 1: Statistical summary of the data set

Stock code	Beginning of the data set (dd/mm/yyyy)	No. of days	Min (TRY)	Max (TRY)	SD (TRY)
AKBNK	05/01/1998	2525	0.11	9.80	2.44
ARCLK	05/01/1998	2525	0.23	11.97	3.25
DENIZ	01/10/2004	810	2.35	17.10	4.51
DOAS	17/06/2004	735	2.84	10.99	1.82
DOHOL	05/01/1998	2525	0.05	3.53	0.89
DYHOL	06/08/1998	2521	0.13	7.50	1.82
EREGL	05/01/1998	2255	0.10	10.25	2.17
FINBN	05/01/1998	2524	0.02	5.80	1.88
FORTS	05/01/1998	2524	0.02	3.74	0.92
GARAN	05/01/1998	2525	0.07	9.55	1.98
GSDHO	11/11/1999	1949	0.25	3.83	0.70
HURGZ	05/01/1998	2525	0.04	5.66	1.47
ISCTR	05/01/1998	1929	0.17	9.16	2.26
ISGYO	09/12/1999	2525	0.37	2.76	0.65
KCHOL	05/01/1998	2519	0.27	6.40	1.46
MIGRS	05/01/1998	2525	0.58	22.10	4.83
PETKM	05/01/1998	2505	0.71	20.33	2.66
PTOFS	05/01/1998	2527	0.24	7.40	1.57
SAHOL	05/01/1998	2525	0.13	7.95	1.84
SISE	05/01/1998	2515	0.17	6.30	1.74
SKBNK	05/01/1998	2518	0.09	4.62	0.92
TCELL	11/07/2000	2525	0.73	9.65	2.29
THYAO	05/01/1998	1782	1.95	18.25	2.58
TOASO	05/01/1998	2510	0.11	6.85	1.54
TSKB	05/01/1998	2506	0.04	3.13	0.82
TUPRS	05/01/1998	2525	0.79	34.25	8.16
ULKER	05/01/1998	2521	0.06	6.96	2.02
VAKBN	18/11/2005	451	2.45	4.34	0.35
VESTL	05/01/1998	2525	0.25	7.00	1.64
YKBNK	05/01/1998	2517	0.11	4.14	0.90

Average number of days for available data of the stocks is 2255. Since number of available trading dates for stocks listed under the name of DENIZ, DOAS and VAKBN are less than 50% of the average number of days; they are not included in this study due to insufficient amount of data, thus, the number of stocks used in this study is dropped to 27.

The period used in the training data sets are between January 5, 1998 (first trading date of 1998) and December 29, 2005 (last trading date of 2005). The period used in the testing data sets are between January 6, 2006 (first trading date of 2006) and August 31, 2007 (last trading date of available data).

As suggested and used in previous studies (Kimoto *et al*., 1990; Freisleben, 1992; Azoff, 1994; Zekic, 1998; Gencay, 1998; Quah and Srinivasan, 1999; Fernandez-Rodriguez *et al*., 2000; Man-Chung *et al*., 2000; Egeli *et al*., 2003; Heaton, 2005), backpropagation ANN model with one hidden layer with eight possible different numbers of neurons for the hidden layer, thus, eight different ANN models have been prepared for seven different PS models. The number of inputs of the ANN models is set to be the number of indicators of the corresponding PS model and the stock price direction, within the boundary values 0 to 1, is set to be the output that follows the below rule:

- Goes down if output is greater than or equal to 0.0 and less than 0.5
- Stays same if output is equal to 0.5
- Goes up if output is greater than 0.5 and equal or less than 0.0

For all of the ANN models, the following network parameters are taken the same:

- Learning rule: Momentum (Momentum factor = 0.5)
- Stopping criteria: 10,000 cycles
- Learning rate: 0.2

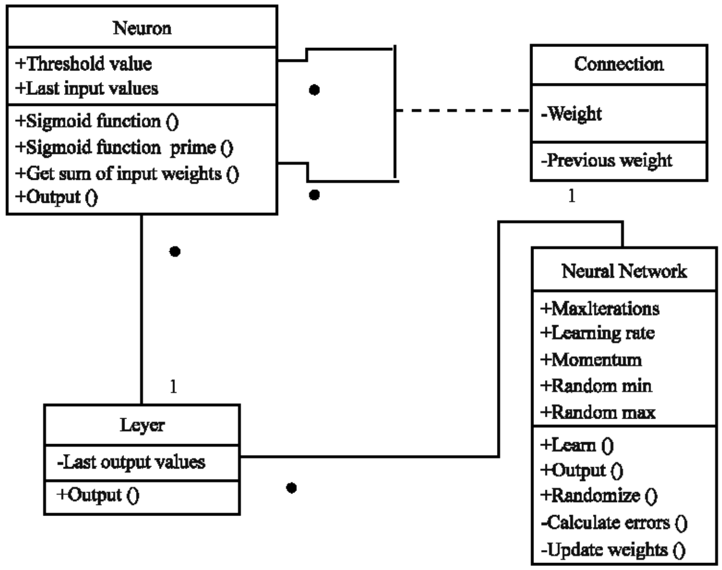

Fig. 1: Class diagram of the developed software library package

- Activation function: Linear Sigmoid
- Initial weight: Randomized

For applying the ANN models to PS models a software library package is developed by object oriented methodology using C#.NET that can easily be integrated to other systems, such as trading applications. The class diagram of the developed software library package is given in Fig. 1.

Using the developed software package library, eight different ANN models are applied to each of the seven PS models for each stock included in ISE-30. Due to the rules that do not yield a possible combination of number of inputs of the PS model with the number of neurons in the hidden layer of the corresponding ANN model, 26 combinations of ANN versus PS models are dropped from the study thus leaving 30 combinations. For each of these 30 combinations, averages of the mean squared errors of training of 27 different stocks are calculated. The ANN and PS models that correspond to the smallest average mean squared error of the trainings are selected to be models of the study and for each of the 27 stocks, the predicting ability of the developed software library package is tested by comparing the predicted outputs of the selected models with actual data.

To check the reliability of the developed software library package, a commercial ANN software is run for the selected ANN and PS models and the outputs are statistically compared with the outputs of the developed software library package.

ANN outputs can also be compared with the results of statistical methods, generally regressive models (White, 1988; Weigend *et al*., 1990; Bernd and Klaus, 1996; Dutta and Shekbar, 1988; Chiang *et al*., 1996). Models which are used in these studies are targeted on forecasting a future stock or index value. Since this study focuses on predicting stock price direction, which is represented by a binary number, a regressive model with a binary output is appropriate for comparison of the outcomes. LR methodology is a statistical method used when the dependent variable is desired to be interpreted as binary (Dreiseitl and Ohno-Machado, 2002), therefore it is an efficient way to measure the accuracy and performance of ANN model when the output is going to be classified as binary (Bell *et al*., 1990; Huang *et al*., 1994; Schumacher *et al*., 1996; Luther, 1998; Dreiseitl and Ohno- Machado, 2002). In this study, the outcomes of ANN approach are compared with the outputs of LR method. For that purpose, the five financial indicators chosen before are used as independent variables and the stock price direction is used as the dependent variable in LR methodology. A commercial statistical analysis software package is used for running up the LR method and the best PS model is determined by taking the significant independent variables into consideration whereas correctness and correlation factors are used for the comparison of outputs of the ANN and LR methodologies statistically.

RESULTS AND DISCUSSION

After applying the ANN models to each system model for 27 stocks included in ISE-30 using the developed software library package, ANN model with three inputs, 11 hidden neurons in the single hidden layer and one output (ANNM.3.11.1) applied to the PS model with the three indicators, R14, K14 and D3 (PSM5) gives the lowest average mean squared error of training. Therefore, these models are selected to be the models of this study. Table 2 gives the success rates of the predicted outputs (price goes down-price stays same-price goes up) of the application of ANNM3.11.1 to PSM5 for 27 different stocks in comparison to the actual price direction data. Average of the success rates is 78.47% and for every stock, the success rate is consistently much higher than 50-50 chance indicating a high predicting capability of the models.

The reliability of the developed software library package is checked by applying the same selected models (ANNM3.11.1-PSM5) to 27 stocks using a commercial ANN software. Table 3 gives the correlations between the predicted outputs and the actual price direction data for the results of both the developed software library package and commercial ANN software. One tailed t-test applied to these correlations shows that in the 95% confidence interval, there is statistically no significant difference ($p = 0.48$) between these sets indicating that the developed software library package is reliable as much as the commercial ANN software.

The results of ANN approach are also compared with the outcomes of the LR method to test if ANN approach outperforms LR method. A commercial statistical analysis software is used to run the LR method. The significant PS model suggested by LR method comes out to be the same as the best performing PS model in ANN approach for each stock (PSM5) determining R14, K14 and D3 as

Table 2: Success rates of the application of ANNM3.11.1 to PSM5

Stock	Success rates (%)
AKBNK	78.47
ARCLK	77.03
DOHOL	76.79
DYHOL	76.79
EREGL	78.47
FINBN	78.71
FORTS	78.47
GARAN	78.71
GSDHO	78.71
HURGZ	77.75
ISCTR	78.23
ISGYO	79.19
KCHOL	79.19
MIGRS	78.95
PETKM	78.95
PTOFS	79.43
SAHOL	78.71
SISE	79.19
SKBNK	77.75
TCELL	77.99
THYAO	78.71
TOASO	79.19
TSKB	78.95
TUPRS	78.23
ULKER	78.23
VESTL	78.47
YKBNK	78.47

Table 3: Correlations between the predicted outputs of ANNM3.11.1-PSM5 models and the actual price direction data for the results of both software

Stock	Developed software library package	Commercial ANN software
AKBNK	0.795942372	0.782258381
ARCLK	0.752443223	0.747107897
DOHOL	0.784575611	0.763939481
DYHOL	0.75943713	0.773711962
EREGL	0.75629375	0.72417282
FINBN	0.651118596	0.616810748
FORTS	0.772380506	0.709701169
GARAN	0.802322539	0.770800673
GSDHO	0.721552783	0.707884941
HURGZ	0.792904571	0.769071176
ISCTR	0.752953771	0.771269508
ISGYO	0.752108926	0.758957729
KCHOL	0.779509645	0.76306179
MIGRS	0.758797487	0.738209417
PETKM	0.741909199	0.733373276
PTOFS	0.744319288	0.721479948
SAHOL	0.808670601	0.785064681
SISE	0.786549946	0.774592048
SKBNK	0.703136272	0.673685488
TCELL	0.75823256	0.778350839
THYAO	0.762541713	0.753298816
TOASO	0.723767644	0.728473442
TSKB	0.733671056	0.659202109
TUPRS	0.754812096	0.748625193
ULKER	0.754490611	0.735967861
VESTL	0.768570727	0.769658901
YKBNK	0.779596555	0.750856765

Table 4: Correlations of outputs of ANN and LR methods with actual values for PSM5

Stock	ANN	LR
AKBNK	0.795942	0.675952
ARCLK	0.752443	0.672036
DOHOL	0.784576	0.701115
DYHOL	0.759437	0.69535
EREGL	0.756294	0.616921
FINBN	0.651119	0.356088
FORTS	0.772381	0.6874
GARAN	0.802323	0.667539
GSDHO	0.721553	0.626456
HURGZ	0.792905	0.680636
ISCTR	0.752954	0.683173
ISGYO	0.752109	0.675791
KCHOL	0.77951	0.639041
MIGRS	0.758797	0.624054
PETKM	0.741909	0.637336
PTOFS	0.744319	0.595209
SAHOL	0.808671	0.673203
SISE	0.78655	0.679835
SKBNK	0.703136	0.59477
TCELL	0.758233	0.695597
THYAO	0.762542	0.635524
TOASO	0.723768	0.652795
TSKB	0.733671	0.660993
TUPRS	0.754812	0.624211
ULKER	0.754491	0.600728
VESTL	0.768571	0.69734
YKBNK	0.779597	0.667241

significant independent variables. Comparison of the correlations of outputs of ANN and LR methods with actual values for the same PS model and for each stock are given in Table 4. Two-tailed t-test applied to correlations show that in the 95% confidence interval, ANN approach method has scored significantly ($p = 0.000020$) higher than the LR method in terms of successful outcomes.

CONCLUSION

This study is aimed at finding the best PS and ANN models for the prediction of the stock price direction using five chosen financial indicators and at showing that ANN model outperforms LR model in prediction. For this purpose; a software library package is developed; a total of 810 sets of predictions, which result from the application of 30 combinations of PS and ANN models to 27 stocks, are produced; developed software package is tested against a commercial ANN software and LR method is applied using the chosen financial indicators as independent variables.

Based on the results of this study it can be concluded that:

- The best PS model comes out to be PSM5 that has R14, K14 and D3 as financial indicators
- The best ANN topology comes out to be ANNM3.11.1 that has three inputs, 11 hidden neurons in single hidden layer and one output
- The developed software package performs statistically same as the commercial software
- Results of LR method also determine PSM5 as the best PS model.
- Comparison of ANN and LR methodologies has shown that ANN methodology statistically outperforms LR methodology
- The results of the study have shown that there is sufficient empirical evidence that ISE-30 is not weak form efficient

ACKNOWLEDGMENTS

Authors would like to thank Dr. Ali Tükel for his valuable suggestions and recommendations.

REFERENCES

Azoff, E.M., 1994. Neural Network Time Series Forecasting of Financial Markets. 1st Edn., John Wiley and Sons, Inc., New York, ISBN: 0471943563 .

Bell, T.B., G.S. Ribar and J. Verchio, 1990. Neural nets versus logistic regression: A comparison of each model's ability to predict commercial bank failures. Proceedings of the 1990 Deloitte and Touche/University of Kansas Symposium on Auditing Problems, May 17-18, Kansas, pp: 29-53.

Bernd, F. and R. Klaus, 1996. Volatility estimation with neural network. Proceeding of the IEEE/IAFE Conference on Computational Intelligence for Financial Engineering, March 24-26, IEEE Press, New York, pp: 177-181.

Chiang, W.C., T.L. Urban and G.W. Baldridge, 1996. A neural network approach to mutual fund net asset value forecasting. Omega, 24: 205-215.

Dietrich, J.R., S.J. Kachelmeier, D.N. Kleinmuntz and T.J. Linsmeier, 2001. Market efficiency, bounded rationality and supplemental business reporting disclosures. J. Accos. Res., 39: 243-268.

Dreiseitl, S. and L. Ohno-Machado, 2002. Logistic regression and artificial neural network classification models: A methodology review. J. Biomed. Inform., 35: 352-359.

Dutt, S.D. and D. Ghosh, 1999. A note on the foreign exchange market efficiency hypothesis. J. Econ. Finance, 23: 157-161.

Dutta, S. and S. Shekhar, 1988. Bond rating: A non-conservative application of neural network. Proceeding of the IEEE International Conference Neural Networks, July 24-27, IEEE Press, California, pp: 443-450.

Egeli, B., M. Özturan and B. Badur, 2003. Stock market prediction using artificial neural networks. Proceedings of the 3rd International Conference on Business, June 18-21, Hawaii, pp: 1-8.

Fernandez-Rodriguez, F., C. Gonzalez-Martel and S. Sosvilla-Rivero, 2000. On the profitability of technical trading rules based on artificial neural networks: Evidence from the Madrid stock market. Econ. Lett., 69: 89-94.

Freisleben, B., 1992. Stock market prediction with backpropagation networks. Proceeding of the 5th International Conference Industrial and Engineering Applications of Artificial Intelligence and Expert System, June 9-12, Springer, Berlin, pp: 451-460.

Gencay, R., 1998. Optimization of technical trading strategies and the profitability in security markets. Econ. Lett., 59: 249-254.

Heaton, J.T., 2005. Introduction to Neural Networks with Java. 1st Edn., Heaton Research, Inc., Missouri, ISBN: 0977320608 .

Hellstrom, T. and K. Holmstrom, 1997. Predicting the stock market. Tech. Report Series IMa-TOM-1997-07. Vasteras, Malardalen University, http://www.e-m-h.org/HeHo98.pdf.

Huang, C.S., R.E. Dorsey and M.A. Boose, 1994. Life insurer financial distress prediction: Neural network model. J. Insur. Regul., 12: 131-167.

Kimoto, T., K. Asakawa, M. Yoda and M. Takeoka, 1990. Stock market prediction system with modular neural networks. Proceedings of the IEEE International Joint Conference on Neural Networks, June 17-21, IEEE Press, California, pp: 1-6.

Luther, R.K., 1998. An artificial neural network approach in predicting the outcome of Chapter 11 bankruptcy. J. Bus. Econ. Stud., 4: 57-73.

Man-Chung, C., W. Chi-Cheong and L. Chi-Chung, 2000. Financial time series forecasting by neural network using conjugate gradient learning algorithm and multiple linear regression weight initialization. Computing in Economics and Finance 2000. (No. 61). Barcelona, Society for Computational Economics, Retrieved from http://fmwww.bc.edu/cef00/papers/paper61.pdf.

Phua, P.K.H., D. Ming and W. Lin, 2000. Neural network with genetic algorithms for stocks prediction. Proceeding of the Fifth Conference of the Association of Asian-Pacific Operations Research Societies within IFORS. July 5-7, 2000. Singapore.

Quah, T.S. and B. Srinivasan, 1999. Improving returns on stock investment through neural network selection. Expet. Syst. Appl., 17: 295-301.

Reilly, F. and K. Brown, 1997. Investment Analysis and Portfolio Management. 1st Edn., Dryden Press, California, ISBN: 0324289039 .

Schumacher, M., R. Robner and W. Vach, 1996. Neural networks and logistic regression part I. Comput. Stat. Data Anal., 21: 661-682.

Tsibouris, G. and M. Zeidenberg, 1995. Testing the Efficient Markets Hypothesis with Gradient Descent Algorithms. In: Neural Networks in the Capital Markets, Refenes, A.P. (Ed.). John Wiley and Sons, Inc., New York, ISBN: 0471943648, pp: 127-135.

Van Eyden, R.J., 1996. The Application of Neural Networks in the Forecasting of Share Prices. 1st Edn., Finance and Technology Publishing, Virginia, ISBN: 0965133203 .

Versace, M., R. Bhatt, O. Hinds and M. Shiffer, 2004. Predicting the exchange traded fund DIA with a combination of genetic algorithms and neural networks. Expet. Syst. Appl., 27: 417-425.

Weigend, A.S., D.E. Rumelhart and B.A. Huberman, 1990. Generalization by weight-elimination with application to forecasting. Proceedings of the 1990 Conference on Advances in Neural Information Processing Systems 3. June 1990, Morgan Kaufmann Publishers Inc., pp: 875-882.

White, H., 1988. Economic prediction using neural networks: the case of IBM daily stock returns. Proceedings of the IEEE International Conference on Neural Networks, July 24-27, IEEE Press, California, pp: 451-459.

Yao, J. and H.L. Poh, 1995. Forecasting the KLSE index using neural networks. Proceedings of the IEEE International Conference on Neural Networks 2, Dec. 1, IEEE Press, pp: 1012-1017.

Yümlü, S., F. Gürgen and N.Okay, 2004. A comparison of global, recurrent and smoothed-piecewise neural models for Istanbul Stock Exchange (ISE) prediction. Pattern Recogn. Lett., 26: 2093-2103.

Zekic, M., 1998. Neural network applications in stock market predictions - a methodology analysis. Proceedings of the 9th International Conference on Information and Intelligent Systems, Sep. 23-25, Varazdin, pp: 255-263.