

Stock market price index return forecasting using ANN

Gunter Senyurt, Abdulhamit Subasi

E-mail : gseyurt@ibu.edu.ba, asubasi@ibu.edu.ba

Abstract

Even though many new data mining techniques have been introduced in prediction estimation, there is still no single best solution to all financial problems. In this study, an artificial neural network (ANN) model is utilized for predicting price index returns through regression. Ten technical market indicators, seven macroeconomic variables, a couple of other international market indices and a sliding window of ten inputs make up the 30 attributes used in this study. Different combinations of attribute sets is experimented with different ANN model parameter values to find the highest forecasting accuracy.

Keywords: Price index return, ANN, Forecasting, Data Mining Techniques.

1. INTRODUCTION

While there are certain techniques to forecast in which direction the market would be moving or what price levels would be expected, empirical evidence shows that some models work better than the others in different cases (Satchell, 2005). It is of utmost importance for investors to estimate the trend of the markets as precisely as possible in order to reach the best trading decisions for their investments, so in this context it is in the investor's best interest to use the most accurate time series forecasting model to maximize the profit or to minimize the risk. All in all, it is a quite challenging job to make accurate predictions of stock market index movements and model the time series data, especially in highly volatile markets such as the Turkish stock market. That is due to the fact that stock markets are in general chaotic and complex mechanisms with dynamic, nonlinear and nonparametric variables (Abu-Mostafa and Atiya, 1996). Moreover, markets are influenced by numerous macroeconomic factors, institutional investor choices, human psychology, political events, company policies, other stock market movements and economic affairs (Tan, Quek, and See, 2007). In this study, the ISE National 100 Index (XU-100) has been chosen for data analysis, since the Turkish stock market is a relatively young emerging market and it has presented an outstanding growth rate since its establishment in the late 80's. There is lots of empirical work available in literature on well established and developed markets such as Dow Jones (USA) or DAX (Germany), whereas little research material is available on new emerging markets such as ISE (Kara, Boyacioglu, and Baykan, 2010). The Istanbul Stock Exchange is highly volatile in terms of market returns, a feature which is attracting many local and

international investors worldwide seeking for high return possibilities (Armano, Marchesi, and Murru, 2005). By means of this study, it is aimed at contributing to the demonstration and verification of the XU-100 index price level predictability through ANN. The related predicting performances are compared based on statistical criteria such as relative absolute error (RAE), root relative squared error (RRSE) and the squared value of the correlation coefficient R^2 . The remaining part of this study is organized into four sections. The next section presents an overview of the theoretical literature while in section 3 the research data and the structure of ANN is described. In section 4, the reports and results of empirical findings from the comparative analysis are given. Finally, the last section contains the concluding remarks.

2. Literature Review

There are various ANN methods that can be used in predicting stock price returns and a great deal of research has been conducted on using ANN to forecast financial time series data outputs suggesting ANN as a powerful tool in predicting stock market return (Avci, 2007; Karaatli, 2005). Chen, Leung and Daouk (2003) used the probabilistic neural network (PNN) which showed strong predictive power over other models such as the GMM-Kalman filter and random walk. Diler (2003) who trained back propagation neural networks, based the input attributes on some technical market indicators like momentum, moving average, moving average convergence divergence (MACD), RSI and stochastic %K and forecasted the ISE 100 index direction with % 60.81 accuracy while Altay and Satman (2005) also used ISE-30 and ISE-ALL indices to see the performances of several neural network models. Cao, Leggio, and Schniederjans (2005) effectively proved that multivariate neural networks could outperform the linear models for stock price movement predictions of Shanghai Stock Exchange listed companies.

3. Materials and Methods

3.1 Research Data

In this study, all experiments were conducted on WEKA software using its MLP built-in tool to make comparisons of prediction performances based on the chosen dataset. The full dataset is comprised of 30 input variables in total. The first 10 input attributes are technical market indicators as used by Kara, Boyacioglu and Baykan (2010) which are 10-day moving average, 10-day weighted moving average, momentum, stochastic %K, stochastic %D, RSI (Relative Strength Index), MACD (moving average convergence divergence), Larry William's %R, A/D (Accumulation/Distribution) Oscillator and CCI (Commodity Channel Index). Another 10 inputs are mainly chosen from macroeconomic variables, consisting of USD(sell)-Turkish Lira exchange rate, gold price (close), monthly interest rate, CPI (consumer price index), WPI (wholesale price index), PPI (producer price index), Industrial

Production Index, DJI (Dow Jones) closing price, DAX (Germany) closing price and BOVESPA (Brazil) closing price. These variables are slightly differently chosen than Boyacioglu and Avci (2010)'s input variables. The final 10 inputs are a sliding window of the last 10 elements of XU-100 closing price index. In Yumlu and Gurgen (2005) an input window size of seven was used but it is preferred to use the last 10 elements in this study. For the regression analysis, 10-fold cross-validation was used as the test option in WEKA.

3.2 Artificial Neural Network (ANN) Model

Artificial neural networks are capable estimation models for financial modeling and prediction (Kara, Boyacioglu, and Baykan, 2010). In this study, a three layered feed-forward ANN structure (a multilayer perceptron) is used to forecast stock market index movements. Multilayer perceptrons (MLP) have one or more layers between input and output layers, called hidden layers, that can approximate any nonlinear relation to any accuracy given sufficiently large number of neurons. The nonlinearity used in the nodes provides MLP with a universal approximation power. "It has been scientifically proved that a three-layered MLP using sigmoidal activation function can approximate well any continuous multivariate function to any accuracy." (Du and Swamy, 2006). MLP shows high efficiency in function approximation for high-dimensional spaces. It has clear advantage over linear regression methods in that the input dimensionality does not affect the error convergence rate, while conventional linear regression methods suffer from the size of dimensionality. The most popular learning rule in supervised learning is the back propagation learning algorithm which is used to train the neural network. In order to minimize a cost function that is equivalent to MSE (mean squared error) between the desired and actual network outputs, a gradient search method is utilized. An input pattern is introduced to the system and the resulting computed output is compared with the actual given output (target output). The error of each calculated output is propagated backward that establishes a closed-loop control system which adjusts weights by a gradient-descend based algorithm (Du and Swamy, 2006).

4. Results and Discussion

The relevance and quality of the data, usually, has a big impact on the performance of the model used. Thus, the choice of data becomes the most important part in forecasting the markets. In this study, all series are real-valued and the input data spans from 02/01/1997 to 31/12/2007. For WEKA testing, the statistical model adequacy metrics relative absolute error (RAE), root relative squared error (RRSE), and the square of the correlation coefficient R^2 are utilized, showing the ability of the model to capture the data. A dataset of 10, 20 and 30 inputs are tested in order to see which attribute set have better predictive power over the others. Table 1 and 2 prove the effectiveness of the sliding window when used together with technical indicator inputs creating much lower error values.

Table 1. MLP regression results (% relative absolute error values - % RAE).

Input Feature Sets	# of neurons in the hidden layer(n)							
	4	7	10	20	40	50	70	90
technical indicators + macro economic variables + last 10	1	0.87	1.06	1.15	1.13	1.24	0.94	1.33
technical indicators + macro economic variables	1.80	1.61	1.70	1.76	1.88	1.90	1.78	1.83
technical indicators	1.71	1.63	1.74	2	2.32	2	2	2.1
technical indicators + last 10	0.39	0.42	0.42	0.6	0.73	0.75	1.84	1.63
macroeconomic variables + last 10	3.46	3.35	3.33	3.41	3.55	3.60	3.41	8.9

Table 2. MLP regression results (% root relative squared error - %RRSE).

Input Feature Sets	# of neurons in the hidden layer(n)							
	4	7	10	20	40	50	70	90
technical indicators + macro economic variables + last 10	1.05	0.95	1.20	1.29	1.24	1.35	1.04	1.53
technical indicators + macro economic variables	1.73	1.91	1.79	1.87	1.95	1.98	1.87	1.90
technical indicators	1.86	1.80	1.91	2.22	2.46	2.1	2.13	2.24
technical indicators + last 10	0.47	0.49	0.49	0.69	0.83	0.87	3.1	1.94
macroeconomic variables + last 10	3.81	3.70	3.70	3.79	3.96	4	3.91	18.9

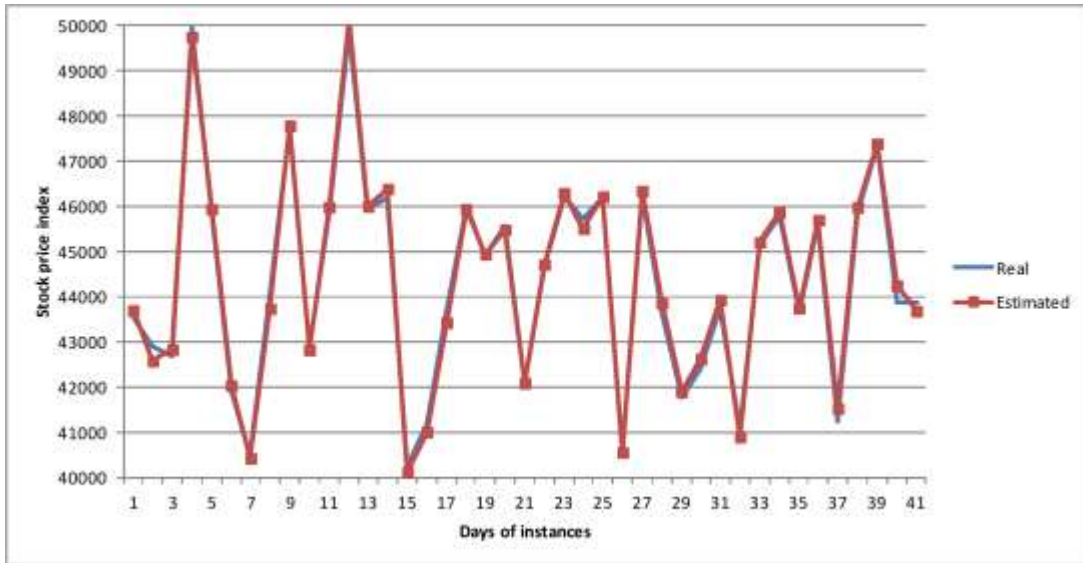


Figure 1. MLP regression result for n=4 (4 neurons in the hidden layer) and 30 features (technical indicators+macroecon. variables+last 10 slid. window).

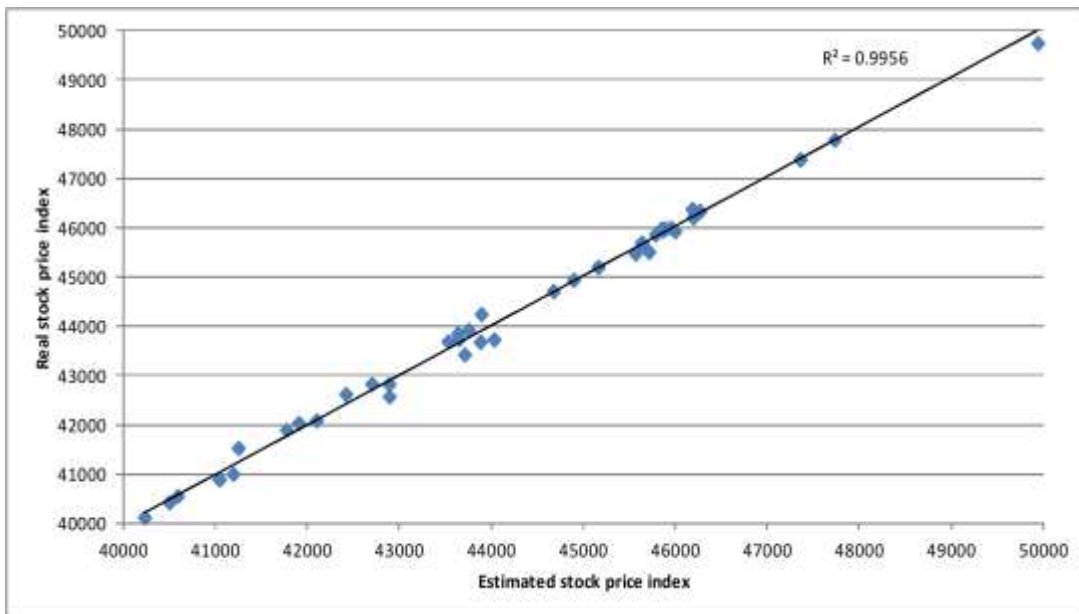


Figure 2. R^2 , MLP regression for n=4 (4 neurons in the hidden layer) and 30 features (technical indicators+macroecon. variables+last 10 sliding window).

5.CONCLUSION

The issue of accurately predicting the stock market price levels is highly important for formulating the best market trading solutions. It is fundamentally affecting buy and sell decisions of an instrument that can be lucrative for investors. This study focuses on predicting the ISE National 100 closing price levels using ANN based on the daily data from 1997 to 2007. The experimental results give us some very important clues. Firstly, ANN shows superior predicting power in forecasting the stock market price level index. MLP presents 0.39 % RAE in its best case, which is a perfectly good outcome. Even though the prediction performance of ANN outperforms studies alike in literature, it is still likely that the forecasting performance of the model can still be improved by doing the followings: Either the model parameters should be adjusted by thorough experimentation or the input variable sets need to be modified by selecting those input attributes that are more realistic in reflecting the market workings. (Kara, Boyacioglu, and Baykan, 2010) had already proved the significance of using ten particular technical market indicators which gave also good results in this study, as well. Besides, the use of a sliding window of the last ten elements of the ISE 100 index proved to be an effective tool in forecasting the market level and direction. However, the seven macroeconomic variables and three other international market indices were not found to be very useful in this study, which means that more appropriate variables has to be found that may improve the forecasting performance of the models employed that can be a further subject of study for interested readers.

Acknowledgement : We sincerely deliver our special thanks to Assist. Prof. Melek Acar Boyacioglu for her graciousness in sharing her knowledge with us.

REFERENCES

- Abu-Mostafa, Y. S., & Atiya, A. F. (1996). Introduction to financial forecasting. *Applied Intelligence*, 6(3), 205–213.
- Altay, E., & Satman, M. H. (2005). Stock market forecasting: Artificial neural networks and linear regression comparison in an emerging market. *Journal of Financial Management and Analysis*, 18(2), 18–33.
- Armano, G., Marchesi, M., & Murru, A. A. (2005). Hybrid genetic-neural architecture for stock indexes forecasting. *Information Sciences*, 170, 3–33.
- Avci, E. (2007). Forecasting daily and sessional returns of the ISE-100 index with neural network models. *Journal of Dogus University*, 8(2), 128–142.
- Boyacioglu M.A., Avci D., (2010). An Adaptive Network-Based Fuzzy Inference System (ANFIS) for the prediction of stock market return: The case of the Istanbul Stock Exchange. *Expert Systems with Applications* 37, 7908–7912.

- Cao, Q., Leggio, K. B., & Schniederjans, M. J. A. (2005). A comparison between Fama and French's model and artificial neural networks in predicting the Chinese stock market. *Computers & Operations Research*, 32, 2499–2512.
- Chen, A. S., Leung, M. T., & Daouk, H. (2003). Application of neural networks to an emerging financial market: Forecasting and trading the Taiwan Stock Index. *Computers & Operations Research*, 30(6), 901–923.
- Colby, Robert W. *The Encyclopedia of Technical Market Indicators*, McGraw-Hill, 2nd. edition, 2003.
- Diler, A. I. (2003). Predicting direction of ISE national-100 index with back propagation trained neural network. *Journal of Istanbul Stock Exchange*, 7(25–26), 65–81.
- Du K.-L., Swamy M.N.S., (2006). *Neural Networks in a Softcomputing Framework*, Springer-Verlag.
- Kara Y., Boyacioglu M.A., Baykan O.K., (2010). Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange. *Expert Systems with Applications* 38, 5311–5319.
- Karaatli, M., Gungor, I., Demir, Y., & Kalayci, S. (2005). Estimating stock market movements with neural network approach. *Journal of Balikesir University*, 2(1), 22–48.
- Satchell, C., (2005). *Pattern Recognition and Trading Decisions*, McGraw-Hill.
- Tan, T. Z., Quek, C., & See, Ng. G. (2007). Biological brain-inspired genetic complementary learning for stock market and bank failure prediction. *Computational Intelligence*, 23(2), 236–261.
- Tan, T. Z., Quek, C., & See, Ng. G. (2007). Biological brain-inspired genetic complementary learning for stock market and bank failure prediction. *Computational Intelligence*, 23(2), 236–261.
- Weka, Waikato Environment for Knowledge Analysis, Version 3.7.3, The University of Waikato Hamilton, New Zealand, 1999-2010.
- Yumlu, S., Gorgen, F., Okay, N., (2005). A comparison of global, recurrent and smoothed-piecewise neural models for Istanbul stock exchange (ISE) prediction. *Pattern Recognition Letters* 26, 2093–2103.