Stock Market Prediction Using Neural Networks: Does Trading Volume Help in Short-term Prediction?

Xiaohua Wang
School of Computing
National University of Singapore
Singapore, 119260
Email: wangxia3@comp.nus.edu.sg

Paul Kang Hoh Phua
School of Computing
National University of Singapore
Singapore, 119260
Email: dispkh@nus.edu.sg

Weidong Lin
School of Computing
National University of Singapore
Singapore, 119260
Email: dislw@nus.edu.sg

***Abstract*— Recent studies show that there is a significant bidirectional nonlinear causality between stock return and trading volume. This research reinforces the results presented in [4] and we further investigate whether trading volume can significantly improve the forecasting performance of neural networks, or whether neural networks can adequately model such nonlinearity. Neural networks are trained with the data of stock returns and trading volumes from Standard and Poor 500 composite index (S&P 500) and Dow Jones Industry index (DJI). The results are used to compared with those networks developed without trading volumes. Daily data is applied to train neural networks in order to test whether trading volumes can help in short-term forecasting. Directional symmetry (DS) and mean absolute percentage error (MAPE) are both employed to test the result of robustness. Empirical results indicate that trading volume has little effect on the performance of direction forecasting. Sometimes it may lead to over-fitting. For forecasting accuracy, trading volume leads to irregular improvements.**

I. Introduction

Using neural networks to model and predict stock market returns has been the subjects of recent empirical and theoretical investigations by academics and practitioners alike. However, several design factors significantly impact on the accuracy of neural network forecasting. Input selection is one of the demanding and intricate tasks of training neural networks. Usually, trading volume is considered as one of fundamental factors, which is beneficial for long-term forecasting [1]. Recently, the relationship between stock price and volume or between volatility and volume becomes the hot topic in theory as well as empirical research [4], [5], [9]. This study aims to find whether trading volume is beneficial to stock market forecasting by using neural networks.

Reference [13] argues that it is more helpful to understand the market through studying the joint dynamics of stock prices and trading volume than by focusing only on the univariate dynamics of stock prices. This argument indicates that trading volume is related to stock price and may be an important determinant of stock prices. [9] investigates the relationship between stock market trading volume and the serial correlation of daily stock returns. Their results suggest that the decrease or increase in stock price has a strong relation with the decrease or increase in trading volume. In other words, trading volume has an impact on the dynamic of stock price. [5] finds evidence of significant bidirectional nonlinear causality between returns and volume. In addition, the authors examine whether the nonlinear causality from volume to return can be explained by volume serving as a proxy for information flow in the stochastic process and they find that there exits nonlinear causality from volume to returns.

Both theoretical and empirical studies have proven a nonlinear relation between stock return and trading volume. However, whether this nonlinear relationship can be of help to improve the forecasting performance is still an open question. [4] explores a number of statistical models (both linear and nonlinear) for predicting the daily stock return volatility of an aggregate of all stocks traded on the New York Stock Exchange market (NYSE). The author finds that lagged volume leads to very modest improvement in forecasting performance. It is inferred that such results are attributed to the transformation method applied in the data, which may lose the important information in trading volume. [4] just focus on the relationship between trading volume and stock volatility. It is still unclear whether trading volume can improve the forecasting performance of stock return. [1] introduces stationary transformations of dividends and trading volume as fundamental explanatory variables to neural network models. Their results indicate that inclusion of nonlinear term in the relation between stock returns and fundamentals can improve the out-of-sample forecasting accuracy. However, this study just focuses on long-term (monthly return) forecasting not short-term, limiting the validation of the conclusion.

This study emphasizes on whether trading volume can improve the forecasting performance of neural networks or whether neural networks can take advantage of such nonlinearity to get more accurate results. The terminate goal of this research is to investigate to which degree trading volume can improve the performance of stock return forecasting and give some guidance in input selection. Three kinds of neural networks with different inputs are trained to simulate financial time series, and the results are obtained from daily financial time series of two popular stock markets, say Standard and Poor 500 composite index (S&P 500) and Dow Jones Industry index (DJI).

This paper is organized as following: Section II presents the details of our experiment design; Section III discusses the results and analyzes the possible reasons; Section IV draws the conclusions.

0-7803-7898-9/03/$17.00 ©2003 IEEE

II. Experiment Design

The data from both stock markets, S&P 500 and DJI, are employed to train neural networks. Three-layer feed-forward neural networks with various number of hidden neurons are employed to model nonlinearity among time series. The details are shown in the following:

A. Data

The financial time series of daily S&P 500 and DJI observations from Oct 1st,1990 to Oct 1st, 2002, including stock index and trading volume, are used in this research. The index returns are calculated as 100 times the first difference of the natural logarithm of the index values, which is commonly used in the literature. Similarly, stationary transformation method used in [1] is also applied to obtain the information about volume values [1]. One advantage of such transformation is stationary, which may not lose the important information from trading volume.

Table I shows some statistical results of the time series for the two markets. Obviously, DJI appears more volatile than S&P 500 as standard deviation of the former is a little higher than that of the latter. Another finding is that the linear correlation coefficient of DJI is bigger than that of S&P 500. Although correlation coefficient represents the linear relationship among the variables, it reflects the intensity of such relationship to some extent. We just wonder whether higher correlation between stock return and trading volume can more help improve the forecasting accuracy.

As stated in previous study [14], it is not true that the more data sets for training, the more accurate results neural networks produce. So, the preliminary experiments are conducted to decide the optimal training sample sets. The data sets with 600, 800, 1000, 1200, 1500, 2000 training patterns are used. And each data set is divided into two parts: 90% are used to train neural networks while the rest 10% are employed to test the generalization ability.

B. Neural Networks

We employ the time delay neural network (TDNN), first proposed by [3], [10], to explore the usefulness of volume information in the explanation of the predictability of stock index returns. Although a variety of different TDNN models have been devised in the literature, the three-layer feedforward network is adopted in the present study as it is in the most widespread use. The sigmoid transfer function and linear transfer function are applied in the hidden layer and output layer respectively. Biases are added into both input layer and hidden layer.

To build a forecaster of time series, the inputs of the networks are carefully selected to not only reflect the internal movement dynamics of the time series by using its time delay values, but also reveal the environmental impacts and interactions among different effective factors by adding explanatory variables. Mathematically, such a model can be expressed as:

$$y_{t+s} = \phi(y_t, y_{t-1}, \dots, y_{t-h+1}, u_{1t}, u_{2t}, \dots, u_{mt}) + \epsilon_t \quad (1)$$

where s is the number of the steps ahead ($s = 1$ at the present paper), h is called the looking back window size and $u_{1t}, u_{2t}, \dots, u_{mt}$ are explanatory variables at time t. We also assume that the size of lagged values is five days, or one week. This choice is somewhat arbitrary, and the model could easily be refined through a more detailed specification approach based on statistical analysis, for example [7]. Additional explanatory variables, such as interest rate differentials, book-to-market values, lagged daily trading activity measures, and etc. could also easily be included. However, this paper only considers one of the daily trading activity measures: volume. To facilitate the comparison, we choose the following three typical models:

- T1: Inputs consists of only five lagged return values (R_t, $R_{(t-1)}, \dots R_{(t-4)}$);
- T2: Inputs are composed of five lagged return values and one daily information for trading volume (T_t);
- T3: Inputs have five lagged return values and three daily information for trading volume (Three delayed trading volumes, T_t, $T_{(t-1)}$, $T_{(t-2)}$).

The values of all weights can be numerically estimated in the training procedure which is equivalent to typically solving nonlinear optimization problem. There are scores of numerical computation algorithms designed for training neural networks such as back propagation algorithms, nonlinear least square techniques based algorithms, nonlinear optimization techniques based algorithms, and so on. Due to the complexity and high-degree nonlinearity of the problem, there is no guaranteed way that the global minimum can be reached. To improve the robustness and convergent speed of training algorithms, we employ the modified BFGS [2] to carry out the computing of all neural network activities.

TABLE I
The Preliminary Statistics of 2000 (total) Training Pattern

Preliminary Statistics	SP 500 Return	SP 500 T.Vol.	DJI Return	DJI T.Vol.
Mean	0.0302	0.0939	0.0360	0.0938
St.D.	1.1708	21.9165	1.1413	24.7828
Skewness	-0.1782	-0.0343	-0.3686	0.1170
Kurtosis	3.3249	54.5670	4.2994	40.8519
Correlation Coefficient	-0.0202		-0.0269	

[1] That is, the natural logarithm of the change in trading volume that are computed as 100 times are also introduced as inputs into neural networks.

[2] This modified BFGS training algorithm is used Levenberg-Marquardt method to initialize the Hessian matrix and used BFGS method to update the Hessian matrix in the following epoches. See [12].

III. Results and Discussions

To enhance the analysis of results, directional symmetry (DS) and mean absolute percentage errors (MAPE)[3] are applied to test the generalization ability and accuracy. Since the forecasts generated by the various models for the same time series are not statistically independent, the comparisons are made using the paired t-test, which is robust to such problems [11].

The stopping criterion is set to mean square error $(MSE) = 1.2$. In order to avoid the effect of initial weights, each training trial is proceeded 10 times with different initial weights and average results are presented in the following tables. The results of preliminary experiments show that 600 data patterns can develop the most appropriate neural networks. So in the following experiments, we apply the 540 data set as training patterns to build the neural models and the rest 60 data set as testing patterns to verify the generalization ability.

A. Convergence Rate

Table II shows the convergence rates [4] of the three network models. First, the iteration numbers in the columns of Tab. II keep decreasing with the number of hidden neurons increasing. Such results show that increasing hidden neurons can greatly accelerate the convergence rate.

Second, it is interesting to find that T2 and T3 converges significantly faster than T1 when the number of hidden neurons is 8,10,12. For DJI, average 66.5 epoches are needed for training network T1 with eight hidden neurons, while T2 and T3 just need 57.8 and 56 epoches respectively at the same situation. However, with more hidden neurons added to the networks, T1 converges faster in comparison with T2 and T3. Such results are perhaps surprising. It is generally considered that the more information neural networks deal with, the more time they need. So, the convergence rate should be slower. However, in our experiments, we find that T2 (6 inputs) and T3 (8 inputs), which deal with more information from trading volume, converge much faster than T1 (only 5 inputs). One possible explanation is that when hidden neurons become less, trading volume can significantly accelerate the convergence rate of neural networks. However, with adding more hidden neurons, both single trading volume and lagged trading volume lose this advantage, because the more complex networks is, the more time they need to deal with time series. At last, T2 comparably converges more rapidly than T3, for T3 deals with more information. Both S&P 500 and DJI have the same phenomena. But DJI seems convergent more quickly than S&P 500.

[3]In this paper, MAPE and DS are taken as robust measurement of generalization ability. And they are calculated as $MAPE = \frac{1}{T}\sum_{i=1}^{T}\left|\frac{Y_{i-obs}-Y_{i-pre}}{Y_{i-obs}}\right| \times 100$, where Y_{i-obs} represents the observed value and Y_{i-pre} presents the predicted value. And DS is the percentage of correctly predicted sign in proportion to total observations.

[4]In this study, convergence rate is simply measured by the number of iterations for training networks

TABLE II

Convergence Rate vs. Hidden Neurons

NNs	SP 500			DJI		
TP.	T1	T2	T3	T1	T2	T3
8	111.4	71.5	56.9	66.5	57.8	56
	(39.1)	$(12.6)^a$	$(8.8)^a$	(11.4)	$(21.7)^a$	$(15.2)^a$
10	71	53.6	50.5	53.3	53	55.4
	(19.1)	$(7.3)^a$	$(8.0)^a$	(7.3)	(11.1)	(29.3)
12	46.8	41.1	40.7	47.4	37	41.6
	(8.8)	$(6.1)^a$	$(2.6)^a$	(8.4)	$(6.9)^a$	$(4.0)^a$
15	33.5	36.3	36.9	33	31.3	36.1
	(5.5)	(6.1)	$(4.2)^b$	(5.1)	(5.8)	$(4.3)^b$
20	26.5	32	34.4	24.5	27.8	33.2
	(2.7)	$(3.9)^b$	$(4.6)^b$	(4.2)	$(3.9)^b$	$(4.5)^b$

Average epoches of 10 trials with different initial weights are shown with their Standard Deviation in the parenthesis.
^{a} represents T2 or T3 outperforms T1 in terms of paired t-tests for 0.05 significant differences. ^{b} represents T1 outperforms T2 or T3 in terms of paired t-tests for 0.05 significant differences.

B. Generalization Ability

We use DS results to test the direction correction when training process is stopped. DS results are also used to test the generalization ability of these networks models. Table III presents these results for training as well as testing.

First of all, no significant differences exits among the results of different hidden neurons for each type network model. A reasonable comment on this result is that the number of hidden neurons seems to have no significant effect on the DS accuracy. Second, although all average DS for training are around 60%, it appears that T3 can greatly improve the direction forecasting for training. And these improvements are quite significant. For S&P 500, when the number of hidden neurons is 20, DS of T3 is 66.1%, which is quite greater than 62.7 of T1 (P-value is less than 0.001). But T2 seldom outperforms T1. It can be concluded that lagged trading volume can significantly improve the direction forecasting while single trading volume cannot improve the performance of neural networks for both markets.

However, for testing, lagged trading volume increases the average direction forecast error, which is extremely different from training. For S&P 500, T1 appears to pronouncedly outperform T2 and T3. All DS values for T1 are better than those for T3. In addition, when the number of hidden neuron equals to 15 and 20, the results of T1 are significantly better than those of T3 (P-value is less than 0.001). In comparison of T1 with T2, T1 performs at least as good as, if not better than (when the number of hidden neurons equals to 15), T2 additionally. Consequently, for S&P 500, trading volume leads to bad generalization. On the other hand, for DJI, the results are a little different. Although when the number of hidden neurons equal to 8, 12, 15, T2 or T3 appears to be better than T1. But such improvement becomes quite modest as p-value is larger than 0.2. In addition, in the case that T1 seems better than T2 and T3, such goodness is also very modest. So, for DJI, although trading volume can significantly improve the direction forecasting results in training process, the results of testing (generalization ability) without trading volume are as

TABLE III

DS VS. HIDDEN NEURONS FOR TRAINING AND TESTING

NNs	Training						Testing					
TP.	SP 500			DJI			SP 500			DJI		
H.N.	T1	T2	T3	T1	T2	T3	T1	T2	T3	T1	T2	T3
8	63.7	62.4	64.2	61.5	59.9	61.8	53.2	48.8	51.2	51.0	54.3	45.7
	(0.02)	(0.01)[b]	(0.03)	(0.02)	(0.02)[b]	(0.03)	(0.06)	(0.08)	(0.06)	(0.03)	(0.06)	(0.02)[b]
10	62.4	62.9	64.7	61.7	61.4	63.1	51.2	53.2	49.8	51.3	50.3	47.5
	(0.02)	(0.02)	(0.02)[a]	(0.02)	(0.02)	(0.03)	(0.07)	(0.04)	(0.04)	(0.06)	(0.07)	(0.04)[b]
12	62.7	62.3	64.5	62.0	60.1	63.1	51.0	49.5	49.5	47.2	49.2	51.2
	(0.02)	(0.02)	(0.02)[a]	(0.02)	(0.02)[b]	(0.01)[a]	(0.05)	(0.06)	(0.05)	(0.07)	(0.05)	(0.08)
15	63.0	61.8	65.7	61.9	60.3	62.3	53.2	49.7	49.2	50.5	50.8	48.5
	(0.02)	(0.01)[b]	(0.02)[a]	(0.02)	(0.02)[b]	(0.02)	(0.06)	(0.04)[b]	(0.03)[b]	(0.04)	(0.06)	(0.06)
20	62.7	62.1	66.1	60.9	60.5	63.3	52.2	52.0	48.3	51.3	51.2	46.5
	(0.02)	(0.01)	(0.02)[a]	(0.02)	(0.02)	(0.02)[a]	(0.03)	(0.05)	(0.06)[b]	(0.08)	(0.02)	(0.06)

Average DS (% percentage)of 10 different initial weights are shown in table. Standard Deviations are shown in the parenthesis.
[a] represents T2 or T3 outperforms T1 in terms of paired t-tests for 0.05 significant differences. [b] represents T1 outperforms T2 or T3 in terms of paired t-tests for 0.05 significant differences.

good as those with trading volume.

Conclusively, it is quite surprising to find that trading volume cannot improve the performance of direction forecasting in out-of-sample sets. Although trading volume can significantly improve the direction forecasting results during training process, unfortunately, such improvement is lost and leads to worse generalization during testing process. All results indicate that trading volume cannot help and more obviously, it is probably inclined to *over-fitting* instead of improving the forecasting performance for S&P 500, especially in T3 models. [2] states that when adding more fundamental factors into neural network to produce better correlation results with observed values, the network is facing the risk of over-fitting. For S&P 500, T3 produces better results in training process than other models but leads to the worst generalization among the three types of models. It is possibly attributed to too more trading volumes included into the networks.

C. Accuracy

In order to draw more reasonable conclusions, another measurement MAPE is also employed, due to its popularity and robustness [8]. Table IV presents the MAPE results. One of the most interesting findings is that the results of MAPE are not consistent with those results of DS. It is also surprising that the results of S&P 500 are quite different from those of DJI. For S&P 500, the results are dynamic: sometimes T1 seems better than T2 and T3, while other times T3 seems better than T1. Whatever, such differences among MAPE are not significant (P-values are always larger than 0.25). In another words, the results of the three models are almost same. Trading volume cannot help improve the forecasting accuracy of S&P 500. On the contrary, for DJI, when the number of hidden neurons equals to 10, 12 and 20, MAPE of T3 are significantly less than those of T1(P-values are less than or equal to 0.004.). In other cases, there are no considerable differences among the results. So, DJI seems to take the nonlinearity of volume and returns in their stride producing reasonable results in some occasions, although such improvements are quite erratic.

TABLE IV

MAPE VS. HIDDEN NEURONS FOR TESTING

NNs	SP 500			DJI		
TP.	T1	T2	T3	T1	T2	T3
8	2.96	2.97	3.07	1.58	1.62	1.37
	(0.06)	(0.64)	(0.62)	(0.28)	(0.21)	(0.28)
10	2.89	3.05	3.33	1.64	1.44	1.43
	(0.97)	(1.15)	(1.28)	(0.28)	(0.21)[a]	(0.19)[a]
12	2.71	3.25	2.58	1.55	1.64	1.37
	(0.78)	(1.62)	(0.84)	(0.25)	(0.28)	(0.20)[a]
15	3.26	3.37	2.89	1.66	1.68	1.58
	(0.86)	(1.32)	(1.37)	(0.22)	(0.28)	(0.34)
20	3.06	3.08	3.23	1.65	1.78	1.38
	(1.08)	(0.85)	(0.91)	(0.23)	(0.21)[b]	(0.11)[a]

Average MAPE of 10 different initial weights are shown in table. Standard Deviations are shown in the parenthesis.
[a] represents T2 or T3 outperforms T1 in terms of paired t-tests for 0.05 significant differences. [b] represents T1 outperforms T2 or T3 in terms of paired t-tests for 0.05 significant differences.

D. Discussion

The above analysis indicates that trading volume cannot significantly improve the direction forecasting accuracy of the proposed neural networks. And the improvements, if any in MAPE, are quite erratic and infrequent by using neural networks. One possible explanation for this conclusion is that such neural networks cannot handle the nonlinearity between stock return and trading volume. Previous studies have shown that neural networks with one hidden layer can approximate a continuous function and achieve the desired accuracy [6]. However, in practice, it may be invalid because no knowledge exists about the optimal structure for a special problem. Structure and training method are the determinant factors in achieving accurate forecasting results for neural networks. Although various number of hidden neurons are tested in the experiments and no significant improvements appear, it may be due to not finding the optimal architecture and available training methods.

Another reasonable explanation is that such nonlinearity cannot improve the performance in stock return forecasting. Many empirical studies have proven that nonlinearity exists between stock return and trading volume. Although causality nonlinearity is found from trading volume to stock return, it may be too weak to be modelled by neural networks. [4] indicated that the relationship is stronger from volatility to volume than the other way around. Likewise, maybe such nonlinearity between return and volume is so weak that it cannot be modelled by neural networks and then it cannot improve the forecasting performance. However, some research find that including fundamental factors into neural networks can significantly improve the performance [1]. Still, such successful researches not only introduce trading volume but many other fundamental factors, such as dividends, interests and so on. Accordingly, it is doubtful whether such nonlinearity between stock return and trading volume may be strengthened by introducing other fundamental factors. No references exists for which reason of the above is more sensible, although the latter seems more rational.

In comparison of single trading volume with lagged trading volume, the results of the latter are little better than the former, for the latter seems to be more informative than the former. In addition, the effect of trading volume in DJI seems stronger than that in S&P 500. It may owe to that the relation between the stock returns and trading volume of DJI is a little stronger, than that of S&P 500.

IV. Conclusions

This study employs three types of neural networks to test whether trading volume can significantly improve the forecasting performance in financial market. S&P 500 and DJI are tested for several architecture neural networks. Contrary to theories and empirical based-exception, the results show that trading volume cannot significantly improve the forecasting performance of S&P 500 and can seldom improve the forecasting performance of DJI except special occasions. It may due to such nonlinearity between stock return and trading volume is too weak to be captured by neural network. Or neural networks cannot simulate such nonlinearity on account of architecture or training methods.

These findings have a number of important implications for future research in this area. The most interesting one is how to use the information in trading volume to improve the forecasting results. In this paper, neither one lagged trading volume nor three lagged trading volume can considerably improve the performance, compared with those without trading volume. One possible way to improve the performance is combining trading volume and other fundamental factors in neural network models, for it perhaps reinforces the nonlinearity between volume and returns. However, when adding more fundamental factors into neural networks, such models are more opt to over-fitting [2]. An alternative is to apply the input pruning method in this area. Sensitive analysis may be used to analyze to what extent trading volume can affect the stock return forecasting. More types of neural networks may be tested, especially recurrent networks and probability neural networks, which seem to be superior to feed-forward network in certain applications.

References

[1] A. Kanas and A. Yannopoulos, "Comparing linear and nonlinear forecasts for stock return", *International Review of Economoics and Finance*, vol.10, pp. 383-398, July, 2001.

[2] A. P. N. Refenes, A. N. Burgess and Y. Bentz, "Neural networks in financial engineering: a study in methodology", *IEEE Trans. on Neural Networks*, vol.8, pp. 1222-1267, November, 1997.

[3] A. Waibel, T. Hanazawa, G. Hinton, K. Shikano and K. J. Lang, "Phoneme recognition using time delay neural networks", *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol37, pp. 328-339, 1989.

[4] C. Brooks, "Predicting stock index volatility: Can Market Volume help?",*Journal of Forecasting*, vol.17, pp. 59-98, December, 1998.

[5] C. Hiemstra and J. D. Jones,"Testing for linear and nonlinear Granger causality in the stock price-volume relation", *The Journal of Finance*,vol.49, pp. 1639-1664, December, 1994.

[6] G. Cybenko, "Approximation by superpositions of a sigmoidal function", *Math. Control, Signals and systems*,vol.2, pp. 303-314, 1989.

[7] J. Lakonishok, E. Maberly, "The weekend effect: trading patterns of individual and institutional investors", *The Journal of Finance*, vol.45, pp. 231-243, March, 1990.

[8] J. S. Armstrong, F. Collopy,"The selection of error measurements for generalising about forecasting methods: emplirical tests", *International Journal of Forecasting*, vol.8, pp. 69-80, June, 1992.

[9] J. Y. Campbell, S. J. Grossman, J. Wang, "Trading volume and serial correlation in stock returns", *The Quarterly Journal of Economics*,vol.108, pp. 905-936, November, 1993.

[10] K. J. Lang and G. E. Hinton, *The development of the time-delay neural network architecture for speech recognition*, Technical Report CMU-CS-88-152,Carnegie-Mellon University, Pittsburgh, PA., 1998.

[11] M. Nelson, T. Hill, W. Temus and M. O'Connor, "Time series forecasting using neural networks: should the data be deseasonalized first?", *Journal of Forecasting.*, vol.18, pp. 359-367, September, 1999.

[12] P. K. H. Phua, X. H. Wang, C. X. Xu, "A comparative study of the least-square methods as training algorithms for feedforward neural networks",working paper, unpublished.

[13] R. Gallant, P. E. Rossi and G. Tauchen, "Stock Prices and Volume", *Review of Financial Studies*, vol.5, pp. 199-242, 1992.

[14] S. Walczak, "An empirical analysis of data requirements for financial forecasting with neural networks", *Joural of Management Inforormation Systems*, vol.17, pp. 203-222, Spring, 2001.