Predicting Abnormal Stock Returns with a Nonparametric Nonlinear Method

Alan M. Safer

California State University, Long Beach

Department of Mathematics

1250 Bellflower Boulevard

Long Beach, CA 90840-1001

Email: asafer@csulb.edu

Abstract

Neural networks (NN) can be applied to the predication of stock market trends based on information from legal insider trading. These data are available because officers of companies are required by law to submit to the Securities Exchange Commission a record of the sales and purchases of their companies stock. Because purchases are more useful in this endeavor than are sales, all smallcap, midcap, and largecap companies that averaged multiple buys made by insiders of companies over a 4 ½ year period (1993 to the middle of 1997) were assessed in relation to the price fluctuation of the company's stock in the forthcoming period (3, 6, 9 and 12 months ahead) as related to the index of stocks and individual stock reaction to the market as a whole. The use of NN has advantages over alternative methods as evidenced by its accuracy without using assumptions involved in other techniques.

Introduction

A vast majority of previous research studies find that insider traders usually make abnormal returns [16] [17]. Outsiders who use insider information can also make increased profits [17] [2] [8]. The ability of outsiders using insider trading information to predict abnormal returns can be increased by focusing on data such as the size of the company and the number of months in the future that are predictive for stock prices [15] [11] [7].

A more mathematically precise analysis using insider trading data for the prediction of abnormal returns is possible with the aid of recent technology, for example using neural networks. NN are a set of non-parametric techniques useful for analyzing nonlinear data sets such as those that characterize stock price information. Indeed, neural networks have commonly been used to analyze stock market data [14] [18] [6]. However, only one previous study has used neural networks to predict abnormal returns of stocks based on insider trading information [12]. That study was limited by having a very small sample of companies (n=36).

This research will increase the number of companies (from 36 to 343), the number of variables used in the prediction of abnormal returns, and the number of previous and future months on which the prediction is based.

Methods

Stock Selection

The insider trading data used in this study are from January 1993 to mid June 1997. The stocks used in the analyses included all stocks in the S&P 600 (small cap), S&P 400 (midsize cap) and S&P 500 (large cap) as of June 1997 that had insider records for the entire period of the study. There were 946 stocks in the three market caps which had available data in January 1993. From the list of 946 stocks, the sample included every stock that averaged at least 2 buys per year, that is, at least 9 total during the 4 ½ year study period. The resultant number of stocks used for the

0-7803-7044-9/01/$10.00 ©2001 IEEE

study was 343. The reason for using insider purchases over sales is that they are more closely aligned to a company's prospects and are therefore more useful for the prediction of abnormal returns. For example, if an insider sells shares of his company, he may think the company is going in the wrong direction and thus the stock will go down. Or, he may need the money for other matters such as house payments or to pay for his children's education or other needs [15] [7]. The rationale for requiring at least 2 purchases per year is that it provides sufficient transaction data for the analyses.

The original data came from the Securities and Exchange Commission (SEC). These data include: company, name of insider, rank, transaction date, stock price, shares traded, type of transaction (buy or sell), and shares held after trade. The report of the data can be delayed up to a maximum of one month and ten days after each transaction. Thus, one important aspect of this study was to see if this delay or reporting would be significant in predicting for months in the future.

Variables

The variables used in the study to predict abnormal returns are shown in Table 1. Variable 1 indicates whether an insider is a new shareholder in the company. Variable 2, the 8 week sell/buy ratio (# selling transactions/# buying transactions), is a very good indicator of the market as a whole. Variables 3 and 4 involve the median number of shares bought and sold relative to the amount the insiders held before the transaction.

Variable 5 retroactively conveys the overall average of all traders' individual average returns from the previous 2 and 4 month periods to the subsequent 3 months of insider buy transactions. Variables 6 and 7 convey average returns as in variable 5, but are based from the previous 2 and 4 month periods to the subsequent 6 and 9 months of transactions. The reason for using these variables (5-7) is to ascertain whether traders who bought stock in the past that resulted in financial gain achieved similar results in subsequent trading. This is done because some insiders have been more aware of their company's prospects than others. The rank of the insider (e.g., CEO, CFO) was not used as a variable in this study because it has had mixed results in regard to predicting abnormal stock price returns [10] [9]. For similar reasons, insiders owning 10% or more of company shares and who were not involved in company decisions were not included [17].

Past and Future Periods of Analysis

In the present study, an important design issue involves finding the optimal length of time in the past from which to analyze buy and sell transaction data. Many studies take an aggregate of insider activities one month before the current date and then predict future returns [17]. Based on the expertise of investigators who followed insider trading for many years (e.g. Moreland [8]), this study uses 2 months (specifically 9 weeks) and 4 months (18 weeks) of insider trading history to appraise past trading patterns. The period of time in the future used to predict abnormal returns was arbitrarily set to 3, 6, 9, and 12 months.

Handling of Lags

The question of what to do with the lagged data is very important. One can take the lagged data up to 18 weeks plus the current week and use principal component analysis to form new inputs. In addition, one may aggregate the lags into one summed group. Another way to attack such a problem includes weighted moving averages. Here, the farther away the week is from the current one, the less weight it has. This approach does not seem to work very well. The reason appears to be that many insiders tend to buy in smaller amounts in order to avoid being noticed by the Securities and Exchange Commission. Another method of handling lags involves grouping. If there were groupings of specific weeks for each variable that seemed logical, then this would help reduce the number of inputs. However, there are no clear groupings of the variables by weeks. Yet another way of handling the data is by forcing equivalent lags of different variables to have the same weight. This can be done by a shared network architecture. This architecture was tried and does not apply to insider trading data.

One final way to handle the lags, which seems very appropriate considering the nonlineariety of the data, is nonlinear principal component analysis. This analysis can be tried using an architecture that involves 5 layers. The input and output layer are the same. The second and fourth layers have the same number of nodes and are smaller than the input/output layer in nodes. The middle layer has a smaller number of nodes than the second and fourth layer. Nonlinear principal components tend to compress a greater amount of the variability into the same number of components than linear principal component analysis. However, there is no easy way to decide on the number of nodes for the second (and thus fourth layer) and the middle layer. In addition, the five layer architecture takes a long time to run. Nonetheless, this is an appropriate way to attack the problem.

Abnormal Returns

In order to control for risk and determine abnormal returns for stocks, an event study similar to the one by Brown and Warner [3] was used. Consequently, the Sharpe-Lintner form of the Capital Asset Pricing Method (CAPM) was used in this study [4]. This method includes a pre-event period starting the day before the event and going back 3 months. The event in this study is not the traditional event, such as an earnings report date [13]. In this study, the event is the current day when an investor decides whether to make a transaction in a particular stock. The 4 different event periods used are the intervals from the event to 3, 6, 9, and 12 months ahead.

Part 1 of CAPM

First, a multiple regression analysis was used to estimate α_i and β_i (a measure of the systematic risk of an asset) based on the pre-event period (t-1 day to t-90 days).
This is done using the following equation:

$$(R_{i,t} - r_{f,t}) = \alpha_i + \beta_i*(R_{m,t} - r_{f,t}) + \varepsilon_{i,t} \quad (1)$$

where
$R_{i,t}$ is the pre-event return on stock i for day t
$r_{f,t}$ is the 3 month daily treasury bill (tbill) rate
$R_{m,t}$ is the pre-event return of the market
$\varepsilon_{i,t}$ is the error for stock i; t is the pre-event period from t-90 to t-1 (t-0 is the event day);

Part 2 of CAPM

To calculate abnormal returns, α_i and β_i from part 1's pre-event period were used.
The outputs are: abnormal returns 3, 6, 9, and 12 months ahead and determined using the following equation:

$$\varepsilon_i = R_{i,T} - r_{f,T} - \alpha i - \beta_i*(R_{m,T} - r_{f,T}) \quad (2)$$

T is the event period (either 3, 6, 9, or 12 months ahead).

Neural Networks Defined

A neural network is a set of computational units (nodes or neurons). A neural network is characterized by its pattern of connections between the neurons (called its architecture), 2) its method of determining the weights on the connections (called its training or learning algorithm), and 3) its activation function(s) [4].

The nodes are connected in a network of layers that appraise the parameters (weights) of complex largely undefined data. The connections between the nodes are used to store knowledge and to make it available for use. In a feedforward network, each connection has a weight, interpreted as the strength of the connection from the previous layer to the current layer. The method of determining the weights on the connections is called its training or learning algorithm. The algorithm used in this neural network study is a variation of the backpropagation method.

A training set consists of inputs and associated outputs used for learning the values of the weights. A validation set is made up of a smaller nonoverlapping group of observations used to better approximate the parameters.

Neural Network Specifications

This study covers a period of 4 ½ years or 232 weeks. When using neural networks with the inputs from Table 1 and abnormal returns as the output from the 9 previous weeks of aggregate data, 223 total weeks were covered. This analysis used 180 of the 223 weeks (80.7%) for the training set and the rest, 43 weeks (19.3%), for the validation set. The 2 sets were randomly selected. For the 18 week set, there were 214 weeks available. For this part of the analysis, 173 of the 214 weeks (80.8%) were used for training and the remaining 41 (19.2%) were used for the validation set.

There was one hidden (middle) layer in the neural network analysis (i.e., 1 input layer, 1 hidden middle layer, 1 output layer). The number of nodes in the hidden layer varied depending on the stock, but usually was between 5 and 9.

For the 9 week and 18 week analyses, different numbers of neurons in the hidden layer were used. The number of neurons selected was the amount in the network with the least means squared error in the validation set.

For the 2 sets of data described, the data were aggregated. That is, the inputs were aggregated from the week of the transaction decision event and included every week up to 9 weeks back and up to 18 weeks back.

Sensitivity Analysis

The relative importance of each input can be determined by using sensitivity analysis. In essence, this procedure tests how the network would perform with each of the 13 inputs taken out individually and leaving the remaining 12 as the predictor variables.
The sensitivity ratio (SR) =

$$\frac{\text{error with omission of a specific variable}}{\text{baseline error with all variables in the model}}$$

For each variable, the greater the SR, the more important it is to the model. A baseline error is used for comparison purposes. When the sensitivity ratio ≤ 1, the network performs better if the variable is taken out of the model. Variables that have a sensitivity ratio much bigger than 1 are clearly the most important variables.

Standard Deviation (S.D.)

The S.D. ratio is very useful in determining the model fit. The

$$\text{S.D.ratio} = \frac{\text{std dev of errors for the output variable}}{\text{std dev of the target output variable}}$$

The explained variance of the model can be found by subtracting the S.D. ratio from 1 (i.e, 1-S.D. ratio).

Results

Explained Variance by Time Period before and after the Transaction Decision

Determining the length of time before and after a stock transaction decision is made is essential. This is necessary so as to obtain the prediction of abnormal returns in stocks achieving the highest explained variance. Using the 12 months future prediction and the 18 weeks back aggregated data result in the highest percentage of stocks with the most explained variance for abnormal returns. With shorter periods of future prediction, the percentage of stocks with the highest explained variance decreases. Likewise, with a shorter duration of data before the event decision (specifically, 9 week back aggregated data), the results show a lower percentage of stocks with high explained variance.

For stocks achieving an explained variance over 60%, the overall percentages meeting this criterion are as follows: 1) 35% (119/343) for the 12 months ahead prediction; 2) 30% (102/343) for the 9 months ahead prediction; 3) 17% (59/343) for the 6 months ahead prediction, and; 4) 3% (11/343) for 3 months ahead. These data make clear that 12 months ahead -- and to a lesser extent 9 months ahead -- are particularly useful periods to maximize predictability.

Sensitivity Analysis Results

For 18 week aggregated back data and 12 month future prediction, those variables that had the highest sensitivity ratios (SR) were more important than others in predicting abnormal returns. These were analyzed for the stocks that had the highest level of explained variance (i.e., 60% and more). The predictor variable SR values that were highest--in rank order of importance--for each stock were: sell volume, number of buy transactions, buy value, buy volume, sell value, and the number of sell transactions. The predictor variable group incorporating the previous buying record was only important for about half of the stocks with a high explained variance. However, when this was the case, they were amongst the most important predictor variables.

Summary and Discussion of the Major Findings

Using neural network technology, the study revealed that the prediction of abnormal returns can be maximized in the following ways**:** 1) extending the time of the future forecast up to 1 year; 2) increasing the period of back aggregated data. The fact that the time of the future forecast up to 1 year is the best shows that the delay in reporting (on average of about 1 month) by the insider to the Securities and Exchange Commission does not affect the prediction for someone using the insider trading data.

Studies have previously reported two of this study's findings involving insider trading data. These are: 1) twelve months in the future is a better predictor for abnormal returns than shorter forecasts [7];

This study has certain advantages over previous insider trading studies. It uses up to 4 months back aggregated data. Furthermore, sensitivity analysis is used to ascertain predictor variable importance. Last, the study analyzes the prediction of insider trading data using a nonlinear technique, neural networks.

The advantage of using neural network analyses for the assessment of insider trading data is they are especially useful for predictions when the data being analyzed are nonlinear in nature, such as is the case for insider trading data used to predict abnormal returns. NN analyses with one small exception [12] have not been previously used in other insider trading research prediction studies.

Future Research

Several ways to extend this study are: 1) including composite industry-wide insider trading as an input variable. 2) increasing the number of years in the study. 3) using lagged data instead of aggregate data. 4) applying types of network architectures other than feedforward neural networks 5) comparing this NN analysis with analyses using other nonlinear techniques.

References

[1] Banz, R. 1981. "The relationship between return and market value of common stocks." *Journal of Financial Economics*, vol. 9, no. 1 (March): 3-18.

[2] Bettis, C., D. Vickrey, and D. W. Vickrey. 1997. "Mimickers of Corporate Insiders Who Make Large-Volume Trades." *Financial Analysts Journal*, vol. 53**,** no. 5 (September/October): 57-66.

[3] Brown, S, and J. Warner. 1985. "Using Daily Stock Returns: The Case of Event Studies." *Journal of Financial Economics*, vol. 14, no. 1 (March): 3-31.

[4] Fausett, Laurene V. (1994) *Fundamentals of Neural Networks*, Upper Saddle River, NJ: Prentice-Hall.

[5] Guo, E., S. Nilanjan, D. Shome. 1995. "Analysts' Forecasts: Low-Balling, Market Efficiency, and Insider Trading." *The Financial Review*, vol. 30, no. 3 (August): 529-539.

[6] Kryzanowski, L., M. Galler, and D. Wright. 1993. "Using Artificial Neural Networks to Pick Stocks" *Financial Analysts Journal*, Vol. 49, No. 4 (July/August): 21-27.

[7] Lakonishok, J. and I. Lee. 1998. "Are Insiders' Trades Informative?", Cambridge, MA: National Bureau of Economic Research, Inc. Working Paper 6656.

[8] Moreland, J. 2000. Profit from Legal Insider Trading: Invest Today on Tomorrow's News. Chicago, IL: Dearborn Publishing.

[9] Nunn, K. P., G. P. Madden, and M. Gombola. 1983. "Are Some Insiders More Inside Than Others?" *Journal of Portfolio Management*, vol. *9*, no. 3 (Spring): 18-22.

[10] Pescatrice, D., V. Calluzzo, and M. Fragola. 1992. "Insider Trading Characteristics Offering Superior Investment Returns" *American Business Review*, vol. 10, no. 2 (June): 73-77.

[11] Rozeff, M. S and M. A. Zaman. 1988. "Market Efficiency and Insider Trading: New Evidence" *Journal of Business*, vol. 61, no. 1 (January): 25-44.

[12] Safer, A. M., B. M. Wilamowski, and R. Anderson-Sprecher. 1998. "Neural Networks for Prediction Using Legal Insider Stock Trading Data" *Intelligent Engineering Systems Through Artificial Neural Networks* **8** ANNIE'98 (Artificial Neural Networks in Engineering), St. Louis, MO., Nov. 1998: 683-689.

[13] Safer, A. M., and B. M. Wilamowski. 1999. Using Artificial Neural Networks to Predict Abnormally High Stock Returns Around Quarterly Earning Reports *IJCNN'99 (International Joint Conference on Neural Networks)* Washington, D.C., #302: 1-8,July 1999

[14] Schoneburg, E. 1990. "Stock Price Prediction Using Neural Networks: A Project Report" *Neurocomputing*, vol. 2: 17-27.

[15] Seyhun, H. N. 1986. "Insiders' profits, costs of trading, and market efficiency." *Journal of Financial Economics*, vol. 16, no. 2 (June): 189-212.

[16] Seyhun, H.N. 1988. "The Information Content of Aggregate Insider Trading." *Journal of Business*, vol. *61*, no. 1 (January): 1-24.

[17] Seyhun, H. N. 1998. *Investment Intelligence from Insider Trading.* Cambridge, Mass: MIT Press.

[18] Swales, G., and Y. Yoon. 1992. "Applying Artificial Neural Networks to Investment Analysis" *Financial Analysts Journal*, Vol. 48, No. 5 (September/October): 78-80.

Table 1. Predictor Variables Used in This Study

	Variable Name	**Description Of Variable**
1	# new holders	Number of new shareholders
2	8 week ratio of sells/buys	8 week ratio of # selling transactions/#buying transactions of the market as a whole
3	Median buy	Median individual insider shares bought relative to holdings(# shares bought/#shares held before trade)
4	Median sell	Median individual insider shares sold relative to holdings(# shares sold/#shares held before trade)
5	Avg pct increase (3 months)	look 3 months ahead and see avg of pct increase from past insiders who bought
6	Avg pct increase (6 months)	look 6 months ahead and see avg of pct increase from past insiders who bought
7	avg pct increase (9 months)	look 9 months ahead and see avg of pct increase from past insiders who bought
8	# buys	# buy transactions in period
9	# sells	# sell transactions in period
10	buy volume	# shares bought in period
11	sell volume	# shares sold in period
12	buy value	dollar value of buy transactions
13	sell value	dollar value of sell transactions