

NEURAL NETWORKS FOR PREDICTION USING LEGAL INSIDER STOCK TRADING DATA

ALAN SAFER

Department of Statistics, University of Wyoming

BOGDAN M. WILAMOWSKI

Department of Electrical Engineering, University of Wyoming

RICHARD ANDERSON-SPRECHER

Department of Statistics, University of Wyoming

ABSTRACT:

Legal insider trading is a useful source of information for forecasting stock prices. Artificial neural networks can be used to help determine relationships between insider information and future prices.

INTRODUCTION

Illegal insider trading on the stock market has received much attention in recent years. Many individuals close to high-level officers and executives in American companies have made fortunes based on their privileged information, and those who have been caught in improprieties of this type have inspired both scorn and envy in the public eye. There is, however, another type of such trading—legal insider stock trading—that is written about far less, but which can still be very profitable in its own right.

Legal insider trading is based upon information that must be made public by law. In 1962 Perry Wyson started the first newsletter dedicated to tracking the trades of insiders, *The Consensus of Insiders* (Gabele 1991). As Wyson stated, "You cannot know what they know, but you can know what they do." Now many publications devote their whole journal, or at least a section, to insider trading. Some examples of such publications include, among others, *Vicker's Weekly Insider*, *CDA\Investnet*, *Insider's Chronicle*, *The Wall Street Journal*, *Barron's*, and *Individual Investor*.

How this information is used is left to the discretion of the reader, and the only commonly available guidance for quantitative assessment is in the form of subjectively-based index numbers. The complexity of the problem is such that superior yet simple prescriptions for the use of legal insider trading information are not expected to be forthcoming.

Artificial neural network (ANN) techniques have been used successfully in a variety of applications, including some very complex problems such as speech recognition and identification of cancer cells. In business and finance neural networks have been used for stock portfolio selection and in market research. The purpose of this paper is to explore the use of the ANN approach to incorporating legal insider trading information in the prediction of stock prices. Results of

applications using historical stock data are provided for a selection of companies of various types.

LEGAL INSIDER TRADING

Who exactly is an insider in a company? Insiders consist of CEOs, vice presidents, directors, chief financial officers, and others involved in policy-making at U.S. public corporations. In addition, any shareholder owning at least 10% of a company is also considered an insider. By the 1934 Securities Act, all insiders are required to file their public transactions. Thus, anyone can find out whether the insiders are publicly buying or selling of their own company's stock, and how much they are trading.

The main source of documentation for this trading information is the Securities and Exchange Commission (SEC). Publications such as the Vicker's report contain listings of variables including insider's names and companies; numbers of shares traded, bought, or sold; and numbers of shares remaining after trading. *Vicker's* summarizes this information by creating an index number for each company, using a weighted combination of variable values with weights subjectively chosen based on experience. These index numbers are intended to indicate a rough forecast of increase or decrease in stock price. One difficulty with the Vicker's summaries is that they treat different sizes and kinds of companies in a single way; thus similar values published for two different companies in the Vicker's report could indicate different projected movements in the stock prices.

There have been relatively few statistical studies on the subject of insider trading. One of the earliest such papers involved a multivariate analysis (Finnerty 1976). Finnerty tried to determine relationships between insiders' trading and subsequent announcements of financial and accounting results. Using factor analysis, he concluded that insiders do rely on future financial and accounting information. In addition, he found that the magnitude of that information has a significant effect on buying and selling. Most other studies on insider trading use statistical tests such as t-tests to determine significance. An earlier study (Lorie 1968), which is similar to others, found that companies experiencing intensive insider buying are more likely than others to have subsequent stock price advances. Similarly, the paper found that price drops are more likely to occur following heavy selling. One study (Nunn 1983) tried to determine if the type of the insider (i.e. CEOs, directors, CFO's, etc) had an effect on price. The authors concluded that the particular type of insider is important, but differences exist for buys and sells. Subsequent studies failed to confirm these conclusions, but studies do generally show insiders are using some information that is not known to the public when trading shares in public.

STUDY DESIGN AND APPLICATION

Variable and Stock Selection

Data used in the present analysis consist of measures for various stocks subdivided according to size (market cap, that is stock price times the number of shares), and industry. The Standard and Poor's (S&P) 500, the best known group of stocks,

forms a benchmark against which traders measure their performance. The S&P 500 consists of 500 large, market leading companies. The S&P 400 midcap has leading stocks with market caps smaller than the S&P 500, and the S&P 600 smallcap has the industry leaders in the group of smallest stocks. Types of companies are delineated according to the Standard Industrial Classification (SIC) codes. Companies are broken down into about ten divisions and are subdivided into major groups within divisions.

We sampled pairs of stocks with similar industrial classifications and S&P size groupings. All three size classes and six selected industrial classes were used for this study (see Table 1). The six industrial classes used were, based on the following SIC categories: (A) Oil and Gas Extraction, (B) Food Packaging and Canned Food Companies, (C) General Merchandise Stores and Food Stores, (D) Restaurants, (E) Banks, and (F) Computer and Business Services.

Table 1. Sampling groupings by size and classification

	smallcap S&P 600	midcap S&P 400	largecap S&P 500
classification A	PPP, PROP	APA, VMC	BR, LLX
classification B	GDMK, COKE	IBP, DRYR	CPB, RAL
classification C	WFMI, CASY	FDO, DG	DH, CM
classification D	LUB, SHN	BOCB, CBRL	MCD, WEN
classification E	CFBI, KSTN	BOH, FSCO	KRB, ONE
classification F	ABM, AMSY	OLS, BID	CSC, CEN

Data analyzed consist of weekly totals from January 1993 to June 1997. The response of interest for a particular company is stock price, and the predictor variables used are selected from variables listed in Table 2. For a particular company, predictor variables with coefficient of variation near zero are not used because this condition indicates very little change, if any, of the variable over time. Variables that do not change have negligible predictive value.

Table 2. Predictor variables used for ANN

1. Number of insiders who bought shares
2. Number of insiders who sold shares
3. Number of new shareholders
4. Median buy (number of shares traded/number of shares held before trade)
5. Median sell (as above)
6. Percent of traders who bought shares
7. Total dollar value of shares bought
8. Total dollar value of the shares sold
9-13. Previous success of past buys by insiders (For 1, 2, 3, 4, and 6 months ahead of previous buys by an individual, did the stock go up or down? Record average success rate for previous buys.)
14. Number of buys over \$100,000
15. Number of sells over \$100,000
16. Total dollar value of shares bought by top insiders (i.e., CEO, CB, CFO)
17. Total dollar value of the shares sold by top insiders
18. Closing price of stock
19. S&P 500 Index,
20. Nasdaq Index

The rationale for emphasizing buying variables over selling variables is that an insider's reasons for selling are less obvious than are reasons for buying. When insiders buy, they almost certainly believe in the company to some degree. When an insider sells, the reason may be an expectation that the stock will fall in the future, but it may also be because the individual needs ready cash for personal reasons. In consequence, although selling can provide useful information, it is less reliable information than is that associated with buying, which directly indicates positive belief in the company.

Time Effects

In the present application an important question is how far back in time do data provide relevant information. Expert opinion, solicited from numerous individuals experienced in insider trading, indicates that no more than six months of information will be useful for prediction. A related question is, how far in the future is prediction plausible and interesting. It was judged that predicting stock prices two months ahead was the limit of useful forecasting.

Although the above considerations for variable selection have been nontrivial from an economic standpoint, the actual structure delineated thus far is relatively straightforward. Using a neural network with multivariate time series data involves some less obvious choices. As mentioned, the past six months of information are considered to be potentially important for predicting the direction of future price movements. The main problem with this fact is that the natural model, with lagged values for the past six months, includes far too many parameters for the number of available observations.

A method to overcome this difficulty used in some of the insider publications is to aggregate the predictors over time, that is to sum the variables together for the past six months. Prediction, however, is desired for shorter time frames, so one must shift the grouping over a small period, from one to a few weeks, and form another cumulative grouping. In this method one would wish to shift groupings as close as possible to one week so that all current information can be used in prediction. The amount of overlapping involved between groups is tremendous. A variant of the above method that is less prone to the problems of overlapping uses exponentially weighted moving averages, and assumes decaying importance of the past data the farther away the lag is from the current time period. For each input variable, i , ($i=1\dots v$), at time t , a weighted aggregate variable $x_{i,t}^*$ is created from the predictor's current value and its lagged values for $i = 1, 2, \dots, v$:

$$x_{i,t}^* = x_{i,t} + \alpha_1 x_{i,t-1} + \alpha_1^2 x_{i,t-2} + \alpha_1^3 x_{i,t-3} + \dots + \alpha_1^{L-1} x_{i,t-L+1} + \alpha_1^L x_{i,t-L} \quad (1)$$

The problem with this technique is selection of values for the α 's. The α 's should be between 0 and 1, but arbitrary selection at, for example 0.2 or 0.5 is expected to yield unnecessarily inferior results. A better way is to optimize the α 's in relation to the weights of the neural network.

The procedure used below is an iterative process: First set the α 's to some initial value, random or preselected, between 0 and 1. Then, find the best weights in the ANN when using these α 's, according to minimum mean squared error. Next

optimize the α 's given the weights, again according to least squares. Continue, alternately improving weights and α 's until values stabilize.

After the best α 's and weights are found, future stock prices can be predicted. The number of input/output pairs is still relatively low (about 200, even though the time period is over 4 years). To obtain a measure of reliability for the prediction, different random starting weights are used and the results of training for a number of runs are compared. In our case the training set used was the entire data set except for the final input/output pair. Using this training set, 20 iterations were run for each stock. The repeated runs then provide rough confidence intervals for the forecast, that is, for the conditional expectation of the latest value based on previous values. Note that this interval is merely a confidence interval, not a prediction interval for the next observation.

The aforementioned method for approaching this problem assumed decaying importance of the past data the farther the lag is from the current period. This may or may not be true. Some insiders may see the company doing well in the near, but not immediate, future and believe the best time to get in is early. Consequently, there may be delays between buying activity and positive news, and thus increase in stock price in the company.

Another way to deal with the x_i^* 's is directly through a special type of neural network and without additional optimization techniques (see figure 1). Using the custom neural network shown in the figure, the weights connecting each variable to the next layer are forced to be the same for similar time periods. This then accomplishes the same goal of the former method of having the same weights for each lag period. However, the difference is that there is no assumed pattern for the weights, such as exponential decay, that was forced into the previous equations.

ANN Architecture

The architecture of the first proposed ANN is relatively simple and involves an input layer, an output layer, and one hidden layer in between. The input layer includes N predictor variables (usually around 15 variables, depending on the stock), and the output layer involves the price of the stock 2 months in the future. The hidden layer consists of 4 neurons. The relationship between the inputs and the output is modeled as a nonlinear relationship, and each neuron in the hidden layer thus has a nonlinear transfer (activation) function. The hyperbolic tangent sigmoid transfer function is used in this network.

To select the number of neurons in the hidden layer, several trial stocks were used. The number of neurons was changed from one to eight in order to test the accuracy of the predictions, and, after the trials, it was decided that four neurons were adequate to produce accurate responses without causing overfitting. More neurons in the hidden layer may lead to overfitting of the training set, in which case the network is unable to generalize well to new input data (Cheng 1994).

The output layer involves a linear activation function since the output is continuous. The algorithm used for estimation in the neural network is the Levenberg-Marquardt algorithm. This training method is among the fastest available, but it uses the matrix of second derivatives, so the memory requirements can be quite high. Using Levenberg-Marquardt is another reason for reducing the number of inputs.

The architecture of the second proposed ANN (figure 1) is a modified version of the first architecture. The chief difference is that there is an extra layer of linear transfer functions placed between the input layer and the original hidden layer in architecture 1. In addition, there are now seven neurons in the second hidden layer with sigmoidal transfer functions. Moreover, the weights from the input layer to the first hidden layer are set so that the weights are constant for each time lag across variables. The rest of the architecture stays the same.

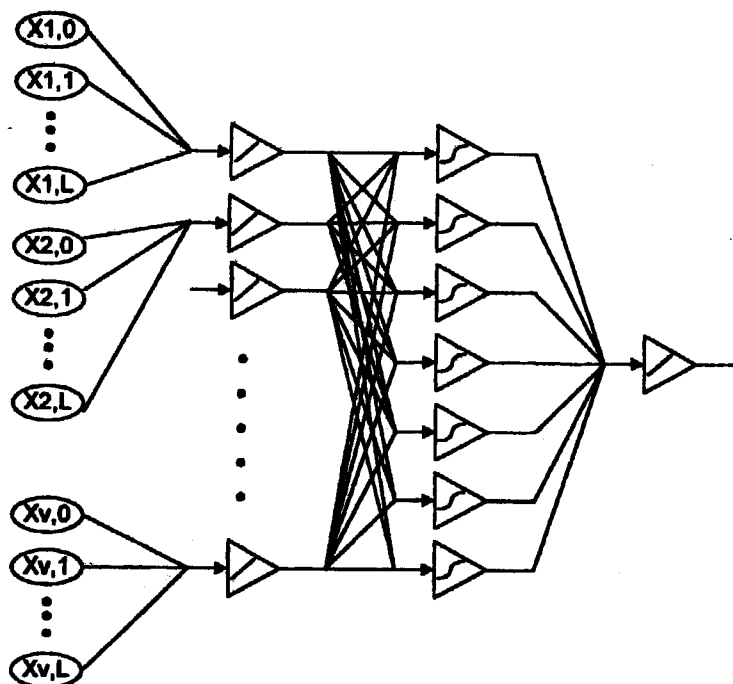


Figure 1: Modified ANN architecture

RESULTS

The 36 stocks were forecast two months from the latest insider trading data. Partial results are shown in Table 3. Three stocks from the S&P 400 midcap were about 10% or less absolute difference from the estimated value. Similarly, there were four S&P 500 largecap stocks and five S&P 600 smallcap stocks that were around this precision of accuracy. Consequently, it seems from this sample of stocks that the market cap size of the stock does not have a major influence on the accuracy of the prediction using the insider trading data and this network. However, when one looks at the classifications, differences do appear. There were six different classifications (A-F) used, each used twice for each of the three market sizes. There were 12 stocks that were around 10% accuracy. If the classifications had an equal chance of occurring, one would expect to see about two of each classification (12 stocks/6 classifications). However, there were five stocks of classification B (Food Packaging and Canned Food Companies). Why this particularly industry classification came out more in the results than the others is not clear. However, some industries are probably easier to forecast with insider trading data than are industries in general.

Table 3. Stock values that were close to 10% or less different from estimate

% difference in estimated output and actual value	average absolute value % difference	Stock Symbol	Classification	S&P GROUP
-0.02	1.05	IBP	B	400
-8.71	8.29	BOCB	D	400
-7.96	7.63	BOH	E	400
-6.77	10.91	BR	A	500
-5.52	5.24	CPB	B	500
-12.14	11.19	RAL	B	500
-2.82	4.12	WEN	D	500
9.13	9.41	PROP	A	600
-5.18	4.97	GDMK	B	800
-5.48	7.82	COKE	B	600
-0.95	0.62	KSTN	E	600
7.27	4.64	AMSY	F	600

CONCLUSIONS

Insider trading stock data is accepted by the financial community as information that can be utilized to predict stock prices. However, most stock analysts don't exactly know how to use the insider trading numbers mathematically and try to use their experience in looking at insider trading data to help them predict stock prices. Neural networks appear to be a technique worth exploring more in this particular time series field. In addition, nonstandard types of neural network architectures, such as the one used in this paper, may help in the prediction. In these results, there appears to be a relationship between industry classification and predication accuracy in forecasting stocks using insider trading data. Market size did not show a significant relationship to forecasting. In terms of the future of insider trading, as the internet expands, information on insider trading will be more accessible. Consequently, more people will be able to take advantage of this useful stock information and can use techniques like neural networks to help forecast.

REFERENCES

- Cheng, B., Titterington, D., (1994). "Neural Networks: A Review from a Statistical Perspective" *Statistical Science*, Vol. 9, No. 1, pp. 2-54.
- Finnerty, J., (1976) "Insiders and Market Efficiency" *The Journal of Finance*, Vol. 31, No. 4, pp. 1141-1148.
- Gabele, B., (1991) "Insider Transactions: Should You Buy When They Buy, Sell When They Sell?" *Barron's*, May 13, pp. 12-13.
- Lorie, J., Niederhoffer, V., (1968) "Predictive and Statistical Properties of Insider Training" *Journal of Law and Economics*, Vol. 11, pp 35-51.
- Nunn, K., Madden, G. and Gombola, M. (1983) "Are Some Insiders More 'Insider' Than Others?" *Journal of Portfolio Management*, Vol. 9, pp. 18-22.