Using Neural Networks to Predict Abnormal Returns of Quarterly Earnings

Alan M. Safer and Bogdan M. Wilamowski
University of Wyoming
safer@uwyo.edu,mailto:safer@uwyo.edu wilam@ieee.org

Abstract

Artificial neural networks are used in conjunction with the Sharpe-Linter form of the Capital Asset Pricing Method (CAPM) to predict when the returns on U.S. stocks will be greater than financial risk models would predict. The advantage of using a nonlinear approach is to model the financial system more accurately than linear techniques. The Sharpe-Lintner form is used to control for risk and determine abnormal returns of stocks. Inputs include ratios of recent to past stock price averages over pre-event time periods, similarly, stock volume ratios, and previous quarter standardized unexpected earnings (SUE). The earnings data is quarterly and runs from the first quarter of 1993 to the second quarter of 1998. Event periods that had the smallest width around the earnings report tended to be easier to predict abnormal returns. In addition, event periods that were closest to the event (the earnings report) were more accurate at predicting the abnormal returns of stocks.

Introduction

Artificial Neural Networks (ANN) are commonly used today in many financial areas such as bankruptcy prediction, credit scoring of loan applications and bond rating analysis [12]. Another very popular way to use ANN's is in helping to predict stocks [13]. Because of the potential financial rewards with being able to accurately predict future stock prices, many people associated with Wall Street have started learning about this relatively recent non-linear technique that has quickly branched beyond it's core applications in engineering. Many mutual funds are now picked using neural networks. In fact, there are several internet sites that use it to help day traders pick their stocks better and quicker using lots of very current information. In this paper, an aspect of predicting stocks is explored using ANN's. Artificial neural networks are used in conjunction with a popular financial technique, Capital Asset Pricing Method, to help predict earnings surprise.

Data

The Data in this study involves all U.S. stocks having quarterly earning reports from the first quarter of 1993 to the second quarter of 1998. I would like to thank First Call Corporation for supplying me with the quarterly earnings data. The data from First Call included the following: information: the reported quarterly earnings date, the stock symbol, the mean of the analysts' quarterly estimates, the actual quarterly earnings, the number of analysts estimating the earnings, the standard deviation of the analysts' estimates, and the Standardized Unexpected Earnings (SUE).

CAPM Method

An event study method similar to the one by Brown and Warner [5] was used. The Sharpe-Lintner form of the Capital Asset Pricing Method (CAPM) was incorporated to control for risk and determine abnormal returns of stocks [8]. The method begins with a pre-event period consisting of an interval of time starting j business days before the earnings announcement (t-j where the earnings announcement is t-0) and going back approximately 1 year before the announcement, or t-265. The event period is the interval from one day after the last pre-event day (t-j+1 or t-k1). includes the earnings day, and goes past it k2 days (t+k2). The procedure is as follows:

Part 1) First, estimate a_i and β_i (a measure of the systematic risk of an asset) using multiple regression from the pre-event period (t-265 to t-j). This is done using the following equation:

$$(R_{it} - r_{ft}) = a_i + \beta_i * (R_{mt} - r_{ft}) + \varepsilon_{it}$$

CAPM Equation 1

$$\left[\ln\left(\frac{P_{i,t-j}}{P_{i,t-j-1}}\right)-\left(\left(\frac{100}{100-\frac{tbill_{t-j-1}}{4}}\right)^{\frac{1}{90}}-1\right)\right]_j =$$

$$a_i+\beta_i\left[\ln\left(\frac{P_{m,t-j}}{P_{m,t-j-1}}\right)-\left(\left(\frac{100}{100-\frac{tbill_{t-j-1}}{4}}\right)^{\frac{1}{90}}-1\right)\right]_j+\varepsilon_{i,t}$$

where:
R_{it} is the pre-event return on stock i for day t, r_{ft} is the 3 month daily tbill rate, R_{mt} is the pre-event return of the market, ε is the error, t is the pre-event period from t-265 to t-j (t-0 is the event day, i.e. earnings report day)

Part 2) Next, use a_i and β_i obtained from analysis of the pre-event period to determine the abnormal returns over the event period
$\varepsilon_i = R_{iT} - r_{fT} - a_i - \beta_i * (R_{mT} - r_{ft})$

CAPM Equation 2

$$\varepsilon_i=\left[\ln\left(\frac{P_{i,T+k2}}{P_{i,T-k1}}\right)-\left(\left(\frac{100}{100-\frac{tbill_{T-k1}}{4}}\right)^{\frac{k1+k2}{90}}-1\right)\right]_k$$

$$-a_i-\beta_i\left[\ln\left(\frac{P_{m,T+k2}}{P_{m,T-k1}}\right)-\left(\left(\frac{100}{100-\frac{tbill_{T-k1}}{4}}\right)^{\frac{k1+k2}{90}}-1\right)\right]_k$$

The event period, T, is from t-k1 to t+k2. The input variables used to try to predict the abnormal returns included four ratios of price averages and similarly four ratios of the stock volumes. The ratio periods of averages for both the stock prices and stock volumes involved the pre-event periods. The four numerators of the ratio averages involve a period of 3, 5, 10 and 15 days before the event period, and the four denominator averages encompass 30 days right before the numerator period. Consequently, there are eight total ratio averages.

In addition to the ratio variables, the Standardized Unexpected Earnings (SUE) of the past quarter are used to predict the abnormal returns of the event period. The SUE relays how far the actual quarterly earnings was away from the estimated earnings value adjusted for the variation in the estimates [3]. Many studies have shown that there is a positive correlation between stocks that had a positive SUE in the previous quarter and the current quarter, and that stocks that had a negative SUE in the last quarter are likely to have a negative one in the current quarter [1,4]. One continuously successful stock, Microsoft, beat its earnings estimate 41 out of its first 42 quarters. The likely increase in stock price after beating an earnings estimate is obviously taken very seriously by companies. Through most of the time involved in this individual study, from the first quarter of 1993 to the first quarter of 1997, a record 16 consecutive quarters saw more S&P 500 companies beat the consensus earnings estimates than did not [7].

In addition to the studies on similar positive or negative earnings in consecutive quarters, one large study showed there is almost a 6% difference in three-month returns between companies with high and low SUE's [10]. In terms of this study, if the previous quarter had just one analyst or had no variation in the estimates, the previous SUE was used. If a similar pattern happened to the previous SUE, then this earnings estimate was excluded from the model. However, this particular occurrence of no variability among the quarterly estimates of the analysts of the stocks occurred only about 10% of the time.

The pre-event periods and their corresponding event-periods were as follows:

Table 1. Pre-event and Event Periods

Period	Pre-event period	Event period
1	t-265 to t-16	t-15 to t+5
2	t-265 to t-11	t-10 to t+5
3	t-265 to t-6	t-5 to t+5
4	t-265 to t-6	T-5 to t+ 3
5	t-265 to t-3	T-2 to t+ 3

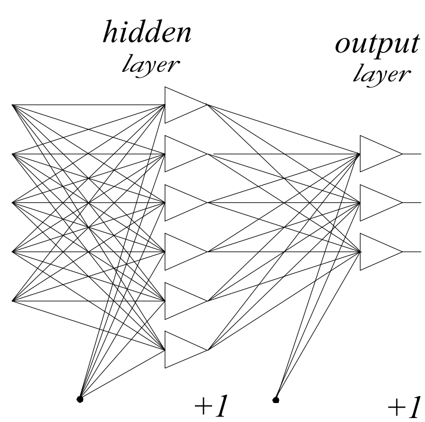

Fig.1. An example of the three layer feedforward neural network, which is also known as the backpropagation network.

Artificial Neural Networks

The simplest and most common neural networks use only one directional signal flow. Furthermore, most of feedforward neural networks are organized in layers. An example of the three layer feedforward neural network is shown in Fig. 1.

The feedforward neural network is used for nonlinear transformations (mapping) of a multidimensional input variable into another multidimensional output variable. In theory, any input-output mapping should be possible if neural network has enough neurons in hidden layers (size of output layer is set by the number of outputs required). Practically, it is not an easy task and presently, there is no satisfactory method to define how many neurons should be used in hidden layers. Usually this is found by trial and error method. In general, it is known that if more neurons are used, more complicated shapes can be mapped. On the other side, networks with large number of neurons lose their ability for generalization, and it is more likely that such network will try to map noise supplied to the input also.

Weights in artificial neurons are adjusted during a training procedure. Various learning algorithms were developed, but only a few are suitable for multi-layer neuron networks. The error backpropagation algorithm (EBP) was a significant breakthrough in neural network research, but it is also known as an algorithm with a very poor convergence rate. The LM algorithm is now considered as the most efficient one [9] . It combines the speed of the Newton algorithm with the stability of the steepest decent method.

The Levenberg-Marquardt learning algorithm [9] is the second order search method of a minimum. At each iteration step, the error surface is approximated by a parabolic approximation and the minimum of the paraboloid is the solution for the step. Simples approach require function approximation by first terms of Taylor series

$$F(\mathbf{w}_{k+1}) = F(\mathbf{w}_k + \Delta\mathbf{w}) + \mathbf{g}_k^T \Delta\mathbf{w}_k + \frac{1}{2}\Delta\mathbf{w}_k^T \mathbf{A}_k \Delta\mathbf{w}_k$$

where $\mathbf{g} = \nabla E$ is gradient and $\mathbf{A} = \nabla^2 E$ is Hessian of global error E.

$$Gradient \Rightarrow \frac{\partial E}{\partial w_1}, \ \frac{\partial E}{\partial w_2}, \ \frac{\partial E}{\partial w_3}, \ \cdots$$

$$Hessian \Rightarrow \begin{matrix} \frac{\partial^2 E}{\partial w_1^2} & \frac{\partial^2 E}{\partial w_1 \partial w_2} & \frac{\partial^2 E}{\partial w_1 \partial w_3} & \cdots \\ \frac{\partial^2 E}{\partial w_2 \partial w_1} & \frac{\partial^2 E}{\partial w_2^2} & \frac{\partial^2 E}{\partial w_2 \partial w_3} & \cdots \\ \frac{\partial^2 E}{\partial w_3 \partial w_1} & \frac{\partial^2 E}{\partial w_3 \partial w_2} & \frac{\partial^2 E}{\partial w_3^2} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{matrix}$$

Steepest decent (error backpropagation) method calculates weights using:

$$\mathbf{w}_{k+1} = \mathbf{w}_k - a \ \mathbf{g}$$

while Newton method uses:

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \mathbf{A}_k^{-1}\mathbf{g}$$

The Newton method is practical only for small networks where Hessian $\mathbf{A}_k$ can be calculated and inverted. In the Levenberg-Marquardt method the Hessian $\mathbf{A}_k$ is approximated by product of Jacobians

$$\mathbf{A} \approx 2\mathbf{J}^T\mathbf{J}$$

and gradient as

$$\mathbf{g} \approx 2\mathbf{J}^T\mathbf{e}$$

where **e** is vector of output errors and Jacobian **J** is

$$Jacobian \Rightarrow \begin{matrix} \frac{\partial E_1}{\partial w_1} & \frac{\partial E_1}{\partial w_2} & \frac{\partial E_1}{\partial w_3} & \cdots \\ \frac{\partial E_2}{\partial w_1} & \frac{\partial E_2}{\partial w_2} & \frac{\partial E_2}{\partial w_3} & \cdots \\ \frac{\partial E_3}{\partial w_1} & \frac{\partial E_3}{\partial w_2} & \frac{\partial E_3}{\partial w_3} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{matrix}$$

It is much easier to calculate the Jacobian than the Hessian. Also, usually the Jacobian is much smaller so less memory is required. Therefore weights can be calculated as

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \left(2\mathbf{J}_k^T\mathbf{J}_k\right)^{-1} 2\mathbf{J}_k^T\mathbf{e}$$

or

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \left(\mathbf{J}_k^T\mathbf{J}_k\right)^{-1} \mathbf{J}_k^T\mathbf{e}$$

To secure convergence the Levenberg-Marquardt introduces μ parameter

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \left(\mathbf{J}_k^T\mathbf{J}_k + m\mathbf{I}\right)^{-1} \mathbf{J}_k^T\mathbf{e}$$

when μ = 0. This method is similar to the second order Newton method. For larger values of μ parameter the Levenberg-Marquardt works as the steepest decent method with small time steps. The m parameter is automatically adjusted during computation process so good convergence is secured. The Levenberg-Marquardt recently became very popular because it will usually converge in 5 to 10 iteration steps. The main disadvantage of this method is a large memory requirement.

The activation function used in the hidden layer was sigmoidal. The final layer had a linear activation function. The mean squared error was used to minimize the error. The data was broken up into two groups. The first random group included 80% of the data and was used for training the network and to come up with weights. The other 20% were used for validation.

In all, there were nine input variables (four price ratio averages, four volume ratio averages, and one previous

SUE) and one output variable (abnormal return for the event period). The number of inputs was held reasonably small because of the so called curse of dimensionality [2]. That is, the number of points needed to model grows in general by 2**N where N is the number of inputs. The number of inputs should still be kept relatively small even though, in this study, the number of input patters was very large. The whole group of input patterns were applied to each of the different five pre-event/event periods that were used in the study. The network was tested with and without the previous SUE's for each of the different pre-event periods.

The number of hidden units in the one hidden layer that seemed to work the best was five. The number of earnings estimates, and consequently, the number of input patterns evaluated, was very large. One of the major advantages of using artificial neural networks is indeed their capacity to model extremely large amounts of information such as used in this study. Most other studies dealing with quarterly earnings deal with only small samples (about 400 to 1,000 stocks) [10]. Consequently, these did not take full advantage of the benefits of neural networks.

Results

The results show that for the shorter interval width of the last day of the pre-event period to the actual earnings date (t-0), the lower the mean squared error. In addition, the smaller the width of the event period, the mean squared error is also decreased.

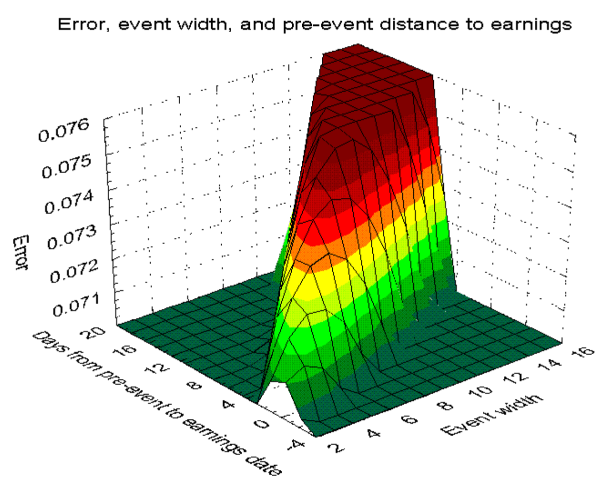

Figure 2. Relationship between the mean square error and the width of the interval and closeness to the event day

In terms of the SUE, running the neural network with and without previous SUE's did not produce significantly less reduction of a mean squared error.

Conclusions

Artificial Neural Networks used in conjunction with a financial technique such as the Sharpe-Litner form of the Capital Asset Pricing Method work well together to increase accuracy of prediction. Other possible study might include using other non-linear non-parametric techniques such as MARS, Multivariate Adaptive Regression Splines [6]. This technique may also work well together with other traditional financial techniques. In addition, including the size of the firm as an input variable may help increase the accuracy of the prediction [9].

References

[1] Ball, Ray. "The Earnings-Price Anomaly" *Journal of Accounting and Economics*, Vol. 15 (1992): 319-345.

[2] Bishop, C. Neural Networks for Pattern Recognition. Oxford University Press (1995).

[3] Brown, Lawrence. "Earnings Surprise Research: Synthesis and Perspectives" *Financial Analysts Journal,* March/April (1997): 13-19.

[4] Brown, Melissa. "The Anatomy of an Earnings Surprise" Handbook of Security Analyst Forecasting and Asset Allocation Contemporary Studies in Economics and Financial Analysis, Vol. 69. (1993) 13-25.

[5] Brown, Stephen J., and Jerold B. Warner. "Using Daily Stock Returns: The Case of Event Studies." *Journal of Financial Economics* Vol. 14 (1985): 3-31.

[6] De Veau, R. D., , D. C Psichogios., and L. H. Ungar. "A Comparison of Two Nonparametric Estimation Schemes: MARS and Neural Networks" *Computers and Chemical Engineering*, Vol. 17, No.8, pp. 819-837 (1991): 819-837.

[7] Fox, Justin. "Learn to Play the Earnings Game" *Fortune, March 31* (1997): 77-80.

[8] Guo, Enyang, Nilanjan Sen and Dilip K. Shome. "Analysts' Forecasts: Low-Balling, Market Efficiency, and Insider Trading." *The Financial Review*, Vol. 30, No. 3 (1995): 529-539.

[9] Hagan M. T. and M. Menhaj, "Training feedforward networks with the Marquardt algorithm," IEEE Transactions on Neural Networks, vol. **5**, no. 6, (1994) 989-993.

[10] Mott, Claudia and Daniel P. Coker. "Earnings Surprise in the Small-Cap World" *The Journal of Portfolio Management, Fall (1993):* 64-75.

[11] Rendleman, Richard J. Jr., Charles P. Jones, and Henry A. Latane. "Empirical Anomalies Based on Earnings'' Yields and Market Values." *Journal of Financial Economics*, Vol. 10, No. 3, (1982): 269-287.

[12] Swales, George S. Jr, and Young Yoon. "Applying Artificial Neural Networks to Investment Analysis" *Financial Analysts Journal*, September-October (1992): 78-80.

[13] Wong, Shee Q., and J. Allen Long. "A Neural Network Approach to Stock Market Holding Period Returns" *American Business Review*, June (1995): 61-64.