Note: This paper has been published in **International Journal of Intelligent Systems in Accounting, Finance, and Management,** vol. 3, 241-252, 1994.

A Comparative Analysis of Artificial Neural Networks Using Financial Distress Prediction.

Kurt M. Fanning and Kenneth O. Cogger

March 1994

School of Business
University of Kansas
Summerfield Hall
Lawrence, Ks. 66045-2003, USA.

Tel:(913)-864-7554
Fax:(913)-864-5328
Internet:cogger@kuhub.cc.ukans.edu

A Comparative Analysis of

Artificial Neural Networks Using

Financial Distress Prediction.

Kurt M. Fanning[1] and Kenneth O. Cogger[2]

School of Business, University of Kansas, Lawrence, Ks. 66045-2003. Tel(913)-864-7554, Fax: 913-864-5328, Internet: cogger@kuhub.cc.ukans.edu

Abstract

This paper examines the efficiency of a generalized adaptive neural network algorithm (GANNA) processor in comparison to earlier model based methods, a back propagation artificial neural network, and logistic regression approaches to data classification. The research uses the binary classification problem of discriminating between failing and non-failing firms to compare the methods. The results indicate the potential in time savings and the successful classification results available from a GANNA processor.

[1]Doctoral student in accounting
[2]Professor of Management Science.

I. INTRODUCTION

The literature on classification techniques is vast. One subsection of this area is the classification of failed and non-failed firms. The various ways of determining the potential for failure of corporations using past financial data is a well-documented topic. Foster (1978), Altman (1983), Zavgren (1983), Jones (1987), and Boritz (1991) provide comparisons and summaries of the major extant methods and the empirical results of tests of these methods. The documented performance of a wide variety of competing models for distress prediction makes this arena suitable for testing new methods. This paper will examine the potential of a generalized adaptive neural network algorithm (GANNA) processor in a financial distress prediction context.

Most proposed models for failure prediction have done reasonably well for distinguishing (in sample) between failed and non-failed firms. Hamer (1983) in comparing several suggested models found that although the models differed, their discriminating power was comparable. However, no method has been totally effective and no satisfactory theory has been developed to explain differences between failed and non-failed firms. Therefore, this issue remains of academic interest and the results of this study indicate an additional tool for this research.

It must be recognized also that non-failed firms statistics are far more plentiful than those for failed firms. This results in most published studies using validation samples with higher proportions of failed firms than extant in practice in order that a statistically meaningful number of firms of both kinds be subject to prediction. To assess the relative performance of all model most completely, we have utilized a wide variety of performance measures. These include total misclassification rates when firms are matched by industry as well as unmatched and are further compared separately for Type I and Type II errors. These results are further broken down by number of years to failure. Also, to account for differential costs of the two types of error as well as a prior probabilities and validation sample composition, we report the relative resubstitution risks for each model under varying assumptions regarding the relative costs of Type I errors.

The most popular method of bankruptcy prediction has been subject to criticism for methodological reasons. Certain aspects of the use of linear discriminant analysis have been criticized in Joy and Tollefson (1975), Eisenbeis (1977), Ohlson (1980), Pinches (1980), Richardson and Davidson (1983), Zmijewski (1984), Zavgren (1985) and Lo (1986). Common forms of linear discriminant analysis implicitly assume normality of variables. Another common critical assumption is that the group dispersion (variance-covariance) is equal across groups. Eisenbeis (1977) suggests that both assumptions are probably violated in most samples used in empirical research since many measurements are of nominal or ordinal nature at best .

Logistic regression (Logit) does not suffer from many of these problems but Hamer (1983) suggests its ability to discriminate between failed and non-failed firms has not proven statistically superior to linear discriminant analysis.

Zmijewski (1984) suggests that the matched pairs design of most of the prior empirical work on bankruptcy prediction may lead to biased results. This is the result of sampling a larger proportion of failed firms than exist in the general population.

It has been proposed by Bell et al. (1990), Coleman et al. (1991), Coats and Frank (1991-92), Salchenberger et al. (1992) and Tam and Kiang (1992) that Artificial Neural Networks (ANN's) may be able to overcome these difficulties and differentiate between failed and non-failed firms more effectively than logistic regression and linear discriminant analysis. ANN's have long been recommended for tasks of pattern recognition and categorization. ANN's success in other fields such as speech recognition, control systems and visual pattern recognition may indicate its potential for use in economic research. For a summary of ANN's success in other fields see Schalkoff (1992).

Another major issue in the use of ANN's in research is the effectiveness of certain types of ANN's in comparison to others. The number of competing ANN's are numerous and growing rapidly. For a summary of many of the different types see Gallant (1993) or Zahedi (1993). In Cogger (1992) the benefits of the generalized adaptive neural network algorithm (GANNA) are discussed. It is suggested that GANNA minimizes the labor involved in constructing a neural network while achieving results comparable to more traditional ANN's. Much of the difficulty in the use of ANN's lies in the trial and error approach required of most current ANN's. The GANNA method does not suffer from this requirement. It also provides a specific ending point that indicates an optimum configuration for the ANN. Both factors contribute to its potential use in classification and other endeavors. The architecture of the neural network is determined by the nature of the data and software based on the conceptual framework of GANNA provides a fast and simple means of constructing networks.

This paper compares a standard back propagation ANN with a GANNA processor using financial distress prediction. Performance of these two networks is measured relative to both model based predictors and logistic regression. The next section will review the prior literature specifically related to bankruptcy prediction. This will be followed by a discussion of ANN's. The design of the empirical study is then described. The results of five different approaches (Gambler's ruin, Wiener process, Logistic regression, ANN, and GANNA) to the classification of failed and non-failed firms are presented next. Finally, a discussion of the results, conclusions, and prospects for future research is provided.

II BANKRUPTCY PREDICTION

In the bankruptcy prediction literature several methods have been used successfully in the past. First, there are model based methods. Among the more successful of these are the approaches of Wilcox (1971, 1973), based on the Gambler's Ruin model, and

Emery and Cogger (1982), based on a Wiener process. These approaches utilized results from the theory of stochastic processes. Under certain assumptions, mathematical expressions may be derived for the probability of financial distress occurring over a particular time span. Such probabilities require the estimation of critical parameters such as the level of liquid assets, the volatility of cash flows, the mean cash flow rate, the time span of interest, etc.

There are advantages and disadvantages to such model based approaches. The advantage of such approaches is that they specify a priori the mathematical function to be employed. Criticisms of "data mining" and "fishing expeditions" cannot be made because the data on bankruptcy per se is not used to develop the mathematical expression for the probability of financial distress. The models may therefore be validated in a true ex ante fashion. A potential disadvantage of such models is that their supporting assumptions may not be satisfied. As an example, the stochastic process models cited above assume that successive (e.g., daily) cash flows are independent random variables. Some empirical evidence for this assumption is given in Cogger and Emery (1983) and some practical reasons for this property to exist in a well-managed firm can also be given. The literature on stochastic processes is vast, but readers interested in further reading in this area might peruse the works by Karlin and Taylor (1975), Cox and Miller (1980), and Heyman and Sobel (1982).

Another method employed in financial distress prediction is statistical analysis, using a variety of financial variables and ratios for failed and non-failed firms. Discriminant analysis is a common statistical analysis for determining which variables are most significant in predicting financial distress. Many studies, for example Altman (1968), Deakin (1972) and Altman et al. (1977), used linear discriminant analysis. The advantage of such approaches is that a stochastic model is not imposed a priori. Relevant variables, such as cash flows, volatilities, etc. are identified by the analysis if they are included by the investigator in the analysis. A disadvantage is obvious: if a variable is not thought to be important, or not even envisioned as being potentially relevant, the appropriate discriminant function cannot be discovered. Another disadvantage is that the functional form of the discriminant function must be chosen by the investigator. Usually linear functions are employed, creating a potential for model misspecification. A final limitation arises from the ex post nature of these analyses. The estimation and validation sample both have firms that are known to have failed on particular dates. A true ex ante analysis of predictive ability would be based on currently non-failed firms with predictions about the timing as well as the occurrence of financial distress. For detailed treatment of discriminant analysis, the reader may refer to Lachenbruch (1975).

It should be noted that linear and quadratic discriminant functions are special cases of GANNA as described in Cogger (1992) and their performance is not studied in this paper. Our comparison of GANNA with model based approaches implicitly subsumes these two common discriminant functions.

Logistic Regression (Logit) is another method that has been applied to the problem of bankruptcy prediction. Logit does not suffer from the problems associated with linear discriminant analysis such as the assumption of normality and similar dispersion measures. Press and Wilson (1978), Ohlson (1980), Zavgren (1985) and Lo (1986) suggest that Logit should be used when dealing with qualitative variables. A discussion of the Logit and Probit models can be found in Maddala (1983, 1991). The Logit model was used for comparison in this paper since it is the most prevalent and Gessner et al. (1988) found that Logit and Probit results were similar across various conditions. This similarity is not surprising given that the logistic and cumulative standard normal functions used in Logit and Probit, respectively, are nearly identical except in the extreme tails. The relative computational simplicity of the logistic function thus favors its choice.

Many other methods have been suggested for bankruptcy prediction. Messier and Hansen (1988) and Tam and Kiang (1992) recommend Inductive Dichotomizer 3 (ID3). Tam and Kiang (1992) propose the *k* nearest neighbor (*k*NN) approach. Marais et al. (1984) and Fryman et al. (1985) recommend the Recursive Partitioning Algorithm. Several mathematical programming methods, summarized in Erenguc and Koehler (1990), can also be applied to classification problems.

Thus many methods have potential as classification approaches. However, since the purpose of this paper is to suggest the possible potential of GANNA, a thorough examination of all methods and variables is unnecessary. For comparative purposes, we have chosen several methods with very high reported accuracy in bankruptcy prediction. Based on reported results in the literature, it is unlikely that methods not chosen for this study would achieve significantly higher levels of accuracy in our sample. The computational simplicity of GANNA further promotes its consideration relative to more complex methods in future research endeavors.

III. ARTIFICIAL NEURAL NETWORKS

In this paper, we will compare two artificial neural networks (ANN's) methods to established methods using the financial distress arena. Use of ANN's in accounting and financial research has been growing rapidly over the last several years. ANN's have been used successfully in many accounting and financial applications. Some applications include the following studies. Dutta and Shekhar (1988) and Surkan and Ying (1991) have applied ANN's to bond ratings. Collins et al. (1988) applied an ANN to mortgage underwriting judgments. White (1988) examined the efficient market hypothesis using an ANN. Hansen et al. (1992) used an ANN to distinguish between qualified and unqualified audit opinions and litigated and non-litigated firms. Yoon et al. (1993) found information using an ANN in forecasting stock price movements. Bell et al. (1990) and Salchenberger et al. (1992) employed an ANN in examining thrift failure. Coats and Frank (1991-92), Tam and Kiang (1992) and Fletcher and Goss (1993) used ANN's to predict bankruptcy. All these papers indicate that ANN's have potential to achieve greater discriminating power than alternative methods.

The ANN method may be defined as an approach to computing based on mathematical models inspired by biological theories of information processing. Typically, computing is accomplished by an ANN with simple processing elements connected in several layers. Each processing element is analogous to a neuron, the fundamental cell of any central nervous system. The neuron was first mathematically modeled by McCulloch and Pitts (1943). Thus, ANN's are not new. However, recent developments have encouraged their use in a variety of applications, and this paper explores their performance in the financial distress area relative to the traditional approaches described in the previous section. ANN's are often suggested for problems in pattern recognition, and in areas where a model based theory is not available. Hawley et al. (1990) point out that ANN's are most effectively applied to classification, associative memory and clustering. Financial distress prediction, if one is unwilling to accept stringent assumptions about stochastic processes, is arguably such an area. ANN's are constructed using learning paradigms that are analogous to statistics. The commonly employed backward propagation algorithm described in Rumelhart et al. (1986) is an iterative least squares procedure applied to the connections in an ANN. The back propagation network is normally a feed forward network that modifies its errors backwards. The training data is presented to the network repeatedly with the model adjusting the errors using the gradient descent method to minimize the errors.

For our purposes, it is convenient to think of an ANN as a statistical model that does not impose an a priori model on the data; unlike linear discriminant analysis, no limitations are imposed on the complexity of the predictive model. One disadvantage of this approach is that the investigator must decide upon the physical architecture of the network. This is often done by trial and error, varying the number of layers, the number of processing elements in each layer, the nature of the connection patterns, etc. Some heuristics, see for example Caudill (1991) and Zahedi (1993), exist to guide the network architect. Simply connected networks of two or three layers of sigmoidal (logistic) processing elements combined with standard back propagation estimation of connection weights have performed quite well in many settings. This was our choice for the present study. There are many good references to ANN's. Perhaps the single most comprehensive treatment is Rumelhart et al. (1986).

The fifth method examined in this paper is generalized adaptive neural network architectures (GANNA). This method develops an ANN with a particular type of processing element where the architecture of the network is itself learned by trial and error. Described in detail in Cogger (1992), these networks grow to fit the problem at hand and do not require a priori specification of the number of layers, numbers of nodes, etc. The particular implementation of GANNA employed in this paper uses processing elements which compute quadratic functions of their inputs. With many such elements connected together, of course, there is no restriction on the functional form of the input-output function. Linear and quadratic functions, for example, are special cases of a GANNA network. Briefly, this approach uses an evolutionary mechanism to grow networks. New layers are added provided they offer improved performance, measured in squared output error. To protect against overspecialization, performance is validated on a hold-out sample. When performance in this sample declines, network evolution ceases.

All computing was performed on standard i486 processors. We used NeuralWare Professional(C) from NeuralWorks, Inc. for development of the back propagation ANN. There are many other commercial packages available, although this appears the most popular at this time. We used AutoNet(C) from Peak Software, Inc. for the development of adaptive neural network architectures. AIM(C) from Abtech, Inc. is the only other similar software package that we are aware of. Cogger (1992) develops the conceptual framework for generalized adaptive neural network architectures. AutoNet(C) was developed following this framework. AIM(C) is quite similar but detailed descriptions of its algorithm are unavailable. Since these approaches are novel, they are included in these study to introduce them to a wider audience. The Statistical Analysis System (SAS) provided the logistic regression results.

IV STUDY DESIGN

One of the few proposed models for bankruptcy prediction is the gambler's ruin model. This was adapted from Feller (1968). A summary of this model is found in Altman (1983). The basic premise is that a company's future liquidity position is a function of its current liquidity, its cash flow, and the variance of the cash flow. In Wilcox (1973), the author tested this theory on a sample of matched pairs of 52 failed firms with 52 non-failed firms from one to five years prior to failure. Firms were matched in terms of size and industry characteristics. This sample was also used in Emery and Cogger (1982), who proposed a different decision criterion (Lambda), based on Wiener processes, for differentiating between failed and non-failed firms. Results of Wilcox (1973) and Emery and Cogger (1982) are our standards of comparison for model based methods in the present paper. It should be emphasized that our selection of these studies was not based on a view that they are without methodological critics. Rather, they describe models with very high predictive ability, giving us difficult levels of accuracy to match or exceed.

The three inputs required for all prediction models in this paper are X, the mean adjusted cash flow divided by its standard deviation, N, the firm's adjusted cash position divided by its standard deviation, and T, the number of years prior to failure for the failed firm. The values for X ranged from .99 to -.93, while the values for N ranged from 34.79 to -26.30. The values for X and N were provided over a five year period, so T ranged from one year before bankruptcy to five years to bankruptcy. Not all firms had five years worth of data thus the total sample of 380. The definition of a failed firm was a filing for Chapter X or XI bankruptcy. Details of these variables can be found in Appendix B in Wilcox (1973). Emery and Cogger (1982) also used these variables to calculate their proposed decision criterion, Lambda.

It should be recognized that the current study compares several methodologies using the data of Wilcox (1973) and, subsequently, Emery and Cogger (1982). To fairly compare all methods, only the three variables, X, N, T, can be used. This limits the predictive ability of the neural network models, which could profit from additional variables if they were available. In future studies, with a larger variable sets, it is reasonable to expect that the relative performance of neural network models would improve over the results reported in this paper.

The first step in the study design was to establish a training sample and a hold-out sample. This allows the testing of any network on a sample not used in its design. This will help address some of the ex post criticism of discriminant model design. The complete sample consisted of the 190 matched pairs reported in Wilcox (1973). Since Joy and Tollefson (1975) recommend the use of inter-temporal validation, the first 75 matched pairs (150 firms) were selected in chronological order for the training sample. The last 115 matched pairs (230 firms) defined the hold-out sample. One matched pair reported in Wilcox (1973) and Emery and Cogger (1982) was not used in this study due to incomplete data. This matched pair has been eliminated in reporting the results. The training sample covers the period from 1942 to 1951. The hold-out sample covered the period from 1948-1965. The industry mix of the two samples was similar in composition.

In Wilcox (1973) an elaborate series of rules decided correct choice in the matched pairs comparisons. These results are continued in this paper as the "Wilcox" results. Emery and Cogger (1982) based their results on a particular statistic, Lambda, which is a function of X, N, and T. This terminology is preserved in this paper.

A Logistic Regression was performed on the independent variables X, N, T. The results for the logistic regression are reported as Logit. The back propagation network employed in this paper is displayed in Figure 2. There are six nodes in the first hidden layer and seven in the second hidden layer. The results for the back propagation network are reported as ANN.

Research has shown that any continuous function can be approximated by one hidden layer (Cybenko (1989), Hornik et al. (1989), Hecht-Nielsen (1990), Hertz et al. (1991)), while at most two hidden layers are necessary (Cybenko (1988), Lapedes and Farber (1988) and Zwietering et al. (1991)). Kolmogorov's Mapping Neural Network Existence Theorem (Hecht-Nielsen (1990)) states that a network should be able to describe any continuous function using $2n + 1$ nodes with n being the number of inputs. Of course, these theoretical minima may in practice be achieved only at the expense of a large number of hidden units. We found two hidden layers, as described, to be an efficient representation in our problem domain.

All nodes used sigmoidal (logistic) transfer functions and were fully connected to the nodes in the subsequent layer. We examined many other back propagation network architectures with training at 4,000 epochs. These included networks with a single hidden layer with up to 75 nodes and networks with two hidden layers, each with as many as 20 nodes. None of these alternatives were superior to the network reported here. The decision criterion was success on the matched pairs comparison. As was expected, the best net fell within the Kolmogorov's Mapping rule.

The ANN models were based on the commonly chosen back propagation (bkpstd) with a heteroassociative network type. Training of the network revealed that connection weights were well stabilized by 4,000 epochs or presentations. To avoid the problems of over-fitting (Hecht-Nielsen (1990), Zahedi (1993)) the models which discriminated best at 4,000 epochs were tested over intervals of 500 epochs starting at 500. The model selected is the result of training at 3,500 epochs since this was the best result. The development of the back propagation network required a couple of orders of magnitude of additional labor and time than the other methods. Some of the labor and computer time was related to preprocessing of input data, since the sigmoidal (logistic) transfer function only works well when inputs are scaled to the interval (0,1). Also, since the models tested had all nodes fully connected and only certain intervals were examined there is no guarantee that the best back propagation model was found.

GANNA has the advantage of constructing the network without this manual trial and error. Network architecture is entirely data-driven and the process is very fast. Typically a few seconds, or minutes depending on sample size, are needed to construct a network with the GANNA approach. Preprocessing of the input variables was also not required, as it often is for a standard neural network. AutoNet(C) produced a GANNA network in less than a minute. Also, since GANNA is automated, a new network can be easily constructed when new data is obtained.

Figure 1 represents the network architecture selected as the best discriminator using GANNA. All three inputs are used in this network which has a single hidden layer with two nodes. It should be noted that GANNA implicitly examined larger networks, but found the structure in Figure 1 to be optimal.

Figures 1 and 2 about here

Both networks produced quite similar output patterns. We suspect the different network architectures are due to the nature of the transfer functions employed. The GANNA network used quadratic functions while the back propagation network was based on sigmoidal (logistic) functions.

V RESULTS

Tables 1 to 6 display the results of the five methods . The Tables indicate that generally, the neural networks are producing results that are competitive with, and often superior to, the predictive ability of model based predictors and Logit. The results for GANNA are strongly supportive of its usefulness. The results in predictive ability are comparable to the back propagation ANN in most cases and are better in most instances. Probably the strongest argument for considering GANNA is its ease of use and speed. While achieving very good predictive results, very little was required of the investigators in terms of specifying architectural parameters, preprocessing inputs, etc.

It should be noted that Wilcox, Lambda, Logit, GANNA and ANN achieved 92, 93, 96, 95 and 95 percent accuracy, respectively, in matched pair comparisons in the training sample. Also, both neural networks as well as Logit were able to achieve 100 percent accuracy in matched pair comparisons within two years of failure. That is, within two years of the bankruptcy of the failed firm, GANNA, ANN and Logit achieved perfect prediction. These results, of course are biased upward since they are based on the training sample. The results reported in Tables 1-6 are for the hold-out sample only.

The results in Table 1 are the misclassification rates occurring when making a comparison within each matched pair. The GANNA processor compares well to the other methods, with only a 6% error rate within one year of failure. Predictions based on Wilcox and ANN fare poorly by comparison, with error rates doubled, at 12%, which increased to as high as 28% as the time to failure increased to five years. As expected, error rates for all methods increased as the prediction lead time increased, with quite high error rates for all methods at the five year point.

Matched pair comparisons are easier to make since it is known a priori that one of the two firms will fail; it is often easy to decide which of the two firms is more solvent. Thus predictions on a non-paired basis are of more importance in practice.

The results presented in Tables 2 to 4 are the result of using sensitivity analysis to determine a cutoff point with the objective of minimizing total overall misclassifications in the training sample. This cutoff score is then used to classify all firms in the hold-out sample. The predictions made reflect how they would be made in practice, since the training sample would be the data a user of these methods would have. Sensitivity analysis could not performed on the Wilcox model since its rule-based decision criteria are not compatible with such cut-off scores.

Tables 2 to 4 must be interpreted with appropriate caution. We note that Joy and Tollefson (1975), Altman and Eisenbeis (1978), Ohlson (1980), Pinches (1980), Frydman et al. (1985), and

Tam and Kiang (1992), among others warn that the a priori probabilities of failure and the relative costs of Type I and II prediction errors are not equal. Tables 2 to 4 make implicit assumptions about these quantities. The cost of a Type I error, classifying a failed firm as non-failed, denoted by C_{12}, is greater than the cost of a Type II error, denoted by C_{21}. Also, the prior probabilities for failure, $\_{}_{2}$, are normally much less than for non-failure, $\_{}_{1}$. The relative values of these four parameters influences how Tables 2-4 can be interpreted. For more details on this issue see Altman et al. (1981).

To facilitate a comparison of the methods (Tam and Kiang (1992)) equal expected misclassification costs are implicitly being assumed in Tables 2 to 4. This represents the case of C_{12_2} = C_{21_1} where C_{12_2} represents the cost of misclassifying a bankrupt company as non-bankrupt times the probability of a bankrupt company and C_{12_2} represents the cost of misclassifying a non-bankrupt company as bankrupt times the probability of a non-bankrupt company.

Tables 2 to 4 show the misclassification rates occurring when comparing each firm's score with the cutoff value as opposed to the success in the matched pair comparisons reported in Table 1. The cutoff value is the one that minimizes total misclassifications for Table 2. Tables 3 and 4 break down these errors by type. It is important to note, therefore, that Tables 3 and 4 do not report the lowest achievable Type I and II errors, since cutoff scores could be fine-tuned to emphasize one error type or the other.

Tables 1 to 4 about here

Again the GANNA processor compares well to the other methods in Tables 2 to 4. In terms of total error, Table 2 indicates an error rate of only 15% within one year of failure for the best three methods, which again includes GANNA. As before, as the prediction lead time increases, error rates increase to as high as 40%. However, GANNA continues to compare favorably with the best of the other methods at all lead times.

In terms of the most costly Type I error, Table 3 reveals that GANNA error rates are the lowest by large margins up to three years before failure, with Logit predictions being more accurate for lead times exceeding four years. Type II errors are not usually thought of as being particularly costly, but the comparisons in Table 4 are of some interest. There, predictions based on Lambda dominate except five years from failure, where interestingly the GANNA predictor is again superior with an error rate of 28%.

Tables 2 to 4 implicitly assume equal expected misclassification costs. To examine the relative performance under different conditions, we present Table 5 which breaks down misclassification errors by type, under a variety of assumptions about the relative cost of Type I and Type II errors. Assuming the cost of a Type I error is 1, 5, 10, 20, 30, 40, 50, 60, 70, and 100 times as large as that of a Type II error, which is assigned a value of 1.0, and choosing a cut-off score to minimize resubstitution costs in the training sample, Table 5 reports misclassifications results for the validation sample.

Table 5, as expected, reveals a decrease(increase) in Type I(II) errors as the relative cost of a Type I error increases, and this is uniform across all methods. For moderately large cost ratios, which are probably realistic, Lambda and Logit predictions appear superior in total misclassification percentage. In terms of Type I error, however, this superiority is not as dominant; at a cost ratio below 70, for example, GANNA begins to produce results superior to Lambda and Logit, while the traditional ANN predictor becomes superior for intermediate cost ratios.

Tables 5 and 6 about here

Table 6 summarizes the results in terms of the resubstitution risk. As described in Frydman et al. (1985) and Tam and Kiang (1992), the resubstitution risk is the observed expected cost of misclassification and is defined as

[--- Unable To Translate Graphic ---]

where n_i is the number of Type i misclassifications and N_i is the number of firms of each type in the training sample. $\__i$ are the a prior probabilities of failure/non-failure and C_{ij} are the costs of misclassification. Following Fryman et al. (1985) and Tam and Kiang (1992) we set $\__1$ = 0.98 and $\__2$ = 0.02, C_{21} = 1 and C_{12} = 1, 5, 10, 20, 30, 40, 50, 60, 70, and 100 in Table 6. These results are perhaps easier to interpret graphically in Figure 3.

Figure 3 about here

No single method dominates for all values of C_{12}. For larger values, Logit is best, followed by GANNA, ANN, and Lambda, with a slight reversal at C_{12} = 100. For C_{12} _ 10 all four methods yield comparably low risks, with no consistent dominance.

VI CONCLUDING REMARKS

The results are encouraging for the use of ANN's in financial distress prediction. It is clear from our results that ANN's and GANNA processors rival the predictive ability of the two model-based approaches using this set of data. Potential reasons for this performance are that ANN and GANNA are capable of handling nonlinear response functions automatically. Further, GANNA does not impose a priori network architecture, thus reducing architecture misspecification as a possibility.

While no single approach studied in this paper was uniformly superior across all comparisons and statistics, it should be recalled that we were necessarily restricted to using only three variables in this comparative study. This biases downward the performance of ANN, GANNA, and Logit since all are capable of profiting from additional explanatory variables.

Since ANN's are suggested for "model scarce" / "data rich" problems of classification, the excellent comparative performance of both an ANN and the GANNA processor is not surprising. This lends itself to two interpretations. First, ANN's should be viewed as serious competitors in the financial distress prediction area, especially when large data sets are encountered with many potential variables. They can be very useful as data exploration devices when little is known about data structure. Second, their nearly equivalent predictive ability may provide indirect evidence for the validity of the model based approaches, or, at a minimum, that the proposed models of Wilcox (1973), Emery and Cogger (1982), and the ANN's may all be measuring the same underlying phenomena. The use of these models and networks with only three inputs is just the start of possible future research in the area of financial distress. There have been many other samples of failed versus non-failed firms in the literature which could be tested. There are also many other possible input variables incorporated in other models for prediction of financial distress that could also be examined using ANN's.

The area of financial distress is not the only area that ANN and GANNA show potential for beneficial use. The authors are currently examining whether ANN's and GANNA processors can be helpful in detecting fraudulent financial statements. There are many other topics that also could benefit from the potential of neural networks.

Finally, the automated GANNA approach performed very well with respect to a standard back propagation network. Given the latter's trial and error nature, the data processing required, and the architectural experience needed, GANNA seems preferable in practice.

In terms of guidelines for future investigations in this area, it seems prudent to offer several suggestions and comments. The potential importance of nonlinear relationships, and the difficulty of modelling these in a trail and error fashion with Logit, Discriminant Analysis, etc. favors a continuing interest in GANNA and ANN approaches. This is even more so in a large variable set environment. We would expect the architectural specifications required by traditional artificial neural networks to be daunting to inexperienced investigators. The data-driven architecture of GANNA favors its consideration in future work.

REFERENCES

Altman, E. 1968. Financial Ratios, Discriminant Analysis, and the Prediction of Corporate Bankruptcy. *Journal of Finance*. September:589-609.

_________. 1983. *Corporate Financial Distress: A Complete Guide to Predicting, Avoiding, and Dealing with Bankruptcy*. John Wiley & Sons. New York.

_________, R. Avery, R. Eisenbeis, and J. Sinkey. 1981. *Application of Classification Techniques in Business, Banking, and Finance.* JAI Press.

_________, and R. Eisenbeis. 1978. Financial Applications of Discriminant Analysis: A Clarification. *Journal of Financial Analysis.* March:185-200.

_________, R. Haldeman and P. Narayanan. 1977. Zeta Analysis. *Journal of Banking & Finance.* June:29-54.

Bell, T., G. Ribar. and J. Verchio. 1990. Neural Nets versus Logistic Regression: A Comparison of Each Model's Ability to Predict Commercial Bank Failures. *Proceedings of the 1990 Deloitte & Touche/University of Kansas Symposium on Auditing Problems.* 29-53.

Boritz, J. 1991. *The Going Concern Assumption.*. Canadian Institute of Charted Accountants. Toronto.

Caudill M. 1991. Neural Network Training Tips and Techniques. *AI Exp*

Coates, P. and L. Fant 1991-2. A Neural Network Approach to Forecasting Financial Distress. *The Journal of Business Forecasting*. Winter:9-12.

Cogger, K. 1992. A Learning Paradigm for Neural Network Architecture. *Euro XII/TIMS XXXI Joint International Conference, Helsinki*. June. Working Paper No.237, School of Business, University of Kansas, Lawrence, KS, 66045.

_________. and G. Emery. 1981. A Determination of the Risk of Ruin: Comment and Reply. *Journal of Financial and Quantitative Analysis.* 16. 5. December:759-772.

Coleman K., T. Graettinger and W. Lawrence. 1991. Neural Networks for Bankruptcy Prediction: The Power to Solve Financial Problems. *AI Review.* Summer:48-50.

Collins, E., S. Ghosh and C. Scofield. 1988. An Application of a Multiple Neural-Network Learning System to Emulation of Mortgage Underwriting Judgements. *Proceedings of the IEEE International Conference on Neural Networks.* 2. 459-466.

Cox, D. and H. Miller. 1980. *The Theory of Stochastic Processes*. Sci

Cybenko, G. 1988. Continuous Valued Neural Networks with Two Hidden Layers are Sufficient. *Technical Report, Department of Computer Science, Tufts University.* Medford Ma.

_________. 1989. Approximation by Superpositions of a Sigmoidal Function. *Mathematics of Control, Signals, and Systems.* 2. 303-314.

Deakin, E. 1972. A Discriminant Analysis of Predictors of Business Failure. *Journal of Accounting Research.* Spring:167-179.

Dutta, S. and S. Shekhar. 1988. Bond Rating A Non-Conservative Application of Neural Networks. *Proceedings of the IEEE International Conference on Neural Networks.* 2. 443-450.

Eisenbeis, R. 1977. Pitfalls in the Application of Discriminant Analysis in Business, Finance, and Economics. *The Journal of Finance*. June:875-900.

Emery, G. 1981. Some Empirical Evidence on the Properties of Daily Cash Flow. *Financial Management*. Spring:21-28.

_________. and K. Cogger. 1982. The Measurement of Liquidity. *Journal of Accounting Research*. Autumn:290-303.

Erenguc S. and G. Koehler. 1990. Survey of Mathematical Programming Models and Experimental Results for Linear Discriminant Analysis. *Managerial and Decision Economics.* 11. October:215-225.

Feller, W. 1968. *An Introduction to Probability Theory and its Application*. Volume 1, 3rd ed. John Wiley & Sons. New York.

Fletcher D. and E. Goss. 1993. Forecasting With Neural Networks: An

Foster, G. 1978. *Financial Statement Analysis*. Prentice Hall. Englewood Cliffs.

Frydman, H., E. Altman and D. Kao. 1985. Introducing Recursive Partitioning for Financial Classification: The Case of Financial Distress. *The Journal of Finance.* March:269-292.

Gallant, S. 1993. *Neural Network Learning and Expert Systems.* The MIT Press. Cambridge.

Gessner G., W. Kamakura, N. Malhotra, and M. Zmijewski. 1988. Estimating Models with Binary Dependent Variables: Some Theoretical and Empirical Observations. *Journal of Business Research.* 16. 1. 49-65.

Hamer, M. 1983. Failure Prediction: Sensitivity of Classification Accuracy to Alternative Statistical Methods and Variable Sets. *Journal of Accounting and Public Policy*. 2. 289-307.

Hansen J., J. McDonald and J. Stice. 1991. Artificial Intelligence and Generalized Qualitative-Response Models: An Empirical Test on Two Audit Decision Domains. *Decision Sciences.* 23. 3. May/June:708-723.

Hawley, D., J. Johnson, and D. Raina. 1990. Artificial Neural Systems: A New Tool for Financial Decision-Making. *Financial Analysis Journal*. November-December:63-72.

Hecht-Nielsen, R. 1988. Neurocomputing:Picking the Human Brain. *IEEE Spectrum.* March:36-41.

__________. 1990. *Neurocomputing.* Addison-Wesley. New York.

Hertz, J., A. Krogh and R. Palmer. 1991. *Introduction to the Theory of Neural Computation.* Addison-Wesley. New York.

Heyman, D. and M. Sobel. 1982. *Stochastic Models in Operations Research*. I&II. McGraw Hill. New York.

Jones, F. 1987. Current Techniques in Bankruptcy Prediction. *Journal of Accounting Literature*. 6. 131-164.

Joy M. and J. Tollefson. 1975. On the Financial Application of Discriminant Analysis. *Journal of Financial and Quantitative Analysis*. December:723-739.

Karlin, S. and H. Taylor. 1975. *A First Course in Stochastic Processes*. 2d edition. Academic Press. New York.

Lachenbruch, P. 1975. *Discriminant Analysis*. Hafner Press. New York.

Lapedes, A. and R. Farber. 1987. *Nonlinear Signal Processing Using Neural Networks: Prediction and System Modeling.* Technical Report LA-UR-87-2662. Los Almos National Laboratory. Los Almos, NM.

Lo, A. 1986. Logit versus Discriminant Analysis. A Specification Test and Application to Corporate Bankruptcies. *Journal of Econometrics*. 31. 151-178.

Maddala, G. 1983. *Limited Dependent and Qualitative Variables in Ec*

_________. 1991. A Perspective on the Use of Limited-Dependent and Qualitative Variables Models in Accounting Research. *The Accounting Review.* 66. October:788-807.

Marais, M., J. Patell, and H. Wolfson. 1984. The Experimental Design of Classification Models: An Application of Recursive Partitioning and Bootstrapping to Commercial Bank Loan Classification. *Journal of Accounting Research.* Supplement:87-114.

McCulloch, W. and W. Pitts. 1943. A Logical Calculus of the Ideas Imminent in Nervous Activity. *Bulletin of Mathematical Biophysics*. 5. 115-133.

Messier, W. and J. Hansen. 1988. Inducting Rules for Expert Systems Development: An Example Using Default and Bankruptcy Data. *Management Science.* 34. 12. 1403-1415.

Ohlson, J. 1980. Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research.* Spring:109-131.

Pinches, G. 1980. Factors Influencing Classification Results from Multiple Discriminant Analysis. *Journal of Business Research.* 8. 429-456.

Press S. and S. Wilson. 1979. Choosing Between Logistic Regression and Discriminant Analysis. *Journal of the American Statistical Association.* 73. December:699-705.

Richardson F. and L. Davidson. 1983. An Exploration Into Bankruptcy Discriminant Model Sensitivity. *Journal of Business Finance & Accounting.* 10. 2. 195-207.

Rumelhart, D. (ed). 1986. *Parallel Distributed Processing, I & II*. MIT Press. Cambridge.

Salchenberger L., E. Cinar. and N. Lash. 1992. Neural Networks: A New Tool for Predicting Thrift Failure. *Decision Sciences.* 23. 899-916.

Schalkoff, R. 1992. *Pattern Recognition: Statistical, Structural anc*

Surkan A. and X. Ying. 1991. Bond Rating Formulas Derived Through Simplifying a Trained Neural Network. *1991 IEEE International Joint Conference on Neural Networks.* 2:1566-1570.

Tam K. and M. Kiang. 1992. Managerial Application of Neural Networks: The Case of Bank Failure Predictions. *Management Science.* 38. 7. July:926-947.

White H. 1988. Economic Prediction Using Neural Networks: The Case of IBM Daily Stock Returns. *1988 IEEE International Conference on Neural Networks.* 2. 451:458.

Wilcox, J. 1971. A Simple Theory of Financial Ratios as Predictors of Failure. *Journal of Accounting Research*. Autumn:389-395.

________. 1973. A Prediction of Business Failure Using Accounting Data. *Empirical Research in Accounting: Selected Studies. Supplement to Journal of Accounting Research*. 11. 163-179.

Yoon Y., G. Swales Jr. and T. Margavio. 1993. A Comparison of Discriminant Analysis Versus Artificial Neural Networks. *Journal of Operational Research Society.* 44. 1. 51-60.

Zahedi, F. 1993. *Intelligent Systems for Business: Expert Systems with Neural Networks.* Wadsworth. Belmont, Ca.

Zavgren, C. 1983. The Prediction of Corporate Failure: The State of the Art. *Journal of Accounting Literature*. 2. 1-37.

_________. 1985. Assessing the Vulnerability to Failure of American Industrial Firms: A Logistic Approach. *Journal of Business Finance & Accounting.* Spring:19-45.

Zwietering P., E. Aarts and J. Wessels. 1991. *The Classification Ca*

Zmijewski, M. 1984. Methodological Issues Related to the Estimation of Financial Distress Prediction Models. *Journal of Accounting Research (Supplement)*. 59-82.

TABLE 1
The Percentage of Misclassifications
Using the Matched Pairs in the Hold-Out Sample[1]

Model	Year Before Failure 1	2	3	4	5	Total
Wilcox	12	13	12	12	24	15
Lambda	6	4	12	12	24	12
Logit	6	9	20	16	24	16
GANNA	6	9	20	20	28	17
ANN	12	13	28	20	28	21
Number of Matched Pairs in Each Year	17	23	25	25	25	115

[1] Assuming $C_{12\_2} = C_{21\_1}$

TABLE 2
The Percent of Total Misclassifications
in the Hold-Out Sample[1]

Model	Year Before Failure 1	2	3	4	5	Total
Lambda	15	24	22	30	40	27
Logit	15	22	28	30	30	26
GANNA	15	24	30	34	38	29
ANN	18	33	32	38	36	32
Number of Firms in Each Year	34	46	50	50	50	230

[1] Assuming $C_{12\_2} = C_{21\_1}$

TABLE 3

Misclassification Rates of Type 1
Errors in Hold-Out Sample[1]
(Predicting that a Bankrupt Firm is a Non-Bankrupt Firm)

Model	Year Before Failure					Total
	1	2	3	4	5	
Lambda	12	39	28	32	44	32
Logit	6	26	24	24	20	21
GANNA	6	13	24	36	48	27
ANN	6	30	24	36	40	29
Number of Firms in Each Year	17	23	25	25	25	115

[1] Assuming $C_{12\_2} = C_{21\_1}$

TABLE 4
Misclassification Rates of Type 2 Errors
in Hold-Out Sample[1]
(Predicting that a Non-Bankrupt Firm is a Bankrupt Firm)

Model	Year Before Failure					Total
	1	2	3	4	5	
Lambda	18	9	16	28	36	22
Logit	24	17	32	36	40	30
GANNA	24	35	36	32	28	31
ANN	29	35	40	40	32	36
Number of Firms in Each Year	17	23	25	25	25	115

[1] Assuming $C_{12\_2} = C_{21\_1}$