
Using Neural Networks to Predict Corporate Failure

Daniel E. O'Leary*

University of Southern California, USA

ABSTRACT Predicting corporate failure or bankruptcy is one of the most important problems facing business and government. The recent Savings and Loan crisis is one example, where bankruptcies cost the United States billions of dollars and became a national political issue. This paper provides a 'meta analysis' of the use of neural networks to predict corporate failure. Fifteen papers are reviewed and compared in order to investigate 'what works and what doesn't work'. The studies are compared for their formulations including aspects such as the impact of using different percentages of bankrupt firms, the software they used, the input variables, the nature of the hidden layer used, the number of nodes in the hidden layer, the output variables, training and testing and statistical analysis of results. Then the findings are compared across a number of dimensions, including, similarity of comparative solutions, number of correct classifications, impact of hidden layers, and the impact of the percentage of bankrupt firms. © 1998 John Wiley & Sons, Ltd.

Keywords: neural networks; bankruptcy prediction

INTRODUCTION

Many business decision-making problems involve the classification of data into different groups. A special case of this problem is predicting if firms will fail (become 'bankrupt' or experience 'financial distress') or not fail. If firms fail then there can be substantial costs. The recent Savings and Loan (S&L) crisis led to the closing of many S&Ls, with considerable cost to their patrons. The cost to the United States government was in the billions. The disaster in the S&L industry became a national political issue.

In addition, the bankruptcy prediction problem has application to a broad base of decision makers (e.g. Boritz *et al.*, 1995; Tam and Kiang,

1992; Bell *et al.*, 1990, Bell, 1997). For example, bankers considering loaning money are interested in determining if the firm to which the loan is being made will be able to repay the loan. The ability to predict bankruptcy could be used to alert auditors about the likelihood of a firm failing, possibly influencing potential litigation against the auditor. As noted by Coleman *et al.* (1991) California sought to have the license of one of the largest auditing firms (Ernst & Young) removed because of their role in the well-publicized collapse of Lincoln Savings & Loan Association. Further, regulators could use a bankruptcy prediction model to decide whether a particular bank should be closed or at least receive increased attention and guidance. Accordingly, the prediction of bankruptcy probably is one of the most important business decision-making problems facing auditors, consultants, management and government policy makers.

* Correspondence to: Daniel E. O'Leary, University of Southern California, 3660 Trousdale Parkway, Los Angeles, CA 90089-1421, USA
E-mail: oleary@rcf.usc.edu

Research in the Analysis of Bankruptcy

There has been substantial research into the prediction of corporate failure (e.g. Zavgren, 1983). However, Odom and Sharda (1990) and Bell *et al.* (1990) were probably the first two papers published investigating the ability of neural networks (nns) to predict bankruptcy. Since those two studies there have been a number of other studies, whose efforts have crossed into a broad range of disciplines including artificial intelligence/computer science, accounting, decision/management sciences, finance and information systems.

Purposes of this Paper

This paper has the advantage of reviewing what is the virtual universe of published research on using nns to estimate bankruptcy in order to provide a 'meta analysis' of the process. As a result, it uses these studies as data to draw inferences about:

- How can nns be used to analyze bankruptcy decision data?
- What nn characteristics seem to be the most effective for bankruptcy models?
- Are there any interesting or unusual behaviors exhibited by nns used to solve bankruptcy problems?
- How do nns compare to other types of decision-support system tools in generating solutions for bankruptcy-type decisions?

Accordingly, the purpose of this paper is to summarize what has worked (or not worked) in the past and how it worked. In addition, by reviewing the data of all these studies, can we develop new conclusions? As a result, this paper should be of interest to those concerned about predicting bankruptcy, or those responsible for developing models to support critical problems facing management, such as predicting management fraud and a variety of other conceptually similar classification decisions.

Further, an examination of the previous research papers finds that there has been little direct cross-fertilization of the work of different researchers. In most papers there is only an

occasional reference to previous research in the area. Accordingly, this paper also provides a vehicle to bring together research on this important topic.

Outline of this Paper

This paper proceeds as follows. The next section investigates each of the studies for a number of characteristics, e.g. what data was used, what types of nn models were used, what software has been used, what kind of network structure was used, and a variety of other issues. The following section discusses the comparative quality of solutions that were found, comparing nns to a number of other approaches. The final section briefly summarizes the paper.

COMPARISON ACROSS STUDIES

Since each study uses different data, different software, different variables, different training and testing and a variety of other factors, it is difficult to directly compare the fifteen studies using nns to predict bankruptcy. Accordingly, the purpose of this section is to analyze the methodologies used across each of the previous studies, drawing out similarities and tendencies where they are apparent.

Data

Data in each of the fifteen studies has varied across industry and time. As a result, data has varied in homogeneity, providing a number of different testbeds for the use of nns. The data that derived from the recent S&L failures in the mid-1980s has been the source for six of the studies to date (Bell *et al.* 1990; Tam and Kiang 1992; Salchenberger *et al.* 1992; Chung and Tam, 1993; Bell, 1997; Etheridge and Sriram 1997). Although the sample source was similar, there were no uniform criteria for matching and the time periods were not the same in the studies. Bell (1997), Bell *et al.* (1990) and Etheridge and Sriram (1997) used a sample of 233 failed banks from the years 1984 and 1985 that was matched against 1834 banks that did

not fail. The most homogeneous data probably was used by Chung and Tam (1993) and Tam and Kiang (1992), since they focused on thrift failures in a single location (the state of Texas), during the time period 1985–7. Tam and Kiang (1992) used a sample of about 80 failed and 80 nonfailed firms (matched on asset size, number of branches, age and charter status) for one and two years before failure. Chung and Tam used a sample of about 100 failed and nonfailed matched firms. Salchenberger *et al.* (1992) used a sample of 100 savings and loans that failed from 1986 and 1987. They matched those banks against 100 nonfailed banks, based on geographic location and total value of assets.

However, most studies included data from a variety of industries. Odom and Sharda (1990) and Wilson and Sharda (1994) used data that consisted of 129 firms drawn from different industries between 1975 and 1982. Sixty-five firms went bankrupt and 64 nonbankrupt firms that were matched to the bankrupt firms, based on industry and year. Coates and Fant (1993) focused on manufacturing firms, but about one-half of their firms were nonmanufacturing firms. They generated a sample of 94 distressed firms and 188 viable firms that had data available for three years prior to assessment of the firm as distressed. In the study with the smallest sample, Fletcher and Goss (1993) used data that compared 18 bankrupt and 18 nonbankrupt firms, from different industries, from 1970 to 1981. The nonbankrupt firms were matched based on sales and asset size. Fanning and Cogger (1994) compared a total of 190 matched bankrupt and nonbankrupt firms from a variety of industries, from 1942 to 1965. That time horizon was the largest of any of the studies. Raghupathi (1994) used 102 companies (51 bankrupt and 51 healthy) drawn from the *Wall Street Journal Index* for 1980–88 and from *Moody's Industrial Manual*, and matched on industry and size. In a study with substantial data, Boritz *et al.* (1995) used a sample of data with 171 bankrupt companies over the time period 1971–84. They compared those 171 companies with subsamples of companies from 6153 companies over the same time period. Barniv *et al.* (1997) drew on a sample of publicly traded firms which filed for bankruptcy

between 1980 and 1991, resulting in 237 bankrupt firms.

Generally, the data for predicting bankruptcy has been drawn from a sample of bankrupt and nonbankrupt firms, one year prior to the firm going bankrupt, although there have been a few studies that drew on information from more than just one year prior to bankruptcy. For example, Tam and Kiang's (1992) and Chung and Tam's (1993) samples included both one and two years before bankruptcy. Salchenberger *et al.* (1992) included data that allowed analysis of 6, 12 and 18 months before failure. Coates and Fant (1993) sample allowed data ranging from the year that financial distress was recognized to three years later. Fanning and Cogger (1994) used data from one to five years prior to bankruptcy. Etheridge and Sri-ram (1997) drew data from one to three years prior to bankruptcy.

Impact of Relative Samples of Failed and Nonfailed Firms

Although most of the studies employed equal quantities of failed and nonfailed firms, one finding from the studies that did not use equal quantities is that the ratio of failed to nonfailed firms has a substantial impact on the performance of virtually all techniques of predicting bankruptcy, including nns. Low proportions of bankrupt firms result in models that predict that substantially all firms are not bankrupt. Accordingly, we might expect that models developed with a small percentage of bankrupt firms may not be as effective in populations with larger percentages of bankrupt firms. We can examine this finding since we have data on models trained and tested in different populations. Boritz *et al.* (1995) experimented with proportions that varied from 1% to 50% bankrupt firms (1%, 5%, 10%, 20%, 33% and 50%). Wilson and Sharda (1994) tested the same conclusion with proportions that varied from 10% to 50% bankrupt firms (10%, 20% and 50%).

Neural Network Models

The research studies used three nn models. Most studies used backpropagation as the stan-

dard. As a result, although optimal estimation theory (Boritz *et al.*, 1995) and a heuristic 'Generalized Adaptive Neural Network Algorithm' (GANNA) (Fanning and Cogger, 1994) were used, the focus is on classic backpropagation.

Software

NeuralWorks was the most frequently used shell across this set of studies, although some of the analysis was done with author-programmed software. Since most studies used packaged software, the results provide a sort of lower bound on the quality of the networks generated. Author-developed software did appear to provide additional flexibility. In one case where the authors programmed their own software it was extended to accomplish additional functions. Tam and Kiang, (1992) and Chung and Tam (1993) developed and integrated different cost functions into the models.

NeuralWorks was used in six of the studies, including Bell *et al.* (1990), Coleman *et al.* (1991), Salchenberger *et al.* (1992), Coates and Fant (1993), Fanning and Cogger (1994) and Bell (1997). NeuroShell was used by Fletcher and Goss (1993) and Odom and Sharda (1990). Brainmaker was employed by Wilson and Sharda (1994) and Barniv *et al.* (1997). In addition, Fanning and Cogger also used AutoNet for the development of the adaptive nns. Tam and Kiang (1992) and Chung and Tam (1993) implemented the backpropagation in Pascal. Boritz *et al.* (1995) programmed the OET algorithm using FORTRAN and the backpropagation using C.

Input Variables

Choosing the variables for bankruptcy prediction is a critical issue. If the goal of the prediction is to anticipate government policy then using those variables that government uses to make decisions should be used. On the other hand, if the goal is to anticipate when a company will cease operations, then other empirical models of the process could be. A model of the world could influence the quality of the results generated.

The variables in the bankruptcy models generally come from either promulgated government policies and procedures (Tam and Kiang, 1992; Chung and Tam, 1993); a detailed analysis of the variables and their statistical relationships to each other (Bell *et al.*, 1990; Salchenberger *et al.*, 1992; Bell 1997; Etheridge and Sriram 1997); or previous well-known bankruptcy studies (Odom and Sharda, 1990; Coates and Fant 1993; Fanning and Cogger, 1994; Wilson and Sharda, 1994; Raghupathi 1994; Boritz *et al.*, 1995; Barniv, 1997). In particular, one set of five variables (the 'Altman' variables) was analyzed by (Odom and Sharda, 1990; Coates and Fant 1993; Wilson and Sharda, 1994; Boritz *et al.*, 1995; Barniv *et al.*, 1997). Another set of nine variables that also received attention were the 'Ohlson' variables (Boritz *et al.*, 1995; Barniv *et al.*, 1997). Fanning and Cogger (1994) used the smallest number of variables, two financial measures and time. In only one case was there a seemingly arbitrary choice, with no explanation for the source of the variables that was used (e.g. Fletcher and Goss, 1993). The variables used by these authors were not consistent with the other studies.

Finally, Coleman *et al.* (1991) indicate that they extended the results of Odom and Sharda (1990), but provide no further details. Accordingly, for the remainder of this section that study is not discussed.

Hidden Layers

There are at least three questions regarding hidden layers. First, how many hidden layers should there be? Second, what should be the ratio of input variables to hidden layer variables? Third, should the hidden variables be fully connected?

Only one study provided results for a model without a hidden layer (Tam and Kiang, 1992). Fourteen of the fifteen studies ultimately generated models that used a single hidden layer, while two studies used two hidden layers (Fanning and Cogger, 1994; Raghupathi, 1994). Further, in general, the hidden units were fully connected to the input units, because no study claimed to have prior information as to how the units should be otherwise connected.

There are a number of heuristics that can be used to determine a starting point for the number of nodes in a hidden layer. However, based on the findings presented here, it appears that the number of nodes in the hidden layer ultimately is a function of both the data set and the input variables.

In the studies that analyzed the S&L data, the ratios of number of input nodes to number of hidden layer nodes were, 2:1, 1.9:1 and 1.66:1. Bell *et al.* (1990) and Bell (1997) used 12 input nodes and a single hidden layer with 6 nodes. Tam and Kiang (1992) and Chung and Tam (1993) used 19 input nodes and a single hidden layer, with 10 nodes. Salchenberger *et al.* (1992) used 5 input nodes and a single hidden layer with 3 nodes.

As noted above, four studies used the Altman variables. Three of those studies (Odom and Sharda, 1990; Coates and Fant, 1993; Boritz *et al.*, 1995) matched the 5 input variables with 5 nodes in a single hidden layer (1:1). A fourth study (Wilson and Sharda, 1994) used a structure with 5 input variables and 10 nodes in the hidden layer (0.5:1). In a study of the Ohlson variables, Boritz *et al.* (1995) used 9 input variables, with a single hidden layer of 9 nodes (1:1). In addition, some of the networks generated by Coates and Fant (1993) developed up to 8 nodes in the hidden layer (0.6:1). Accordingly, for those studies based on previously researched variables, there typically was a ratio of input variables to hidden variables of one to one. However, that ratio dropped as 0.5 or 0.6. This is substantially less than those studies done with the S&L data.

Generally, only the findings for a single configuration are reported. However, at least three studies presented data for multiple configurations of hidden layers. Fletcher and Goss (1993) generated performance data for their three input nodes using a single hidden layer, ranging from 3 to 7 nodes (1:1 to 0.43:1). They found the best performance with 4 nodes (0.75:1). Using a model with five input variables, Salchenberger *et al.* (1992) found that a hidden layer with 3 nodes performed better than 4 or 5 nodes. Given 13 input variables, Raghupathi (1994) found that one hidden layer

of 15 variables outperformed 10 and 20 variable hidden layers.

Only two studies used two or more hidden layers. Fanning and Cogger (1994) used a configuration of two hidden layers, where the first had 6 nodes and the second had 7 nodes. They report that they examined a broad base of configurations, but that none of those alternatives were 'superior'. They also developed a GANNA model with a configuration of 3 input variables and a single hidden layer with two nodes. Raghupathi (1994) found that for 13 input variables, two hidden layers with 15 and 2 variables were able to generalize better.

Output Variables

Typically, data for the output variable was coded as a '1' (0) if the firm did not go bankrupt and a '0' (1) if the firm did (or conversely). In terms of prediction, if the value of the output variable was less than (greater than or equal to) 0.500 that would indicate a prediction of bankrupt (not bankrupt). The 0.500 is referred to as a 'cut-off point'. Different cut-off points yield different types of errors.

All but two studies used a single output variable. Wilson and Sharma (1994) used two output variables where one was coded as a '1' if the firm did not go bankrupt and the other a '0'. Similarly, if the firm did go bankrupt then the first node was coded as a 0 and the second as a 1. The firm was predicted as being bankrupt (not bankrupt) if the first node was greater than or equal to 0.5 and second one was less than 0.5. If both were greater than 0.5 or both were less than 0.5 then they were counted as incorrect classifications. Barniv *et al.* (1997) used three output variables, arguing that firms in financial distress are likely to be acquired by other firms, emerge from bankruptcy filings as independent firms or are liquidated.

Training and Testing

All but two of the studies (Fletcher and Goss, 1993; Wilson and Sharda, 1994) used different samples for training and testing. Bell *et al.* (1990) and Bell (1997) trained using the 1008

firms (102 failed and 906 nonfailed) from 1985 and tested their model using a separate holdout sample of 1059 firms (131 failed and 928 nonfailed) for 1986, 1987 and 1988. Using the same data, Etheridge and Sriram (1997) used 863 (749 healthy and 114 failed), 867 (752 healthy and 115 failed) and 892 (776 healthy and 116 failed) for 1988, 19987 and 1986. In addition, they used a holdout sample of 215 firms for each year (192 healthy and 23 failed) for each year.

Using a training sample of 38 bankrupt and 36 nonbankrupt firms, Odom and Sharda (1990) tested their model on a separate holdout sample of 27 bankrupt and 28 nonbankrupt firms. Tam and Kiang (1992) also used a holdout sample. They used 59 failed and 59 nonfailed for both one year prior to bank failure and two years prior to bank failure. Then they tested the results on a holdout sample of 22 failed and nonfailed banks one year prior to failure and 20 failed and nonfailed banks two years prior to failure. Chung and Tam (1993) used 59 failed and 59 nonfailed for both one year prior to bank failure and two years prior to bank failure. Then they tested the results on a holdout sample of 44 banks in the one-year period and 40 in the two year period prior to bankruptcy. Salchenberger *et al.* (1992) used a matched set of 100 failed and nonfailed banks to develop the model and two separate holdout samples. One sample included 58 (47 and 24) failed and nonfailed firms six (twelve and eighteen) months prior to failure. The other sample did not have equal numbers of failed and nonfailed firms. Instead, there were 75 failed and 329 nonfailed firms. The second sample was done to test the impact of different ratios of failed to nonfailed firms. Coates and Fant (1993) used different training samples and testing samples, each with 47 distressed firms and 94 viable firms. Fanning and Cogger (1994) used 75 matched pairs in chronological order for the training sample and 115 match pairs for the holdout sample. Boritz *et al.* (1995) used a training set of 115 (56) bankrupt and 4100 (2053) nonbankrupt firms for training (testing). The proportion of bankrupt firms varied from 50% to 1% of the total. Barniv *et al.* (1997) used both estimation and holdout samples.

Unfortunately, the use of a holdout sample is not unbiased. Accordingly, Tam and Kiang (1992) and Fletcher and Goss used a 'jack-knife' technique to test the nn approach. Tam and Kiang (1992) consolidated the training and holdout samples to develop a sample with 162 (158) pairs where bankruptcy occurred in one year (two years) for the failed firm. Using 161 (157) pairs they developed a model and then categorized the remaining observation. This approach results in an unbiased estimate. Fletcher and Goss (1983) had a small data set. As a result, they generated 18 training sets of 30 observations (15 bankrupt and 15 not bankrupt), and a test set of the three remaining pairs, from their data of 18 observations, resulting in a test set of 54 pairs of observations.

Wilson and Sharda (1994) used Monte Carlo resampling to generate multiple subsamples from the original sample. Using this approach allowed them to generate samples of different proportions of bankrupt and nonbankrupt firms, e.g. 50% (80% or 90%) nonbankrupt and 50% (20% or 10%) bankrupt.

The Monte Carlo and jackknife approaches are useful in those situations where there is limited data. Generally, those studies with larger amounts of data can use holdout sample and similar approaches.

Prior Probabilities and Misclassification Costs (Type I versus Type II)

The probability of bankruptcy ($P(B)$) is roughly 0.02 in the United States, while the probability of not going bankrupt ($P(NB)$) is about 0.98. Similarly, the cost of errors differs depending on whether there is classification of a bankrupt firm as a nonbankrupt firm (type I) (cost = CB), or a nonbankrupt firm as a bankrupt firm (type II) (cost = CNB). Typically, the highest costs are associated with the type I error, e.g. since this means that a loan will be granted and lost. The costs of type II errors are smaller, since they represent the lost benefit of not making a loan because we predicted the firm would go bankrupt.

These cost differences are accounted for using two different approaches. First, these

costs can be built directly into the algorithm (e.g. Tam and Kiang, 1992; Fanning and Cogger, 1994). They argue that the best classification is when $P(B) \cdot CB = P(NB) \cdot CNB$. Accordingly, they have modified the basic backpropagation model to explicitly account for the prior probability of the occurrence of bankruptcy and the relative cost of bankruptcy. Second, the costs can be assessed by analyzing the relationship between the type I and type II errors (e.g. Bell *et al.*, 1990; Bell 1997). Different cut-off values of the output variable can be used to find different proportions of errors. Third, empirical studies of different ratios can be done, e.g. Etheridge and Sriram (1997) studied a wide range of different cost ratios.

Unfortunately, the importance of these costs depends on who is impacted. As a result, there is no one way to assess the quality of a solution by the classification errors. However, we generally assume that it is more important to find type I errors, i.e. correctly classifying bankrupt firms is more important.

Statistical Analysis of Results

Results from a comparison of nns to other approaches generally should use statistical analysis to determine if the differences are 'statistically significant'. At least three approaches have been used in the fifteen studies. Each test is aimed at measuring the difference in correct classification percentages (because of differences in costs of type I and II errors), e.g. percent of total predicted correctly or percent of bankrupt firms predicted correctly. Salchenberger *et al.* (1992) used a chi-squared test for the equality of proportions. Coates and Fant (1993) tested with the normally distributed z-test of equality of proportions. I also use a similar approach later in the paper to analyze the results in some of the studies that did not include a statistical analysis of their data. Wilson and Sharda (1994) employed the nonparametric Wilcoxon test for paired observations. Barniv *et al.* (1997) used both a chi-squared test and a matched pairs ranked test.

FINDINGS

Each of these studies referenced in this paper have compared nn models to a variety of other methodologies. In general, nns have outperformed or performed as well as all comparative models across all the reported studies. However, nn have not dominated the results in all studies.

Comparative analysis can be made by examining a number of solution characteristics. If two methods classify firms in similar manners then there is little difference between the two methods. Number of correct classifications, either bankrupt, nonbankrupt, or total is another issue. However, as noted above, the correct classification of firms that ultimately go bankrupt may be the most critical issue. Alternatively, a broader-based view of relative costs of both errors is another means of assessing the quality of solutions.

Intervening variables that are found to influence the quality of the results also are a concern. Finally, the percentage of bankrupt firms in both the training and test data are found to have a major impact on the percent classified as correct.

Basis of Comparison

Some research studies have compared nn models to a single alternative approach. Odom and Sharda (1990), Coates and Fant (1993) and Wilson and Sharda (1994) compared nns to discriminant analysis. Bell *et al.* (1990), Salchenberger *et al.* (1992), Fletcher and Goss (1993), Fanning and Cogger (1994) and Bell (1997) compared nns to logistic regression.

Other studies compared nns to multiple approaches. Tam and Kiang (1992) and Chung and Tam (1993) compared nn models to discriminant analysis, logistic regression, ID3 and heuristic models. Boritz *et al.* (1995) compared nn models to discriminant models, logit and probit. Barniv *et al.* (1997) and Etheridge and Sriram (1997) compared nn models to discriminant models and logit models.

Similarity of Comparative Solutions

Generally, the similarity of solutions using different approaches, i.e., how particular firms are classified, has received little attention. Typically, discussions do little more than note, e.g. that there were some differences.

Salchenberger *et al.* (1992) was one of the few papers to statistically compare the similarity of the results between different methodologies. They used a nonparametric approach (Cohen's *K*) that allowed them to compare whether the categorizations made by logit and the nn were 'in agreement'. They find, in one case, that there is significant disagreement between the misclassifications made by nns and discriminant analysis.

Analysis of Number of Correct Classifications

Perhaps the most frequently used assessment of the quality of nn solutions is the measurement of the number of correct classifications or number of correctly predicted bankruptcies. Generally, such comparisons should employ statistical analysis to establish their conclusions.

Bell *et al.* (1990, p. 43) and Bell (1997) found no significant differences between logit and nns. As they note '... most of the differences between the logit model and the nn are not greater than two holdout sample banks'. Using the same data, but breaking it down to one, two and three years before bankruptcy, Etheridge and Sriram (1997) found that, one year before failure, discriminant analysis performed the best, and for two and three years out a naïve model outperformed nn, logit and discriminant analysis.

Odom and Sharda (1990) and Wilson and Sharda (1994) both found that the nn had at least as many correct classifications as discriminant analysis for samples of nonbankrupt to bankrupt of 50/50, 80/20 and 90/10. Using a z-test of proportions, I found that the Odom and Sharda (1990) model had one out of three data sets where the total number of classifications was statistically significantly different (better) than discriminant analysis at about the 0.06 level. I also found that two of the three

data sets had a proportions of bankrupt firms, statistically significantly different (better), at the 0.04 and 0.08 levels. Coleman *et al.* (1991) claim to have further improved on the results of Odom and Sharda (1990). Wilson and Sharda (1994) found that six of the nine sets of data that they tested provided results where the proportion correct using the nn was statistically significantly different (better) than the discriminant analysis results, at the 0.05 level or better.

Salchenberger *et al.* (1992) found that the nn resulted in at least as many correct classifications as logistic regression for each scenario tested. In addition, they used a test of equality of proportions to test the probability that there is a difference between the findings using logit and the findings using a nn. In six out of thirty-two cases, the number of correct classifications by the nn was statistically different (better) than that for logit at the 0.10 level or better. Coats and Fant (1993) found that for two, one and no years prior to the recognition of distress, the nn model had a higher percentage of correct classifications. Barniv *et al.* (1997) found nns performed better than logit and discriminant analysis for the entire sample and the estimation sample, whereas, for the holdout sample, the results were equivocal.

I found that Tam and Kiang (1992) had results that were statistically significant using a z-statistic approach to compare the proportions. In particular, the total number of classifications errors were statistically significantly (better) for the nn model as compared to the ID3 model, at the 0.10 level or better, for the holdout sample in the second year. In addition, the number of correctly classified bankrupt firms by the nn were statistically significantly different (better) than those by discriminant analysis (0.01 or better), logit (0.10 level or better) and ID3 (0.002 or better). I also found that the Fletcher and Goss (1993) nn model was statistically significantly different (better) than the discriminant model at the 0.03 level, using the z-statistic.

Impact of Different Cut-off Points on Number of Correct Classifications

Reducing the cut-off point reduces the probability of a type I error. As the cost of a type

I error increases there is a relative decrease (increase) in the number of type I (II) errors. As a result, rather than limiting the analysis to the type I and II errors associated with a cut-off point of 0.5, researchers have examined some alternatives.

Salchenberger *et al.* (1992) found that at cut-off points of 0.2 and 0.5 the nn made no more, and generally fewer, classification errors than logistic regression for time periods ranging from 6 months to 18 months before failure. In addition, Salchenberger *et al.* (1992) found that the nn was less sensitive to changes in the cut-off point than logistic regression.

Fletcher and Goss (1993) found that for each of the cut-off points 0.35, 0.45, 0.65 and 0.75 the nn had a higher percentage of correct classifications than logit. For 0.25 they found that there was a small difference in favor of logit.

Minimizing a Cost Function of Type I and II Errors

Since the cost of a type I error is different than a type II error, the studies were examined for evidence in terms of the type I and II errors generated by the different solution methodologies. Odom and Sharda (1990), Coats and Fant (1993) and Wilson and Sharda (1994) found that nn models had fewer type I errors than discriminant models for all samples tested. Wilson and Sharda (1994) found that difference was statistically significant at the 0.002 level or better in two out of the three cases studied. Coats and Fant (1993) found that the difference between the number of bankrupt firms were statistically significantly different at the 0.05 level or better.

Tam and Kiang (1992) and Fanning and Cogger (1994) explicitly integrated the cost function into their results by studying a number of different ratios of CB to CNB, ranging from 1:1 to 100:1. Tam and Kiang (1992) found that the nn dominated discriminant analysis, once the cost figures were embedded in the algorithm. However, Fanning and Cogger (1994) found no single method dominated, but that for larger values, logit was best.

Etheridge and Sriram (1997) found little difference between different prediction methods

when costs of type I and II errors were not considered. However, when they did consider those costs in the methods, a nn approach dominated.

Impact of Number of Hidden Layers

Deviation from a single layer to zero layers had a negative impact on the relative results. Tam and Kiang's (1992) comparison of a nn with no hidden layer to one with a hidden layer found no statistical difference in the proportions of total correct classifications, although the nn with the hidden layer did perform better. However, for the prediction of the number of bankrupt firms the nn with the hidden layer had proportions that were statistically significantly different (better) for the first and second year of data, at the 0.04 level or better.

Two studies employed multiple hidden layers. Fanning and Cogger (1994) was the only study with multiple hidden layers and a comparison to nns. I found that logit outperformed the two hidden layer nn with statistically significantly different (better) results four out of five years for total errors and two out of five years for number of correctly classified bankrupt firms. Raghupathi (1994) found that a nn with two hidden layers outperformed single-layer models.

Impact of Different Percentages of Bankrupt Firms

One of the most extensive computational studies testing nn models is Boritz *et al.* (1995), who found substantial variation in which methodology produced the best results when they varied the percentage of bankrupt firms. They trained networks and other techniques based on 1%, 5%, 10%, 20%, 33% and 50% bankrupt firms using two sets of variables and corresponding networks ('Altman' and 'Ohlson'). Then they tested those networks on networks of all six percentages. In general they found that for networks trained with the smaller percentages of bankrupt firms (1%, 5% and 10%) linear discriminant analysis, logit, probit and backpropagation all generated similar results.

For both the 'Altman' and 'Ohlson' models

I found the data in Boritz *et al.* (1995) provides some evidence that backpropagation generalizes better 'downward' rather than 'upward'. The backpropagation model trained on 33% bankrupt firms was the best model for test proportions of 1%, 5% and 10% bankrupt firm populations. Similarly, the backpropagation model trained on 20% bankrupt firms was the best model for test proportions of 1%, 5% and 10% bankrupt firm populations, while the backpropagation model trained on 10% bankrupt firms was the best model for test proportions of 1%, and 5% bankrupt firm populations. However, the backpropagation model based on 20% (10%) bankrupt population has one of the worst classification accuracies when tested on populations with 33% or 50% (20%, 33% or 50%) bankrupt firms.

The data in Wilson and Sharda (1994) also support the generalization downward proposition. In the case where 10% of the firms were bankrupt in the training sample and the tests were made of samples with 20% and 50% bankrupt firms, there was not a statistically significant difference between the nn and the discriminant analysis in the number of correctly classified bankrupt cases found. Accordingly, these results suggest that the percentage of bankrupt firms that is used to train should be roughly greater than or equal to that in the test sample.

Impact of Time

The comparative quality of bankruptcy prediction using nns appeared to decline with time away from bankruptcy. Etheridge and Sriram (1997) found that a naïve model outperformed nns in prediction accuracy 2 and 3 years prior to bankruptcy. Although they generally had very positive results in the use of nns, three years before the recognition of financial distress, Coates and Fant (1993) found that the discriminant approach had a higher percentage correct than nn. I found statistically significant results in Fanning and Cogger (1994)'s results where the number of correctly classified bankrupt firms using logit was better than nns for years 2, 3, and 5.

Solution Time/effort

A common finding was that nn approaches commonly take substantially longer to execute than statistical approaches. Odom and Sharda (1990) indicate that it took an IBM PC-XT an average of 24 hours to develop nn models. Fletcher and Goss (1993) found that their 3-variable formulation, using an Intel 486-based system (~25 megahertz), yielded solutions to the logistic regression problem in 0.3 minute, while the nn approaches required between 9.8 to 18.8 minutes.

There are a number of computing environments and since not all researchers reported the computing time perhaps a better measure of computation effort is the average number of iterations required to generate the nn. In most cases, the reported computations were large. Bell *et al.* (1990) and Bell (1997) used 300,000 iterations in training. Odom and Sharda (1990) report that convergence was achieved after 191,400 iterations. Salchenberger *et al.* (1992) achieved convergence after 40,000 iterations. Fletcher and Goss (1993) indicate that they stopped computation at over 700,000 iterations.

However, in a few cases computational effort was significantly less. Tam and Kiang (1992) and Chung and Tam (1993) allocated a maximum of 2000 iterations per run, while Coates and Fant (1993) allowed a maximum of 1400 iterations. Raghupathi (1994) reported convergence after about 6000–8000 iterations.

Despite the higher cost of computation, the real issue is the potential benefit of improved prediction. Computation is cheap compared to the alternative of a loan going bad or a bank failing.

SUMMARY

Nns are an effective tool for supporting managerial decision making. Nn formulations used for bankruptcy have generated results that are at least as good as those generated by discriminant analysis, logit, probit and ID3. However, in some settings, e.g. over time, other approaches seemed to perform quite well relative to nns.

Few researchers have used statistical approaches to try to understand the quality of nn solutions, e.g. in comparison to other approaches. However, a number of approaches were reviewed and used to understand existing studies.

A number of characteristics in the formulation of the nn models influenced the quality of the findings. The training proportion of bankrupt firms in the data influenced the quality of the results in the training and testing, resulting in lack of 'upward' generalization. Further, deviation from a single hidden layer apparently had a negative impact on the relative quality of the models. Finally, time seems to negatively influence the relative quality of the nn models.

Acknowledgements

An earlier version of this paper was presented at the American Accounting Association National Meeting, August 1996. The author acknowledges the anonymous helpful referee comments.

References

- Barniv, R., Agarwal, A. and Leach, R., 'Predicting the outcome following bankruptcy filing: a three-state classification using neural networks', *International Journal of Intelligent Systems in Accounting, Finance and Management*, **6**, No. 3, 1997, 177-194.
- Bell, T., 'Neural nets or the logit model: a comparison of each model's ability to predict commercial bank failures', *International Journal of Intelligent Systems in Accounting, Finance and Management*, **6**, No. 3, 1997, 249-264.
- Bell, T., Ribar, G. and Verchio, J., 'Neural nets vs. 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*, 1990, pp. 29-58.

- Boritz, E., Kennedy, D. and Albuquerque, A., 'Predicting corporate failure using a neural network approach', *International Journal of Intelligent Systems in Accounting, Finance and Management*, **4**, No. 2, 1995, 95-112.
- Chung, H-M. and Tam, K., 'A comparative analysis of inductive learning algorithms', *International Journal of Intelligent Systems in Accounting, Finance and Management*, **2**, No. 1, 1993, 3-19.
- Coates, P. and Fant, L., 'Recognizing financial distress patterns using a neural network tool', *Financial Management*, Autumn 1993, 142-155.
- Coleman, K., Graettinger, R. and Lawrence, W., 'Neural networks for bankruptcy prediction: the power to solve financial problems', *AI Review*, **4**, Summer 1991, 48-50.
- Etheridge, H. and Sriram, R., 'A comparison of the relative costs of financial distress models', *International Journal of Intelligent Systems in Accounting, Finance and Management*, **6**, No. 3, 1997, 235-248.
- Fanning, K. and Cogger, K., 'A comparative analysis of artificial neural networks using financial distress prediction', *International Journal of Intelligent Systems in Accounting, Finance and Management*, **3**, No. 3, 1994, 241-252.
- Fletcher, D. and Goss, E., 'Forecasting with neural networks: an application using bankruptcy data', *Information and Management*, **24**, 1993, 159-167.
- Odom, M. and Sharda, R., 'A neural network model for bankruptcy prediction', *Proceedings of the International Joint Conference on Neural networks*, June 1990, pp. II-163-II-168.
- Raghupathi, W., 'A neural network approach to bankruptcy prediction', *Journal of Information Science and Technology*, **3**, No. 4, July 1994. First presented at the 24th Annual Hawaii International Conference on System Sciences, January 1991.
- Salchenberger, L., Cinar, E. and Lash, N., 'Neural networks: a new tool for predicting thrift failures', *Decision Sciences*, **23**, No. 4, July/August 1992, 899-916.
- Tam, K. and Kiang, M., 'Managerial applications of neural networks: the case of bank failure prediction', *Management Science*, **38**, No. 7, July 1992, 926-947.
- Wilson, R. and Sharda, R., 'Bankruptcy prediction using neural networks', *Decision Support Systems*, **11**, 1994, 545-557.
- Zavgren, C., 'The prediction of corporate failure: the state of the art', *Journal of Accounting Literature*, **2**, 1983, 1-37.