Neural Networks: A New Tool for Predicting Thrift Failures*

Linda M. Salchenberger
Department of Management Science, Loyola University of Chicago, 820 N. Michigan Avenue, Chicago, IL 60611

E. Mine Cinar
Department of Economics, Loyola University of Chicago, 820 N. Michigan Avenue, Chicago, IL 60611

Nicholas A. Lash
Department of Finance, Loyola University of Chicago, 820 N. Michigan Avenue, Chicago, IL 60611

ABSTRACT

A neural network model that processes input data consisting of financial ratios is developed to predict the financial health of thrift institutions. The network's ability to discriminate between healthy and failed institutions is compared to a traditional statistical model. The differences and similiarities in the two modelling approaches are discussed. The neural network, which uses the same financial data, requires fewer assumptions, achieves a higher degree of prediction accuracy, and is more robust.

Subject Areas: Banking and Finance, Computer Applications, and Statistical Decision Theory.

INTRODUCTION

Recently, there has been considerable interest in the development of artificial neural networks (ANNs) for solving a variety of problems. Neural networks, which are capable of learning relationships from data, represent a class of robust, nonlinear models inspired by the neural architecture of the brain. Theoretical advances, as well as hardware and software innovations, have overcome past deficiencies in implementing machine learning and made neural network methods available to a wide variety of disciplines. Financial applications that require pattern matching, classification, and prediction such as corporate bond rating [10], credit evaluation, and underwriting [9] have proven to be excellent candidates for this new technology.

In this paper, we present a neural network developed to predict the probability of failure for savings and loan associations (S&Ls), using the financial variables that signal an institution's deteriorating financial condition. We compare its performance with a logit model, since logit has frequently been used to discriminate between failed and surviving institutions. For all cases examined, the neural network performs as well or better than logit in classifying institutions as failed or nonfailed.

Even a moderate improvement in the ability to correctly classify insolvent institutions represents a significant contribution, given the magnitude of the current crisis in the thrift industry and the enormous costs of resolution [14]. Early identification

*We wish to thank Dr. Elijah Brewer, III, of the Federal Reserve Bank of Chicago, and the anonymous referees for their helpful comments. We also appreciate the help we received from Wichai Saenghirunwattana, Jane Bajc, and Sandy Knight with data management. This project was partially funded by a research grant from Loyola University of Chicago.

of financial decline provides the opportunity for close monitoring of the problem institution and the ability to take immediate corrective actions.

The paper is organized into four major sections. First, a selected group of past studies of thrift failures is presented. This is followed by a discussion of backpropagation neural networks and the relationship of the backpropagation learning to statistics. Next, we present a backpropagation neural network which uses financial data to predict thrift institution failures. Finally, we evaluate the ability of the neural network to predict thrift failures using a logit model as a performance benchmark.

PAST STUDIES OF THRIFT FAILURES

Past studies of thrift institution failures were typically devoted to either building early warning systems or explaining failures. The discriminatory variables included a set of financial ratios, determined from internal call report data and organized into a CAMEL framework. CAMEL is an acronym for the five major classes of financial data: capital adequacy (*C*), asset quality (*A*), management efficiency (*M*), earnings quality (*E*), and liquidity (*L*). Capital adequacy, determined largely by the extent of compliance with regulatory capital requirements, is important not only to absorb losses, but also to deter management from taking inordinate risks. Asset quality, measured by the volume and severity of problem loans, is often considered the most critical determinant of an institution's soundness and receives much of the examiners' attention. Other critical factors are the technical competence, leadership ability, and integrity of management. However, assessing management is highly subjective, and difficult to capture directly with specific financial ratios. In empirical studies, management efficiency is often evaluated in terms of the ability to control costs and expenditures.[1] Earnings are judged on the basis of level, trend, stability, and source. The final component refers to liquidity, and is gauged by the institution's liquid assets and ability to tap funding sources.

A review of past studies of thrift institution failures reveals that multiple discriminant analysis, logit, and probit models are frequently developed to classify institutions as failed or surviving. Altman [1, p. 446-447] used a quadratic discriminant model to categorize 212 S&Ls as "serious problem," "moderate problem," and "no problem" using data from 1966 through 1973 (see Table 1). Barth, Brumbaugh, Sauerhaft, and Wang [3] developed a logit model using semiannual data for 318 closed and 588 solvent institutions for the period 1981 through 1985. Logit was also used by Benston [4] in a study of 178 closed and 712 solvent S&Ls for the years 1981 through 1985. An early warning system based on multiple discriminant analysis was developed by Pantalone and Platt [23] for the S&Ls in the Boston district of the Federal Home Loan Bank System. Recently, a robust multivariate procedure, based on the evaluation of statistical outliers, was developed by Booth, Alam, Ahkam, and Osyk [6] to predict savings and loans failures.

Expert systems have also been used to predict bankruptcy in the thrift and other industries, as well. Elmer and Borowski [11] developed a rule-based expert system to compute an index which was a weighted average of measures of capital, asset quality, earnings, and liquidity. A system that incorporates this index performs

[1]In some studies of S&Ls, the management component is ignored, resulting in a CAEL framework [1] [3] [4] [11].

Table 1: Financial variables found to be significant in selected studies.

Author	Statistical Technique	Significant Financial Ratios
Altman [1]	MDA	Net worth/Total assets (*C*) Net operating income/Gross operating income (*E*) Real estate owned/Total assets (*A*)
Barth, et al. [3]	logit	Net worth/Total assets (*C*) Interest sensitive funds/Total funds (*E*) Net income/Total assets (*E*) Loans/Total assets (*L*) Liquid assets/Total assets (*L*)
Benston [4]	logit	Net worth/Total assets (*C*) Net income/Total assets (*E*) Change in interest and fee income/earning assets (*E*) Change in interest and depositor's dividends/Earning assets (*E*)
Pantalone and Pratt [23]	MDA	Net worth/Total assets (*C*) Cash and securities/Total savings and short-term borrowing (*L*) Operating expense/Gross operating income (*M*)

Note: MDA = multiple discriminant analysis,
A = measure of asset quality,
C = measure of capital adequacy,
E = measure of earnings quality,
M = measure of management efficiency,
L = measure of liquidity.

as well as statistical models in identifying problem institutions six months before failure and achieves greater prediction accuracy in identifying problem institutions twelve and eighteen months before failure.

A direct comparison of the performance of the statistical predictor models is difficult for a number of reasons. Each study differs with respect to modelling technique, a priori categories, and classification criteria. However, some common results do emerge from an examination of these past studies. First, predicting thrift institution failure is often formulated as a classification problem in which a group of independent variables are used to predict failure. Second, the relationship between failure or nonfailure and the financial variables is frequently assumed to be nonlinear. Finally, the financial predictor variables may be highly correlated. For classification problems in which the dependent variable is a nonlinear function of correlated independent variables, neural networks provide a promising tool. These observations motivated our decision to develop a neural network to discriminate between surviving and failed institutions.

NEURAL NETWORKS

Inspired by studies of the brain and nervous system, neural networks are composed of neurons, or processing elements and connections, organized in layers. These layers can be structured hierarchically, and the first layer is called the input layer, the last layer is the output layer, and the interior layers are called the middle or hidden layers. Feedforward networks map inputs into outputs with signals flowing in one direction only, from the input layer to the output layer. A two layer neural network consisting of an input layer and an output layer is shown in Figure 1. Each connection between neurons has a numerical weight associated with it which models the influence of an input cell on an output cell. Positive weights indicate reinforcement; negative weights are associated with inhibition. Connection weights are learned by the network through a training process, as examples from a training set are presented repeatedly to the network.

Each processing element has an activation level, specified by continuous or discrete values. If the neuron is in the input layer, its activation level is determined in response to input signals received from the environment. For cells in the middle or output layers, the activation level is computed as a function of the activation levels on the cells connected to it and the associated connection weights. This function is called the transfer function or activation function and may be a linear discriminant function (i.e., a positive signal is output if the value of this function exceeds a threshold level, and 0 otherwise). It may also be a continuous, non-decreasing function. For example, the sigmoidal (logistic) function

$$f(\theta) = (1 + \exp(-\theta))^{-1} \tag{1}$$

(see Figure 2) which assigns values between 0 and 1 (or −1 and 1) to inputs is often used in backpropagation networks.

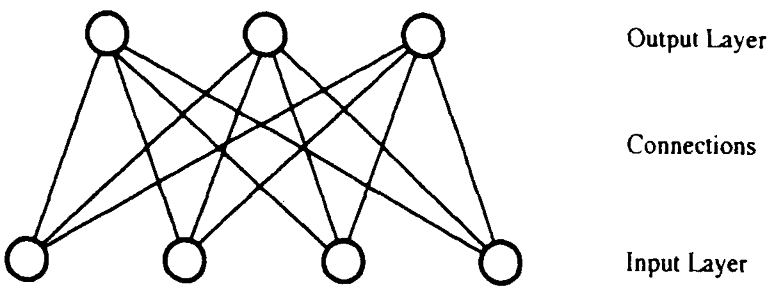
While basically an information processing technology, neural networks differ from traditional modelling techniques in a fundamental way. Parametric statistical models require the developer to specify the nature of the functional relationship between the dependent variable and the independent variables (e.g., linear, logistic). Once an assumption is made about the functional form, optimization techniques are used to determine a set of parameters that minimizes a measure of error. Neural networks with at least one middle layer use the data to develop an internal representation of the relationship between the variables so that a priori assumptions about underlying parameter distributions are not required. As a consequence, better results might be expected with neural networks when the relationship between the variables does not fit the assumed model. Nevertheless, many decisions regarding model parameters and network topology can affect the performance of the network.

Two-layer neural networks do not have the ability to develop internal representations. They map input patterns into similar output patterns. While these networks have proved useful in a variety of applications, they cannot generalize or perform well on patterns that have never been presented.

Neural networks with hidden layers have the ability to develop internal representations. The middle layer nodes are often characterized as feature detectors which combine raw observations into higher order features, thus permitting the network to make reasonable generalizations.

A two-layer neural network has an input layer which can be represented by a vector $x=(x_1, x_2, \ldots, x_n)$ of features and an output layer which can be represented

Figure 1: A two-layer neural network.



by an output vector $\mathbf{y}=g(\mathbf{x})$. We assume, for the remainder of this discussion, that there is a single output node $y=g(\mathbf{x})$, the connection weights are represented by β_i, $i=0, \ldots, n$, and we have a linear transfer function. It follows that

$$y = g(\mathbf{x}) = \Sigma\beta_i x_i. \tag{2}$$

Thus, the familiar linear regression model is similar in form to a two-layer neural network which has a linear transfer function.

In the two-layer perceptron model with a linear discriminant transfer function, neurons are not activated until some threshold level θ_0 is reached, that is,

$$y = F(\Sigma\beta_i x_i) \tag{3}$$

where $F=1$ when $\Sigma\beta_i x_i > \theta_0$ and 0 otherwise. Since the transfer function F can be any continuous, nondecreasing function, F can represent a cumulative distribution frequency (cdf). When F is the normal cumulative distribution function, $F(\Sigma\beta_i x_i)$ is the conditional expectation of a Bernoulli random variable generated by a probit model. When F is the logistic cdf, $F(\Sigma\beta_i x_i)$ is the conditional expectation generated by the logistic model [20]. Therefore, in a two-layer neural network, the network output function can be compared to the familiar logit and probit regression models.

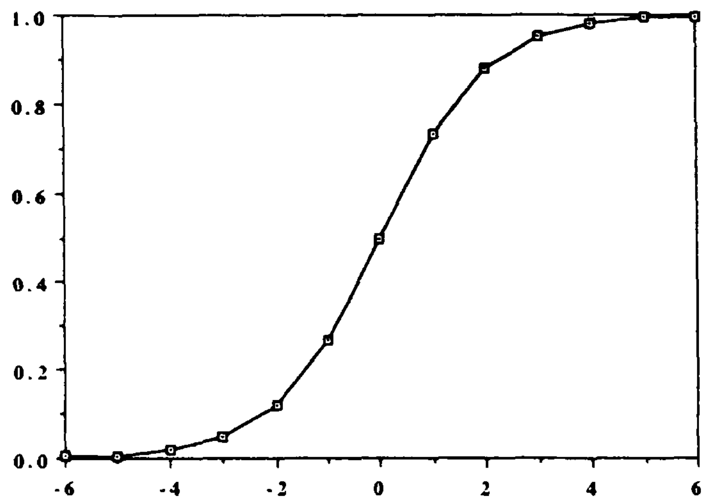
We now focus attention on multi-layered neural networks. Consider a single middle layer, feedforward network with a single output cell, k middle layer nodes, and n input nodes. Any middle layer cell receives the weighted sum of all inputs and a bias term and produces an output signal

$$m_j = f(\Sigma w_{ij} x_i), \qquad j = 1, \ldots, k, \; i = 0, \ldots, n, \tag{4}$$

where f is the transfer function, x_i is the ith input signal, w_{ij} is the strength of the connection from the ith input neuron to the jth middle layer cell. The activation levels from the middle layer cells are transferred to the output layer cells in the same way so that the output cell sees the weighted sum of the outputs of the middle layer cells and produces a signal

$$y = F(\Sigma v_j f(\Sigma_{w_{ij}} x_i)) = g(\mathbf{x},\boldsymbol{\theta}) \tag{5}$$

Figure 2: Graph of sigmoidal function.



where $\mathbf{x}$ is the vector of inputs, and $\boldsymbol{\theta}$ is the vector of network weights. We can interpret (5) as a nonlinear regression function which represents a single middle layer, feedforward neural network.

The ability of multi-layer networks to represent nonlinear models has been established through an application of an existence theorem proved by Kolmogorov [18] [19]. This theorem has been used to show that, for any feedforward network with a single middle layer, there exists a network output function

$$g(\mathbf{x},\boldsymbol{\theta}) = \Sigma v_j f(\Sigma w_{ij} x_i) \tag{6}$$

which can provide an accurate approximation to any function of $(x_1, x_2, \ldots, x_n)$, if the inputs are scaled to be within [0,1]. The number of middle units required by the theorem is $2n+1$ where n is the number of input nodes, although representations with fewer middle nodes may also exist [27]. The implication of the Kolmogorov theorem is that classification problems which are not linearly separable can be solved with multi-layered neural networks.

It is interesting to note that research in neural networks was abandoned for almost twenty years when it was discovered that only linearly separable problems could be solved with two-layer networks. The linear perceptron network, a two-layer network using a linear threshold response function, was proposed by Rosenblatt [22] in the 1950s. While some theoretical limitations of the linear perceptron, including the requirement that data points be linearly separable for perfect classification, were recognized by Rosenblatt, a more complete analysis of its computational limitations was developed by Minsky and Papert [21]. They made extensive use of geometrical arguments to prove that properties like connectedness and parity could not be computed with perceptrons.[2]

[2]For a more complete discussion of the development of the perceptron by Rosenblatt, and the contributions of Minsky and Papert, see Anderson and Rosenfeld [2].

Backpropagation Networks and Learning Rules

Backpropagation is an approach to supervised network learning which permits weights to be learned from experience, based on empirical observations on the object of interest. Training consists of repeatedly presenting the network with examples that can be viewed as input/output vectors. Supervised learning methods require that for each input pattern, an appropriate response or classification of the output be presented to the network during training. These networks cannot learn from an input pattern for which no correct response has been provided.

In an approach to learning which does not require a teacher, no correct response is provided. These unsupervised learning methods take advantage of natural groupings that appear within the data by using a variety of approaches to learning. In the simplest form of competitive learning, an output node with the greatest net input is denoted winner and its weights are updated, using a learning rule [26]. Adaptive resonance theory (ART) overcomes some of the limitations and instability associated with competitive learning [7]. In self-organizing feature maps, the concept of neighbor is added to competitive learning and a group of associated nodes responds to an input pattern [17].

Although the term backpropagation can be used to refer to the dynamic feedback of errors propagated backward through a network to adjust the weights, it is also commonly used to describe learning by the generalized delta rule. Rumelhart, Hinton, and Williams [24] described the generalized delta rule with the following three steps. First, the derivative of the square error with respect to the outputs and the target values of the network is computed. The chain rule is applied in the second step to calculate the derivatives of error with respect to outputs and weights within the network. Finally, the weights are updated using

$$\theta_t = \theta_{t-1} + \alpha \blacktriangledown \mathbf{f}(X_t, \theta_{t-1})(Y_t - f(X_t, \theta_{t-1})) \qquad (7)$$

$$t = 1, 2, \ldots,$$

where α is the learning rate, Y_t is the target, θ_0 is a random set of small initial weights, and $\blacktriangledown\mathbf{f}$ is the gradient (vector of partial derivatives with respect to the weights θ). Thus, the learning process updates the current set of weights with a function of the difference between the system's response to an input vector and the associated correct category. The steepest descent algorithm is used in which changes are made in the direction of the gradient (i.e., direction of the largest change in the error).

Better results may be achieved by replacing the steepest descent algorithms with response surface methods. Response surface optimization does not require any functional form assumption, in contrast to the specification function required in current methods and thus represents a future research area which may improve neural network performance.

Under appropriate conditions, the learning rule given in (7) yields weights converging to a vector θ^* that solves

$$E[\blacktriangledown \mathbf{f}(X_t, \theta)(Y_t - f(X_t, \theta))] = 0 \qquad (8)$$

where the mathematical expectation is taken with respect to the joint distribution of the random variables X_t, Y_t. A solution θ^* to this equation satisfies the necessary conditions for a local solution to the least squares problem

$$\min_{\theta} E[(Y_t - f(X_t, \theta))^2]. \tag{9}$$

A set of weights that solves (7) yields a network output which is mean square optimal for Y_t; it minimizes expected squared error as a prediction for Y_t. Also, the weights θ^* guarantee a network that is a mean square optimal approximation to the conditional expectation $E(Y_t|X_t)$. Thus, backpropagation learning is an approach for which convergence to a local optimal set of weights is guaranteed [27].

The application of the generalized delta rule involves a forward pass through the network during which errors are accumulated by comparing actual outputs with the targets. This is followed by a backward pass during which adjustments in connections weights are made based on the errors, using a recursive rule such as (7).

Backpropagation is a gradient descent method that minimizes the mean squared error of the system by moving down the gradient of the error curve. The error surface is multi-dimensional and may contain many local minima. As a result, training the network often requires experimentation with starting position, adjusting the weights during training, and modifying various learning parameters. In particular, the learning rate α is usually adjusted downward during training and a momentum term may be increased to avoid getting stuck in local optima.

A NEURAL NETWORK TO PREDICT THRIFT INSTITUTION FAILURES

A backpropagation neural network that forecasts the probability of failure of thrift institutions has been developed using five financial variables as inputs. Backpropagation learning was selected because it has been successfully used to solve many pattern recognition and classification problems.

Selection of Variables

In past studies, the usual procedure was to select a rather large group of independent variables and reduce that to a smaller group of statistically significant variables. We wished to see how well the neural network would perform, when measured against the best logit model we could formulate. To reduce the dimensionality of the model, we experimented with 29 variables (see Table 2), and performed stepwise regression which resulted in the identification of five variables. Each variable selected represents one of the CAMEL categories. Although this is a rather small set of variables, we obtained the best results using logit with this group and consequently developed our models with these variables.

The predictor variables representing the categories of capital adequacy, asset quality, management efficiency, earnings, and liquidity were: GNWTA (GAAP net worth/Total assets), RETA (Repossessed assets/Total assets, NIGI (Net income/Gross income), NITA (Net income/Total assets), and CSTA (Cash securities/Total assets), respectively.

Table 2: Financial ratios tested.

Capital	GNWTA*	GAAP net worth/Total assets
	ESTA	Earned surplus/Total assets
	RAPTA	RAP net worth/Total assets
Assets	RETA*	Repossessed assets/Total assets
	RLTA	High risk loans/Total assets
	REOTA	Real estate owned/Total assets
	ORATA	Other risky assets/Total assets
	TLS	Total loans/Savings
	ITA	Direct investments/Total assets
Management	NIGI*	Net income/Gross income
	TEOI	Total operating income/Other income
	OETA	Other expenses/Total assets
	OHAOI	Overhead/Adjusted operating income
Earnings	NITA*	Net income/Total assets
	NIM	Net interest margin
	ROA	Return on assets
Liquidity	CSTA*	Cash+securities/Total assets
	VLTA	Volatile liabilities/Total assets
	LATA	Liquid assets/Total assets
	ADRAP	FHLBB advances/RAP net worth
	ADGAP	FHLBB advances/GAAP net worth
	OBMTF	Other borrowed money/Total funds
	CSSB	Cash+securities/Savings+borrowed money
	BDS	Brokered deposits/Savings
	ADTA	FHLBB advances/Total assets
	ISFTF	Interest sensitive funds/Total funds
	BFTA	Borrowed funds/Total assets
	IBLEA	Interest bearing liabilities/Earning assets
Size	LOGTA	Log total assets

Note: GAAP = generally accepted accounting principles,
RAP = regulatory approved principles.
*Used in results reported in Tables 3 through 5.

Since regulators use RAP (Regulatory Accounting Principles) net worth to close institutions, we experimented with models utilizing both GNWTA (GAAP net worth/Total assets) and RAPTA (RAP net worth/Total assets) as measures of capital adequacy. These variables were highly correlated for all our data sets and the best prediction rates on the training set were obtained with GNWTA.

The Data Set and Sampling Techniques

The data set consists of financial data on 3,479 S&Ls for the period January, 1986 to December, 1987. The data are taken from Federal Home Loan Bank Board quarterly tapes. The training sets, which were used to generate the logit model and to provide examples to the neural network during its training process, were developed from call report data for June, 1986. For the first neural network and logit models, the 100 failures from January, 1986 to December, 1987 were matched with 100 nonfailed S&Ls, based on geographic location and value of total assets. For each failure, all the surviving institutions located in the same state were identified. Next, the absolute percentage differences in asset size between each failed institution and the survivors in the state were computed. These were ranked, and those that most closely matched the asset size of the failure were included in the appropriate sample.

For testing the predictive capabilities of logit and the neural network, a second sample consisting of call report data for each failed institution six, twelve, and eighteen months prior to failure was used. Data were available for 58 failed and 58 surviving institutions six months prior to failure, 47 failed and 47 surviving institutions twelve months prior to failure, and 24 failed and 24 surviving institutions eighteen months prior to failure.

A third sample was also tested, in which 75 failures were matched with 329 non-failed institutions. Due to mergers and regulatory actions, data were available for the two-year period for only 75 of the 100 failures. The reasons for diluting this sample with surviving institutions were to develop a larger sample that more closely resembles the true population, and to test the robustness of our models with respect to sampling rates.

The Development of the Neural Network

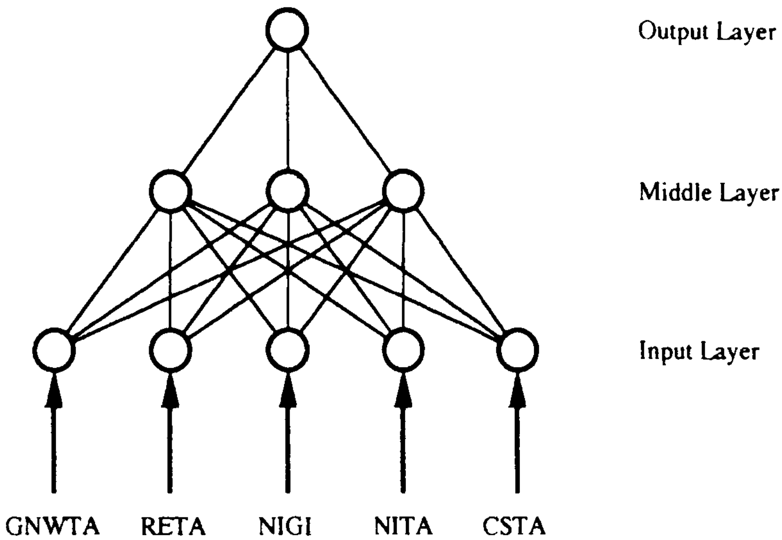
The first step in the development of a neural network model is to select an appropriate neural network paradigm by matching the application requirements with the paradigm capabilities. The application is heteroassociative (i.e., it requires mapping one set of patterns onto a different set). Supervised learning was used since the network would be trained using examples which included the result. For this application, a single middle layer, feedforward, backpropagation neural network consisting of five input nodes, three middle layer nodes, and one output node was developed (see Figure 3). The input nodes represent the financial ratios selected which measure capital adequacy (C), asset quality (A), management efficiency (M), earnings (E), and liability (L), and the output node is interpreted as the probability that an institution was classified as failed or surviving.

After experimentation with two middle layers did not result in better prediction rates, a single middle layer was adopted. This was consistent with other results with classification problems that demonstrated no improvement with more than one middle layer [9] [10].

Determining the proper number of nodes for the middle layer is more art than science and experimentation and heuristics assisted in making this choice. Our initial network was constructed with three nodes in the middle layer, based on a rule of thumb which suggests that the number of nodes in the middle layer should be 75 percent of the number of nodes in the input layer. Other models using four and five nodes in the middle layer were also tested. Generally speaking, too many nodes in the middle layer, and hence, too many connections, produce a neural network that memorizes the input data and lacks the ability to generalize. Therefore, increasing the number of nodes in the middle layer will improve the network's ability to classify institutions in the training set, but degrade its ability to classify institutions outside the training set. This proved to be true for our application.

The required output layer in this model consists of a single node which would be interpreted as a classification node, indicating insolvency or solvency of the

Figure 3: Neural network for predicting thrift failures.



institution. Initially, we set a threshold of .5 (i.e., if the output value is greater than .5, failure is predicted, otherwise the institution is classified as surviving). We also used a threshold of .2 to test the sensitivity of the network to changes in this value.

The generalized delta rule was used with the backpropagation of error to transfer values from internal nodes. (For a more detailed explanation of backpropagation learning and the generalized delta rule, see [25].) The sigmoidal function is the activation function specified in this neural network and is used to adjust weights associated with each input node. This is the same transformation employed in logit analysis in which the dependent variable is assumed to be a logistic function of the independent variables and a constant. This function was the best choice for our application since its steepest slope occurs at .5. Therefore, the result of applying the sigmoidal function to a weighted sum of the independent variables will be greatest at the midpoint and the transfer function will have less effect when an output transferred from the input layer is close to the extreme values (see Figure 2).

The network was implemented using the software package NeuralWorks Explorer running on a 386-based microcomputer with a math co-processor. NeuralWorks Explorer, developed by NeuralWare, Inc., was selected because it can be used to implement over a dozen network paradigms [15]. It allows the user to easily alter learning parameters during training and to view the weights and output values associated with trained networks. Automatic scaling of input parameters and randomization of the training set reduce the development effort. The major limitation encountered when using NeuralWorks Explorer for this application was its inability to support the numerous experiments required to find a satisfactory combination of network architecture and set of learning parameters.

Training the Neural Network

Supervised learning was conducted with training sets consisting of the five CAMEL ratios and the corresponding result (failure/surviving) for each S&L. For the input nodes in which the data was not in ratio form, the values were scaled to be within a range of 0 to 1. Automatic scaling was performed by the network software which computes the range of a set of values, and the difference between the input value and the minimum, and then divides the latter by the former. This minimizes the effect of magnitude among the inputs and increases the effectiveness of the learning algorithm. The selection of the examples for the training set focused on quality and the degree to which the data set represented the population. The size of the training set is important since a larger training set may take longer to process computationally, but it may accelerate the rate of learning and reduce the number of iterations required for convergence.

All the weights in the fully connected network were randomized before training. The learning rate and momentum were set initially at .9 and .6, respectively, and the learning rate was adjusted downward and the momentum was adjusted upward during training to improve performance. The learning rate is a constant of proportionality (see (7)) which determines the effect of past weight changes on the current direction of movement in the weight space. In backpropagation, it is usually set close to 1.0 for early iterations when larger changes take place as the optimum is approached. The learning rate should be set as high as possible, while avoiding oscillation. A momentum term can be added to the backpropagation learning rule (7) and this value is increased as an optimum is approached. This effectively provides the impetus to help the algorithm avoid becoming lodged in local minima. The NeuralWorks software allows the user to adjust these parameters dynamically, to maximize learning.

The training examples were presented to the network in random order to maximize performance and to minimize the introduction of bias. During each iteration, a training vector is presented to the network, the network error is computed, the error is propagated back through the network, and the weights are updated using this new information. In practice, the number of presentations made before the weights are updated is called the epoch size and is under the control of the user. Convergence was achieved after 40,000 iterations when the network errors remained relatively unchanged for subsequent iterations. In NeuralWorks, the user specifies the number of iterations to be performed and the training process is stopped when an appropriate measure of network error is achieved. The training process can be interrupted at any point in time and measures of network error can be displayed to assist the user in determining whether to stop or continue.

RESULTS

The performance of the neural network that was trained using the 100 failures and 100 surviving institutions from January, 1986 to December, 1987 is compared to the logit model. The regression coefficients, significant variables, and the log likelihood function for the logit model are given in Table 3. Four of the five variables, representing capital adequacy, asset quality, management efficiency, and earnings quality, are significant at the 1 percent error level, as is the log likelihood ratio. The explanatory variables were checked for linear dependence by examining the correlation matrices. None were found to be linear combinations of any others, and this was confirmed by a principal components analysis. Therefore, multicollinearity does not pose a serious problem for this model with these data [13] [16].

Table 3: Logit results for matched sample.

Intercept	-.0103
GNWTA	-55.7485*
RETA	28.7573*
NIGI	-1.2828*
NITA	-79.3499*
CSTA	6.4446
Log Likelihood Ratio	-51.000*

*Significant at the 1 percent error level.

For both the logit and the neural network models, we classify the institutions as failed if the output values exceed a cutoff point. While some previous studies set a cutoff point of .5, this overlooks the fact that the costs associated with a Type I error (misclassifying a failed institution) are usually greater than those with a Type II error [3]. Thus, we present the results for cutoff points of .5 and .2, since lowering the cutoff point reduces the probability of committing a Type I error.

Results with the Matched Sample

In Table 4, we report the number of institutions correctly classified by each model and the p-values from the nonparametric test of equality of proportions [12]. The null hypothesis that the proportion of institutions correctly classified by each method is the same is tested using a nonparametric test since the data is categorical. The chi-square test statistic for equality of k proportions is

$$Q = \sum_{j=1}^{k} \frac{(f_j - n_j p)^2}{n_j p(1 - p)} \tag{10}$$

where p is the proportion of successes, f_j is the observed frequency of success, and n_j is the number of observations.

For the training set data and cutoff points of .5 and .2, each failed and nonfailed institution correctly classified using logit is also correctly classified by the neural network. In addition, the network commits fewer total classification errors for each cutoff point and the difference is significant at the 10 percent level for a cutoff point of .5. For the six-, twelve-, and eighteen-month predictions, the number of correct classifications made by the neural network is greater than or equal to those made by logit regardless of the cutoff point. For a cutoff point of .5, significant differences at the 10 percent level were indicated for the predictions made eighteen months prior to failure. There was one institution, a failed S&L, which was correctly classified by logit and incorrectly classified by the neural network.

Differences between the misclassifications for the training set and eighteen-month forecasts made by the models were observed. Tables were developed to further analyze these misclassifications, with the rows as the misclassified failures and nonfailures made by the logit model and the columns as the misclassifications made by the neural network. We used Cohen's measure of agreement [5] to determine whether the categorizations developed by the models were in agreement. The maximum likelihood estimate of Cohen's K is computed as

Table 4: Comparison of logit and the neural network January, 1986 to December, 1987 (matched sample).

	Correctly Classified S&L's								
	All			Failed			Non-failed		
Cutoff=.5	Logit	Neural Network		Logit	Neural Network		Logit	Neural Network	
Training Set	187/200 (93.5%)	194/200 (97.0%)	p=.10*	90/100 (90.0%)	96/100 (96.0%)	p=.10*	97/100 (97.0%)	98/100 (98.0%)	p=.60
Six Months Before Failure	102/116 (87.8%)	107/116 (92.2%)	p=.25	56/58 (96.6%)	56/58 (96.6%)	p=1.0	46/58 (79.3%)	51/58 (87.9%)	p=.20
Twelve Months Before Failure	79/92 (85.9%)	85/92 (92.4%)	p=.15	43/46 (93.5%)	44/46 (95.6%)	p=.99	36/46 (78.3%)	41/46 (89.1%)	p=.15
Eighteen Months Before Failure	40/48 (83.3%)	44/48 (91.7%)	p=.25	18/24 (75.0%)	22/24 (91.7%)	p=.10*	22/24 (91.7%)	22/24 (91.7%)	p=1.0
Cutoff=.2									
Training Set	166/200 (83.0%)	186/200 (92.5%)	p=.01**	95/100 (95.0%)	96/100 (96.0%)	p=1.0	71/100 (71.0%)	90/100 (90.0%)	p=.001**
Six Months Before Failure	101/116 (87.1%)	104/116 (89.7%)	p=.60	56/58 (96.6%)	56/58 (96.6%)	p=1.0	45/58 (77.6%)	48/58 (82.8%)	p=.5
Twelve Months Before Failure	79/92 (85.9%)	85/92 (92.4%)	p=.15	44/46 (95.7%)	45/46 (97.8%)	p=.60	35/46 (76.1%)	40/46 (87.0%)	p=.05*
Eighteen Months Before Failure	41/48 (85.4%)	44/48 (91.7%)	p=.40	19/24 (79.2%)	22/24 (91.7%)	p=.25	22/24 (91.7%)	22/24 (91.7%)	p=1.0

*Significant difference at the 10 percent level.

**Significant difference at the 5 percent level or less.

$$K = \frac{N\Sigma x_{ii} - \Sigma x_{i+}x_{+i}}{N^2 - \Sigma x_{i+}x_{+i}}$$

where N=total number of observations, x_{ij} is the observed value in the (i,j)th cell, x_{i+} is the sum of the observed values in row i, and x_{+i} is the sum of the observed values in column i.

Since we observed a large number of cases in which the methods agreed, we computed K for the cases in which the models disagreed and using the asymptotic variance [5], confidence intervals for K were developed. The 95 percent confidence interval for the true value of K was (−.7,.25) for the training set and (−.9,.58) for the eighteen-month forecasts. Since a value of 1 would indicate perfect agreement, these results lead to the conclusion that there is significant disagreement between the misclassifications made by the models at the 5 percent error level. When combined with the results of the test of equality of proportions, this shows that, for these two samples, the neural network did provide a better forecast of failures.

For the training set, changing the cutoff point had a greater effect on the number of correct classifications made by logit, because the neural network assigned fewer output values between .2 and .5. For example, when the cutoff point is changed from .5 to .2, logit correctly classifies 5 more failures, but the total number of nonfailures misclassified increases from 3 to 29. In contrast, the number of nonfailures misclassified by the neural network increases from 2 to 10. Changing the cutoff point had less effect on the prediction set for both models; this may be a consequence of the smaller sample sizes.

Results with Diluted Sample

The results of the logit and neural network models which included 75 failed and 329 nonfailed institutions are given in Table 5. The coefficients for the logit model, the number of S&Ls correctly classified, and the p-values for the nonparametric test for differences [12] are reported. The variables representing capital adequacy, management efficiency, and earnings quality are statistically significant at the 1 percent level. Logit correctly classifies 54 of 75 failures and the neural network correctly classifies 64 of 75 with this difference significant at the 5 percent level. The 95 percent confidence interval for the value of K are (.23, .55), indicating mild disagreement on misclassification errors. A significant difference is also observed for the classification of the nonfailed institutions when the cutoff is .2. Also, when the cutoff point changes from .5 to .2, the number of failures misclassified by logit is reduced from 21 to 10, and the number of misclassified nonfailures increases from 2 to 21. The number of failures misclassified by neural network is reduced from 11 to 7, and the number of misclassified nonfailures increases from 2 to 10. As with the matched sample, the neural network is less sensitive to reducing the cutoff point. Finally, each S&L correctly classified by logit was also correctly classified by the neural network.

CONCLUSIONS AND SUGGESTIONS FOR FURTHER RESEARCH

In our study, we evaluated the ability of a neural network to predict thrift institution failures by comparing it with the best logit model we could develop with

Table 5: Comparison of logit and the neural network January, 1986 to December, 1987 (diluted sample).

	Correctly Classified S&Ls					
	All		Failed		Non-failed	
Cutoff=.5	Logit	Neural Network	Logit	Neural Network	Logit	Neural Network
	381/404 (94.3%)	391/404 (96.8%)	54/75 (72.0%)	64/75 (85.3%)	327/329 (99.4%)	327/329 (99.4%)
	p=.10*		p=.05**		p=1.0	
Cutoff=.2	373/404 (92.3%)	387/404 (95.8%)	65/75 (86.7%)	68/75 (90.7%)	308/329 (93.6%)	319/329 (96.9%)
	p=.05**		p=.50		p=.05**	

Logit Results	
Intercept	-1.0501**
GNWTA	-49.1956**
RETA	11.5391
NIGI	-1.3257**
NITA	-105.8590**
CSTA	-.1824
Log Likelihood Ratio	-80.6950**

*Significant difference at the 10 percent level.
**Significant difference at the 5 percent level or less.

our data. Since, for each data set examined in our study, the neural network has performed as well or better than logit, neural networks may offer a competitive modelling approach for failure prediction. We also observe that, in some cases, when the cutoff point was lowered, the reduction in Type I errors committed was accompanied by greater increases in Type II errors for the logit model than for the neural network. This may be an important result when examiners factor in the cost of committing Type I and Type II errors. An examination of Table 5 shows a significant difference between the total number of correct classifications made by the two models when the sample is diluted with healthy institutions, for both cutoff points. Since the diluted sample more closely resembles the total population of thrift institutions, the neural network may yield more consistent results when used with the data sets available to regulators. Finally, our results are consistent with those obtained in other studies [9] [10] in which neural network technology is determined to be a promising tool for classification problems. For this application, the three-layer neural network gains some predictive power over logit, which can be viewed as a two-layer model. While model specifications such as the choice of activation function and learning parameters are required in neural network models, benefits may be derived when there is insufficient information available to make assumptions about population distributions.

There are several limitations which may restrict the use of neural network models for prediction. There is no formal theory for determining optimal network topology; therefore, decisions such as the appropriate number of layers and middle layer nodes must be determined using experimentation. The development and interpretation of neural network models requires more expertise from the user than traditional statistical models. Training a neural network can be computationally intensive and the results are sensitive to the selection of learning parameters. Poor results can also occur if the wrong activation function is selected. Finally, backpropagation neural network models seem to be most successful when solving pattern recognition and classification problems; more research is required to determine if there are other types of problems which may be good candidates.

The inability of neural networks to provide explanations of how and why conclusions may restrict the use of this modeling technique. This is in contrast to expert systems, which can provide explanations to the user about how inferences are made. One approach used to determine the relative importance of individual input variables is to design a special data set which exaggerates the values of the input variables to be tested. The activation levels of hidden nodes and the output nodes are examined as each observation is processed by the network [8]. Another approach which applies nonlinear statistical methods for misspecified models is under investigation [27]. Finally, some hybrid systems have been developed which include an expert system to provide explanations for the behavior of the neural network [8].

There are many opportunities for conducting research in this area. Neural networks may be extended to other financial applications, particularly those requiring classification such as credit approval and bond rating. Further investigation into the relationship between the backpropagation network and traditional nonlinear statistical models may yield benefits to both areas of study. [Received: April 2, 1991. Accepted: January 29, 1992.]

REFERENCES

[1] Altman, E. I. Predicting performance in the savings and loan industry. *Journal of Monetary Economics*, 1977, 443-466.

[2] Anderson, J. A., & Rosenfeld, E. *Neurocomputing: Foundations of research*. Cambridge, MA: MIT Press, 1989.

[3] Barth, J. R., Brumbaugh, R. D., Sauerhaft, D., & Wang, G. H. K. Thrift institution failures: Estimating the regulator's closure rule. In G. G. Kaufman (Ed.), *Research in financial services* (Vol. 1). Greenwich, CT: JAI Press, 1989.

[4] Benston, G. J. An analysis of the causes of savings and loans failures. *Monograph Series in Finance and Economics*, 1986, Series 1985 (4).

[5] Bishop, Y., Fienberg, S., & Holland, P. *Discrete multivariate analysis*. Cambridge, MA: MIT Press, 1975.

[6] Booth, D. E., Alam, P., Ahkam, S. N., & Osyk, B. A robust multivariate procedure for the identification of problem savings and loan institutions. *Decision Sciences*, 1989, *20*, 320-333.

[7] Carpenter, G., & Grossberg, S. ART2: Self-organization of stable category recognition codes for analog input patterns. *Applied Optics*, 1987, *26*, 4919-4946.

[8] Caudill, M., & Butler, C. *Naturally intelligent systems*. Cambridge, MA: MIT Press, 1990.

[9] Collins, E., Ghosh, S., & Scofield, C. An application of a multiple neural-network learning system to emulation of mortgage underwriting judgments. *Proceedings of the IEEE International Conference on Neural Networks*, 1988, 2, 459-466.

[10] Dutta, S. & Shekhar, S. Bond-rating: A non-conservative application of neural networks. In *Proceedings of the IEEE International Conference on Neural Networks*, 1988, 2, 443-450.

[11] Elmer, P., & Borowski, D. An expert system approach to financial analysis: The case of S&L bankruptcy. *Financial Management*, 1988, 66-76.
[12] Gibbons, J. *Nonparametric methods for quantitative analysis*. New York: Holt, Rinehart, & Winston, 1976, 259-262.
[13] Intrilligator, M. D. *Econometric models, techniques, and applications*. Englewood Cliffs, NJ: Prentice-Hall, 1978, 151-156.
[14] Kane, E. J. *The S&L insurance mess: How did it happen?* Washington, D.C.: The Urban Institute, 1989.
[15] Klimasauskas, C. C. *NeuralWorks™*: An introduction to neural computing. Sewickley, PA: NeuralWare, Inc., 1988.
[16] Kmenta, J. *Elements of econometrics*. New York: MacMillan, 1971, 389-391.
[17] Kohonen, T. *Self-organization and associate memory*. New York: Springer-Verlag, 1988.
[18] Kolmogorov, A. N. On the representation of continuous function of many variables by superposition of continuous functions of one variable and addition. *Doklady akademii nauk SSSR*, 1957, *144*, 679-681.
[19] Kolmogorov, A. N. On the representation of continuous function of many variables by superposition of continuous functions of one variable and addition. *American Mathematical Society Translation*, 1963, *28*, 55-59.
[20] Maddala, G. S. *Limited dependent and qualitative variables in econometrics*. Cambridge: Cambridge University Press, 1983, 12-41.
[21] Minsky, M., & Papert, S. *Perceptrons*. Cambridge, MA: MIT Press, 1969.
[22] Rosenblatt, F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 1958, *65*, 386-408.
[23] Pantalone, C., & Platt, M. Predicting failure of savings and loan associations. *AREUEA Journal*, 1987, *15*(2), 46-64.
[24] Rumelhart, D. E., Hinton, G., & Williams, R. Learning internal representation by error propagation. In D. Rumelhart & J. McClelland (Eds.), *Parallel distributed processing*. Cambridge, MA: MIT Press, 1986, 318-362.
[25] Rumelhart, D. E., & McClelland, J. L. *Parallel distributed processing*. Cambridge, MA: MIT Press, 1986.
[26] Rumelhart, D. E., & Zipser, D. Feature discovery by competitive learning. *Cognitive Science*, 1985, *9*, 75-112.
[27] White, H. Some asymptotic results for learning in single hidden layer feedforward network models. *Journal of the American Statistical Association*, 1989, *84*, 1003-1013.

Linda M. Salchenberger is Assistant Professor of Management Science at Loyola University of Chicago. She received her Ph.D. from Northwestern University in managerial economics and decision sciences. Her research interests include applications of artificial intelligence, including neural networks and expert systems, and she has published in *Information and Management* and the *Journal of Neural Network Computing*.

E. Mine Cinar is Associate Professor of Economics at Loyola University of Chicago. She received her Ph.D. in economics from Texas A&M University. Her fields of interest and research are econometrics and labor economics.

Nicholas A. Lash is Chairperson and Professor, Department of Finance, Loyola University of Chicago. He received his Ph.D. in economics from Wayne State University. He is the author of *Bank Laws and Regulations: An Economic Perspective* and has published articles in *Review of Economics and Statistics* and *Journal of Money, Credit, and Banking*.