EVALUATING CONSUMER LOANS USING NEURAL NETWORKS

Rashmi Malhotra
Assistant Professor of Management Information Systems
E. Haub School of Business
St. Joseph's University
5600 City Avenue
Philadelphia, PA 19131
Phone: (610) 660-3497
E-mail: Rmalhotr@sju.edu

D.K. Malhotra
Associate Professor of Finance
School of Business Administration
Philadelphia University
School House Lane and Henry Avenue
Philadelphia, PA 19144-5497
Phone: (215) 951-2813
Fax: (215) 951-2652
E-mail: MalhotraD@philau.edu

January 2001

EVALUATING CONSUMER LOANS USING NEURAL NETWORKS

Abstract

A number of credit-scoring models that accurately classify consumer loan applications have been developed to aid traditional judgmental methods. This study compares the performance of multiple discriminant analysis and neural networks in identifying potential loan. While there is not a significant improvement in the performance of neural network over discriminant analysis model in identifying good credit loans, the neural network models consistently perform better than the multiple discriminant analysis models in identifying potential problem loans. To alleviate the problem of bias in the training set and to examine the robustness of neural network classifiers in identifying problem loans, we cross-validate our results through seven different samples of the data.

EVALUATING CONSUMER LOANS USING NEURAL NETWORKS

I. INTRODUCTION

According to the Federal Reserve Board, consumer credit includes short- and intermediate-term credit that is extended through regular business channels to finance the purchase of commodities and services for personal consumption, or to refinance debt incurred for such purposes. Automobile loans, home equity loans, secured and unsecured revolving credit, and mobile home loans are among the major types of consumer credit. The total consumer credit outstanding in the United States has grown from $9.8 billion in 1946 to $1,815.31 billion in September 2000.

As consumer credit has soared in recent years, so too have consumer bankruptcies and delinquencies. According to the American Bankers Association, about 1.28 million consumers filed for bankruptcies in 1999—an almost 50 percent increase since 1995. Despite an increase in consumer bankruptcies, competition in the consumer loan market is getting intense everyday. Lenders are using various types of consumer credit risk management systems to evaluate consumer loans and shave off loan losses. These systems employ a variety of techniques: human judgment, credit scoring, decision tree analysis, mathematical programming, and neural networks.

The use of a technique depends on the complexity of the institution, and the size and the type of the loan. Analytical models, such as empirically derived credit scoring systems (based on historical data), use the probability of default to predict the relative creditworthiness of a loan applicant. But, the credit-scoring model does not completely eliminate the human element. The selection of cutoff scores is a subjective decision. Moreover, the evaluation of applicants that have scores between the accept-scores and the reject-scores is quite subjective. Many institutions are exploring the use of artificial intelligence techniques such as artificial neural systems and fuzzy logic.[1] In response to the concern for classification accuracy in consumer loan applications, research is now focused on the use of neural network classification models, and particularly backpropagation neural network (BPN). Neural networks are specialized hardware or software that emulate the processing patterns of the biological brain.[2] Neural networks are non-linear models that classify based on pattern recognition capabilities. Although various studies have highlighted the uses of neural networks in financial analysis for improving forecasting, detecting fraudulent activities, running credit evaluations, and engaging in securities trading, there is recent evidence that the backpropagation may not be the most appropriate neural architecture to use for classification decisions.

Using a pooled data set of twelve credit unions, this study evaluates the classification accuracy of the backpropagation model with adaptive learning to differentiate between "good" loans and "bad" loans.[3] We also compare the performance of the neural network model with the multiple discriminant (MDA) models. The objective of this study is to evaluate the effectiveness of neural networks and/or statistical techniques to assist the loan officer in screening out potential loan defaulters in the credit

---

[1]Expert systems are knowledge-based systems that mimic the behavior of an expert. Many institutions have developed expert systems as an aid to the decision-making process. Unlike standard credit-scoring models, a knowledge-based system aids the lending officer to interactively evaluate the loan applications; However, expert systems lack robustness and flexibility. Hence, the applicability of expert systems to managerial problems is currently highly limited. Furthermore, expert systems, based on the prior knowledge of a few alternatives, are unsuitable to the problem that managers face on a daily basis. Shaw and Gentry (1988) illustrated an expert system called MARBLE (managing and recommending business loan evaluation) for analyzing commercial loans. Duchessi, Shawky, and Seagle (1988) described expert systems called CLASS (COMMERCIAL LOAN ANALYSIS SUPPORT SYSTEM) for evaluating commercial loans.

[2] There are two ways to construct a neural network as an analog to the human brain: a physical analog or a logical analog. A physical analog involves specialized machines that are hardwired to form a neural network system. These systems are expensive, experimental, and largely confined to laboratories. A much less expensive approach is to simulate a neural network on a traditional computer. The logical analog of the neural network involves using the neural network algorithm (software) to solve business problems.

union environment. Secondly, we also investigate the superiority of the neural network models over statistical techniques. Further to check the robustness of the neural network model and to cross-validate our results, we test the neural network classifier over seven different random cross-sections of the data sample.

Our analysis indicates that, although, there is no statistically significant difference in the performance of multiple discriminant analysis and neural networks in classifying good loans, neural networks do a better job of identifying problem loans. With seven different cross-sections of the data sample, the statistical technique of discriminant analysis always under-performed the accuracy of neural network models in identifying problem loans. Our research is also motivated by the fact that a neural network analysis of the same input variables as used by MDA, for the same objective, is possible without any of the circumscription that binds MDA. Neural networks are not subject to such restrictive assumptions of MDA as bias of extreme data points; multivariate normality assumption; and equal group covariance assumptions.

This study is organized into seven parts. Section II reviews previous studies on neural networks in finance. Section III describes the characteristics of the data set used in this study. In addition, section III also details the justification and basis for the methodology used in this study. Section IV discusses the statistical and neural network models used in this study. Section V explains the training of the neural networks. Furthermore, section VI explains the empirical results for loan classification and prediction using the backpropagation model with adaptive learning. Finally, section VII concludes and summarizes the study.

---

[3] We use backpropagation model for identifying potential loan defaulters, because it is the most widely used neural network and is readily accessible to everyone through standardized software packages.

II. PREVIOUS STUDIES

Many studies highlight the use of artificial neural systems in financial analysis. Tam and Kiang (1992) compare the artificial neural system approach with a linear classifier, the logistic regression model, kNN model, and ID 3 model to predict bank failures. They conclude that neural networks are more accurate, adaptive, and robust in comparison to other methods. Swales and Yoon (1992) apply artificial neural networks to differentiate among stocks that perform well and stocks that perform poorly. The neural network model performs significantly better than the linear multiple discriminant models. Coats and Fant (1993) show that neural networks are more effective than multiple discriminant analysis for early detection of financial distress developing in firms. Lacher, Coats, Sharma, and Fant (1995) also use neural networks to estimate the future fiscal health of a corporation. Salchenberger, Cinar, and Lash (1993) use a neural network model to predict the financial health of thrift institutions. The study compares neural network results with the traditional statistical models. The study concludes that the neural network models require fewer assumptions, achieve a higher degree of prediction accuracy, and are more robust. Also, studies by Dutta and Shekhar (1988), and Surkan and Singleton (1991) illustrate the use of neural networks to generate improved risk ratings of bonds. Kimoto, Asakawa, Yoda, and Takeoka (1990) determine optimum buy time and sell time for an equity index using neural networks. Trippi and DeSieno (1992) outline a specific neural network-based intra-day trading system for S&P 500 futures contracts. Hutchinson, Lo and Poggio (1994) price options via learning networks and reported that in many cases the network pricing formula outperforms the

Black-Scholes model. Altman, Marco, and Varetto (1994) compare the performance of artificial neural systems to multiple discriminant analysis to predict the health of a corporation. They find the performance of neural networks inferior to multiple discriminant technique. Using the data from three credit unions, Desai, Crook, and Overstreet (1996) show that the performance of discriminant analysis is comparable to the performance of backpropagation networks in classifying loan applicants into good and bad credit. They indicate that more customized architectures might be necessary for building effective generic models to classify consumer loan applications in the credit union environment. Franses and Van (1998) report that artificial neural networks do not perform well in forecasting the daily exchange rate return relative to Dutch guilder. Plasmans, Verkkooijen, and Daniels (1998) used feedforward artificial neural network specification to investigate the prediction performance of structural and random walk exchange rate models. They do not find any nonlinearity in the monthly data of U.S. dollar rates in Deutsche marks, Dutch guilders, British pounds, and Japanese yen. Anders, Korn, and Schmitt (1998) use statistical inference techniques to build neural network models to explain the prices of call options on the German stock index DAX. They show that statistical specification strategies lead to parsimonious networks that have a superior out-of-sample performance when compared to the Black-Scholes model. Ntungo and Boyd (1998) report that out-of-sample neural network trading returns for corn, silver, and Deutsche mark futures contracts are positive and at about the levels as the returns with ARIMA models. Desai and Bharati (1998) test the efficacy of neural networks in predicting returns on stock and bond indices. They find that the neural network forecasts are conditionally efficient with respect to linear regression models for large stocks and corporate bonds, whereas the evidence is not statistically significant for small stocks and intermediate-term government bonds. Zhang and Hu (1998) illustrate the use of neural networks in forecasting UK pound/U.S. dollar exchange rate. They report that neural networks outperform linear models, particularly when the forecast horizon is short. Zhang, Hu, Patuwo, and Indro (1999) show that neural networks are significantly better than discriminant analysis models in bankruptcy prediction. Indro, Jiang, Patuwo, and Zhang (1999) show that neural networks outperform linear models in forecasting the performance of mutual funds that follow value, blend, and growth investment styles. Thus, all the above-mentioned studies provide mixed evidence regarding the potential of neural networks to analyze, evaluate, and predict many financial systems.

III. DATA AND METHODOLOGY

The data for this study is a pooled data set of loans made by twelve different credit unions with a total of 1078 observations.[4] The data set is divided randomly into a training set of 700 observations and a predict/test set of 380 observations.[5]

The applicants can be categorized into two major groups: applicants who were accepted, and were good credit (Group 1); applicants who were accepted, but defaulted on their loan obligations (Group 2). Further, the data set also includes information such as the applicant's age, housing, address time, income from different sources, total debt,

---

[4] The data set includes loans made by the following twelve credit unions: Jefferson County Teachers; Jefferson County Employees; Family Security; Steering; Etowah Steel Workers; Washington Hill Federal; Riverdale; Delchamps; Lister Hill; Alabama Teachers; Alabama Central; Tenco.

[5] No definite guidelines have been established for dividing the sample into analysis and holdout groups. Between the analysis and holdout samples, some researchers advocate a 60-40 split whereas others prefer a 75-25 split. See *Multivariate Data Analysis* by Joseph F. Hair, and Rolph E. Anderson, and Ronald L. Tatham, second edition, McMillan Publishing, Inc.

monthly rent, mortgage payments, credit union credit card, and credit rating of each applicant. These variables are used to calculate the total income and the total payment of the applicant.[6] In addition, on the basis of the information submitted by the applicant, the ratio of total payment to the total income and the ratio of total debt to the total income are calculated. If the ratio of total payment to the total income of the applicant is high, the risk of loss due to default by the borrower is high. Similarly, the higher the ratio of debt to the total income of the applicant, the higher will be the applicant's credit risk. On the other hand, low ratios of total payment to total income and total debt to total income indicate a good credit applicant. Besides these two ratios, the variable credit rating also reflects the creditworthiness of a loan applicant. Credit unions in the data set assign loan applicants into four credit groups—excellent (1), good (2), marginal (3), and poor (4).

Based on the information supplied by the loan applicant, this study uses six input variables:

1. Home ownership
2. Length of time at current residence (years)
3. Credit card
4. The ratio of the total payment to the total income (ratio 1)
5. The ratio of debt to the total income (ratio 2), and
6. The credit rating of the applicant as the factors that can discriminate between a good and a bad loan.

---

[6] Total payments include payments for mortgage, rent, automobile loan, and other payments. Total income includes income from all sources

To cross-validate the robustness of the neural network classifier, we examine five different cross-sections of the data. In most neural network research, a simple cross-validation scheme that uses a training sample and a test sample is typically utilized. The classification error rate on the test set is then reported as the estimate of the classifier's true error rate. The main problem with this single training and test partition is that the partition may be uncharacteristic of the true underlying population. Multiple partitions of training and test samples are one way to ameliorate this problem. In this study, we employ seven different cross-sections of the data to validate our results. Figure 1 displays the plot of the data space showing the two categories: group 1 (good credit) and group 2 (bad credit).

<Insert Figure 1 about here>

As illustrated by Figure 1, the observations are grouped as two major clusters. There are no distinct clusters for the two categories of the applicants. Each cluster has data points from all classes. Therefore, with overlapping classes, we cannot expect the discriminant analysis model to show a very high prediction rate. Corresponding to two classes, a linear regression model divides the data space linearly into two parts. Thus, a linear regression model cannot efficiently discriminate between the two classes. On the other hand, a neural network model divides the data space into multiple subspaces, and learns the mapping of input parameters and output parameters for these subspaces. The next section illustrates the use of the discriminant analysis model and the neural network models to discriminate between good and bad loan applications. The two models were trained with the training sample, and their performance was tested with the test sample.

IV. MODEL STRUCTURE

A. Discriminant Analysis Model

Discriminant analysis involves the linear combination of the two (or more) independent variables that differentiate best between the a priori defined groups. This is achieved by the statistical decision rule of maximizing the between-group variance relative to the within-group variance; this relationship is expressed as the ratio of between group to within group variance. The linear combinations for a discriminant analysis are derived from an equation that takes the form of equation 1.

$$Z = W_1X_1 + W_2X_2 + W_3X_3 + \ldots\ldots\ldots\ldots + W_nX_n \qquad (1)$$

Where

Z = the discriminant score

W = the discriminant weights

X = the independent variables

Discriminant analysis is the appropriate statistical technique when the dependent variable is categorical and the independent variables are metric. Discriminant analysis helps in testing the hypothesis that the group means of two or more groups are equal. To do so, discriminant analysis multiplies each independent variable by its corresponding weight and adds these products together. The result is a single composite discriminant score for each individual in the analysis. By averaging the discriminant scores for all of the individuals within a particular group, we arrive at the group mean. This group mean is referred to as centroid. When the analysis involves two groups (good credit or bad credit), there are two centroids. The test for the statistical significance of the discriminant function is a generalized measure of the distance between the group centroids. It is computed by comparing the distribution of the discriminant scores for the two groups. If the statistical test indicates that the function discriminates significantly, the next step is to construct classification matrices to provide a more accurate assessment of the discriminating power of the function. To construct classification matrices, we need an optimum cutting score, also called critical Z-value. The cutoff point is selected to minimize the risk of misclassification. In this study, a Type I error occurs when bad credit is classified as a good credit. A type II error occurs when a good credit is classified as a bad credit. For lending decisions, a Type I error is more critical than Type II error. The assumptions for deriving the discriminant function are multivariate normality of the distribution and unknown (but equal) dispersion and covariance structures for the groups. Discriminant analysis minimizes the expected misclassification cost, provided the normality and equal dispersion assumptions are satisfied.

B. Neural Network Models

A neural network model involves constructing computers with architectures and processing capabilities that mimic certain processing capabilities of the human brain. A neural network model is composed of neurons, the processing elements. These elements are inspired by biological nervous systems. Each of the neurons receives inputs, processes the inputs, and delivers a single output. Thus, a neural network model is a collection of neurons that are grouped in layers such as the input layer, the hidden layer, and the output layer. Several hidden layers can be placed between the input and the output layers. The studies illustrated in the previous section have used one to three hidden layers to illustrate the application of artificial neural systems.

A neural network model feeds input variables to the input layer neurons. For this study, the input variables can be attributes such as the applicant's income, debt, and the monthly payments out of the applicant's income. The output of the network is the solution to a problem. For example, in this study, the output can be a good loan or a bad loan. The network assigns numeric values, such as 0 for the good loans and 1 for the bad loans.

To calculate output, an artificial neural network uses weights. Weight numerically represents a connection between two neurons. The weights express the relative importance of each input to a processing element. It is through the repeated adjustment of weights that a neural network learns. The process of learning in a neural network involves computing output, and adjusting the weights in proportion to the difference between the outputs and the desired targets, and repeating the process. By repeating the learning process a number of times, the network identifies the correct values for the weights. The difference between the actual and desired output for a given set of inputs is an error called delta. The objective is to minimize the delta. The network changes the weights to reduce delta. Different artificial neural networks compute the delta in different ways, depending on the learning algorithm/network architecture that is being used.

The backpropagation model can discern patterns in vast pools of data that can help the financial industry predict the performance of equities, corporate bond ratings, credit rating, or corporate bankruptcies. It can also help the industry analyze loan applications and detect credit card fraud. The architecture of backpropagation model with adaptive learning is based on the generalization of the Widrow-Hoff learning rule to multiple-layer networks and non-linear differentiable transfer functions.

From a user's viewpoint, as illustrated in Figure 2, a simple backpropagation network model consists of three layers: the input layer, the hidden layer, and the output layer.

<Insert Figure 2 about here>

The input layer processes the input variables, and provides the processed values to the hidden layer. The hidden layer further processes the intermediate values, and transmits the processed values to the output layer. The output layer corresponds to the output variables of the backpropagation neural network model. The network is trained with a training sample. The training involves repeatedly presenting the input layer with the training sample until the network is able to remember the output of most of the sample items. The backpropagation model develops memory by identifying the relationship between the input variables and the output variables. If the network commits a mistake, the backpropagation algorithm starts at the output layer and propagates the error backward through hidden layers. Therefore, besides the input variables and output variables, an optimal backpropagation model also includes the optimal number of hidden neurons.

To evaluate consumer loans, a backpropagation model with the following features was developed:

a. Six input variables: Whether the applicant owns a house or not, address time, owns a credit card, the ratio of total payment to the total income of the applicant, the ratio of total debt to the total income of the applicant, and the credit rating of the applicant.

b. Hidden nodes, varying as discussed in the training section and the summary of results.

c. One output variable that takes two values: (1) Good Loan (2) Bad Loan.

C. Cross-Validation

To determine the robustness of models, we applied two cross-validation methodologies. Typically, the cross-validation methodology is employed to test the effect of sampling variation on the model performance. In our design of the experiment, we apply a simple validation technique by dividing the data set into a training sample and a holdout sample that tests the predictive effectiveness of the fitted model. As the best model is tailored to fit one sub sample, the model often estimates the true error rate over optimistically. Therefore, to get a true estimate of the error rate, we employ the five-fold cross-validation [Zhang et. al., 1999]. In this study, we adopt two cross-validation methods as suggested by Zhang et. al. (1999). Firstly, we perform a comprehensive analysis of the out-of-sample performance of the two models using a small test set. Secondly, to study the overall predictive capability of the classification models, we use the whole data set as a test set (large test set). To implement the first cross-validation methodology, we divided the data sample into five mutually exclusive and equal sub samples. We designed the neural network and the discriminant classifier models with four sub samples, and tested the models with the fifth remaining sub sample. The experiment was repeated five times with a different testing sub sample each time and the remaining four sub samples as the training data. Therefore, out-of-sample prediction of each sub sample gives us the average classification rate of all the observations in the data set. Secondly, to test the overall predictive capability of the unknown population comprehensively, we should use the entire data sample [Zhang, et. al., 1999]. The best representative of a population is the complete data set as opposed to only one fifth of the population used in the small test methodology. In addition, by using the entire data set as the test sample, we can reduce the sampling variation in the test design because the same sample is tested five different times. Further, the variability across the five test results reflects only the effect of different training samples. Finally, we apply statistical tests to test the two models. We used paired t-test to test the difference between the means of the two classes of models in experimental and cross-validation stages.

The next section illustrates the application of the linear discriminant model and the neural network model to discriminate between good and bad loan applications. The two models were trained with the training sample, and their performance was tested with the test sample.

V. TRAINING THE NETWORK

To adequately train the network, the training sample should be a good representative of the population under study. Neural network models are good at generalization, but cannot extrapolate well. Thus, the training data should cover the entire expected input data space. Further, the network should not be trained completely with input vectors of one class, and then switched to another class; the network will forget the original training. Thus, in accordance with these guidelines, the network was trained with a sample of 700 observations. The training set is an unbiased sample with

350 vectors from the good loan class and 350 vectors from the bad loan class. The statistical details of this sample have been discussed in the previous section. Further, to ensure that the training data covered the entire input space (i.e. learn different characteristics of good and bad loan applicants), the observations were selected from all the credit unions. This prevented the network from learning the characteristics of only one credit union that can be misleading. Moreover, to ensure that the network is not trained with vectors from one class or one single credit union, the observations were intermingled randomly (i.e. they did not show an observable pattern).

Finally, for a network to be efficiently trained, besides a good training sample, a sufficient number of hidden neurons are also essential. Usually, the number of hidden nodes used is a relatively small fraction of the size of the input layer. Although many rules of thumb can be applied to select the size of the hidden layer, but in most of the cases trial and error is the best guide.[7] If the network fails to converge to a solution, then more hidden nodes may be required. And if it does converge, one might try fewer hidden nodes, and settle on a size based on the performance of the system. Therefore, we experimented with neural network models with six neurons in the input layer, one neuron in the output layer by varying the number of hidden neurons. The network was trained with the training sample, and tested with the hold out sample to determine the optimal design. In addition to having an optimal number of hidden neurons, neural network should also be trained with optimal number of training cycles. A network if undertrained, results in underfitting, but if overtrained, results in overfitting. Thus, we also

[7] The following rules are suggested for calculating the number of hidden neurons:
a. Number of hidden neurons = training facts × error tolerance, or
b. Number of hidden neurons = 5 percent to 10 percent of training facts, or
c. Number of hidden neurons = (sum of inputs + sum of outputs)/2.

experimented with different number of training cycles to design the optimal neural network model. The next section discusses the result of our empirical analysis.

VI. EMPIRICAL ANALYSIS[8]

A. Multiple Discriminant Analysis

Table 1 lists the results of the discriminant analysis model and the proportion of application classified correctly by the discriminant analysis model into good and bad loan categories for six different cross-sections of the data set.

<Insert Table 1 about here>

For the training sample, the discriminant model accurately classified 70% to 71% of the 700 applications for all the five data samples. Groupwise, the model correctly classified 75% to 78% of the 350 applications of good loan applicants and 62% to 65% of the 350 applications of bad loan applicants. Similarly, for the hold out sample, the discriminant model accurately classified, on an average, 67% of the 380 applications. Groupwise, the model correctly classified 74.39% of good loan applicants and 59.56 of the bad loans for all the seven data samples. Thus, although well trained, the discriminant analysis model is not very efficient in identifying potential loan defaulters when applied to the test sample.

B. The Backpropagation Model with Adaptive Learning

Table 2 summarizes the results of the empirical analysis of the credit union environment using a one hidden layer in a backpropagation neural network model.

<Insert Table 2 about here>

[8] Discriminant analysis routines were performed using SPSS program. Neural network routines were performed using MATLAB.

The single hidden layered backpropagation network was trained on 700 applicants (350 good credit and 350 bad credit) for five different cross sections of the data for varying cycles and varying hidden neurons.[9] After training, the network screened 72 to 85 percent of the good loan applicants and 61 to 76 percent of the bad loan applicants accurately, with an overall classification accuracy of 70 to 77 percent. In terms of generalizations, the network provided a prediction accuracy of 75.4% on an average for good loan applicants and 64.80 percent on an average for bad loan applicants for all the seven data samples. The overall classification accuracy is 70.10% on an average (190 good loan and 190 bad loan applicants). Therefore, as in the case of discriminant analysis model, the single hidden layered backpropagation network trains well with an overall prediction rate of 80 percent in identifying good and bad loans. However, with the holdout sample, there is an improvement in the performance of the neural network over the traditional statistical method, in identifying problem loans. In all the seven data samples discussed in this study, neural network model performed consistently better than the discriminant model in identifying problem loans.

C. Discriminant Analysis versus Neural Networks

We performed statistical tests to determine if the average classification accuracy of discriminant models is the same as the average classification accuracy of the neural network models. Table 3 displays the results of the paired t-test for the two models. We do not find any statistically significant difference between the average predictive performance of artificial neural systems and the discriminant analysis models. For "bad" loans the average performance of the neural network model is 64.79%, which is higher

[9] The network was also trained with varying degrees of hidden neurons and different number of training cycles. Optimum number of hidden neurons and training cycles were determined through trial and error approach.

than the discriminant analyzer's classification rate of 59.56%. A paired t-test shows that there is a statistically significant difference in the superiority of artificial neural network model over discriminant analysis in identifying problem loans. On the other hand, discriminant analyzer's average performance in classifying "good" loans is 74.39%, marginally lower than the neural network classifier's rate of 75.4%.

<Insert Table 3 about here>

As shown in Table 3, on an average, the overall performance of neural networks is better than the average performance of discriminant analysis and the difference is statistically significant at 1% rejection level in favor of neural networks. In addition, neural network is a better technique, because the analyst does not need the restrictive assumption of multivariate normality of the distribution and unknown (but equal) dispersion and covariance structures for the groups. Discriminant analysis minimizes the expected misclassification cost provided the normality and equal dispersion assumptions are satisfied.

D. Cross-Validation

As illustrated before, we used two cross-validation methodologies: cross-validation using small test sets on five equal and mutually exclusive sub samples and cross-validation using the large test set on the five experimental models. Table 4 lists the results of the cross-validation experiment using small test sets.

<Insert Table 4 about here>

As illustrated in Table 4, neural network models are more robust in contrast to the discriminant analysis models in predicting the overall classification rate in the credit union environment. Across the five small test sub samples, the classification rates of the two models show a clear pattern in favor of neural network models. For the entire sub-samples neural network models are better than the discriminant analysis models. Table 5 displays the results of the paired t-test for the two models. We find statistically significant difference between the average predictive performance of neural network models and the discriminant analysis models. For "bad" loans the average performance of the neural network model is 66.73%, which is positively higher than the discriminant analyzer's classification rate of 61.14% and the difference is statistically significant. Furthermore, the discriminant analyzer's average performance in classifying "good" loans is 74.58%, lower than the neural network classifier's rate of 75.51%; however the difference is not statistically significant. In addition, there is statistically significant difference between the overall performance of neural networks and discriminant analysis model in favor of neural networks.

<Insert Table 5 about here>

Table 6 displays the results of the cross-validation experiments for the large test set. For each experiment we used the entire data sample as the large test set. Therefore, we computed the classification rates in Table 6 by combining the results for the training samples and test samples of Table 4.

< Insert Table 6 about here>

For the large test set, neural network models have higher classification rates for "bad" loans and the overall sample. For "good" loans, the "bad" loans and the combined data set (overall) neural network models perform consistently better the discriminant analysis models. Table 7 shows the results of the paired t-test to compare the average performance of the neural network models with the discriminant analysis models. The average performance of the neural network models is greater than the discriminant analysis models in all the three categories: "good" loans (77.36% vs. 75.64%), "bad" loans (66.61% vs. 62.75%), and overall sample (71.98% vs. 69.32%). Table 7 displays the results of the paired t-test to compare the average performance of the two models. For the overall sample and the "bad" loans, there is a statistically significant difference between the performance of the neural network models and the discriminant analysis models. There is statistically significant difference between the two models in predicting "good" loans at 5% level of rejection in favor of neural network models.

<Insert Table 7 about here>

The performance of the two models over small test samples varies from the large test samples due to many reasons. Firstly, for each sub sample in the small test experiment we have a different test sample, whereas for the large test sample experiment, we use the same test sample. Thus, the variability in the predictive performance across the five large test samples is much smaller than that of the small test samples. Hence, the variability in the test results reflects the differences in the training set. Secondly, the neural network models have higher classification accuracy for the large test sample runs as compared to the small test sample runs, while the performance of the discriminant analysis models is the same for the small test samples as well as the large test samples. The rationale behind these differences is that neural network models train better than the discriminant analysis models. Table 8 presents the learning rates of the five training sessions. The learning rate of the neural network models is higher than the discriminant analysis model for "good" loans, "bad" loans, and the overall sample. Table 9 presents the results of the paired-t test to compare the average performance of the neural network model and the discriminant analysis model for the training sessions. The difference in the classification rates of the neural network models and the discriminant analysis models for "bad" loans and the “good” loans is statistically significant at the 0.02 levels. For the overall sample, the performance of neural network is superior to the discriminant analysis model at 0.001 levels.

<Insert Tables 8 & 9 about here>

VII. SUMMARY AND CONCLUSIONS

Contribution of this Study

Although mathematically derived credit-scoring systems (analytic) are a definite improvement over subjective judgmental (intuitive) methods to evaluate consumer loans, we can improve the method of credit scoring considerably through the use of artificial intelligence techniques such as expert systems, neural networks, fuzzy logic, and neural network systems. While expert systems have been extensively used by many organizations, neural networks have captured the attention of the business community only recently. In this study, we applied the neural network models to the credit union environment. This study explored the performance of neural network model with adaptive learning algorithm to identify problem loans from among the applications the loan officer has initially screened. Field data from 12 case organizations (credit unions) was analyzed to evaluate the performance of the neural network models against the multiple discriminant analysis models.

To assess the performance of neural network and discriminant analysis models in distinguishing potentially "good" credit loans from "bad" credit loans, we trained both set of models with a training sample of 700 observations, and tested them with a holdout sample of 380 observations. The neural network models outperformed the discriminant model in identifying potential loan defaulters and, therefore, in minimizing the Type I error. Between the two models, the single-layered Backpropagation model with adaptive learning showed better performance in comparison to multiple discriminant analysis in identifying problem loans.

Although the study reported here is far from sufficient to generate any conclusive statements about the applicability of neural network models in general, it does provide some insights into their potentials and limitations. Based on the comparison reported above, the neural network approach offers a comparative alternative to classification techniques especially under the following conditions:

Multi-modal distribution: The nonlinear discrimination function represented by a neural network model provides a better approximation of the sample distribution, especially when the distribution is multi-modal. A loan officer may have higher classification accuracy than the linear logistics model because of his/her ability to relate variables and loan outcome in a nonlinear manner. Therefore, human judgements can better be approximated by nonlinear, multi-modal neural network methods.

Adaptive model adjustment: The neural network models can be easily adjusted in an adaptive manner by modifying network weights and the learning rate. Therefore, neural network models are able to respond swiftly to changes in the real world.

Robustness: A neural network does not assume any probability distribution or equal dispersion. Further, besides continuity and differentiability, there is no rigid restriction on the use of input/output functions in a neural network. Therefore, simple and universal training (learning) algorithms function independently of the number of inputs. Data having repeated, wrong, or missing value(s) is easily accommodated. The size of input data arrays being considered in parallel is not subjected to any artificially imposed limits. Very large problems can be solved in the absence of the kind of detailed information that is essential to support traditional divide-and-conquer or rule-based approaches. Using neural networks, it is not necessary to first make an analysis based on a detailed knowledge of a problem or a system's internal structure. Thus, neural network models are more robust than the traditional discriminant analysis model.

Limitations and Implications for Researchers and Practitioners

In our study, we evaluated the robustness of neural network models to identify potential loan defaulters in the credit union environment by comparing them with the multiple discriminant analysis model. Since the neural network model has performed as well or better than the logistics model, neural networks may offer a competitive modeling approach for loan evaluations. But, there are several limitations that may restrict the use of neural network models for classifications. The following are the limitations of neural network methods:

Network Topology: There is no formal theory to determine optimal network topology for a given classification application. Therefore, decisions such as the appropriate number of layers and the size of the middle layer nodes must be determined using experimentation. In addition, there is a possibility of underfitting or overfitting the network, and the results are sensitive to the selection of learning parameters. Poor results can also occur if the wrong transfer function is selected. Thus, the development and interpretation of neural network models requires more expertise from the user than traditional statistical models.

Computational Efficiency: Training a neural network can be computationally intensive. In our study the computation time of the neural network models was a few minutes to 1 hour on Pentium II mini workstation. On the other hand, all statistical methods took at most half a minute on the same machine.

Symbolic Expression: The discriminant capability of a neural network model is difficult to express in symbolic form. This may not be a serious drawback if one is concerned with predictive accuracy only. However, a neural network is limited if one wants to test the significance of individual inputs.

Explanatory Capability: Neural networks cannot explain how and why they identified a "bad" loan application. Hence, this inability to explain conclusions may restrict the use of the neural network modeling technique. This is in contrast to expert systems that can provide explanations to the user about how inferences are made.

Finally, in the light of the case analysis carried out, neural networks are a very interesting tool and have great potential capacities that undoubtedly make them attractive to the field of business classification. The following are the implications of this study for the academicians and practitioners:

- Loan evaluation is a non-trivial and important problem. A better understanding of the causes will have tremendous financial and managerial consequences. We have presented a general framework for understanding the role of neural networks for this problem. While traditional statistical methods work well for some situations, they may fail miserably when the statistical assumptions are not met. Neural network models are a promising alternative tool that should be given much consideration when solving problems like loan evaluation

- Neural network models offer a useful data mining and analytical capability through which financial practitioners can discover meaningful relationships and hidden patterns in the data. Neural network models, unlike the statistical methods, are not constrained by any statistical assumptions about the characteristics of the data. Furthermore, the designer of the network has the flexibility to add any number of layers between the input and output layers or to change the size of the layers (number of hidden neurons) depending on the optimal solution of the problem. Besides, easily available, off-the-shelf neural network software such as Smartlearn, Brainmaker, Neurolab, etc., offer friendly graphical user interfaces to discover meaningful patters in the data.
- In this study we used the existing theoretical neural network models to screen loans in the credit union environment. Researchers can further explore the process of designing neural network models to develop new neural network models (combination of the theoretical neural network models) most suitable for the credit union environment.
- Another area of research, researchers and practitioners should focus on is the application of genetic algorithms to the design of network configurations to differentiate between "good" credit loans and "bad" credit loans. Genetic algorithms adopt an evolutionary approach where a pool of networks, called the population, is continuously being modified by using genetic operators such as crossover and mutation [Goldberg, 1989]. Each synthesized network, which corresponds to a possible configuration, is evaluated using the backpropagation algorithm. Genetic algorithms have a built-in-bias towards retaining and combining good configurations

in the next generation. The evolutionary nature of the algorithm enables the search for good configurations to proceed in a parallel fashion, thus reducing the possibility of trapping in local optimal configuration.

- Our implementation of the neural network models is computer simulation software that runs on a double processor Pentium II workstation. As the intrinsic parallel processing capability of the network was not exploited, the computation time is quite long, especially for the Backpropagation with Levenberg-Marqurdt approximations. Managers should use neural network models etched on silicon to reduce computation time.
- As explained above, neural networks are better for credit loan evaluation only if we are not interested in knowing how a particular conclusion was made. Thus, the limited explanatory capability of neural networks can be improved by linking them with fuzzy logic to form neurofuzzy systems. Therefore, another area for conducting research is the application of neurofuzzy systems to the credit union environment. Fuzzy logic provides a means of combining symbolic and numeric computations in inference processing. The linkage between neural networks and symbolic reasoning can be established through the membership function of fuzzy logic. The membership function measures the degree of possibility of a concept as related to a numeric quantity. A neural network can be used to synthesize a membership function by training it with instances of the relation [Jang, Sun, & Mizutani, 1997].
- For loan evaluation, a neural network may represent the membership function of the concept "high risk of default." Such a representation can easily be combined with other symbolic conditions appearing in the rules of an expert system. By utilizing

neural networks as frontends in rules definition, one can take advantage of the explanatory capability of expert systems as well as the subsymbolic computational capability offered by neural networks.

- Another area that researchers and practitioners can explore is the design and development of a decision support system that embeds the subjective (judgmental) models and objective (analytic) models to aid the loan officer in making a decision on a loan application in a credit union environment.
- Another area that academicians and practitioners should pursue is the integration of neural networks and statistical techniques. Neural networks have great potential as a forecasting tool. Thus, evaluating the performance of integrating neural networks, with statistical techniques to address finance problems, is likely to provide fruitful opportunities in the future.

Finally, neural networks have shown enough promising features to provide an incentive for more thorough and creative testing. As advancements are made in AI technologies and computer-based related systems, there should be new opportunities to apply neural network technology for finance research.

REFERENCES

Altman, E., G. Marco, and F. Varetto. "Corporate Distress Diagnosis: Comparisons Using Linear Discriminant Analysis and Neural Networks." *Journal of Banking and Finance*, 18, May 1994, 505-529.

Anders, U., O. Korn, and C. Schmitt. "Improving the Pricing of Options: A Neural Network Approach." *Journal of Forecasting,* 17, 5, September/November 1998, 369-388.

Coats, P. and F. Fant. "Recognizing Financial Distress Patterns Using a Neural Network Tool." *Financial Management,* 22, 3, Autumn 1993, 142-155.

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

Desai, V. and R. Bharati. "The Efficacy of Neural Networks in Predicting Returns on Stock and Bond Indices." *Decision Sciences,* 29, 2, 405-425.

Desai, V., J. Crook, and G. Overstreet. "A Comparison of Neural Networks and Linear Scoring Models in Credit Union Environment." *European Journal of Operations Management,* 95, 1, November 1996, 24-37.

Duchessi, P., H. Shawky, and John P. Seagle. "A Knowledge-Engineered System for Commercial Loan Decisions." Financial Management, 17, Autumn 1988, 57-65.

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

Franses, P. and P. Van. "On Forecasting Exchange Rates Using Neural Networks." *Applied Financial Economics,* 8, 6, December 1998, 589-596.

Freeman, J. and David M. Skapura. *Neural Networks: Algorithms, Applications, and Programming Techniques*, Addison-Wesley Publishing Company, 1991.

Hinton, G., and T. Sejnowski 1987. *Neural Network Architectures for AI.* Tutorial No. MP2, AAAI87, Seattle, WA, July 1987.

Hutchinson, J., A. Lo, and T. Poggio. "A Nonparametric Approach to Pricing and Hedging Derivative Securities Via Learning Networks." *Journal of Finance,* 49, 3, July 1994, 851-889.

Indro, D., C. Jiang, B. Patuwo, and G. Zhang. "Predicting Mutual Fund Performance Using Artificial Neural Networks." *Omega,* 27, 3, 373-380.

Kimoto, T., K. Asakawa, M. Yoda, and M. Takeoka. "Stock Market Prediction System with Modular Neural Networks." *Proceedings of the International Joint Conference on Neural Networks*, Vol. 1, San Diego, IEEE Network Council, 1990, pp. 1-6.

Lacher, R., P. Coats, S. Sharma, and F. Fants. "A Neural Network for Classifying the Financial Health of a Firm." *European Journal of Operations Management,* 85, 1, August 1995, 53-65.

Ntungo, C. and M. Boyd. "Commodity Futures Trading Performance Using Neural Network Models versus ARIMA Models." *Journal of Futures Markets,* 18, 8, December 1998, 965-983.

Plasman, J. W. Verkkooijen, and H. Daniels. "Estimating Structural Exchange Rate Models by Artificial Neural Networks." *Applied Financial Economics,* 8, 5, October 1998, 541-551.

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

Shaw, M. and J. Gentry. "Using an Expert System with Inductive Learning to Evaluate Business Loans." *Financial Management*, 17, 3, Autumn 1988, 45-56.

Surkan, A., and J. Singleton. "Neural Networks for Bond Rating Improved by Multiple Hidden Layers." *Proceedings of IEEE International Conference on Neural Networks*, July 1991, pp. II-157-162.

Swales, G., and Y. Yoon. "Applying Artificial Neural Networks to Investment Analysis." *Financial Analysts Journal*, September/October 1992, pp. 78-80.

Tam, K. Y. and M. Kiang. "Managerial Applications of Neural Networks: The Case of Bank Failure Predictions." *Management Science*, Vol. 38, No. 7, July 1992, 926-947.

Trippi, R. and D. DeSieno. "Trading Equity Index Futures with a Neural Network." *Journal of Portfolio Management*, Fall 1992, pp. 27-33.

Turban, E. and J. Aronson. *Decision Support Systems and Intelligent Systems.* Prentice Hall, New Jersey, 1998.

Yam, K. and Melody Y. Kiang. "Managerial Applications of Neural Networks: The Case of Bank Failure Predictions." *Management Science*, Vol. 38, July 1992, pp. 926-947.

Yoon, Y. and G. Swales. "Predicting Stock Price Performance: A Neural Network approach." *Proceedings of the 24th Annual Hawaii International Conference on* *systems Sciences*, Vol. IV. Hawaii: IEEE Computer Society Press, 1991, pp. 156-162.

Zhang, G., M. Hu, B. Patuwo, and D. Indro. "Artificial Neural Networks in Bankruptcy Prediction: General Framework and Cross-Validation Analysis." *European Journal of Operations Research,* 116, 1999, 16-32.

Zhang, G. and M. Hu. "Neural Network Forecasting of the British pound/U.S. dollar Exchange Rate." *Omega,* 26, 4, January 1999, 495-506.

Table 1
Classification of Consumer Loan Applications Using Discriminant analysis
The independent variables included the housing, address time, credit card, ratio of total payments to the total income of the applicant, the ratio of total debt to total income of the applicant, and the credit rating of the applicant.

Actual Group	No. of Cases	Accuracy of Predicted Group Membership						
A: Training Set								
		Sample 1	*Sample 2*	*Sample 3*	*Sample 4*	*Sample 5*	*Sample 6*	*Sample 7*
Good Loan	350	77.1%	78.3%	77.7%	76.6%	75.2%	84.6%	77.4%
Bad Loan	350	65.7%	62.3%	62.3%	64.6%	66.5%	57.7%	64.5%
Overall	700	71.4%	70.3%	70.0%	70.6%	70.8%	71.1%	71.0%
B: Test Set								
Good Loan	190	73.6%	74.1%	79.8%	75.1%	73.9%	84.1%	70.3%
Bad Loan	190	61.7%	51.8%	60.6%	63.2%	62.2%	51.3%	66.1%
Overall	386	67.6%	63.0%	70.2%	69.2%	68.1%	67.7%	68.2%

Table 2
Classification of Consumer Loan Applications Using Backpropagation with Adaptive Learning Neural Network Model.
Six input variables included housing, address time, credit card, ratio of total payments to the total income of the applicant, ratio of total debt to the total income of the applicant, and the credit rating of the applicant. The learning rate was 0.01 and the momentum was 0.02.

Actual Group	No. of Cases	Accuracy of Predicted Group Membership						
A: Training Set								
		Sample 1	*Sample 2*	*Sample 3*	*Sample 4*	*Sample 5*	*Sample 6*	*Sample 7*
Good Loan	350	78.57%	74.29%	84.57%	74.29%	71.71%	80.86%	78.86%
Bad Loan	350	74.86%	66.57%	60.86%	70.00%	76.00%	60.86%	70.0%
Overall	700	76.72%	70.43%	72.72%	72.15%	73.86%	70.86%	74.43%
B: Test Set								
Good Loan	190	72.53%	74.09%	74.09%	74.61%	71%	82.01%	69.31%
Bad Loan	190	67.35%	63.73%	67.35%	65.29%	67%	53.4%	69.31%
Overall	380	69.94%	68.91%	70.72%	69.73%	69%	67.71%	69.31?%

Table 3: Pairwise Comparison between Neural Network Model and Discriminant Analysis Model for Seven Samples

Statistics	Overall		Good Loans		Bad Loans	
	Neural Networks	Discriminant	Neural Networks	Discriminant	Neural Netwroks	Discriminant
Mean	70.10	66.97	75.4	74.39	64.79	59.56
t-Statistic	3.42		1.03		4.06	
p-value[c]	0.007		0.17		0.00	

c. one-tail

Table 4: Cross-Validation Results on the Predictive Performance of Neural Networks and Multiple Discriminant Analysis for Small Sample Tests

Method	Sub sample 1[b]			Sub sample 2			Sub sample 3			Sub sample 4			Sub sample 5		
	Good	Bad	Overall	Good	Bad	Overall	Good	Bad	Overall	Good	Bad	Overall	Good	Bad	Overall
Neural Network Model	79.44%	62.62%	71.03%	75.7%	66.36%	71.03%	72.9%	65.42%	69.16%	73.83%	70.09%	71.96%	75.7%	69.16%	72.43%
Discriminant Analysis	79.4%	59.8%	69.6%	76.6%	58.9%	67.8%	72.0%	63.6%	67.8%	74.8%	64.5%	69.6%	70.1%	58.9%	64.5%

a. The models were trained on 816 observations ("Good" Loans: 408 and "Bad" Loans: 408) and tested on a total of 216 observations ("Good" Loans: 108 and "Bad" Loans: 108). The number in the table is the number of correctly classified; percentage is given in the brackets.
b. "Good" loans belong to Group 1and "Bad" loans belong to Group 2.

Table 5: Pairwise Comparison between Neural Network Model and Discriminant Analysis Model for Small Sample Test

Statistics	Overall		Good Loans		Bad Loans	
	Neural Networks	Discriminant	Neural Networks	Discriminant	Neural Networks	Discriminant
Mean	66.73	61.14	75.51	74.58	71.12	67.86
t-Statistic	3.64		0.77		2.68	
p-value[c]	0.01		0.24		0.02	

c. one-tail

Table 6: Cross-Validation Results on the Predictive Performance of Neural Networks and Multiple Discriminant Analysis for Large Sample Test

Method	Sub sample 1[b]			Sub sample 2			Sub sample 3			Sub sample 4			Sub sample 5		
	Good	Bad	Overall	Good	Bad	Overall	Good	Bad	Overall	Good	Bad	Overall	Good	Bad	Overall
Neural Network Model	81.95%	61.75%	71.85%	76.97%	66.63%	71.8%	74.65%	67.55%	71.1%	77.31%	66.41%	71.86%	75.93%	70.69%	73.31%
Discriminant Analysis	78.95%	60.35%	69.65%	77.05%	61.75%	69.4%	73.6%	65.25%	69.43%	76.65%	63.05%	70.08%	71.95%	63.35%	67.65%

a. The number in the table is the number of correctly classified; percentage is given in parentheses.
b. "Good" loans belong to Group 1and "Bad" loans belong to Group 2.

Table 7: Pairwise Comparison between Neural Network Model and Discriminant Analysis Model for Large Sample Test

Statistics	Overall		Good Loans		Bad Loans	
	Neural Networks	Discriminant	Neural Networks	Discriminant	Neural Networks	Discriminant
Mean	71.98%	69.32%	77.36%	75.64%	66.61%	62.75%
t-Statistic	3.53		2.27		3.69	
p-value[c]	0.01		0.04		0.01	

c. one-tail

Table 8: Comparison of Neural Network Model and Discriminant Analysis Model for the Training Sample[a]

Method	Sub sample 1[b]			Sub sample 2			Sub sample 3			Sub sample 4			Sub sample 5		
	Good	Bad	Overall	Good	Bad	Overall	Good	Bad	Overall	Good	Bad	Overall	Good	Bad	Overall
Neural Network Model	84.49%	60.88%	72.68%	78.24%	66.90%	72.57%	76.39%	69.68%	73.04%	80.79%	62.73%	71.76%	76.16%	72.22%	74.19%
Discriminant Analysis	78.5%	60.9%	69.7%	77.5%	64.6%	71.1%	75.2%	66.9%	71.1%	78.5%	61.6%	70.0%	73.8%	67.8%	70.8%

a. The number in the table is the number of correctly classified; percentage is given in the brackets.
b. "Good" loans belong to Group 1and "Bad" loans belong to Group 2.

Table 9: Pairwise Comparison between Neural network Model and Discriminant Analysis Model for the Training Sample

Statistics	Overall		Good Loans		Bad Loans	
	Neural Networks	Discriminant	Neural Networks	Discriminant	Neural Networks	Discriminant
Mean	72.85%	70.54%	79.21%	76.7%	66.48%	64.36%
t-Statistic	6.21		2.72		2.82	
p-value[c]	0.001		0.03		0.02	

c. one-tail