Information Technology and Management 2, 5–26, 2001

Supporting Ordinal Four-State Classification Decisions Using Neural Networks

ANURAG AGARWAL

Department of Decision and Information Sciences, Warrington College of Business Administration, University of Florida, Gainesville, FL 32611-7169, USA
E-mail: aagarwal@ufl.edu

JEFFERSON T. DAVIS

School of Accountancy, College of Business & Economics, Weber State University, Ogden, UT 84408-3803, USA
E-mail: jtdavis@weber.edu

TERRY WARD

Department of Accounting, College of Business, Middle Tennessee State University, Murfeesbro, TN 37132, USA
E-mail: tward@frank.mtsu.edu

Abstract. Many accounting and finance problems require ordinal multi-state classification decisions, (e.g., control risk, bond rating, financial distress, etc.), yet few decision support systems are available to aid decision makers in such tasks. In this study, we develop a Neural Network based decision support system (NN-DSS) to classify firms in four ordinal states of financial condition namely healthy, dividend reduction, debt default and bankrupt. The classification results of the NN-DSS model are compared with those of a Naïve model, a Multiple Discriminant Analysis (MDA) model, and an Ordinal Logistic Regression (OLGR) model. Four different evaluation criteria are used to compare the models, namely, simple classification accuracy, distance-weighted classification accuracy, expected cost of misclassification (ECM) and ranked probability score. Our study shows that NN-DSS models perform significantly better than the Naïve, MDA, and OLGR models on the ECM criteria, and provide better results than MDA and OLGR on other criteria, although not always significantly better. The effect of the proportion of firms of each state in the training set is also studied. A balanced training set leads to more uniform (less skewed) classification across all four states, whereas an unbalanced training set biases the classification results in favor of the state with the largest number of observations.

Key Words: decision support systems, neural networks, ordinal multi-state classification, financial distress

1. Introduction

Although a number of decision problems in accounting and finance require ordinal multi-state classification (e.g., bond rating, control risk assessment and financial distress), few decision support systems (DSSs) exist in the literature to aid such decision problems. Most extant studies in classification deal with two-state classification. For the financial distress problem, for example, studies abound [1–3,5,7,10,11,16,19,23,24] for classification into bankrupt and non-bankrupt firms. Few studies, however, attempt classification on multiple, rank-ordered states of financial health [4,14,21]. The multi-state classification task is clearly more complex than two-state classification for a number of reasons. First, identifying well-defined, non-overlapping states presents a problem. Second, as the number of states increases, the quantitative and qualitative difference between adjacent states decreases, hence finding significant financial variables that discriminate between adjacent states becomes difficult. Third, data collection for building and testing models becomes more complex, and finally, measuring the performance of a model for a multi-state classification task is non-trivial.

Two-state classification models of financial distress do not provide much value to investors, as these models do not distinguish between firms in the non-bankrupt category. All non-bankrupt firms are treated as being equally healthy, which is not true. Managers benefit little from such models since these models fail to provide an early warning of deteriorating financial health. On a spectrum of financial health, with very healthy firms on one end and firms filing for bankruptcy on the other, several states of financial health can be identified. Classification of firms into these multi-states will provide valuable information to both managers and investors.

Lau [14] identified five ordinal states of financial condition and developed a Logit model to classify firms in those states. The five states, in the order of healthy to unhealthy, are:

(1) financial stability,

(2) omitting or reducing dividend payments,

(3) defaulting on loan principal or interest or debt accommodation such as extension of cash payment schedules, reduction in principal, or reduced interest rates,

(4) filing for protection under Chapter 11, and

(5) filing for liquidation under Chapter 7.

Ward [21,22] used four states of financial condition for classifying firms using Ordinal Logistic Regression (OLGR). He simply collapsed the bankruptcy and liquidation filings (Lau's states (4) and (5)) into one state of bankruptcy (state (4)). It may be argued that companies omitting dividends may be growth companies and in fact healthy. However, according to Lau [14] a firm that reduces dividends is typically encountering some financial difficulty. This correlation between dividend reduction and financial distress is further supported by empirical studies [6,10]. Therefore, Lau [14], Ward [21] and Ward and Foster [22] used "dividend omission or reduction" to represent a financial condition between states (1) and (3).

In recent years, several Neural Networks (NNs) models have been successfully used for developing two-state bankruptcy prediction models [19,20,23]. These studies have shown NNs models to outperform the traditional statistical models, such as MDA and Logit, in terms of their classification accuracy. Two studies have used NNs for multi-state classification. Zurada et al. [25,26] report results of NNs multi-state models in comparison with OLGR and find that NNs provide better results in terms of classification accuracy. Barniv et al. [4] also report better NNs results than OLGR and Non-parametric Discriminant Analysis models for a three-state post-bankruptcy filing prediction problem.

This study develops a NNs based decision support system, henceforth NN-DSS, for four-state classification of financial distress, using the same four states as used by Ward [21]. NN-DSS predicts the probability for each of the four states for a given firm. The state with the highest probability is the most likely state for the firm. We compare the classification results of NN-DSS with those of Multiple Discriminant Analysis (MDA), OLGR, and Naïve models using four different evaluation criteria and find that NN-DSS outperforms other models on most criteria. In addition, we study the effect of proportion of firms in each state in the training sample on model performance. We develop each of our four models (NN-DSS, Naïve, MDA and OLGR) twice, once with an unbalanced training sample and once with a balanced training sample. In the unbalanced sample the proportion of firms in each state is unequal whereas in the balanced training sample they are equal.

The present study extends the work of Zurada [25] and Barniv et al. [4] in several ways. First, it takes into account the ordinality of the four decision states, whereas Zurada et al.'s [25,26] did not. Second, NN-DSS provides state-probabilities for the four states. Previous studies either did not develop them or developed them for three-states. Since ordinal MDA and OLGR provide state-probabilities, having these probabilities for NN-DSS makes it easier to compare its results with those of MDA and OLGR. Third, we develop a scheme for misclassification costs for the four-state problem, which has not been done in the past. Fourth, we compare NN-DSS results with OLGR, MDA and Naïve models. Previous studies have not developed MDA models for multi-state problems. Fifth, this study uses four different evaluation criteria for comparing each of the models, which not all of the previous studies did. In addition to the simple classification accuracy, the study uses distance-weighted classification accuracy, expected cost of misclassification, and ranked probability scores. These evaluation criteria will be explained in detail in a later section. Finally, this research tests the effect of balanced verses unbalanced training sets for model development, something not previously studied for multi-state classification tasks.

We use the same initial sample as in Ward [21] and Zurada et al. [25,26] for our study. Although this study uses a financial distress problem, the issues addressed are applicable to all multi-state decision tasks in general. To summarize, the issues addressed are (i) using appropriate NN architectures, (ii) developing state probabilities for multi-state settings, (iii) using various evaluation criteria, and (iv) observing the effect of proportion of firms in each state on model performance.

The rest of the paper is organized as follows. Section 2 discusses issues that arise with multi-state classification problems. These issues need to be understood to better understand the nature of evaluation criteria, which are also discussed in the same section. The model development is discussed in section 3, and the results are presented in section 4. Section 5 provides a summary, highlights limitations of this study and discusses future research ideas.

2. Issues with ordinal multi-state classification

Several issues arise in developing and evaluating multi-state classification models. These issues include: (1) obtaining state-probabilities, (2) describing types of misclassifications, and (3) developing evaluation criteria for the performance of the models.

2.1. State-probabilities

Lau [14] suggested that for multi-state problems with more than two states, it seems appropriate to provide state-probabilities in addition to state classification. A firm is most likely to be in the state with the highest state-probability, although other states have some likelihood of occurrence. A model that produces state-probabilities can be viewed as a decision aid to the manager. For example, if state-probabilities for a firm are almost equal for two or more states, the decision maker can look further into that firm. On the other hand, if state-probability for a firm is clearly high for a particular state, the decision maker may simply use that state as the predicted state.

Most two-group classification NNs models have, in the past, produced classifications without giving state-probabilities. One recent NNs study [4] used NNs for three-group classification and developed state-probabilities. We modify the state-probability formulas to fit our four-state model. The output of a NN, which represents the value of the dependent variable, in this case the state, needs to be transformed into state-probabilities. Since the output score is a continuous variable, it seldom equals the exact state value. Suppose the state values are 1, 2, 3 and 4, and suppose the generated score is 2.3. This score of 2.3 can be transformed into state probabilities using formulas derived in appendix and presented below. The state-probabilities are based on the assumption that they are inversely proportional to the relative distance of the output score from state-centroids and that the combined probability for all four states should add to one. For example, a score of 2.5 would have an equal probability of being in states 2 and 3. It will also have a smaller, but equal probability of being in states 1 and 4 such that the total of four probabilities adds up to 1. The following formulas are used for state-probabilities:

$$P(1) = \frac{1}{1 + d_1/d_2 + d_1/d_3 + d_1/d_4},$$
$$P(2) = \frac{1}{1 + d_2/d_1 + d_2/d_3 + d_2/d_4},$$
$$P(3) = \frac{1}{1 + d_3/d_1 + d_3/d_2 + d_3/d_4},$$
$$P(4) = \frac{1}{1 + d_4/d_1 + d_4/d_2 + d_4/d_3},$$

where $P(i)$ is the probability that a firm belongs to state i (i.e., state-probability), d_i is the relative distance from the centroid of state *i*. See appendix for a derivation of the above formulas. Barniv et al. [4] used similar state-probability formulas for their three-state problem.

2.2. *Types of misclassifications*

Type-I and Type-II errors for the two-state classification problem are well known. If a bankrupt firm is misclassified as healthy, a Type-I error occurs. Conversely, if a healthy firm is misclassified as bankrupt, a Type-II error occurs. In a four-state model, misclassifications can occur in twelve different ways because each of the four states can be misclassified in three possible ways. From an investor's point of view, for the two-state problem, a Type-I error is more costly than a Type-II error. Similarly, for the four-state problem, each of the twelve error types will have different relative costs. If misclassifications are denoted as Type$(i|j)$ errors, where i is the predicted state, j is the actual state, and i is not equal to j, then in general, the cost of misclassification will be a function of (i) the distance between i and j, and (ii) the sign of (i *minus* j). For the case of two-state classification, cost ratios of $10:1$, $20:1$, $50:1$, etc. have been used in the literature [2,3], although there is little or no consensus among researchers as to the relative costs of misclassifications between Type-I and Type-II errors. This study presents three possible sets of relative costs of the twelve possible misclassifications, based on (a) intuitive arguments and (b) what is known about the cost ratios of Type-I and Type-II errors for the two-state problem.

A Type-I error of two-state classification problem corresponds to Type(1|4) error of four-state, while a Type-II error of two-state corresponds to Type(4|1) of four-state. Denoting misclassification cost of Type$(i|j)$ as $C(i|j)$, $C(i|j)$ is a function of distance between i and j and the sign of (i *minus* j). For cases where the distance between i and j are the same but the sign is different, misclassification cost $C(i|j)$ is higher when the predicted state value is less than actual state (i.e., if $i < j$), than if the actual state is less than the predicted state (i.e., if $j < i$). Using these arguments, the following relationships are derived:

$$C(4|1) > C(3|1) > C(2|1), \tag{1}$$

$$C(1|2) > C(3|2) < C(4|2), \tag{2}$$

$$C(1|3) > C(2|3) > C(4|3), \tag{3}$$

$$C(1|4) > C(2|4) > C(3|4), \tag{4}$$

$$C(4|1) > C(4|2) > C(4|3), \tag{5}$$

$$C(3|1) > C(3|2), \tag{6}$$

$$C(1|4) > C(1|3) > C(1|2), \quad \text{and} \tag{7}$$

$$C(2|4) > C(2|3). \tag{8}$$

Using inequalities (1)–(8) and assuming ratios $C(4|1) : C(1|4)$ to be $1:10$, $1:20$ and $1:50$ (ratios used in past studies), possible sets of relative misclassification costs

Panel 1: For C(4|1):C(1|4) = 1:10

		Predicted State *(i)*			
		1	**2**	**3**	**4**
	1	NA	1	2	3
Actual	**2**	5	NA	1	2
State	**3**	15	10	NA	1
(j)	**4**	30	20	10	NA

Panel 2: For C(4|1):C(1|4) = 1:20

		Predicted State *(i)*			
		1	**2**	**3**	**4**
	1	NA	1	2	3
Actual	**2**	10	NA	1	2
State	**3**	20	15	NA	1
(j)	**4**	60	30	15	NA

Panel 3: For C(4|1):C(1|4) = 1:50

		Predicted State *(i)*			
		1	**2**	**3**	**4**
	1	NA	1	2	3
Actual	**2**	25	NA	1	2
State	**3**	50	30	NA	1
(j)	**4**	150	100	50	NA

Figure 1. Cost of misclassification for all types of errors, $C(i \mid j)$.

are presented in figure 1. Theoretically, there are infinite sets of feasible numbers that can fill each of the panels of figure 1. We have chosen one set of feasible numbers for each cost ratio in order to be able to compute the expected cost of misclassification, as discussed later. A different set of numbers would not significantly change the values of expected cost of misclassification. To further elaborate on the meaning of these numbers, let us take a few examples. In figure 1, panel 1, $C(4|1) = 3$ and $C(1|4) = 30$, which implies that the ratio $C(4|1) : C(1|4) = 1 : 10$ which means it is 10 times more costly to predict a state 4 firm as state 1 than it is to predict a state 1 firm as state 4. Also, in the same panel, $C(2|3) = 10$ while $C(2|4) = 20$. This means that it is twice as costly to misclassify a state 4 firm as state 2 than it is to misclassify a state 3 firm as state 2 given that $C(4|1) : C(1|4) = 1 : 10$. In panel 2, $C(4|1) : C(1|4) = 1 : 20$. $C(1|2) = 10$ and $C(2|3) = 15$. This means that it is one and a half times as costly to predict a state 3 firm as state 2 than it is to predict a state 2 firm as state 1 given that $C(4|1) : C(1|4) = 1 : 20$.

2.3. Evaluation criteria

Evaluating the performance of a model in the four-state scenario is more complex than two-state. The following four evaluation criteria are used in this study:

1. Simple Classification Accuracy (overall and by state).
2. Distance-Weighted Classification Accuracy (overall and by state).
3. Expected Cost of Misclassification.
4. Ranked Probability Score (overall and by state).

Simple Classification Accuracy (SCA) by state is obtained by counting the number of firms correctly classified in each state. Overall SCA is given by a count of the total number of firms correctly classified (in their respective state). These counts by states (or overall) can also be expressed as a percentage of total number of firms in each state (or all states combined). This is a very simple, and most commonly used evaluation criteria, but suffers from some limitations. For example, it ignores the degree, the direction and the relative costs of misclassification. In a multi-state task, these limitations become more serious than in a two-state task. The SCA criterion should therefore be used with caution.

Distance-Weighted Classification Accuracy (DWCA) criterion takes into account the degree of misclassification and is thus a richer criterion than SCA. Misclassifications are penalized based on how far the predicted state is from the actual state. A correct classification gets a score of 1. Assuming equal distances between the four states, if the predicted state is off by one state in either direction, the prediction gets a score of 0.66, if off by 2 states, a score of 0.33 and if off by 3 states, a score of 0. Results can be tabulated for each state, to give DWCA by state. The sum of each state's DWCA provides overall DWCA. Again, these scores can be expressed as raw values or as a percentage of total possible value. Past studies have not used this criterion. For the case of two-state classification, the distance of misclassification is always one and therefore the degree of misclassification is not an issue.

While the DWCA criterion takes into account the distance of misclassification, it ignores the direction and relative costs of misclassification. As already discussed in the subsection on misclassification costs, the direction and the relative costs of misclassification matter. Using the misclassification cost matrix of figure 1, a metric called *Expected Cost of Misclassification* (ECM) is computed. Hopwood et al. [11] first used ECM for a two-state model. Barniv et al. [4] used ECM for a three-state model. The ECM for a four-state model is as follows:

$$\text{ECM} = \frac{1}{N}\sum_{i=1}^{4}\sum_{j=1}^{4} n_{i|j} C(i|j),$$

where

- N: total number of firms in the sample,

- $n_{i|j}$: number of state j firms predicted to be in state i,
- $C(i|j)$: misclassification cost of predicting a state j firm to be in state i.

ECM is not given for each firm, but is computed for the entire model. A lower ECM indicates a better model. ECM cannot be expressed as a percentage. ECM for a perfect model will be 0. ECM takes into account the degree, the direction of misclassification and the relative costs of misclassification. This criterion is therefore the richest of all criteria for evaluating and comparing classification accuracies of ordinal multi-state classification models.

The *Ranked Probability Score* (RPS) was used by Ward [21] and Lau [14] to evaluate the aggregate accuracy of state-probabilities. To calculate RPS, state-probabilities must be estimated for each firm for each of the four states. Weather forecasting provides an example. Suppose there are four possible weather states namely sunny, partly cloudy, rain, and snow. Then a forecast might look like this (0.7, 0.15, 0.1, 0.05) meaning that there is a 70% chance it will be sunny, 15% chance it will be partly cloudy, 10% chance it will rain, and 5% chance it will snow. Based on the actual outcome, a RPS score can be calculated for this forecast. The highest possible RPS for a single forecast is 1. This occurs, for example, if the forecast is (1.0, 0, 0, 0) and the outcome is state 1. If the forecast is (0.9, 0.1, 0, 0) and the outcome is state 1, the RPS will be less than 1. Since the four states in the weather example are in increasing order of non-desirability, at least from a human value system viewpoint, the RPS takes into account the distance between actual outcome and the forecast. RPS was in fact first developed by meteorologists to compare different forecasts [9,15]. The four states of financial health are very similar to the weather states because they can be ranked in an order of increasing distress severity. RPS has therefore been used for evaluating four and five state classifications of financial health [4,14,21].

RPS can be tabulated by state and also reported as an overall score. Previous studies have only reported overall RPS. Overall RPS fails to measure if a model is good at classifying a particular state. Given a probability forecast (F):

$$F = (f_1, f_2, f_3, f_4),$$

where f_i is the predicted probability for state i for a given firm, and j is the actual state, then RPS is given by:

$$S = \frac{3}{2} - \frac{1}{2(n-1)} \sum_{i=1}^{n-1} \left[\left(\sum_{k=1}^{i} f_k \right)^2 + \left(\sum_{k=i+1}^{n} f_k \right)^2 \right] - \frac{1}{n-1} \sum_{i=1}^{n} |i-j| f_i,$$

where n is the number of states (in this study, 4). For a derivation of the above formula see Epstein [9]. The last term in the above equation (14) penalizes the score for the absolute distance between the predicted and actual state.

3. Ordinal classification model development

3.1. Variables and data collection

The data for this study is the initial sample of firms as found in Ward [21]. There are a total of 220 firms in the training set and 115 in the holdout set. The distribution of firms in different states for training and holdout samples are reported in table 1. The healthy firms and dividend reduction firms (states 1 and 2, respectively) were obtained from Compustat tapes. For the dividend reduction firms, if the company had a 40% or greater reduction in dividend after a history of constant dividend payments, it was selected. Loan default and bankrupt firms (states 3 and 4, respectively) were obtained from Compact Disc Disclosure. State 2, 3 and 4 firms were selected first and state 1 firms were matched on SIC code and size. More details concerning the data collection, selection of variables and samples can be found in Ward [21]. Ten independent variables are used for classifying firms. These are the same variables used by Ward [21]. Table 2 lists these variables.

Table 1
Distribution of the firm sample by financial state.

State	Training set		Holdout set	
	Count	Proportion	Count	Proportion
1: Healthy	167	75.9%	112	72.3%
2: Dividend Reduction	15	6.8%	14	9.0%
3: Loan Default	22	10.0%	15	9.7%
4: Bankrupt	16	7.3%	14	9.0%
Total	220	100.0%	155	100.0%

Table 2
List of independent variable used in the models.

Name	Description
SIZE	Log of Total Assets
NITA	Net Income/Total Assets
SALESCA	Sales/Current Assets
CACL	Current Assets/Current Liabilities
OETL	Owner's Equity/Total Liabilities
CATA	Current Assets/Total Assets
CASHTA	Cash plus Marketable Securities/Total Assets
CFFO	Cash Flow from Operating Activities/Total Assets
CFFF	Cash Flow from Financing Activities/Total Liabilities
CFFI	Cash Flow from Investing Activities/Total Liabilities

Modeling tools

3.2. Neural networks

Conceptually, neural networks provide a non-linear mapping between independent and dependent variables, using a network of functions defined by an interconnected network of processing elements (PEs). For details of what NNs are and how they work, the readers are referred to Rumelhart and McClelland [18]. A backpropagation NN was trained and tested using Brainmaker™ software. The number of input PEs is 10, one for each independent variable. The number of hidden PEs is set at 15. There is only one output PE. The single output PE is designed to make it possible to obtain state probabilities and at the same time maintain the ordinality of the states. See figure 2 for a graphical representation of NN-DSS for this study. One way of obtaining state probabilities is to have as many output PEs as the number of states and use what is called 1 of *N* code as a representation of the four classes. That is, class one would be represented by (1, 0, 0, 0) and class two would be represented by (0, 1, 0, 0), and so forth. The 1 of *N* code approach was used by Zurada [25,26]. However, this approach does not retain the ordinality or rank order of the classes and is, therefore not appropriate for ordinal multi-state classification problems.

In addition to the architecture of the NN, choice of training parameters (learning rate, training tolerance and stopping rule) can significantly affect the performance of the model. The literature suggests that a gradually decreasing learning rate ensures good convergence. The model was trained starting with a learning rate set at 1.0 for the first

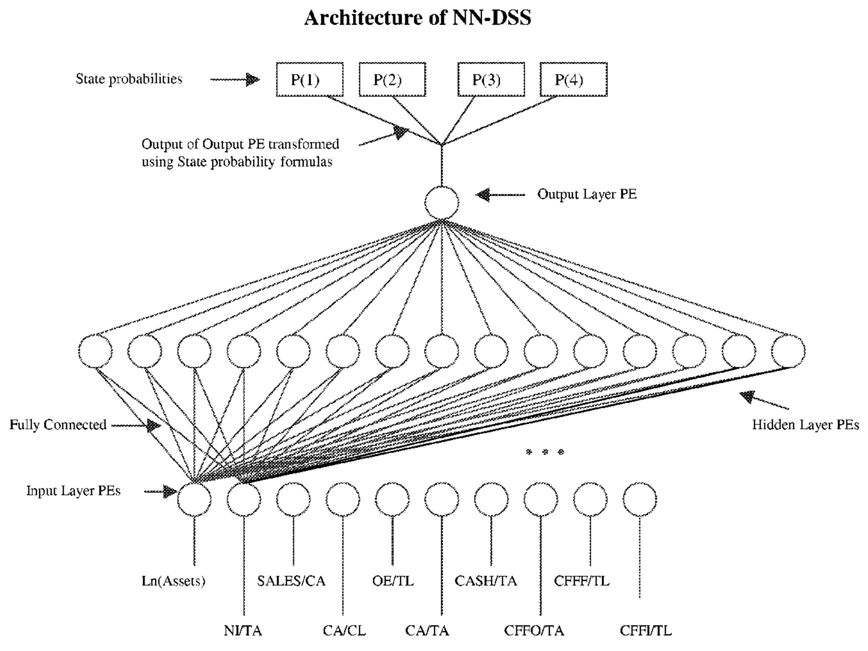

Figure 2. Architecture of NN-DSS.

thousand iterations and subsequently decreased to 0.90, 0.75, 0.50, 0.25, 0.10 for every 1000 runs. An iteration consists of presenting all the observations to the network once. The training tolerance parameter sets the acceptable error value below which no weights need to be updated. A small (or tight) training tolerance can result in over-training or over-fitting and slow network convergence whereas a large training tolerance results in weak training and fast convergence. Neither an over-trained nor a weakly-trained network would perform well on the holdout sample. The NN model was trained using a tolerance of 0.1 a typical value in most classification problems reported in the literature.

The third training parameter, the stopping rule is also to be chosen carefully. A simple stopping rule is to stop when all the examples are within the training tolerance. This approach may not always guarantee that training would stop. For example, if there are two observations with the same independent variables values but different dependent variable value, then training would never stop under this stopping rule. An alternative stopping rule is to stop when a certain percentage of observations are within the training tolerance. The approach used in this study was to stop training when there is no further improvement in training based on simple classification accuracy rate. Since the learning rate strategy influences the rate of convergence, this stopping rule is closely related to the learning rate strategy. Using the training parameters described above, the model converged in about 5000 iterations.

In addition to the issues of training parameters are issues related to the composition of the training data file. Since NNs learn the mapping between the independent variables and the dependent variable by processing example cases in the training sample, the proportion of firms in the training sample can significantly affect learning. This phenomenon is true of the OLGR and MDA models as well. Table 1 indicates that the proportion of state 1 firms dominates the training sample with 75.9% of total firms. This means that the network and other models will “see” state 1 firms more often than other state firms and will therefore learn its characteristics better than other states. If the datasets are more balanced, then the NN has an equal opportunity to learn the characteristics of each of the four states. Effect of unequal (unbalanced) sample on training has been studied by Wilson and Sharda [23] and Patuwo et al. [17]. Although their studies were based on two-group classification, they found that equal (balanced) samples produced better classification accuracy than unequal (unbalanced) samples. In this study, we develop models using both unbalanced and balanced samples. The balanced sample was produced by including state 2, 3 and 4 firms multiple times. This approach was also used by Wilson and Sharda [23]. Although it can be argued that duplicating observations might introduce a bias towards the state of the observation being duplicated, the idea is to balance the bias that already exists in the initial sample. We present the results for both the balanced and unbalanced sample training.

3.3. Multi-state ordered and multinomial logistic models

The OLGR model was developed using SAS. A simple logit assumes no ranking between states whereas ordinal logit assumes a rank order. Elliott and Kennedy [8] and

Kennedy [14] described the unordered and ordered logit procedures. An ordered logit, which has the following cumulative logistic function is used:

$$P(i) = F\big(\alpha_i + X'\beta\big) = \frac{1}{1 + \mathrm{e}^{-(\alpha_i + X'\beta)}},$$

where $i = 1, 2, 3$ and 4 for the four states for the logit model; α is the intercept term, $\alpha_4 > \alpha_3 > \alpha_2 > \alpha_1 = 0$, $X'\beta$ is the vector of coefficients multiplied by the vector of variables, and $P(i)$ is the probability of the outcome of state i.

3.4. Multiple discriminant analysis

Multiple or multivariate discriminant analysis (MDA) has been used in prior classification studies by Altman [1]. MDA uses sample data to develop a functional form in terms of the independent variables to produce a score. Statistica™ software package was used to build the MDA models. MDA does not directly provide state probabilities, however, using Mahalanobis distance from the centroid, state probabilities were obtained from the MDA model.

3.5. The Naïve model

The Naïve model will classify all firms in the state with the maximum proportion in the training sample. For the case of unbalanced sample, since state 1 firms are the largest proportion in the sample (167 out of 220), all 220 firms are classified as state 1 firms. Of course, only 167 will be correctly classified. For the case of balanced sample, the Naïve model will fail to classify firms one way or another since the proportions are the same. For assigning state probabilities, the Naïve model will simply equate the proportion with probabilities.

4. Results

The results of our empirical work are reported in tables 3–8. Table 3 reports the results of SCA (overall and by state) for all models, for the case of unbalanced and balanced sample training, for both training and holdout sets. SCA is expressed as a raw value and as a percentage. For the case of unbalanced sample training, the Naïve model classifies all firms as state 1 since 167 of 220 (or 75.9%) of firms are from state 1. Hence the SCA for the naïve model is 75.9%. MDA, OLGR and NN-DSS, with SCA of 80.9%, 80.0% 84.5%, respectively, clearly outperform the Naïve model on the training sample. On the holdout sample the between models differences were insignificant. It appears that the Naïve model does better (with 72.3% accuracy) on the holdout set than MDA and NN-DSS models (both at 71%). One might question the utility of sophisticated model building when a Naïve model can perform better. It may be recalled that SCA is a very poor evaluation criteria for an ordinal multi-state problem. As we will see later, using other criteria, the Naïve model's results will seem much inferior to those of other

Table 3
Simple Classification Accuracy (SCA) for each model (overall and by state, expressed as raw value and as percentage, for training and holdout sets, for both unbalanced and balanced sample training).

State→	Training set					Holdout set				
	1	2	3	4	Overall	1	2	3	4	Overall
	$n = 167$	$n = 15$	$n = 22$	$n = 16$	$n = 220$	$n = 112$	$n = 14$	$n = 15$	$n = 14$	$n = 155$
Training with unbalanced sample										
SCA raw value										
Naïve	167	0	0	0	167	112	0	0	0	112
MDA	164	0	11	3	178	104	0	4	2	110
OLGR	165	0	7	4	176	108	0	4	1	113
NN-DSS	161	4	17	10	192	101	4	3	2	110
SCA expressed as percentage										
Naïve	100.0%	0.0%	0.0%	0.0%	75.9%	100.0%	0.0%	0.0%	0.0%	72.3%
MDA	98.2%	0.0%	50.0%	18.8%	80.9%	92.9%	0.0%	26.7%	14.3%	71.0%
OLGR	98.8%	0.0%	31.8%	25.0%	80.0%	96.4%	0.0%	26.7%	7.1%	72.9%
NN-DSS	96.4%	26.7%	77.3%	62.8%	84.5%	90.1%	28.6%	20.0%	14.3%	71.0%
Training with balanced sample										
SCA raw value										
Naïve	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
MDA	90	13	15	12	130	67	7	8	3	85
OLGR	102	11	11	9	133	70	8	6	3	87
NN-DSS	113	3	4	16	136	77	4	9	9	99
SCA expressed as percentage										
Naïve	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
MDA	53.9%	86.7%	68.2%	75.0%	59.1%	59.8%	50.0%	53.3%	21.4%	54.8%
OLGR	61.1%	73.3%	50.0%	56.3%	60.5%	62.5%	57.1%	40.0%	21.4%	56.1%
NN-DSS	67.7%	66.7%	18.2%	100.0%	61.8%	68.8%	28.6%	60.0%	64.3%	63.9%

models. While the Naïve model is not able to differentiate between states, the other models are. On the unbalanced sample training, the state-wise SCA results for all four models are biased (or skewed) towards state 1 because the training sample is dominated by state 1 firms. Using the balanced sample, the state-wise SCA for states 2, 3 and 4 are significantly improved at the expense of state 1 accuracy. For example, NN-DSS classifies firms in states 2, 3 and 4 with 66.7%, 18.2% and 100.0% accuracy, respectively, for balanced sample training compared to 26.7%, 77.3% and 62.8%, respectively, for the unbalanced sample. For balanced sample training NN-DSS model outperforms MDA and OLGR models for training as well for holdout set.

It may be noted that for the case of balanced sample training, MDA and OLGR are also better able to identify firms in states 2, 3 and 4 compared to unbalanced sample training. For example, for the balanced sample training, MDA gives SCA of 86.7%, 68.2% and 75.0% for states 2, 3 and 4, respectively, compared to 0.0%, 50.0% and 18.8% SCA for unbalanced sample training. Similar results hold for OLGR. This is an important finding because earlier OLGR studies have not used balanced samples. Since misclassifying states 2, 3 and 4 firms as state 1 is more costly, a cost-weighted evaluation

Table 4
Predicted states for all 220 firms in training set and 155 firms in holdout set for each model.

Model	Actual state	Training set: Predicted state 1	2	3	4	Total	Holdout set: Predicted state 1	2	3	4	Total
					Training with unbalanced sample						
Naïve	1	167	0	0	0	167	112	0	0	0	112
	2	15	0	0	0	15	14	0	0	0	14
	3	22	0	0	0	22	15	0	0	0	15
	4	16	0	0	0	16	14	0	0	0	14
MDA	1	164	0	2	1	167	104	0	5	3	112
	2	14	0	1	0	15	13	0	0	1	14
	3	9	0	11	2	22	10	0	4	1	15
	4	10	0	3	3	16	11	0	1	2	14
OLGR	1	165	0	2	0	167	108	0	4	0	112
	2	14	0	1	0	15	12	0	1	1	14
	3	8	0	7	7	22	9	0	4	2	15
	4	7	0	5	4	16	9	0	4	1	14
NN-DSS	1	161	6	0	0	167	101	4	5	2	112
	2	9	4	2	0	15	8	4	1	1	14
	3	0	3	17	2	22	2	6	3	4	15
	4	0	0	6	10	16	5	3	4	2	14
					Training with balanced sample						
MDA	1	90	59	9	9	167	67	28	8	9	112
	2	0	13	2	0	15	2	7	3	2	14
	3	0	2	15	5	22	0	2	8	5	15
	4	0	0	4	12	16	0	5	6	3	14
OLGR	1	102	51	13	1	167	70	32	6	4	112
	2	2	11	2	0	15	2	8	3	1	14
	3	1	1	11	9	22	1	4	6	4	15
	4	0	0	7	9	16	2	7	2	3	14
NN-DSS	1	113	28	20	6	167	77	17	12	6	112
	2	0	3	10	2	15	1	4	7	2	14
	3	0	0	4	18	22	0	0	9	6	15
	4	0	0	0	16	16	1	2	2	9	14

criteria (such as ECM) would actually make the balanced sample results better than the unbalanced sample results, in spite of a low overall SCA for balanced sample, as will be seen in table 6.

Table 3 only reports the number and percentage of firms correctly classified. As stated above, the misclassification distance must be calculated to obtain DWCA and, the distance and the sign of the misclassifications are necessary to obtain ECM. Therefore, table 4 provides the distribution of correct and incorrect classifications of all firms in training and holdout samples for both unbalanced and balanced sample training for all the models. For example, the training sample consists of 22 firms in state 3. For the unbalanced sample training, NN-DSS does not predict any of those 22 firms to be in

Table 5
Distance-Weighted Classification Accuracy (DWCA) for each model (overall and by state, expressed as raw value and as percentage, for training and holdout sets, for both unbalanced and balanced sample training).

	Training set					Holdout set				
State→	1	2	3	4	Overall	1	2	3	4	Overall
	$n = 167$	$n = 15$	$n = 22$	$n = 16$	$n = 220$	$n = 112$	$n = 14$	$n = 15$	$n = 14$	$n = 155$
Training with unbalanced sample										
DWCA raw values										
Naïve	167.00	10.00	7.33	0.00	184.33	112.00	9.33	5.00	0.00	126.33
MDA	164.67	10.00	15.33	5.00	195.00	105.67	9.00	8.00	2.67	125.33
OLGR	165.67	10.00	14.33	7.33	197.33	109.33	9.00	8.33	3.67	130.33
NN-DSS	165.00	11.33	20.33	14.00	208.67	105.33	10.33	9.99	5.33	130.00
DWCA expressed as percentage										
Naïve	100.0%	66.7%	33.3%	0.0%	83.8%	100.0%	66.6%	33.3%	0.0%	81.5%
MDA	98.6%	66.7%	69.7%	31.3%	88.6%	94.3%	64.3%	53.3%	19.1%	80.9%
OLGR	99.2%	66.7%	65.1%	45.8%	89.7%	97.6%	64.3%	55.5%	26.2%	84.1%
NN-DSS	98.8%	75.6%	92.4%	87.5%	95.8%	94.0%	73.8%	60.0%	38.1%	83.7%
Training with balanced sample										
DWCA raw values										
Naïve	NA[1]	NA	NA	NA	NA	NA	NA	NA	NA	NA
MDA	132.33	14.33	19.67	14.67	181.00	88.33	11.00	12.67	8.67	120.67
OLGR	140.33	13.67	18.00	13.67	185.67	93.33	11.67	11.67	6.67	123.33
NN-DSS	138.33	10.33	16.00	16.00	180.67	92.33	9.66	13.00	11.00	126.00
DWCA expressed as percentage										
Naïve	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
MDA	79.2%	95.5%	89.4%	91.7%	82.3%	78.9%	78.6%	84.5%	61.9%	77.9%
OLGR	84.0%	91.1%	81.8%	85.4%	84.4%	83.3%	83.4%	77.8%	47.6%	79.6%
NN-DSS	82.8%	68.9%	72.7%	100.0%	80.6%	82.4%	69.0%	86.7%	78.6%	81.3%

[1] Not applicable because in a balanced sample, there are equal number of firms in each state and the Naïve model cannot discriminate between any.

state 1, incorrectly predicts 3 firms to be in state 2, correctly predicts 17 firms to be in state 3, and incorrectly predicts 2 firms to be in state 4.

An interesting observation from table 4 is that for the case of unbalanced sample training, the results show that for both training and holdout sets, both MDA and OLGR models fail to classify any firm as state 2, whereas NN-DSS is able to classify firms in all states. This result suggests that NN-DSS provides classification accuracies that are more uniform (less skewed) across all states.

The data in table 4 is used to compute DWCA and ECM results. Table 5 reports the results for DWCA (overall and by state). For the unbalanced sample training, NN-DSS outperforms the Naïve, MDA and OLGR models in the training set and Naïve and MDA in the holdout set. The classification accuracy by state for the holdout sample in the balanced sample case for NN-DSS are 82.4%, 69.0%, 86.7% and 78.6% for states 1, 2, 3 and 4, respectively, giving an overall accuracy of 81.3%. In contrast, OLGR provides an overall accuracy of 79.6% while MDA 77.9%. Table 5 shows that DWCA for

Table 6
Expected Misclassification Cost (ECM) for each model (for three different cost ratios).

Model	Training with unbalanced sample		Training with balanced sample	
	Training set	Holdout set	Training set	Holdout set
	For $C(4\|1):C(1\|4) = 1:10$[1]			
Naïve	4.02	4.61	NA	NA
MDA	2.48	3.72	0.78	1.76
Logit	2.10	3.34	0.89	2.26
NN-DSS	0.66	2.63	0.54	1.10
	For $C(4\|1):C(1\|4) = 1:20$			
Naïve	7.05	8.26	NA	NA
MDA	4.43	6.63	0.91	2.40
Logit	3.67	5.89	1.14	3.38
NN-DSS	1.07	4.43	0.54	1.52
	For $C(4\|1):C(1\|4) = 1:50$			
Naïve	17.61	20.64	NA	NA
MDA	11.18	16.43	1.69	6.41
Logit	9.37	14.92	2.60	8.94
NN-DSS	2.84	11.33	0.54	3.55

[1] This cost ratio suggests that it is 10 times more costly to misclassify a state 4 firm as state 1 than it is to misclassify a state 1 firm as state 4.

the case of balanced sample is slightly lower than for the unbalanced sample for each model. However, in general for the holdout sample, using balanced sample sizes improves the classification accuracies for states 2, 3, and 4. This result is somewhat similar to the result found for SCA, however the difference between the balanced and unbalanced sample accuracies is mitigated somewhat by including the effect of distance of misclassification.

Table 6 reports the results for ECM for three different ratios of $C(4|1):C(1|4)$ ($1:10$, $1:20$ and $1:50$). As discussed before, a lower ECM implies a better model. As discussed before, ECM takes into account the extent of misclassification, the direction of misclassification and the relative misclassification costs. It is therefore a very rich measure of performance. In the best case, if all firms are correctly classified in their respective states, ECM should be 0. For each of the three sets of misclassification costs, discussed in figure 1, NN-DSS gives significantly lower (better) ECM than all other models for both the balanced and unbalanced samples, for both training and holdout sets. For example, for the unbalanced sample, for $1:20$ cost ratio, Naïve model has ECM of 7.05 and 8.26 for the training and holdout sets respectively whereas NN-DSS yields ECM of 1.07 and 4.43. Both MDA and Logit give much worse ECM compared to NN-DSS. The results are especially good on the balanced sample training. For example, for the training set NN-DSS gives an ECM of 0.54 compared to 0.91 and 1.14 for MDA and OLGR, respectively. NN-DSS is also a clear winner for the holdout set giving an ECM of 1.52 compared to 2.40 for MDA and 3.38 for OLGR. These results also signify

Table 7
Ranked Probability Score (RPS) for each model (overall and by state, expressed as raw value and as percentage, for training and holdout sets, for both unbalanced and balanced sample training).

State→	Training set					Holdout set				
	1	2	3	4	Overall	1	2	3	4	Overall
	$n = 167$	$n = 15$	$n = 22$	$n = 16$	$n = 220$	$n = 112$	$n = 14$	$n = 15$	$n = 14$	$n = 155$
Training with unbalanced sample										
RPS raw value										
Naïve	161.81	11.94	12.71	4.69	191.15	108.52	11.15	8.67	4.10	132.44
MDA	164.17	11.73	18.47	9.55	203.92	107.08	10.34	10.5	5.39	133.31
OLGR	163.61	11.96	18.21	11.31	205.09	109.29	10.93	11.32	6.00	137.54
NN-DSS	162.31	12.72	20.26	14.20	209.50	104.27	11.32	12.01	7.70	135.30
RPS expressed as percentage										
Naïve	96.9%	79.6%	57.8%	29.3%	86.9%	96.9%	79.6%	57.8%	29.3%	85.4%
MDA	98.3%	78.2%	84.0%	59.7%	92.7%	95.6%	73.9%	70.0%	38.5%	86.0%
OLGR	98.0%	79.7%	82.8%	70.7%	93.2%	97.6%	78.1%	75.5%	42.9%	88.7%
NN-DSS	97.2%	84.8%	92.1%	88.8%	95.2%	93.1%	80.8%	80.1%	55.0%	87.3%
Training with balanced sample										
RPS raw value										
Naïve	118.29	13.13	19.25	11.33	162.00	79.33	12.25	13.13	9.92	114.63
MDA	143.39	14.24	20.66	14.31	192.60	94.83	12.06	13.49	10.00	130.38
OLGR	146.55	13.80	19.31	13.72	193.38	98.21	12.15	13.03	9.08	132.47
NN-DSS	141.83	12.09	18.92	15.61	188.47	95.31	11.09	13.64	11.24	131.30
RPS expressed as percentage										
Naïve	70.8%	87.5%	87.5%	70.8%	73.6%	70.8%	87.5%	87.5%	70.9%	74.9%
MDA	85.9%	94.9%	93.9%	89.4%	87.5%	84.7%	86.1%	89.9%	71.4%	84.1%
OLGR	87.8%	92.0%	87.8%	85.8%	87.9%	87.7%	86.8%	86.9%	64.9%	85.5%
NN-DSS	84.9%	80.6%	86.0%	97.5%	85.7%	85.1%	79.2%	90.9%	80.3%	84.7%

an important consideration that the performance of all the models improves significantly going from unbalanced sample to balanced sample, a consideration ignored by previous researchers for the MDA and OLGR models. Further, as the cost ratio $C(4|1):C(1|4)$ gets higher, NN-DSS results relative to other models get better.

Table 7 presents the RPS results by state and overall. Higher the RPS, better is the model. For both unbalanced and balanced sample training, all three non-Naïve models perform better than the Naïve model but perform comparable to each other. Looking at the state-wise RPS, it should be noted that NN-DSS has a consistently higher RPS for states 2, 3 and 4 than the other models. This result is consistent with earlier results with DWCA and SCA. Hence NN-DSS seems to be a better discriminator than other models on all criteria.

Table 8 provides one tailed chi-squared test results to test if one model yields statistically better classification accuracy in paired comparisons with other models. No significant differences were observed between the MDA and OLGR models for any evaluation criteria. For SCA on the balanced sample training, NN-DSS is found to be statistically significantly better ($p < 0.05$) than MDA and OLGR on the holdout sample. For

Table 8
Chi-squared values for pair-wise model comparison.

	MDA vs. Naïve	OLGR vs. Naïve	NN-DSS vs. Naïve	OLGR vs. MDA	NN-DSS vs. MDA	NN-DSS vs. OLGR
For simple classification accuracy criteria, on training with unbalanced sample						
Training Set	1.62	1.07	5.17**	0.06	1.02	1.55
Holdout Set	0.10	0.01	0.10	0.18	0.00	0.18
For simple classification accuracy criteria, on training with balanced sample						
Training Set	NA	NA	NA	0.085	0.34	0.09
Holdout Set	NA	NA	NA	0.037	3.77**	3.06**
For distance-weighted classification accuracy criteria, on training with unbalanced sample						
Training Set	2.30*	3.318**	18.04***	0.09	8.16***	6.61***
Holdout Set	0.014	0.57	0.39	0.77	0.56	0.02
For distance-weighted classification accuracy criteria, on training with balanced sample						
Training Set	NA	NA	NA	0.41	0.24	1.24
Holdout Set	NA	NA	NA	0.22	0.89	0.23
For RPS criteria, on training with unbalanced sample						
Training	4.18***	4.95***	8.91***	0.03	0.98	0.65
Holdout	0.019	0.99	0.31	0.74	0.18	0.19
For RPS criteria, on training with balanced sample						
Training	14.01***	14.01***	10.27***	0	0.32	0.32
Holdout	5.59***	7.58***	6.21***	0.19	0.02	0.07

* Significant at 0.10 level.
** Significant at 0.05 level.
*** Significant at 0.01 level.

DWCA, NN-DSS is found to be significantly better ($p < 0.01$) than MDA and OLGR for the training set, but not on the holdout set. MDA and OLGR were not significantly better than NN on any criteria. NN-DSS was found to be significantly better than Naïve model on almost every criteria.

5. Summary, limitations and further research

This study presents a Neural Network based Decision Support System for an ordinal multi-state classification problem. The model serves as a decision aid for classifying firms in four ordinal states of financial distress namely healthy, dividend reduction, debt default and bankrupt. The study compares the results of the neural network model with a Naïve model, a Multiple Discriminant Analysis model and an Ordinal Logistic Regression model. Simple Classification Accuracy, Distance-weighted Classification Accuracy, Expected Cost of Misclassification and Ranked Probability Score (RPS) are used as measures of performance for all models and used for model comparisons. Since RPS requires state probabilities for each state, state probabilities were generated using Neural Networks in ordinal four state decision tasks. RPS is especially suitable for multiple ordinal states, as is the case in this study. The study also included the effect of different proportions of firms (balanced and unbalanced) in the training sample.

NN-DSS models clearly outperform all the other models on the Expected Cost of Misclassification criterion, for both training and holdout sets for both unbalanced sample and balanced sample training. For the balanced sample training, NN-DSS models significantly outperform both the MDA and OLGR on the simple classification accuracy on the holdout set. For the unbalanced sample training, NN-DSS significantly outperforms MDA and OLGR on the distance weighted accuracy criteria on the training set. Neither MDA nor OLGR performed significantly better than NN on any criteria.

This study has some limitations. First, in measuring the distance-weighted classification accuracy, equal distance between the four states was assumed. It can be argued that these distances are in reality unequal. Since there are no universally acceptable estimates of distances between the states in the literature we used equal distances. Average cumulative abnormal returns for each state may be used as the basis for estimating distances between states. Second, the study suffers on account of small number of firms in states 2, 3 and 4 for both training and holdout sets. The equal sample training somewhat overcomes this limitation by duplicating the same observations multiple times. Future research can focus on developing better estimates of distance between states and should verify the effect of different proportion of firms in the training set. The RPS score can also be refined by using cost-weighted RPS.

This research extends the decision support literature by demonstrating principles of model development related to ordinal output for NNs, developing state probabilities for multi-state settings, using various evaluation criteria, and studying the effect of different proportion of firms in the training sample. These principles are generalizeable for developing DSS to support ordinal multi-state decision tasks in other domains as well. Also, these principles can be used for any arbitrary number of states.

Appendix

Derivation of transformation functions to generate state probabilities

For each observation, let

- $P(i)$: probability for state i,
- d_i: relative distance from centroid of state i.

Then

$$P(1) + P(2) + P(3) + P(4) = 1. \tag{A.1}$$

And, assuming probabilities are inversely proportional to relative distances from group centroids,

$$\frac{P(1)}{P(2)} = \frac{d_2}{d_1}, \tag{A.2}$$

$$\frac{P(1)}{P(3)} = \frac{d_3}{d_1}, \tag{A.3}$$

$$\frac{P(1)}{P(4)} = \frac{d_4}{d_1}, \tag{A.4}$$

$$\frac{P(2)}{P(3)} = \frac{d_3}{d_2}, \tag{A.5}$$

$$\frac{P(2)}{P(4)} = \frac{d_4}{d_2}, \tag{A.6}$$

$$\frac{P(3)}{P(4)} = \frac{d_4}{d_3}. \tag{A.7}$$

From (A.2)–(A.7) we have

$$P(1) = \frac{d_2}{d_1} P(2), \tag{A.8}$$

$$P(1) = \frac{d_3}{d_1} P(3), \tag{A.9}$$

$$P(1) = \frac{d_4}{d_1} P(4), \tag{A.10}$$

$$P(2) = \frac{d_1}{d_2} P(1), \tag{A.11}$$

$$P(2) = \frac{d_3}{d_2} P(3), \tag{A.12}$$

$$P(2) = \frac{d_4}{d_2} P(4), \tag{A.13}$$

$$P(3) = \frac{d_1}{d_3} P(1), \tag{A.14}$$

$$P(3) = \frac{d_2}{d_3} P(2), \tag{A.15}$$

$$P(3) = \frac{d_4}{d_3} P(4), \tag{A.16}$$

$$P(4) = \frac{d_1}{d_4} P(1), \tag{A.17}$$

$$P(4) = \frac{d_2}{d_4} P(2), \tag{A.18}$$

$$P(4) = \frac{d_3}{d_4} P(3). \tag{A.19}$$

From (A.1), (A.11), (A.14) and (A.17) we get:

$$P(1) + \frac{d_1}{d_2} P(1) + \frac{d_1}{d_3} P(1) + \frac{d_1}{d_4} P(1) = 1,$$

therefore

$$P(1) = \frac{1}{1 + d_1/d_2 + d_1/d_3 + d_1/d_4}.$$

From (A.1), (A.8), (A.15) and (A.18) we get:

$$\frac{d_2}{d_1}P(2) + P(2) + \frac{d_2}{d_3}P(2) + \frac{d_2}{d_4}P(2) = 1,$$

therefore

$$P(2) = \frac{1}{1 + d_2/d_1 + d_2/d_3 + d_2/d_4}.$$

From (A.1), (A.9), (A.12), (A.19) we get:

$$\frac{d_3}{d_1}P(3) + \frac{d_3}{d_2}P(3) + P(3) + \frac{d_3}{d_4}P(3) = 1,$$

therefore

$$P(3) = \frac{1}{1 + d_3/d_1 + d_3/d_2 + d_3/d_4}.$$

From (A.1), (A.10), (A.13), (A.16) we get:

$$\frac{d_4}{d_1}P(4) + \frac{d_4}{d_2}P(4) + \frac{d_4}{d_3}P(4) + P(4) = 1,$$

therefore

$$P(4) = \frac{1}{1 + d_4/d_1 + d_4/d_2 + d_4/d_3}.$$

How is d_i calculated?

Suppose the NN gives a score of S for an observation. Suppose D_i is the absolute distance difference between S and the state value. Then

$$d_i = \frac{D_i}{\sum_{i=1}^{4} D_i}.$$

References

[1] E.I. Altman, Financial ratios, discriminant analysis and the prediction of corporate bankruptcy, Journal of Finance (September 1968) 589–607.

[2] R. Barniv, Accounting procedures, market data, cash-flow figures, and insolvency classification: The case of the insurance industry, The Accounting Review 65(3) (1990) 578–604.

[3] R. Barniv and A. Raveh, Identifying financial distress: A new non-parametric approach, Journal of Business Finance and Accounting 16(3) (1989) 361–383.

[4] R. Barniv, A. Agarwal and R. Leech, Predicting the outcome following bankruptcy filing: A three state classification using neural networks, International Journal of Intelligent Systems in Accounting Finance and Management 6 (1997) 177–194.

[5] E. Deakin, A discriminant analysis of predictors of business failure, Journal of Accounting Research (1972) 167–179.

[6] T.E. Dielman and H.R. Oppenheimer, An examination of investor behavior during periods of large dividend changes, Journal of Financial and Quantitative Analysis (1984) 197–216.

[7] R.O. Edmister, An empirical test of financial ratio analysis for small business failure prediction, Journal of Financial and Quantitative Analysis (1972) 1477–1493.

[8] J.A. Elliott and D.B. Kennedy, Estimation and prediction of categorical models in accounting research, Journal of Accounting Literature 7 (1988) 202–242.

[9] E.S. Epstein, A scoring system for probability forecasts of ranked categories, Journal of Applied Meteorology 8 (1969) 985–987.

[10] J.A. Gentry, P. Newbold and W. Whitford, Classifying bankrupt firms with funds flow components, Journal of Accounting Research (1985) 146–160.

[11] W.S. Hopwood, J. McKeown and J. Mutchler, The sensitivity of financial distress prediction models to departures from normality, Contemporary Accounting Research 5(1) (1988) 284–298.

[12] C.S. Huang, R.E. Dorsey and M.A. Boose, Life insurer financial distress prediction: A neural network model, Journal of Insurance Regulation 13(2) (1995) 131–167.

[13] D.B. Kennedy, Classification techniques in accounting research: Empirical evidence of comparative performance, Contemporary Accounting Research 8(2) (1992) 419–422.

[14] A. Lau and Hing-Ling, A five-state financial distress prediction model, Journal of Accounting Research Spring (1987) 127–138.

[15] A.H. Murphy, A note on the ranked probability score, Journal of Applied Meteorology 10 (1971) 155–156.

[16] J.A. Olson, Financial ratios and the probabilistic prediction of bankruptcy, Journal of Accounting Research (1980) 109–131.

[17] E. Patuwo, M.Y. Hu and M.S. Hung, Two-group classification using neural networks, Decision Sciences 24(4) (1993) 825–846.

[18] D.E. Rumelhart and J.L. McClelland, *Parallel Distributed Processing – Explorations in the Microstructure of Cognition*, Vol. 1 (MIT Press, 1986).

[19] K.Y. Tam and M.Y. Kiang, Managerial applications of neural networks: The case of bank failure predictions, Management Science 38(7) (1992) 926–947.

[20] K.Y. Tam and M.Y. Kiang, Predicting bank failures: A neural network approach, Applied Artificial Intelligence 4 (1990) 265–282.

[21] T.J. Ward, An empirical study of the incremental predictive ability of Beaver's Naïve operating flow measure using four-state ordinal models of financial distress, Journal of Business Finance and Accounting 21(4) (1994) 547–561.

[22] T.J. Ward and B.P. Foster, An empirical analysis of Thomas's financial accounting allocation fallacy theory in a financial distress context, Accounting and Business Research 26(2) (1996) 137–152.

[23] R.L. Wilson and R. Sharda, Bankruptcy prediction using neural networks, Decision Support Systems 11 (1994) 545–557.

[24] M.E. Zmijewski, Methodological issues related to the estimation of financial distress prediction models, Journal of Accounting Research Supplement (1984) 59–82.

[25] J.M. Zurada, B.P. Foster and T.J. Ward, A comparison of the ability of neural networks and logit regression models to predict levels of financial distress, in: *Information Systems Development: Methods & Tools, Theory & Practice*, eds. S. Wrycza and J. Zupancic (Plenum Press, 1997).

[26] J.M. Zurada, B.P. Foster, T.J. Ward and R.M. Barker, A research note on the relative ability of neural networks and logit regression models to predict dichotomous and multi-state financial distress variables, Working Paper (1997).