

# Improving Bankruptcy Prediction with Hidden Layer Learning Vector Quantization

J. C. NEVES\* AND A. VIEIRA\*\*

*\*ISEG, Lisbon, Portugal and HEC, Paris, France and \*\*ISEP, Oporto, Portugal*

**ABSTRACT** A Hidden Layer Learning Vector Quantization (HLVQ), neural network-learning algorithm is used for correcting the outputs of Multilayer Perceptrons (MLP) for predicting corporate bankruptcy. We call this method HLVQ-C, and it is shown that it outperforms both discriminant analysis and traditional neural networks while significantly reducing type I error, which is the type of error that has the highest costs for banks. Moreover, our approach gives an estimation of the prediction robustness thus providing a useful measure of credit risk, which is of great interest for banks, insurance companies and creditors in general. We also show that unbalanced samples, containing more financially sound firms than bankrupt firms, place a strong bias on the classifiers thus leading to a deterioration of type I error accuracy. Although many studies have been published on bankruptcy prediction using neural networks or discriminant analysis, they used mainly US or UK samples of very limited size. Our study is based on industrial French firms, uses a data-set of 583 bankrupt firms over the period 1998–2000 and tests the effects of different proportions of non-bankrupt firms in the sample. Attention was also given to feature selection to reduce the dimensionality of the problem.

## 1. Introduction

Financial distress prediction is of great importance for banks, insurance firms, creditors and investors. Since the work of Beaver (1966) and Altman (1968) there has been considerable interest in using financial ratios for predicting the

---

*Correspondence Address:* J. C. Neves, ISEG, Rua Miguel Lupi 20, Lisbon 1249-078, Portugal.  
E-mail: [jcneves@iseg.utl.pt](mailto:jcneves@iseg.utl.pt)

financial distress of companies. Using univariate analysis Beaver concluded that 'Cash Earnings to Total Debt' was the best ratio for signalling bankruptcy, while Altman (1968) pioneered the use of multiple discriminant analysis for predicting bankruptcy. Ever since, discriminant analysis has become a standard approach to predict financial distress. However, this method has been criticized (Eisenbeis, 1977) due to its linearity, restrictive assumptions and for treating financial ratios as independent variables.

Non-linear models such as the logit (Martin, 1977; Zavgren, 1983) and the probit (Amemiya and Powell, 1983) have been used, not only to classify but also to estimate the bankruptcy probability (McFaden, 1976; Press and Wilson, 1978; Ohlson, 1980; Lo, 1986). However, these models also have important limitations. First, the choice of the regression function creates a bias that restricts the outcome. Second, these methods are very sensitive to exceptions, which are common in bankruptcy. Third, most conclusions have an implicit Gaussian distribution, which is inappropriate in many cases.

Non-parametric models (Stein and Ziegler, 1984; Srinivasan and Kim, 1987) like Artificial Neural Networks (ANNs) (Charitou *et al.*, 2004), Genetic Programming (GP) (Banzhaf *et al.*, 1989) and Support Vector Machines (SVM) (Fan and Palaniswami, 2000) have also been used for bankruptcy classification.

Initially, Barker (1990), Marose (1990) and Berry and Trigueiros (1993) suggested that ANNs should be used as a complementary tool to classify credit risk. Subsequently, other authors (Coats and Fant, 1993; Wilson and Sharda, 1994; Yang, 1999; Tan and Dihadjo, 2001) found ANNs a promising and robust technique outperforming discriminant analysis in bankruptcy prediction. O'Leary (1998) surveyed the application of neural networks to corporate failure prediction, concluding that the results were at least as good as those generated by other techniques.

On the other hand, Altman *et al.* (1994) criticized the neural networks approach based on their poor generalization. However, in most studies, the data-sets used are very small (sometimes with less than 100 cases) and often highly unbalanced, which does not allow a fair comparison (Grice and Dugan, 2001). Therefore, some authors, such as Laitinen and Kankaanpaa (1999), were unable to find significant differences in accuracy between neural networks and other techniques. Nevertheless, the authors generally agree that further research is required on new network topologies, training algorithms, learning methods and a combination of techniques to achieve higher predictive capabilities, which is the avenue of our research.

In this work, we use a new training algorithm (HLVQ-C) based on the Hidden Layer Learning Vector Quantization to improve the prediction of multilayer neural networks. The main advantages of this algorithm are:

1. It can use a larger set of variables without compromising generalization.
2. It is capable of improving the network predictions for difficult cases and outliers.

3. It gives an easy estimate of the prediction accuracy.

Our database sample, containing the financial records of over 1,000 industrial French companies, is much larger and more complete than most published studies. We also develop an efficiency measure that takes into account the different types of classification error. Finally, we perform statistical tests to evaluate the classification accuracy.

This paper is organized as follows. The next section describes ANNs, and Section 3 presents our method and how it is used to correct the output of a Multi-layer Perceptron. This section also presents the methods used for assessing the neural network's predictive capabilities and introduces a modification to the performance measurement proposed by Korobow and Stuhr (1985) for evaluating the efficiency of early warning models. Methods and criteria used in the multiple discriminant analysis are also discussed. Section 4 describes the data used in the research and Section 5 presents the results. Finally, Section 6 outlines the conclusions.

## **2. Artificial Neural Networks**

Artificial Neural Networks (ANNs) are a set of algorithms inspired by the human brain's distributed architecture and parallel processing capabilities. ANNs are essentially multiple regression machines capable of learning directly from examples and requiring no prior knowledge of the problem. Data classification can be viewed as a regression problem, that is, finding a function that maps an input into the corresponding class whilst minimizing the misclassification rate. ANNs have intrinsic non-linear regression capabilities that make them highly competitive for difficult classification problems (Bishop, 1995).

For the researcher and the financial analyst, the main advantage of ANNs is that there is no need to specify the functional relation between variables. Since they are connectionist-learning machines, the knowledge is directly imbedded in a set of weights connecting arrays of simple processing nodes called neurons. In order to train a neural network properly one needs a large set of 'good quality' examples, that is, representative and error-free data. In the case of bankruptcy problems, the researcher should be cautious when drawing conclusions from neural networks trained with only one or two hundred cases, as observed in most previous studies.

Classifying high dimensional data is a difficult task due to the curse of dimensionality. If, for instance, a data point is characterized by 10 variables, each quantized to 10 states, the number of possible configurations is  $10^{10}$ , which requires a large amount of training to cover this huge search space. Note, however, that this is a pessimistic estimate since most variables are correlated and the regression functions are smooth enough that a reasonable estimation can be achieved from fewer points. Nevertheless, the ANN architecture should be adjusted to the complexity of the problem. As a rule of thumb, the neural network should

have a number of connection weights less than 1/10 of the sample size available for training. In this work we use between 1,000 and 2,000 examples, corresponding to a number of 100 or 200 weights.

The most common type of neural network is Feed-forward Multilayer Perceptrons (MLP) with a single hidden layer and trained with back-propagation. Although other types and training algorithms exist, for most problems MLP are the simplest and most reliable classifier. The error function used is the Sum of Square Error (SSE) with sigmoid transference function for the output. This allows the output  $y$  to be assigned directly to a membership probability (Bishop, 1995). Caution must be used as this interpretation is valid strictly as a limit, when the training data are infinite. In general, this is not a very reliable estimator of the true class membership probability, especially when data are scarce.

For a single output node, classification is based on the following criteria:

$$\begin{cases} y > \theta \rightarrow y = 1 \\ y < \theta \rightarrow y = 0 \end{cases}$$

where  $\theta$  is a threshold parameter that usually takes the value of 0.5. For  $y = 1$  the prediction is bankruptcy while  $y = 0$  is a healthy firm.

### 3. The Proposed Neural Networks Model

In this section we review the Hidden Layer Learning Vector Quantization training algorithm and describe a method where it is used to correct the output of the MLP.

#### 3.1. Hidden Layer Learning Vector Quantization (HLVQ)

Hidden Layer Learning Vector Quantization is an algorithm developed for classification of high dimensional data such as financial accounting data. It is implemented in three steps. In the first step a MLP is trained. In a second step, a supervised Learning Vector Quantization is applied to the outputs of the last hidden layer to obtain the code-vectors  $\vec{w}_{c_k}$  corresponding to each class  $c_k$  in which data are to be classified. Each example,  $\vec{x}_i$ , is assigned to the  $c_k$  class with the smallest Euclidian distance to the respective code-vector:

$$k = \min_j \left\| \vec{w}_{c_j} - \vec{h}(\vec{x}_i) \right\| \quad (1)$$

where  $\vec{h}$  is a vector containing the outputs of the hidden layer and  $\| \cdot \|$  denotes the usual Euclidian distance. In the third step, the MLP is retrained but with two differences. First, the error correction is not applied to the output layer but directly to the last hidden layer. The output layer is therefore ignored from then on. The second difference is in the error correction propagated back to

each hidden node:

$$E = \frac{1}{2} \sum_{i=1}^{N_h} (\vec{w}_{c_k} - \vec{h}(\vec{x}_i))^2 \tag{2}$$

where  $N_h$  is the number of hidden nodes. After retraining the MLP, using an incremental strategy, a new set of code-vectors

$$\vec{w}_{c_k}^{new} = \vec{w}_{c_k} + \Delta\vec{w}_{c_k} \tag{3}$$

is obtained according to the following training scheme:

$$\begin{aligned} \Delta\vec{w}_{c_k} &= \alpha(n)(\vec{x} - \vec{w}_{c_k}) && \text{if } \vec{x} \in \text{class } c_k \\ \Delta\vec{w}_{c_k} &= 0 && \text{if } \vec{x} \notin \text{class } c_k \end{aligned} \tag{4}$$

The parameter  $\alpha(t)$  is the learning rate, which decreases with iteration  $n$  to guarantee convergence. Steps two and three are repeated following an iterative process. The method stops when a minimum classification error is found.

We can calculate the distance of a given example  $x$  to each prototype as

$$d_i = \left\| \vec{h}(\vec{x}) - \vec{w}_{c_i} \right\| \tag{5}$$

which is a measurement of proximity to each class.

### 3.2. The HLVQ-C Method

One of the major drawbacks of neural networks is their poor performance in data space not covered by the training data, particularly frequent in high dimensional data, which is a common problem with the other previously mentioned techniques. In order to alleviate this problem, HLVQ is used to correct the MLP output, a method that we call HLVQ-C, which is implemented according to the following scheme.

After the MLP is trained, each case to be tested,  $x_i$ , is included in the training set, and the neural network retrained by assigning the output of this company as 0 (healthy). The corresponding output,  $y_0(\vec{x}_i) = y_i^0$ , as well as the respective distances to each class prototype obtained by HLVQ are calculated as follows:

$$\vec{d}_i^0 = (d_{c_0}^0, d_{c_1}^0) = (\|h_0(\vec{x}_i) - w_{c_0}\|, \|h_0(\vec{x}_i) - w_{c_1}\|). \tag{6}$$

Then the network is retrained again but now the company is assigned as class 1 (bankrupt). The new output  $y_1(\vec{x}^i) = y_i^1$  and the respective distances to the

prototype are obtained:

$$\vec{d}_i^1 = (d_{c0}^1, d_{c1}^1) = (\|h_1(\vec{x}_i) - w_{c0}\|, \|h_1(\vec{x}_i) - w_{c1}\|). \quad (7)$$

Finally, the most likely output is selected by the following heuristic rule:

$$\begin{aligned} y_i &= y_i^0 && \text{if } d_{c0}^0 < d_{c1}^0 \\ y_i &= y_i^1 && \text{if } d_{c1}^1 < d_{c0}^1 \end{aligned} \quad (8)$$

We call this method Hidden Layer Learning Vector Quantization Correction (HLVQ-C). This method essentially corrects inconsistencies between predictions made by MLP and HLVQ.

### 3.3. Generalization and Benchmarking

The quality of a classifier is measured by its generalization capabilities and robustness. It is very important to avoid over-fitting, a common problem with neural networks, to validate the generalization capabilities of our method properly. Generalization measures network performance on unobserved cases and is often estimated by separating the data-set into two groups: a training set and a test set. The network is trained within the training set and its performance tested on unused data from the test set. Over-fitting is avoided by stopping the training upon reaching a minimum error in the test set. Although this procedure may be adequate for large data-sets, the test set may not be representative when data are scarce.

Cross-validation is the most adequate procedure to evaluate the generalization capabilities of the network. In this case, we used tenfold cross-validation. It consists of dividing the data-set into 10 sets ( $A_1, \dots, A_{10}$ ), using nine ( $A_2, \dots, A_{10}$ ) for training and the remaining  $A_1$  for testing. When training is completed the test error  $e_1$  is recorded and the process is repeated: training with ( $A_1, A_3, \dots, A_{10}$ ), testing with  $A_2$  and test error  $e_2$  is recorded. After completing the 10 cycles, the generalization error, or cross-validation error, is calculated as the average of test set errors.

There are two types of errors in classification problems. Type I error is the number of companies classified as healthy when they are actually bankrupt ( $N_{10}$ ) divided by the total number of bankruptcies in the sample ( $N_1$ ):

$$e_1 = \frac{N_{10}}{N_1}. \quad (9)$$

The double subscript in  $N$  means classified-actual, with 1 standing for bankrupted and 0 for healthy companies.

Type II error is the number of healthy companies classified as bankrupt ( $N_{01}$ ) divided by the number of healthy companies in the sample ( $N_0$ ):

$$e_{II} = \frac{N_{01}}{N_0}. \quad (10)$$

The overall error is calculated as follows:

$$e_T = \frac{N_{01} + N_{10}}{N_0 + N_1}. \quad (11)$$

In credit analysis, type I error implies the loss of capital loan and interest associated with a client that goes bust, when it was classified as healthy. Type II error leads to the loss of business with an existing or potential customer that is healthy, but was classified as bankrupt. Therefore, type I error has a higher cost for banks than type II error. According to Altman *et al.* (1977) type I error costs are 35 times higher for banks than type II error costs. However, other market players may have different approaches. If a government, for example, decided to implement a formal earlier public warning system, type II error costs could be higher for the country as a whole than from type I. This is because, if a healthy company is classified and publicly announced as being at risk of bankruptcy, the market would react negatively thus disrupting the firm's economic relationship with its suppliers and customers, increasing its probability of distress. Unfortunately, misclassification costs are not sufficiently documented in the literature and remain largely unknown.

A common difficulty in bankruptcy prediction is the use of non-random samples leading to biased probability estimations. Zmijewski (1984) studied two common estimation biases: the *choice-based sample bias* that results from 'over-sampling' bankrupt companies; and the *sample selection bias* that results from unavailable or incomplete data for distressed companies, when unavailable or incomplete data occur non-randomly. Zmijewski (1984) shows that the first type of bias can be reduced through adequate weighting of the cost function used in the classification algorithm and that both biases, in general, do not affect the statistical inferences and the overall classification rates. Wilson and Sharda (1994) studied the unbalanced number of distressed firms compared to healthy firms in the sample using three samples with variable proportions of healthy/bankrupt firms: 50/50, 80/20 and 90/10. They concluded that neural networks are more robust and outperformed discriminant analysis in overall classification but did not analyse the effect of type I error and type II error.

Classification performance is commonly measured by the percentage of observations classified correctly (OC). However, this measure is not adequate to evaluate the efficiency of classification since it blends the two types of errors. For example, if a sample consists of 70 healthy firms and 30 distressed firms and

the model classifies all firms as healthy, the overall classification would be 70% despite the fact that it was unable to identify one single bankruptcy.

To circumvent this problem we use a weighted efficiency (WE) that takes into account the two types of errors, independently of their cost to the creditor, defined as follows:

$$WE = \sqrt{OC \cdot BC \cdot BPC}. \quad (12)$$

OC is the overall classification

$$OC = \frac{N_{00} + N_{11}}{N_0 + N_1} \quad (13)$$

where  $N_{00}$  is the number of healthy companies classified correctly and  $N_{11}$  the number of distressed companies classified correctly.

BC is the bankruptcy classification, or the percentage of bankrupt firms classified correctly:

$$BC = \frac{N_{11}}{N_1}. \quad (14)$$

BPC is the bankruptcy prediction classification defined as the number of bankrupt firms to the total of predicted bankruptcies:

$$BPC = \frac{N_{11}}{N_{01} + N_{11}}. \quad (15)$$

This is a modification to the measure of efficiency presented by Korobow and Stuhr (1985) and is sensitive not only to the overall classification, but also to type I and type II errors. For perfect classification, all components are 1 and the efficiency is 100%. The square root was used since the three ratios are not independent. Consider, for instance, a balanced database with type I error equal to type II error and with a small  $N_{01}$ . In this case,  $N_1 = N_0 = N/2$  and if  $N_{11} = N_{00} = x$ , thus

$$WE \leq \sqrt{\frac{2x}{N} \frac{x}{N/2x}} = \sqrt{\left(\frac{2x}{N}\right)^2} = \frac{2x}{N} \quad (16)$$

which is closer to a linear function of the percentage of companies correctly classified,  $x/N$ , than the Korobow and Stuhr formula.

The predictive capability of our method was compared with traditional Multi-layer Perceptrons (MLP) and Multiple Discriminant Analysis (MDA) with two groups. The linear discriminant function was obtained applying a stepwise

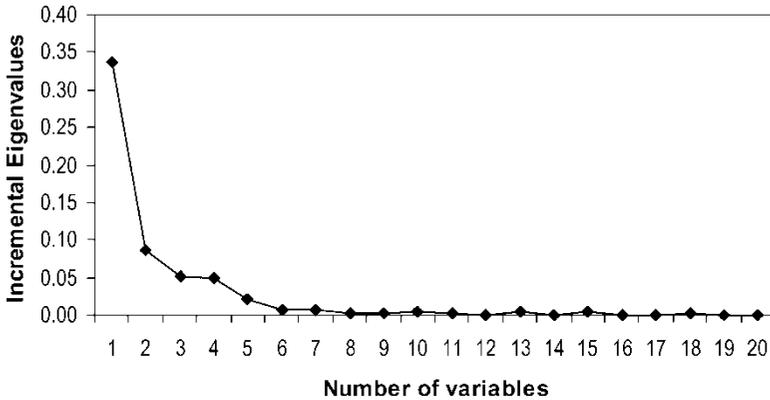


Figure 1. Eigenvalues for MDA.

method using a Wilk's Lambda and the criteria for entry variables was 3.84  $F$ -value and 2.71 for removal for the 20 variables selected by the ANN model. We ran an MDA with a leave-one-out classification and chose the five best discriminators as the eigenvalues show that selecting more than five ratios would offer negligible incremental gains in the MDA (see Figure 1).

Additionally we examined whether the five-variable  $Z$ -score model of Altman (1968) produces equivalent results to the MDA. This model has been used extensively as a standard of comparison by both academics and practitioners, as mentioned by Charitou *et al.* (2004).

#### 4. Data and Sample

The sample was obtained from Diana, a database containing approximately 780,000 financial statements of French companies and of their foreign subsidiaries. The initial sample consisted of financial ratios from 1998, 1999 and 2000 of 2,800 industrial French companies with at least 35 employees. Of these companies, 311 declared bankruptcy in the year 2000 and 272 companies submitted a restructuring plan ('Plan de redressement') to the court for their creditor's approval. We decided not to distinguish between these two categories since both signal companies in financial distress. Consequently, the sample has 583 financially distressed firms, most of them small to medium size with 35–400 employees.

This study's input consists of 30 financial ratios published by Coface,<sup>1</sup> which are available from the Diana database (see the Appendix). Additionally, we include the ratio of Cash Earnings to Total Debt that, according to Beaver (1966), is the best single discriminator of bankruptcy, the five ratios used by Altman (1968) in his  $Z$ -score model and the logarithm of sales as proxy to size.

Since more healthy firms are available than financially distressed firms, we randomly selected healthy companies to obtain the following data-sets with the

respective ratios of bankrupt to healthy firms: data-set 1 (50%/50%), data-set 2 (36%/64%) and data-set 3 (28%/72%).

#### 4.1. Feature Selection

This study uses more variables than other authors, who usually do not use more than 10. The ratios initially selected allow for a very comprehensive financial analysis of the firms including financial strength, liquidity, solvability, productivity of labour and capital, various kinds of margins and profitability and returns. Although, in the context of linear models, some of these variables have small discriminatory capabilities for default prediction, the non-linear approaches used here can extract relevant information contained in these ratios to improve the classification accuracy without compromising generalization.

Feature selection is an important issue in bankruptcy prediction, as in other problems where a large set of attributes is available, since elimination of useless features may enhance the accuracy of detection while reducing the amount of time for processing the data. Due to the lack of an analytical model, the relative importance of the input variables can only be estimated through empirical methods. A complete analysis would require examination of all possibilities, for example, taking two variables at a time to analyse their dependence or correlation, then taking three at a time, etc. This, however, is both infeasible and not error free since the available data may be of poor quality in sampling the full input space.

In the first step, highly correlated variables are automatically excluded. In a second step, elimination is based on the sensitivity of the neural network output to each variable. We thus evaluate the average sensitivity of every variable  $x_i$ , defined as

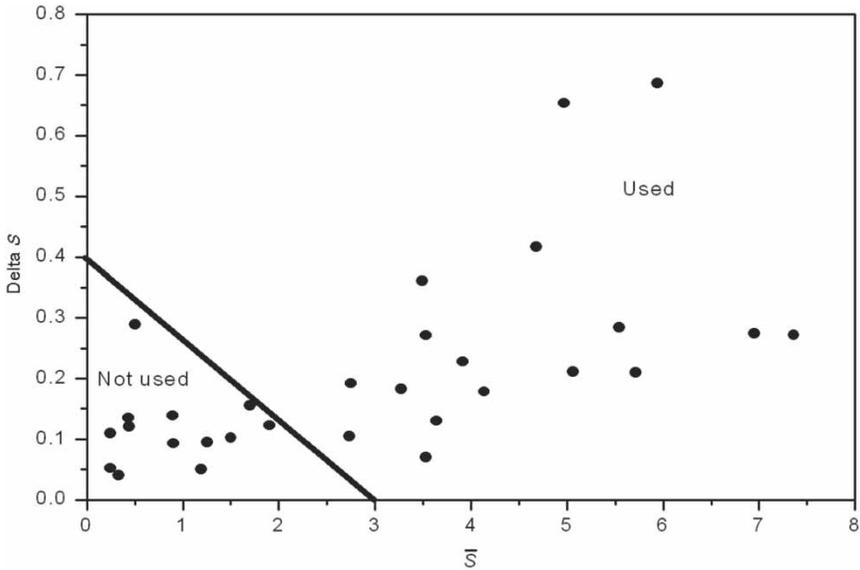
$$\bar{S}_i = \frac{1}{N} \sum_{j=1}^N \frac{y(x_1^j, x_2^j, \dots, x_i^j + \Delta \varepsilon_i, \dots, x_N^j) - y(x_1^j, x_2^j, \dots, x_i^j, \dots, x_N^j)}{\Delta \varepsilon_i} \quad (17)$$

where the sum is taken over all points of the database, and  $\Delta \varepsilon_i = 0.1$  is a small perturbation. The average sensitivity measures the linear and additive influence of the variables on the output. However, this measure does not take into account non-linear influences and possible interactions with other variables. These effects can be estimated by evaluating the standard deviation of the sensitivity  $\Delta S_i$ .

Variables with strong correlations with both small  $\bar{S}$  and  $\Delta S$  were eliminated (see Figure 2).

Thirteen variables were discarded: variables number 4, 6, 8, 9, 14, 16, 17, 22, 26, 27, 28, 29 and 30 (see Appendix).

Some of the selected ratios have large annual variations, sometimes greater than 20 or 30%, especially for companies that subsequently became financially distressed. Therefore, bankruptcy prediction based on financial ratios in a single year may be inaccurate, as it does not take these annual fluctuations into



**Figure 2.** Sensitivity analysis for variable selection

account. To circumvent this difficulty and to include relevant information from previous years, without overloading the neural network, we added the annual variation of the most important variables: Debt Ratio, Percentage of Value Added for Employee and Margin before Extra Items and Taxes. The final set had 20 inputs normalized to zero mean and unity variance.

## 5. Results and Discussion

We tested several neural networks with a single hidden layer, using between five and 20 nodes. We chose a hidden layer of 15 neurons, a learning rate of 0.1 and a momentum term of 0.25. Although smaller networks achieve slightly lower generalization errors, we had to establish a compromise since HLVQ-C performs better on larger hidden layers.

Some firms in the database have a financial record that clearly contradicts their actual financial status. For these evident cases, we decided to invert their output state artificially. This is the case for companies with negative shareholder equity, known as 'technical bankruptcy', which we always assigned to the financial distress category independently of their original category. Although this accounts for less than 3% of bankruptcies in the sample, some improvements were achieved on the training and testing error. The number of these cases found in our database shows that French managers and creditors do not initiate legal bankruptcy or restructuring processes immediately after the occurrence of financial distress.

Table 1 shows the results obtained in the balanced and unbalanced data-sets for 1999, one year prior to the bankruptcy announcement. The training error is considerably smaller than the generalization error, indicating that training data are insufficient. As expected, type I error is higher than type II since distressed companies are more heterogeneous and harder to identify. The results clearly show that using unbalanced samples is inadvisable, as type I error may increase considerably while type II error only improves slightly.

We tested the statistical significance of our results on the percentage of firms correctly classified applying the *t*-test to data-set 1 for 1999 and found statistical significance differences. When comparing HLVQ-C with MLP, we found a *t*-value of 2.16 whereas the *t*-value for the comparison of HLVQ-C and MDA is 4.30.

In Table 2, we compare the weighted efficiency achieved by the four methods. For all samples, HLVQ-C is clearly superior to the discriminant analysis and the non-corrected MLP. The efficiency of discriminant analysis deteriorated more noticeably than neural networks on unbalanced samples, and HLVQ-C is again the method with the highest stability.

**Table 1.** Results for a set sample of balanced and unbalanced sub-samples for 1999

Bankrupt/ healthy	Training error		Generalization error	
	Type I error (%)	Type II error (%)	Type I error (%)	Type II error (%)
<i>50/50</i>				
MLP	15.9	10.6	25.7	13.1
HLVQ-C	10.1	8.2	11.1	10.6
<i>36/64</i>				
MLP	13.8	4.6	30.9	8.8
HLVQ-C	5.8	2.1	18.7	7.3
<i>28/72</i>				
MLP	14.2	2.6	35.8	7.1
HLVQ-C	11.2	1.8	29.0	6.3

HLVQ-C means the output of the MLP is corrected by the method using HLVQ distances.

**Table 2.** Weighted efficiency for 1999

Sample	50/50	36/64	28/72
<i>Discriminant:</i>			
Best discriminant variables (%)	66.1	60.2	59.3
Z-Score variables (%)	62.7	52.1	47.5
<i>Neural networks:</i>			
MLP (%)	71.4	68.5	65.0
HLVQ-C (%)	84.1	78.9	71.0

**Table 3.** Weighted efficiency for 1998

Sample	50/50	36/64	28/72
<i>Discriminant analysis:</i>			
Best discriminant variables (%)	66.4	59.5	47.3
Z-Score variables (%)	61.1	50.9	32.0
<i>Neural networks:</i>			
MLP (%)	67.7	69.5	60.1
HLVQ-C (%)	76.5	74.3	69.5

We repeated the analysis for 1998, two years before bankruptcy.

As expected, predictions from all models deteriorate. The results from HLVQ-C are, once again, clearly superior to the traditional non-corrected MLP (Table 3) with statistically significantly lower error type I (Table 4). It can be seen that neural networks are less biased than discriminant analysis on unbalanced samples.

We also compared the efficiency of the neural network with the five ratios used in the discriminant model (Debt Ratio, Logarithm of Sales, Value Added per Employee, Accumulated Depreciation Ratio and Return on Assets), with a neural network – Table 5. We used only five hidden nodes in order to reduce the complexity of the network since a smaller set of variables is used. The generalization error, as expected, is slightly higher than with the full 20 inputs. However, HLVQ-C was unable to correct efficiently these errors since it does not have sufficient degrees of freedom in the hidden layer.

From sensitivity analysis applied to the neural network, we found that the most significant ratios signalling a financially distressed company (positive sensitivities) are Debt Ratio, Percentage of Value Added for Employees and Debt Ratio Variation. On the other hand, the most relevant ratios for a healthy

**Table 4.** Type I error for balanced and unbalanced samples

	Sample 50/50		Sample 36/64		Sample 28/72	
	1998	1999	1998	1999	1998	1999
<i>Discriminant:</i>						
Best discriminant variables	24.9	26.4	44.5	44.6	68.4	51.5
( <i>t</i> -Test against HLVQ-C)	(2.7)*	(6.0)*	(8.0)*	(8.9)*	(8.8)*	(5.4)*
Z-Score variables	31.6	26.8	57.1	54.5	83.2	66.0
( <i>t</i> -Test against HLVQ-C)	(4.6)*	(5.5)*	(11.8)*	(11.8)*	(12.2)*	(8.7)*
<i>Neural networks:</i>						
MLP	24.9	25.7	30.9	30.9	44.9	35.8
HLVQ-C	16.0	11.1	16.0	18.7	27.8	29.0
( <i>t</i> -Test against HLVQ-C)	(2.1)	(4.8)*	(3.2)*	(3.6)*	(4.2)*	(1.9)

\*Statistically significant at 1% level.

**Table 5.** Neural networks trained with the five inputs chosen by the discriminant analysis, using the balanced database (in 1999)

	Training error		Generalization error	
	Type I error	Type II error	Type I error	Type II error
MLP	20.1	15.6	25.3	17.1
HLVQ-C	14.3	10.5	23.7	14.8

company (negative sensitivities) are Valued Added per Employee, Margin before Extra Items and Taxes, and Accumulated Earnings to Assets.

Financial distress companies tend to have higher debt ratios and lower value added. For these companies, the percentage of value added attributed to employees is higher, and in some cases, employees absorb the total value added generated by the company. It is also worth noting that salaries in France tend to be a fixed cost because of labour legislation.

Financially healthy companies are characterized by high productivity and, consequently, higher value added per employee and profitability, with larger margins before extra items, taxes and higher accumulated profits.

### 5.1. Robustness

We can use HLVQ-C to estimate the robustness of the prediction. For simplicity, consider a trained MLP with a single hidden neuron and a single output  $y$ . The correction to  $y$  introduced by adding a new element  $(\vec{x}, t)$  in the training set is

$$\Delta y = \frac{\partial y}{\partial t} \Delta t = \frac{\partial y}{\partial w} \frac{\partial w}{\partial t} \Delta t \quad (18)$$

where  $t$  is the target,  $w$  the last layer network weight and  $\Delta t = (t - y)$  the output correction. Following the back-propagation algorithm, for a sigmoid transference function, we have

$$|\Delta w| = \eta y(1 - y)h\Delta t \quad (19)$$

where  $\eta$  is learning rate and  $h$  the hidden layer output. The correction is therefore

$$|\Delta y| = \eta y(1 - y) \frac{\partial y}{\partial w} h\Delta t \quad (20)$$

where the term  $\partial y/\partial w$  is evaluated numerically. If the company is classified beyond doubt, then  $y \sim 1$  or  $y \sim 0$  and corrections are small. Major corrections occur when  $y \sim 0.5$ , that is, points closely located to the MLP separation

hyperplane, and points located in regions poorly covered by the training data where  $\partial y/\partial w$  is important.

The output variation

$$\varepsilon = |\Delta y| = |y_1 - y_0| \tag{21}$$

measures the robustness of the prediction against variations in target values. This estimate is of great importance for the analyst since it quantifies the reliability of predictions made by the neural network.

5.2. Some Examples of HLVQ-C Corrections

Finally, the capabilities of HLVQ-C to correct the output of the MLP are demonstrated by analysing some cases (Table 6). Cases 1–3 are bankrupt companies and case 4 is a healthy company. Note that in all these cases the Z-score fails the prediction.

Case 1 is a small bankrupt company with 44 employees. Although it has a low debt ratio, the percentage of value added absorbed by employees is high and the margin before extra items and taxes is low. The MLP predicts a bankruptcy probability of 0.19 only, but HLVQ-C estimated 0.31, pushing the probability in the right direction. The financial situation of case 2 is worse, with higher debt ratio, higher value added absorbed by employees and a negative margin before extra items and taxes. The MLP probability of bankruptcy is 0.49, which is hard to classify but the HLVQ-C adjusts the probability to 0.70, which is clearly a bankrupt company. Case 3 is a less clear situation. The debt ratio is relatively high but with substantial improvement. The profitability is however low and the value added absorbed by employees high. The HLVQ-C corrects the probability of bankruptcy from 0.36 to 0.55, moving the probability in the right direction.

Case 4 is interesting because, from a financially difficult situation in 1998, the company was able to achieve a remarkable improvement in economic

**Table 6.** Some examples with significant corrections from HLVQ-C

Case study	1	2	3	4
Employees	44	37	51	60
Debt ratio	62.8	90.8	78.7	66.3
Debt ratio previous year	51.8	74.1	91.8	91.6
Value added per employee	20	28	40	31
Value added for employees	93.7	98.7	84.6	72.3
Margin before interest and taxes	-1.5	-5.6	0.1	9.0
Margin before interest and taxes in previous year	2.4	-4.9	1.08	-10.6
Z-Score	3.24	2.63	3.08	1.79
MLP	<b>0.19</b>	<b>0.49</b>	<b>0.36</b>	<b>0.90</b>
HLVQ-C	<b>0.31</b>	<b>0.70</b>	<b>0.55</b>	<b>0.62</b>
Variation ( $\varepsilon$ )	0.11	0.25	0.26	0.38

performance and financial position in 1999. Based on the previous year's performance, the neural network assigns a high bankruptcy probability (0.9). However, HLVQ-C is able to correct this value to 0.62. Moreover, the large  $\varepsilon$  parameter (0.37) is also an indication that this is an uncommon situation and the prediction has a high level of uncertainty.

## **6. Conclusions**

We applied neural networks to bankruptcy prediction using a new method called HLVQ-C. This method can handle a higher number of financial ratios and is able to substantially improve the predictions of MLP without compromising generalization.

A measure of classification efficiency was introduced to evaluate the method's performance. We found that HLVQ-C clearly outperformed linear discriminant analysis, the Z-score model of Altman and traditional neural networks.

Predictions based on the financial ratios of a single year can be misleading as some ratios show large yearly variations. When one-year variations of the major ratios were included, the predictions improved substantially. These results show that it is desirable to include more years in the analysis, as two years is not enough to determine a trend.

We also concluded that it is desirable to use balanced data-sets, containing the same number of healthy and bankrupt companies, in order to constrain type I errors.

Banks and creditors in general can benefit from the method that shows a significant statistical reduction of type I error and that consequently reduces the costs of default. For their credit strategy, banks could also use the output of a HLVQ-C model for targeting sectors that tend to have a lower probability of bankruptcy. They could also use the model to estimate the credit at risk by applying the borrower's probability of default. Credit risk management is 'a critical component of a comprehensive approach to risk management and essential to the long-term success of any banking organization' (Basle Committee on Banking Supervision, 2000). Therefore, banks using discriminant analysis for credit risk management could benefit by incorporating this model in their systems.

## **Acknowledgements**

The authors would like to acknowledge Duarte Trigueiros, Andrew H. Sung and Srinivas Mukkamala for their comments on an earlier version of this paper, the helpful comments of two anonymous referees, and the improvements suggested by Srinivas Mukkamala in a final revision as well as the financial support granted by the Fundação para a Ciência e Tecnologia (FCT) and the POCTI programme.

## Note

<sup>1</sup>Coface is a credit risk provider in France that offers the Conan–Holder bankruptcy score, a score based on a discriminant analysis developed by the authors (Conan and Holder, 1979).

## References

- Altman, E. I. (1968) Financial ratios, discriminant analysis and the prediction of corporate bankruptcy, *Journal of Finance*, 23(4), pp. 589–609.
- Altman, E. I. *et al.* (1977) Zeta analysis. A new model to identify bankruptcy of corporations, *Journal of Banking and Finance*, 1, pp. 29–54.
- Altman, E. I. *et al.* (1994) Corporate distress diagnosis: comparisons using linear discriminant analysis and neural networks (the Italian experience), *Journal of Banking and Finance*, 18, pp. 505–529.
- Amemiya, T. and Powell, J. L. (1983) A comparison of the logit model and normal discriminant analysis when independent variables are binary, in: S. Karlin *et al.* (Eds) *Studies in Econometric, Time Series, and Multivariate Statistics*, pp. 3–30 (New York: Academic Press).
- Banzhaf, W. *et al.* (1989) *Genetic Programming: An Introduction on the Automatic Evolution of Computer Programs and its Applications* (San Francisco, CA: Morgan Kaufmann).
- Barker, D. (1990) Analyzing financial health: integrating neural networks and expert systems, *PC AI*, 4(3), pp. 24–27.
- Basle Committee for Banking Supervision (2000) Principles for the management of credit risk, *Basel Committee Publications*, No. 75 (September).
- Beaver, W. H. (1966) Financial ratios as predictors of failures. Empirical research in accounting: selected studies, *Journal of Accounting Research*, 5(Suppl.), pp. 71–102.
- Berry, R. H. and Trigueiros, D. (1993) Applying neural networks to the extraction of knowledge from accounting reports: a classification study, in: R. R. Trippi and E. Turban (Eds) *Neural Networks in Finance and Investing* (Chicago: Probus).
- Bishop, C. M. (1995) *Neural Networks for Pattern Recognition* (Oxford: Oxford University Press).
- Charitou, A. *et al.* (2004) Predicting corporate failure: empirical evidence for the UK, *European Accounting Review*, 13(3), pp. 465–497.
- Coats, P. K. and Fant, L. F. (1993) Recognising financial distress patterns using a neural network tool, *Financial Management*, Autumn, pp. 142–155.
- Conan, D. and Holder, M. (1979) Variables explicatives de performance et contrôle de gestion dans les P. M. I., Thèse d'Etat, CERG, Université Paris Dauphine.
- Eisenbeis, R. A. (1977) Pitfalls in the application of discriminant analysis in business, finance and economics, *Journal of Finance*, 32(3), pp. 875–900.
- Fan, A. and Palaniswami, M. (2000) Selecting bankruptcy predictors using a support vector machine approach, *International Joint Conference on Neural Networks*, Como, Italy, pp. 24–27.
- Grice, J. S. and Dugan, M. T. (2001) The limitations of bankruptcy prediction models: some cautions for the researcher, *Review of Quantitative Finance and Accounting*, 17(2), pp. 151–159.
- Korobow, L. and Stuhr, D. (1985) Performance of early warning models, *Journal of Banking and Finance*, 9, pp. 267–273.
- Laitinen, T. and Kankaanpää, M. (1999) Comparative analysis of failure prediction methods: the Finnish case, *European Accounting Review*, 8(1), pp. 67–92.
- Lo, A. W. (1986) Logit versus discriminant analysis, a specification test and application to corporate bankruptcies, *Journal of Econometrics*, 31, pp. 151–178.
- Marose, R. A. (1990) A financial network application, *AI Expert*, May, pp. 50–53.
- Martin, D. (1977) Early warning of bank failure: a logit regression approach, *Journal of Banking and Finance*, (1), pp. 249–276.
- McFadden, D. (1976) A comment on discriminant analysis versus logit analysis, *Annals of Economic and Social Measurement*, (5), pp. 511–524.

- Ohlson, J. A. (1980) Financial ratios and the probability of bankruptcy, *Journal of Accounting Research*, 18, pp. 109–131.
- O’Leary, D. (1998) Using neural networks to predict corporate failure, *International Journal of Intelligent Systems in Accounting, Finance & Management*, (7), pp. 187–197.
- Press, S. J. and Wilson, S. (1978) Choosing between logistic regression and discriminant analysis, *Journal of the American Statistical Association*, 73, pp. 699–705.
- Srinivasan, V. and Kim, Y. H. (1987) Credit granting: a comparative analysis of classifications procedures, *Journal of Finance*, 42, pp. 665–683.
- Stein, J. and Ziegler, W. (1984) The prognosis and surveillance of risks from commercial credit borrowers, *Journal of Banking and Finance*, 8, pp. 249–268.
- Tan, C. and Dihadjo, H. (2001) A study on using artificial neural networks to develop an early warning predictor for credit union financial distress with comparison to the probit model, *Managerial Finance*, 27(4), pp. 56–77.
- Wilson, R. L. and Sharda, R. (1994) Bankruptcy prediction using neural networks, *Decision Support Systems*, 11, pp. 545–557.
- Yang, Z. R. (1999) Probabilistic Neural Networks in bankruptcy prediction, *Journal of Business Research*, 44, pp. 67–75.
- Zavgren, C. (1983) The prediction of corporate failure: the state of the art, *Journal of Accounting Literature*, 2, pp. 1–38.
- Zmijewski, M. (1984) Methodological issues related to the estimation of financial distress prediction models, *Journal of Accounting Research*, 22(Suppl.), pp. 59–86.

## Appendix

**Table A1.** Mean and standard deviation of all indicators for bankrupt and healthy companies in 1999

Ratio	Definition	Bankrupt ( $\bar{x}_i, \sigma_i$ )	Healthy ( $\bar{x}_i, \sigma_i$ )
1	Number of employees	85, 83	129, 118.3
2	Financial equilibrium ratio	3.2, 47.5	2.0, 11.2
3	Equity to stable funds	–3.0, 115.6	40.9, 18.1
4	Debt to stable funds	33.5, 72.3	14.0, 12.8
5	Financial autonomy	7.2, 32.4	38.3, 18.3
6	Accumulated depreciation rate (%)	66.1, 20.6	68.5, 14.3
7	Current ratio	1.3, 0.6	0.4, 0.3
8	Quick ratio	0.9, 0.5	1.4, 1.3
9	Inventory days of sales	57.0, 66.8	56.1, 63.5
10	Collection period	62.1, 42.9	74.6, 33.0
11	Interest to sales (%)	1.9, 5.3	1.2, 1.9
12	Debt ratio	89.2, 31.9	58.4, 18.4
13	Financial debt to cash earnings	4.7, 68.7	1.9, 52.2
14	Cash earnings to sales (%)	–0.08, 13.8	4.8, 6.1
15	Working capital in sales days	19.0, 77.5	77.5, 81.3
16	Working capital requirements in sales days	14.1, 77.2	56.6, 67.0
17	Exportation (%)	14.5, 21.7	18.8, 23.9
18	Value added per employee	32.8, 15.8	48.9, 29.0
19	Value added to assets	0.6, 0.5	0.4, 0.2
20	EBITDA margin	1.4, 13.3	7.3, 7.5
21	Margin before extra items and taxes	–2.6, 18.3	4.0, 7.0

(continued)

Table A1. Continued

Ratio	Definition	Bankrupt ( $\bar{x}_i, \sigma_i$ )	Healthy ( $\bar{x}_i, \sigma_i$ )
22	Net margin	-2.1, 18.2	2.1, 6.3
23	Return on equity	-44.0, 116.3	12.0, 34.1
24	Value added margin	42.3, 19.4	39.2, 15.3
25	Percentage of value added for employees	91.5, 32.4	76.1, 31.1
26	Sales ('000 euro)	9,093, 17,350	25,217, 40,412
27	Working capital to current assets	0.04, 0.5	0.4, 0.3
28	Payment period	89.0, 121.3	77.0, 31.1
29	Debt on sales days	209.0, 193.2	142.0, 73.3
30	Return on equity before extra items and taxes	-40.1, 135.4	12.4, 34.0
<i>Altman ratios (Z-score)</i>			
31	Working capital to assets	0.067, 0.32	0.288, 0.19
32	Accumulated earnings to assets	-0.071, 0.36	0.234, 0.21
33	Return on assets	0.001, 0.14	0.066, 0.09
34	Equity to debt	0.205, 0.42	0.944, 1.39
35	Asset turnover	1.948, 0.92	1.711, 0.73

