See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/222492052

# Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience)

Article <i>ir</i>	n Journal of Banking & Finance · May 1994 /0378-4266(94)90007-8		
CITATIONS 974		READS 5,449	
3 author	s, including:		
0	Edward Altman New York University 188 PUBLICATIONS 31,476 CITATIONS SEE PROFILE	(i)	Franco Varetto IRCrES-CNR Turin (in the past: Politecnico di Torino) 7 PUBLICATIONS 1,234 CITATIONS SEE PROFILE

#### Some of the authors of this publication are also working on these related projects:

Project

International Risk Management Conference - IRMC View project

# Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience)\*

# Edward I. Altman

New York University, Stern School of Business, New York, NY 10012-1126, USA and Centrale dei Bilanci, Turin, Italy

# Giancarlo Marco and Franco Varetto

Centrale dei Bilanci, Turin, Italy

Received November 1993, final version received January 1994

This study analyzes the comparison between traditional statistical methodologies for distress classification and prediction, i.e., linear discriminant (LDA) or logit analyses, with an artificial intelligence algorithm known as neural networks (NN). Analyzing well over 1,000 healthy, vulnerable and unsound industrial Italian firms from 1982–1992, this study was carried out at the Centrale dei Bilanci in Turin, Italy and is now being tested in actual diagnostic situations. The results are part of a larger effort involving separate models for industrial, retailing/trading and construction firms.

The results indicate a balanced degree of accuracy and other beneficial characteristics between LDA and NN. We are particularly careful to point out the problems of the 'black-box' NN systems, including illogical weightings of the indicators and overfitting in the training stage both of which negatively impacts predictive accuracy. Both types of diagnostic techniques displayed acceptable, over 90%, classification and holdout sample accuracy and the study concludes that there certainly should be further studies and tests using the two techniques and suggests a combined approach for predictive reinforcement.

Key words: Distress diagnosis; Discriminant models; Neural networks; Corporate bankruptcy risk; Financial ratio analysis

JEL classification: G33; C49; C88

Correspondence to: Franco Varetto, Centrale dei Bilanci s.r.l., Società per gli Studi Finanziari, Corso Vittorio Emanuele II, 93, 10128 Torino, Italy. Tel. 39-11-562 7366, Fax 39-11-562 7490.

\*We would like to thank Professor P. Coats of Florida State University for the documents she kindly supplied us, and Professor L. Saitta of the University of Turin for the numerous and profitable discussions about neural networks and the applications of artificial intelligence. We have profited by the comments of Professor Piero Terna and Professor Francesco Borazzo of the University of Turin. An earlier version of this paper was presented at the 'International Seminar on European Financial Statement Data Bases: Methods and Perspectives', Bressanone,

# 1. Introduction

The Centrale dei Bilanci (CB) is an organization established in 1983 by the Banca d'Italia, the Associazione Bancaria Italiana and over forty leading banks and special credit institutions in Italy. In 1993, the 'Sistema Informativo Economico e Finanziario' (the Economic and Financial Information System of the CB which monitors Italian businesses) included approximately seventy members.<sup>1</sup>

One of the 'products' of the CB is a system designed to provide banks with a tool to quickly identify companies that are in financial trouble. The development of this System commenced in 1988 with the creation of an initial version based on a pair of linear discriminant functions, working parallel to one another and adapted to the industrial sector. The functions were estimated from a sample of 213 unsound (distressed) companies compared to a sample group of the same number of healthy companies; the estimation was made on the second year prior to the time that the state of distress was recognized.<sup>2</sup> This system correctly classified, in the year immediately prior to distress, 87.6% of healthy companies and 92.6% cases of unsound companies. For a description of the features of this initial System, see Varetto (1990). In 1989, the System was distributed to half of the banks belonging to CB for actual application in credit analysis at their head offices. The result of the experiment confirmed the System's soundness. In practical terms, automatic diagnosis systems can be used to preselect businesses to examine more thoroughly, quickly and inexpensively, thereby managing the financial analyst's time efficiently. These systems can also be used to check and monitor the uniformity of the judgements made about businesses by the various branches of the bank, without replacing credit analyst personnel.

On the basis of the experiments performed and making use of an extended data base, the CB created a second version of the Diagnostic System that was completed and distributed to the banks belonging to Centrale's information system during 1991. In the same year, initial tests were conducted into the use of neural networks for the identification of businesses showing economic and financial distress.

The aim of this paper is to illustrate the results achieved with neural networks, comparing them with discriminant analysis results and its appli-

Italy, September 16-17, 1993.

<sup>&</sup>lt;sup>1</sup>In addition to the management of data bases with information on the financial statements of over 37,000 companies collected every year, the Centrale dei Bilanci is actively engaged in several lines of operation: the development of financial analysis methodology, production of user software, management education and financial and industrial economic research.

 $<sup>^{2}</sup>$ At the time that the distressed state is recognized, there is a break in the historical series of the annual reports in the data base. For companies subject to the law governing bankruptcy and failure, a period of time passes between the suspension of the availability of balance sheets and the moment of the final declaration of bankruptcy or composition.

cations. The next section gives a brief description of the existing version of the Diagnostic System obtained using what is now recognized as traditional statistical discriminant analysis methodology. The third section examines the essential aspects of the neural network approach. The main conclusions that can be drawn from the experiments in the use of the neural networks (NN) may be summed up as follows:

- a) Neural networks are able to approximate the *numeric* values of the scores generated by the discriminant functions even with a different set of business indicators from the set used by the discriminant functions;
- b) Neural networks are able to accurately classify groups of businesses as to their financial and operating health, with results that are very close to or, in some cases, even better than those of the discriminant analysis;
- c) The use of integrated families of simple networks and networks with a 'memory' has shown considerable power and flexibility. Their performance has almost always been superior to the performance of single networks with complex architecture;
- d) The long processing time for completing the NN training phase, the need to carry out a large number of tests to identify the NN structure, as well as the trap of 'overfitting' can considerably limit the use of NNs. The resulting weights inherent in the system are not transparent and are sensitive to structural changes;
- e) The possibility of deriving an illogical network behavior, in response to different variations of the input values, constitutes an important problem from a financial analysis point of view;
- f) In the comparison with neural networks, discriminant analysis proves to be a very effective tool that has the significant advantage for the financial analyst of making the underlying economic and financial model transparent and easy to interpret;
- g) We recommend that the two systems be used in tandem.

Perhaps the main conclusion of this study is that neural networks are not a clearly dominant mathematical technique compared to traditional statistical techniques, such as discriminant analysis. The tendency for recently published articles on the use of NN approaches in financial distress classification (a number of references to these studies follows shortly) is that this 'new' technique is clearly superior. We find that a more balanced conclusion is appropriate, indicating advantages and disadvantages of the 'black-box' NN technique.

In addition, our study is one that is being applied and tested within an operation that has the potential for being implemented in an actual business and financial context by concerned practitioners. Finally, our samples, consisting of over 1,000 Italian firms, is by far the largest of any distressed prediction study to date – including those using discriminant analysis or NN approaches.

## 2. Centrale dei Bilanci's system of diagnostic risk of distress

Distressed firm risk analysis is one of the CB's permanent projects aimed at developing analytical methodologies concerning business credit. This project allows for the periodic updating of the discriminant functions to maintain or enhance their diagnostic capabilities. The integral parts of the project are the construction and maintenance of a specific data base of unsound companies and the development of research on the companies' dynamics of economic decline leading to distress and bankruptcy.

The System is based on the application of the traditional linear discriminant analysis methodology on the basis of two samples of businesses representative of healthy and unsound companies.<sup>3</sup> A numerical score is obtained from the discriminant function that expresses the 'risk profile' of the business.

Unlike the first version of the System, the new release includes special models each for trading and construction companies as well as the industrial model developed earlier.<sup>4</sup> Work discussed in this study only refers to the existing model for industrial companies.

The essential points are as follows:

- a) The Diagnostic System has been designed and set up to be applied to the medium and small sized businesses in Italy. For this reason, companies with sales of more than 100 billion liras (i.e. 60 million U.S. dollars) have been excluded from the sample. Our tests involve data covering the period 1985–1992.
- b) We have utilized a balanced sample of healthy and unsound companies, rather than to consider all the collected companies in the files of the CB (around 37,000 companies a year) since our sample is quite large in and of itself. This methodological line is common to other models of discriminant analysis.
- c) The discriminant models had only modest ex-post accuracies while using large samples of 'healthy' (non-bankrupt) businesses due to the fact that these companies are broken down into at least three large subsets: 'outstanding', 'normal' and 'vulnerable' companies. And, the breadth of these categories, just as their features, varies over time. The discriminant analysis model seems limited in its ability to differentiate between unsound companies and companies that are 'live' but belong to the vulnerable subset. Certainly, it is far more difficult to discriminate between two 'sick' firm samples (unsound and vulnerable) than between the clearly healthy vs. unsound firms. Consequently, with the

<sup>&</sup>lt;sup>3</sup>For a description of the methodological aspects of discriminant analysis and the main models available in different countries, see Altman (1993).

<sup>&</sup>lt;sup>4</sup>The trading and building sector models are still being tested and will be reported on in a subsequent publication.



Fig. 1. Diagnostic system flow. This chart indicates the basic progression of discriminant analysis models performed within the corporate monitoring system at the Centrale dei Bilanci, Torino, Italy.

increase in the size and industrial scope of the sample, rates of recognition decrease because of the increase in the variability of possible situations. The accuracy does not improve even if use is made of more sophisticated bayesian-type statistical methodologies.

- d) To tackle this problem, it was decided to take another path in the revised version of the System. The Diagnostic System was broken down into two sub-models working in sequence. The first model (F1) was estimated on samples of 404 unsound companies and 404 healthy companies: the former were identified from the entire population of companies collected in the files of Centrale dei Bilanci which underwent (1) some form of bankruptcy proceeding, (2) were wound up in temporary receivership or (3) had stated they were in dire straits with regard to their payments to the banks. The sound-firm sample was obtained from the 'live' company file excluding 'vulnerable' businesses identified through the use of tests on a restricted number of business ratios over the span of a few years. These ratios are not part of those included in the discriminant functions. The sample of businesses 'running normally' was obtained by matching with similar distressed companies by size (in terms of net assets), industry and location. The first model consisted of a nine ratio linear function (F1) that distinguished between 'healthy' businesses and 'unsound' or 'vulnerable' businesses (Fig. 1).
- e) The second phase of the model (F2) comes into play after F1 has diagnosed the business to be 'unsound'. The second discriminant function was estimated from two balanced samples, again each of 404 businesses, of unsound (the same ones used for the F1 estimation) and 'vulnerable' companies. The latter were extracted from among the 'live' sample but found to be diagnostically 'unsound' by F1. Both functions

		Healthy firms	Unsound firms
Estimation sample (404 companies			
in each group)			
Estimation period	T-3	90.3%	86.4%
Control period	T-1	92.8%	96.5%
Holdout sample (150 companies			
in each group)	<b>T-</b> 1	90.3%	95.1%

# Table 1 Rate of successful recognition (F1 discriminant function)

Table 2

Rate of success	ful recognition (F	2 discriminant function)	
Estimation sample (404 cos.)		Vulnerable firms	Unsound firms
Estimation period	T-3	99.0%	60.1%
Control period	<b>T-1</b>	97.8	82.7
Holdout sample			
(150 cos.)	T-1	96.8	81.0

have been estimated based on ratio values from the annual report of the third year prior to the distress date.

- f) All variables of the F1 and F2 models which contained coefficients with counter-intuitive signs were eliminated (even if they were statistically significant). Also, variables with unstable behavior were eliminated and only those that increased the capacity to classify the unsound companies as the time prior to distress approached and maintained (or increased) the capacity to classify healthy businesses were retained.
- g) Estimations were made using logit as well as discriminant analysis but no significant progress was made on ex-post classification. Therefore, we retained the discriminant functions.
- h) The discriminatory capacity of the principal function (F1), on which most of the experiments with neural networks are compared, is shown in Table 1.

The percentage of correct ex-post classification improves as distress approaches; for the unsound companies it goes from 86.4% in T-3 (estimation period) to 96.5% in period T-1. The accuracy of the classification was checked with a holdout sample of 150 unsound businesses and 150 healthy ones, obtaining results that were similar to the estimation sample (90% and 95% in period T-1).

The second function, as expected, has a lower discriminant capacity, especially for the unsound firms. Table 2 lists the F2 function results showing 82.7% correct classification of the unsound firms in the control period (T-1) and 81.0% in the holdout sample for that group (vs. about 95\% in the F1 Function).



Fig. 2. Top: Function F1 – example of unsound company. Bottom: Function F2 – Example of unsound company.

To make it easier to interpret the results, the scores of the functions are represented on graphs where the business under examination is positioned on the two different reference systems: Fig. 2a is an example of an unsound business monitored over the last five years of its life (1985–1989). From F1, the firm is identified as a distressed company in the fifth year prior to failure. At this stage, the system does not yet distinguish if the unsound business is simply vulnerable (with a greater or lesser degree of vulnerability) or if it belongs to the set of unsound companies. Fig. 2b shows the diagnosis of the same firm made by F2 and places the business in the uncertain area between vulnerability and risk of bankruptcy in the first two years of the series and then signals a rapid decline into the higher risk bankruptcy zone. As can be seen, the diagnosis of the company is carried out on the basis of a joint analysis of the two functions with additional reference points supplied by quartile comparisons with the entire CB data base of comparable companies.

The classificatory space described by F1 has been divided into five zones on the basis of the distribution of healthy, vulnerable and unsound companies. These include: (a1) high security; (b1) security; (c1) uncertainty between security and vulnerability; (d1) vulnerability; and (e1) intense vulnerability. Function F2 is calculated as soon as F1's score falls in one of the zones (c1),

b

(d1) or (e1): F2 has been split into zones of: (a2) high vulnerability; (b2) vulnerability; (c2) uncertainty between vulnerability and risk; (d2) risk; and (e2) high risk of bankruptcy.

Score values separating the different zones constitute the ordinates of the fixed classification system shown on the graphs.<sup>5</sup>

# 3. Neural networks

For many years, neural network models have been analyzed both by academics and practitioners, including those efforts outside the circle of artificial intelligence experts.<sup>6</sup> It is too early to say whether the use of experimental N.N. is simply a fad or it will result into something more permanent. Some aspects of the neural networks, however, do seem promising in the area of business and finance applications.<sup>7</sup>

The application of the NN approach to company distress prediction, although relatively new, has seen a number of researchers attempt to improve upon the traditional discriminant analysis technique. An interesting procedure by Coats and Fant (1993), used a limited number of financial ratios to duplicate the 'going-concern' determination by accounting auditors. They utilize the cascade-correlation NN approach (Fahlman and Lebiere, 1992) to duplicate the auditor-expert conclusion on a sample of 94 manufacturing and non-manufacturing failed firms and conclude that it clearly dominates the LDA method in this application.<sup>8</sup> In addition, studies by

<sup>6</sup>For an introduction to the theory of neural networks and the operating mechanisms, see Rumelhart and McClelland (1989); Cammarata (1990); Khana (1991); Freeman and Skapura (1991); and Hertz et al. (1991).

<sup>7</sup>In the area of finance, there have been a number of recent attempts to apply NN. Cadden (1991) has applied neural networks to insolvency analysis by adopting a Boolean transformation of the financial ratios divided into quartiles; Chung and Tam (1993) have compared the performance of the neural networks with that of other inductive learning algorithms for bankruptcy forecasting in the banking industry; Bell et al. (1990) have compared neural networks with logistic regression for the prediction of bank failures. The networks have also been assigned to the rating of bonds (Dutta and Shelber, 1992) to the prediction of the progress of historical series of company data to the selection of investments and to operations on the financial market (Swales and Yoon, 1992), (Wong et al., 1993), and (Trippi and de Sieno, 1992), and the recognition of accounting data patterns (Liang et al., 1993). Kryzanowski, Galler and Wright (1993) applied NN for positive vs. negative common stock return prediction and Kryzanowski and Galler (1994) have analyzed the financial statements of small businesses using neural nets. For a partial list of applications in the financial field, see Pau and Gianotti (1993).

<sup>8</sup>While the Coats and Fant (1993) analysis is of relevance, we must point out that the auditors' qualification is itself an inexact and subjective process and as we have shown in an earlier study (Altman and McGough, 1974), that the discriminant analysis Z-score approach was far more accurate in predicting the actual bankruptcy of a sample of failed firms than was the

<sup>&</sup>lt;sup>5</sup>The coefficients of all the functions are protected by secrecy for the purpose of safeguarding the investments of the *CB*'s owners made in research, testing and data base creation. This latest version of the two-function system has been inserted in a procedure on the PC and distributed to around thirty of the member banks. Actual application in the field is underway and has already given significant, favorable signs.



Fig. 3. General scheme of neural unit.

Karels and Prakash (1987), Odom and Sharda (1990), Ragupathi et al. (1991) and Rahimian et al. (1992) have all assessed NN for bankruptcy prediction. Interestingly, at least three of the above studies utilized the same five financial variables found in Altman's (1968) study.

This paper will explore the basic theory of neural networks but we do not plan to discuss in detail the reasons that inspired the connectionist approach. Connectionist processing models (neural networks) consist of a potentially large number of elementary processing units; every unit is interconnected with other units and each is able to perform relatively simple calculations. The network's processing result derives from their collective behavior rather than from the specific behavior of a single unit. The links are not rigid but can be modified through learning processes generated by the network's interaction with the outside world or with a set of symbolic signals.

The individual units and the connections linking them can be shown as in Fig. 3: each unit (i) receives an input  $(x_i)$  from the outside, or from other neurons with which it is linked, with an intensity (weight) equal to  $w_{ji}$ . The overall input that the *i*th neuron receives equals an assumed potential  $(P_i)$  equal to:

$$P_i = \sum_j n W_{ji} * X_i - S_i$$

where  $S_i$  represents an excitation threshold value that limits the neuron's degree of response to the stimuli received: for example the neurons give a response signal in the 'jump-type' response function only if the total input arriving from outside and/or other neurons is greater than  $S_i$ . It is possible to eliminate the  $S_i$  threshold and replace it with a dummy input (k) of a value

so-called accountant-expert. Still, the auditing disclaimer report is an unambiguous, although possibly incorrect, indicator of distress. And, the Coats and Fant sample of distressed firms were those that discontinued operations after receiving a going-concern qualification.

equal to 1  $(X_k=1)$  and by setting  $W_{kl}=-S_i$ , obtaining the general expression.

$$P_i = \sum_j W_{ji} \cdot X_j$$
, where k is included in  $X_j$ 

The neuron's response  $(Y_i)$  depends on the transfer of potential  $(P_i)$  to the output function. One of the most widely used functions in the literature and used in our tests is the logistic function, according to which:

$$Y_i = \frac{1}{1 + e^{-\mathbf{P}_i}}$$

Generally, the response function determines values between a minimum and a maximum; in our case  $Y_i$  is included between 0 and 1. Output  $(Y_i)$  of the neuron can be either a total response value of the network (if it is the final output value) or an input for further neuron units. The network, made up of many elementary units of the *i*th type, can have different degrees of complexity. The simpler networks consist of a single neuron layer (in extreme cases by a single neuron) each of which is in direct contact with the outside stimuli *i* and generates output from the network directly. A slightly more complicated network has two layers: an intermediate, hidden layer, that receives stimuli from outside the network, and an output layer that generates the network's responses. Networks can be constructed with circuits for feedback between neurons from one layer to those of previous levels, just like self-connecting links.

Considerable limitations of a single-layer network have been shown. Networks with one layer, in addition to the input layer, can only perform linear separations of the input space. Two-layer networks can generate convex geometrical shapes, while networks with at least three layers enable the input space to be separated into shapes of any configuration (the complexity of the regions is determined by the number of neurons). There are no general rules to establish the optimal degree of network complexity.

The crucial aspect of neural networks lies in the fact that the weightings of the connections are not fixed but can be modified on the basis of a learning procedure derived from the comparison of the network responses with those required by actual results. The network, in other words, behaves as an adaptive dynamic system that reacts to response differences.

The network is given a set of inputs generating a response that is compared with the response required; the weightings are not changed if the response obtained corresponds with the response required. If the difference exceeds a certain tolerance level, revisions are introduced into the weightings and learning starts again; then a new case is input. The analysis of all the cases supplied constitutes the maximum extension learning cycle. After the interaction of a large number of cycles, the error is reduced to acceptable levels and, once the holdout set accuracy has been exceeded, the learning ends and the weightings are locked. The network has achieved a stable equilibrium configuration that represents 'its capacity to solve a problem'.

The learning mechanism involves a number of problems; however:

- a) The learning stage can be very long (slow learning);
- b) The system might not achieve a stable absolute minimum configuration (optimal error reduction) but might lock on local minimums without being able to move to the optimum;
- c) The system might give rise to oscillating behavior in the learning phase, i.e., when the minimum point is reached and then exceeded. Hence, it then returns to the previous point;
- d) When the actual situation is significantly different, or changes, compared to the situation implicit in the training examples; it is then necessary to repeat the learning phase. The same applies when the set of examples is not representative of the reality of the problem or concept to be learned;
- e) The analysis of the weightings is complex and difficult to interpret. There is, in other words, little network transparency as far as the examination of the system's logic is concerned. This makes it difficult to identify the causes of the errors or defective responses.

The algorithm determining the network's learning is of fundamental importance for the final performance of the network itself.<sup>9</sup>

Neural networks do not require the pre-specification of a functional form, nor the adoption of restrictive assumptions about the characteristics of statistical distributions of the variables and errors of the model. Moreover, by their nature, NN make it possible to work with imprecise variables and with changes of the models over time, thus being able to adapt gradually to the appearance of new cases representing changes in the situation. As noted earlier, the price to be paid using networks of neurons is their lack of transparency in the use of the variables within the network connections. While one is able to identify the explanatory importance of each variable with the usual estimation techniques, signs of influence on the endogenous variables and the degree of their mutual correlation with each neuron remains unclear.

We know many things about how companies can fall into economic distress, about crisis processes and company decline, but we do not have a complete theory. One of the ways of tackling this problem in operative terms is that of using company classification techniques making use of the tools that statistical methodology supplies. Multiple discriminant analysis is one of

<sup>&</sup>lt;sup>9</sup>The method considered here is the well-known Error Back Propagation Algorithm by Rumelhart et al. (1986).



Fig. 4. Discriminant analysis and multilayer neural network (NN).

the tools most often used and was described earlier in our F1 and F2 functions.

Results obtained appear to be very promising. Linear discriminant analysis can be considered equivalent to a network made up of a single neuron that receives signals from the set of indicators and generates an output with a linear transfer function without transformation,  $Y_i = P_i$ . To exploit the advantages offered by the network of neurons, we have used a three layer network based on a combination of simple (two-layer) elementary networks in a 'cascade' fashion. Fig. 4 illustrates the differences between Discriminant Analysis and a multi-layer Neural Network System.

The experimental program is subdivided into four parts:

*Part 1*: Check the capacity of a neural network to reproduce the numeric values of the scores obtained using linear discriminant analysis, receiving, as input, the signals of ratios *different* from those employed in discriminant analysis. Note that in this first experiment, the multi-layer network has been forced to behave linearly, not exploiting its wealth of descriptive potential. Nonetheless, within this constraint, we can verify the network's capacity to approximate the discriminant analysis' linear functions using a different set of ratios.

*Part 2*: Check the capacity of the neural network to separate the samples between bankrupt and healthy companies. The network's output unit is not the value of a score as in the previous section, but simply the binary values 0 (=healthy) and 1 (=unsound). The network's training stage was carried out in period T-3 while the test of its correct recognition was done in either period T-1 of the training sample or on an independent sample.



Fig. 5. Networks with memory of input.

*Part 3*: This section considers the change in company performance over time. One of the problems involved in identifying distressed companies is that of making the classificatory functions sensitive to the passing of time, and the changes of the companies' business patterns. See the work of Theodossiou (1993) for an analysis of the time series properties of distressed prediction. An attempt was made to capture these aspects by constructing complex networks divided into three segments.

The output of the first sub-network summarizes the conclusions about the economic and financial profile observed in period T-3; these are linked to the conclusions relating to period T-2. If, during this period, the profile follows a trend that is consistent (inconsistent) with the trend in the prior period, then the conclusions come out reinforced (weakened). The same applies to the pattern of period T-1. An alternative way of tackling the problem of time pattern analysis lies in using networks with 'memories'; the simplest network of this type is that of including among the input data the change in value of the variables. Fig. 5 illustrates an example of such a network 'with memory'.

From an economic-logic point of view, it is as if there has been an attempt to reproduce the reasoning of the financial analyst when he examines a historical series of business data. The analyst forms an opinion on the state of business by observing how it has evolved over the entire time span available.

Part 4: The aim of this section is to check the capacity of networks to

separate the three categories of company: healthy, vulnerable and unsound. The networks used for this section have an output level comprising two output neurons: the first distinguishes healthy businesses from the unsound ones while the second separates the unsound businesses from the vulnerable and unsound businesses. The two neurons have the same role as the two functions F1 and F2.

Two types of experiment were carried out for this purpose: in the first case, network training was carried out over period T-3 while the check of the capacity for generalization was conducted on period T-1 and on an independent sample. In the second case, networks with memories were used over the whole three-year monitoring period, while the control sample was limited to healthy and vulnerable companies extracted from the continuing company data base.

All the experiments were carried out on the same samples used to fine tune the discriminant functions: an initial sample involved 1212 businesses: 404 each of healthy, unsound and vulnerable firms. A second independent sample of 453 companies was used, 151 of each type, with data limited to the last year prior to bankruptcy. A final sample, independent of the other two, was analyzed comprising 900 healthy and 900 vulnerable companies for three years of historical series. These were taken from the files of 'live' companies.

#### 4. Results

#### 4.1. Healthy vs. unsound firms

The first tests were conducted to estimate the accuracy of the numeric values of the linear discriminant function. We limited the analysis to approximating the function that separates the healthy from unsound companies (F1) for period T-3. If these approximations can be obtained with a smaller set of indicators (input signals) than what was used for the estimation of the discriminant function, it will be a direct check of the neural network's capacity for adaptation and simplification. The experiments were conducted using networks of varying complexity in terms of the number of input indicators, number of layers and the number of connections.

The best results were obtained with a three-layer network: one initial hidden layer of ten neurons, a second hidden layer with four neurons and an output layer consisting of a single neuron. The input comprised ten financial ratios: four relative to the firms' financial structure and indebtedness, two to liquidity, and four representative of company profitability and internal-financing.

The network neurons are totally interconnected. This means that each neuron on a layer is connected to all the others on the next level, including the input signals which are connected to all of the neurons on the first layer. Training was interrupted after 1000 learning cycles; each of which examined

	Score required	Score calculated
High security	15.2	10.2
Security	34.5	37.2
Uncertainty	11.3	12.0
Vulnerability	23.8	25.4
High level of vulnerability	15.2	15.2
Total	100.0%	100.0%

#### Table 3

Distribution of companies by score intervals

808 companies, adjusting the weighting after each cycle. The resulting profile was extremely close to the desired level.

Another measure of the network results is summarized in Table 3. This shows the distribution of the categorization of company credit worthiness by score intervals. The classification differences based on the scores and the actual categories seem small and concentrated mainly towards positive values near 1 (best credits).

Results obtained after 1000 learning cycles are quite encouraging and lead one to believe that if the learning phase lasted longer the error could be reduced still further. It should be noted that the network built to replicate the discriminant function is comprised of completely different indicators from those included in the functions. The latter's selection required a significant number of man-hours. In the case of the neural network, machine-hours were used more, while the selection of indicators, albeit careful and well thought out, required a tiny fraction of the total time. This is a clear indication of the network's capacity for adaptation.

# 4.2. Multi-layer networks

Networks with varying degrees of complexity were trained using ratios from period T-3 followed by testing in period T-1 from the same sample and also an independent sample. The most satisfactory results were obtained with a three-layer network, comprising fifteen neurons in the first hidden layer, six neurons in the second hidden layer and one neuron in the output layer. The fully interconnected network is fed with the numeric values of fifteen business ratios; these are a broader set than the one in the ten-ratio network described in the previous section. Not having observed substantial benefits with using a random selection of the cases, we used a sequential ordering of the observations.

Although the network training used slightly greater than 2000 cycles, the analysis of the error sequence indicated a typical oscillating phenomenon.<sup>10</sup>

<sup>&</sup>lt;sup>10</sup>Training was conducted with a 0.75 learning rate and null momentum: these values were obtained on the basis of the results of experiments with alternative learning and momentum rates.

At the end of the training period, the network was able to recognize correctly 97.7% of healthy and 97.0% of unsound companies. All the other networks which used a lower degree of complexity, even if trained with a higher number of cycles, did not achieve the same recognition capability as obtained using the 15, 4, 1 network.<sup>11</sup> This compares favorably with the recognition rates obtained by the linear discriminant function F1 in period T-3: 90.3% of healthy companies and 86.4% of unsound ones.

The network's identification capability is clearly greater than the discriminant function's although it is obtained with a higher number of indicators: fifteen as opposed to nine. This aspect is important since the network is more complicated and uses a large number of learning cycles. The results of the learning, however, behave erratically. There are great and rapid improvements in the capacity to identify the two groups with the first cycles; nevertheless, as the cycle procedure continues the convergence becomes slower with frequent oscillations and jumps backward and with deterioration in the recognition rates that are sometimes significant. As can be seen, the network had already achieved recognition levels that were not far off the final results, especially in the healthy group, in the earlier cycles. The unsound firm errors were reduced considerably as the number of cycles increased until the last 560 cycles, when the classification accuracy became erratic.

This network, trained in period T-3, showed a lower recognition capacity than the one in the training period using period T-1; period T-1's identification error was 10.6% for healthy businesses and 5.2% for unsound firms. Compare these rates of error with those obtained with the discriminant functions: 7.2% for healthy and 3.5% for unsound companies (Table 1).

This neural network shows a *lower* capacity for generalization than the traditional discriminant function's. This conclusion is reinforced by the results obtained on the independent samples of 302 companies for period T-1: rates of error are 15.9% for the healthy and 9.5% for the unsound companies as opposed to the 9.7% and 4.9%, respectively, obtained with the discriminant functions.<sup>12</sup>

The simpler network's results are more modest than the ones obtained from traditional discriminant analysis, but show a greater capacity for

<sup>&</sup>lt;sup>11</sup>Experiments were also conducted, among others, using Cascade-Correlation but we did not obtain superior results; for the methodological aspects relating to Cascade-Correlation see Fahlman and Lebiere (1992).

<sup>&</sup>lt;sup>12</sup>Increased generalization was achieved with the simpler, 10, 4, 1 type networks, fed with the ten ratios used in the first section of experiments. After two thousand learning cycles, this network showed a recognition capacity of 93.3% for healthy companies and 84.7% for unsound companies in period T-3, far lower than results obtained with the more complex 15, 6, 1 network. Nonetheless, the simpler network was able to limit the errors on the T-1 sample to 8.2% and 3.7%, respectively, and, on the independent sample, to 14.6% and 6.8% for the two samples of firms.

Sample size $= 404$		Neural net	work	Linear discriminat function (F	nt 71)
in each group		Healthy	Unsound	Healthy	Unsound
Estimation period	T-3	89.4%	86.2%	90.3%	86.4%
Control period	T-1	91.8%	95.3%	92.8%	96.5%

# Table 4 Comparison of recognition rates: NN vs. LDA

generalization than the more complex networks. This confirms what others have shown, i.e., the network judged to be most effective at the end of the learning cycle might not be as suitable with other sets of independent cases. The network is the victim of a phenomenon known as 'overfitting'. We encountered a similar phenomenon when we observed the holdout sample accuracy of quadratic discriminant functions vs. the less complex linear function, see Altman et al. (1977).

# 4.3. Multi-layer networks with discriminant function ratios

The results obtained in the previous section use networks fed with ratios different from the ones used in discriminant functions. The reason for this choice is the need to estimate the classification capacity of the networks using a standard information base (ratios) such as are normally available in financial analysis reports published by the CB. In a related test, nine of the 11 F1 discriminant function's indicators are utilized with networks of differing complexity. The intention was to check the networks' capacity to reproduce the 'knowledge' built into the discriminant functions and convert it into knowledge distributed over the neural connections.

The best result was obtained with a 9, 5, 1 network after 4030 learning cycles with a 0.75 learning rate and 0.30 momentum.<sup>13</sup> Table 4 shows the rates of recognition of businesses in the T-3 period (network estimation) and period T-1 (control period). The results are not dissimilar, although slightly lower, from those obtained using the discriminant function. It is not, however, certain that the formalization of the knowledge built into the network is totally equivalent to the knowledge of the linear function since companies that the network recognized incorrectly were, in part, different from those incorrectly recognized by the discriminant function. Moreover, while the discriminant function always behaves in the same way when the values of the exogenous variables vary, with the use of the network we have seen behavior that is not always consistent when the input changes. We will

<sup>&</sup>lt;sup>13</sup>These values were obtained from the results using alternative parameter experiments.

postpone this discussion until the next section where it can be treated more thoroughly.

# 4.4. Simple network connections

Neural networks can have difficulty when tackling particularly complex problems. In our case, the complexity derives from the nature of the problem and from the wide range of observations. While the wealth of data makes it possible to construct a model that is general and robust, it also tends to make the training of the network more difficult. Complex networks with numerous inputs and neurons are perhaps better able to classify a more heterogeneous sample of firms but has the disadvantage of making the time required (and the expense) for training sometimes prohibitive. Moreover, these networks tend to adopt oscillating or non-convergent behavior as well as often being the victim of the overfitting trap.

One methodology for tackling these problems might be breaking down the total network into simpler networks connected to each other. We carried out experiments along these lines starting with the generation of elementary networks which were then connected to each other in a second level network. Fig. 7 shows how the networks used in this section were organized; we illustrate this with three of the eight simple structure networks (with one hidden and one output layer apiece). Every elementary network (e.g., leverage) is fed with a number of ratios that are representative of that characteristic.

The second level network coordinates the results of the eight elementary networks in order to generate the system's final response. That is, it is 'trained' to combine the conclusions reached by the elementary networks. This is equivalent to a multivariate discriminant or logit analysis with the same potential benefits over a univariate structure.

The results obtained are shown in Table 5. As expected, the classification accuracy of the elementary networks differs greatly from network to network and generally it is not very high. Furthermore, it does not always increase when passing from the estimation period (T-3) to the control period (T-1). The second level network, however, which generates the system's overall responses, performs very well, with correct classification of healthy companies of 98.3% and unsound companies of 92.5% in the estimation period (T-3). This result is considerably better than that achieved through discriminant analysis (cf. Table 1). In addition, the system of networks applied to the control period (T-1) gives outstanding results in this case, too, with 92.8% of healthy companies and 94.5% of unsound companies classified correctly.

On the whole, the test of the simple network system's capacity for generalization on the independent sample of 302 companies in period T-1 is also good. In the face of significant reduction in first level elementary rates of

	output
	one
	with
5	networks
Table	neural
	simple
	of
	onnection
	C

Variable types analyzed	Number	% Correct	classification				
	0t Iearnina	Samula of 1	200 companies			Unldant a	
	ical IIII B	o in Aidillac	ooo companico				unpic
	cycles	Period T-3		Period T-1		Of 302 Co	mpanies
		Estimate		Control		Period T-1	•
		Healthy	Unsound	Healthy	Unsound	Healthy	Unsound
Specialized elementary NN							
1) Leverage; asset/liab. struc.	4640	76.98	82.21	73.76	91.04	76.80	90.48
2) Ability to bear fin. debt	3800	74.50	79.20	69.06	92.04	35.20	97.28
3) Liquidity	4950	69.31	84.96	71.29	90.80	36.00	97.96
4) Profitability, internal fin.	4770	87.87	85.71	84.90	92.29	56.80	96.60
5) Profit accumulation	4050	81.44	75.94	88.61	71.14	59.20	85.03
6) Ability to bear cost of debt	5000	75.25	87.97	76.73	95.77	7.20	99.32
7) General efficiency	1260	62.38	60.40	65.10	69.65	63.20	57.82
8) Trade indebtedness	006	59.90	69.92	60.15	79.60	73.60	66.67
Second-level NN	2850	98.27	92.48	92.82	94.53	93.60	89.12
NN trained of T 2 with locaning	* ***** 0 75	and moments	- 0.15 The				

NN trained at T-3 with learning rate = 0.75 and momentum = 0.25. The revision of the weight is done for each company; in each cycle there are 808 revisions.

identification, second level networks generate correct classification rates of 93.6% for the healthy companies and 89.1% for unsound companies. Compared with results obtained with discriminant functions, there is a significant drop in the number of unsound companies identified correctly (-6 percent) but not enough to cancel out the effectiveness of the system.

These results are encouraging for the use of neural networks. Note that the system's classification was obtained with a small effort on the part of the analyst and by using annual report ratios that are not particularly complex.

### 4.5. Analysis of simple network systems – some concerns

As mentioned earlier, a problem that arises in the use of neural networks concerns the low level of intelligibility of the knowledge base spread over the network and built into the weighting of the connections. We carried out an analysis of elementary and second level networks to try to better understand how they work. Distributed knowledge mapping, see Hinton, et al., (1986), implicit in the weighting values of the network, can be studied from various points of view including the identification of the significance assumed by the various neurons and the analysis of the network's behavior when input conditions change.

The significance of the neurons can be identified either by examining the weightings matrix, in the simplest cases, or by studying the role that individual neurons play in determining the output.<sup>14</sup> By modifying the initial weightings and repeating the learning process, the weightings matrix is modified which might also change the role of the various neurons.

Generally speaking, the behavior of a network can be studied on the basis of the derivatives (or elasticities) of the output compared with the individual inputs. For the analysis in the case where more than one input is changed at the same time, it is possible to refer to the total differential of the output compared with the inputs.

From a mathematical point of view, the neural network is a non-linear system. Its input/output derivatives depend on the input value configuration vector. Therefore, the network's capacity to react to input changes is not always the same but strictly depends on the starting position. This factor makes the identification of the individual contributions of the inputs to the formation of the output complex and uncertain.

Even in its simplicity, the second-level network shows peculiar behavior patterns. We examined the values of the partial derivatives between the

<sup>&</sup>lt;sup>14</sup>The activation patterns that the hidden-layer neurons assume in response to different input value configurations can be observed to try to understand how the network has formed responses. An alternative method consists of causing voluntary 'damages' inside the network by deactivating certain of its connections or removing entire groups or by altering the size of its values.

output and the individual inputs in the case of different starting configurations. The calculation of the partial derivatives showed significant dependence on the base conditions, with sudden changes of sign. This feature of the networks is particularly awkward for the financial analyst because the behavior of the network may be unpredictable and contrary to business logic. For example, for a business to be considered unsound by the network, it only needs to have a very modest general efficiency, a relatively high leverage and an uncertain ability to bear financial indebtedness while having outstanding ratings in the other inputs. If the level of liquidity is worsened under this profile, the network shows an improvement in the output and, under some conditions, the company goes from being unsound to healthy, thus altering the initial conclusions! Such behavior does not occur if the profitability is worsened but it reappears in the case of increased commercial indebtedness.

#### 4.6. Interconnection of simple networks with memories

The next experiment made use of the logic of networks with memories on the inputs (see Fig. 5 for a general description of such a structure). The inputs of this type of network include the entire three-year historical series of the indicators used. The network is trained to consider all the data available about the company at the same time. This is like a financial analyst examining the historical time series of financial statements.

The correct recognition rates of healthy and unsound companies is high, even in several elementary networks, rising to over 99% in the second-level network. The overall accuracy of the interconnected system of elementary networks with memories commits errors of 4 healthy companies (out of 404) and 1 unsound company (out of 404).<sup>15</sup>

We did analyze the overall functioning of the system on simple, interconnected networks with memories. We found, in the second level network, the same non-acceptable behavioral problems already identified above with a frequent inversion of the output value when the inputs are uniformly modified either individually or in limited subsets.

#### 4.7. Multi-group analysis

The last set of experiments is aimed at the generation of a two-output network for the simultaneous, not sequential, separation of healthy, vulnerable and unsound companies. This test was very severe because of the

<sup>&</sup>lt;sup>15</sup>The price paid for these performance levels has been the high number of elementary firstlevel network learning cycles. Consider that with 5000 cycles there are over four million changes made to the weightings via the backward propagation algorithm.

difficulty in identifying in a single solution the characteristics separating the three groups of business. The best results were achieved with a network using families of simple NNs with memory consisting of three layers: one hidden layer with fifteen neurons, a second hidden layer with twelve neurons and one output layer with two neurons. This is somewhat analogous to the three group simultaneous distress S & L analysis of Altman (1977). Since these results are still experimental and not germane to our overall conclusions, we refer the reader to a lengthier working paper, Varetto and Marco (1993).

# 5. Conclusion

In the light of the experiments carried out, neural networks are a very interesting tool and have great potential capacities that undoubtedly make them attractive for application to the field of business classification. The networks assessed on our samples have shown significant capacities for recognizing the health of companies, with results that are, in many cases, near or superior to the results obtained through discriminant analysis. The results of the two-output networks trained to simultaneously recognize the three types of company performance: healthy, vulnerable and unsound, also proved to be very interesting. Nonetheless, taking into account the results obtained in the control periods and in the holdout samples, discriminant analysis was deemed to be better, on the whole, than the networks trained in our experiments.

The greatest problem concerns the existence of non-acceptable types of behavior in the network. These are intrinsic to the non-linear nature of the mathematical model underlying the network, combining a large number of variables several times over in a complex fashion. These behavior patterns are characteristic of networks of any complexity that have at least two inputs.

The extent and frequency of illogical types of behavior (in the judgement of the financial analyst) grow with the increase in the complexity of the network architecture. Only extremely simple networks limit the probability of meeting these unacceptable results. The construction of ultra simplified networks cannot be a solution, however, because the problem is only delayed. It does in fact crop again as a result of the need to coordinate simple networks with others of higher level.

The problem of understanding these types of behavior and how to remedy them is not an easy one to solve. As well as using real examples, it would be possible, for example, to train the network with artificial cases constructed to represent other possible combinations. Given the high number of artificial cases required, the network's capacity for analyzing real cases could be totally distorted if errors are committed at this stage.

On the whole, linear discriminant analysis compares rather well when

compared to neural networks. The fine-tuning of the discriminant function does take longer, but the greater estimation speed makes it possible to carry out careful tuning at relatively low cost. Furthermore, the linear form, albeit with the limitations of its ability to perform well, ensures consistent behavior for any type of variable. This makes it possible to interpret the model's operating logic on the basis of the coefficients.

In the neural network, it is not possible to ascertain whether a particular variable comes into the interpretative model with the wrong sign and to eliminate or replace it, as can be accomplished with traditional econometric and statistical tools. With discriminant analysis, it is possible to learn what the most important variables are for explaining the differences between the companies in the sample. In the network fine-tuning process, the length of the training process, the rate of mean error decline, and the results on the recognition rates are the keys for estimating the soundness of the variables, ex-post.

Our conclusions on the use of neural networks are not straightforward and they recognize the undoubted advantages such networks have. A path we intend to adopt in the future is to integrate networks and discriminant functions, applying the former to less clear and more complex problems of classification in which the flexibility of networks and their capacity for structuring into simple, integrated families could prove to be very useful. The key determinant as to whether neural networks, in conjunction or not with traditional classification procedures, will be integrated into practitioner decisions is the accuracy, logic, and understandability of the process and its components. It must be emphasized again, however, that we have found illogical behavior patterns in all of the many NN systems tried in our research. These results have not been emphasized in previous applications of NN systems to business related problems.

Neural networks have shown enough promising features to provide an incentive for more thorough and creative testing. Analysts' fascination with artificial intelligence models will, no doubt, motivate continued firm related investigations.

#### References

- Altman, E., 1968, Financial ratios, discriminant analysis and the prediction of corporate bankruptcy, Journal of Finance (Sept.) 589-609.
- Altman, E., 1993, Corporate financial distress and bankruptcy, 2nd ed. (John Wiley and Sons, New York).
- Altman, E. and T. McGough, 1974, Evaluation of a company as a going concern, The Journal of Accountancy (Dec.) 50–57.
- Altman, E., R. Haldeman and P. Narayanan, 1977, ZETA analysis, a new model to identify bankruptcy risk of corporations, Journal of Banking and Finance 1, 29–54.
- Altman, E., 1977, Predicting performance in the S&L industry, Journal of Monetary Economics, October.

- Altman, E., E. Kahya and P. Theodossiou, 1994, A time series business failure prediction model for U.S. firms, Working Paper (NYU Salomon Center) 1994.
- Bell, T., G. Ribar and J. Verchio, 1990, Neural networks vs. logistic regression in predicting bank failures, in: P. Srivastava, ed., Auditing Symposium X (University of Kansas).
- Cadden, D., 1991, Neural networks and the mathematics of chaos an investigation of the methodologies as accurate predictors of corporate bankruptcy (IEEE).
- Cammarata, S., 1990, Reti neuronali (Etas Kompass).
- Chung, H. and K.Y. Tam, 1993, A comparative analysis of inductive learning algorithms, Intelligence Systems in Accounting, Finance and Management.
- Coakley, J. and C. Brown, 1993, Artificial neural networks applied to ratio analysis in the analytical review process, Intelligent System in Accounts, Finance and Management (Jan.).
- Coats, P. and L. Fant, 1993, Recognizing financial distress patterns using a neural network tool, Financial Management (Nov.), 142–155.
- Coleman, G., T. Graettinger and W. Lawrence, 1992, Neural networks for bankruptcy prediction: the power to solve financial problems, AI Review, July/August 1991, pp. 48–50; Reprinted in Trippi and Turban, 1992.
- Dutta, S. and S. Shekhar, 1992, Generalization with neural networks: an application to the financial domain, Working Paper 92/30 (INSEAD, Fontainebleau, France).
- Fahlman, S. and C. Lebiere, 1992, The cascade-correlation learning architecture technical report: CMU-90-100 (Carnegie Mellon University) February.
- Freeman, J. and D. Skapura, 1991, Neural networks (Addison Wesley).
- Hertz, J., A. Krogh and R. Palmer, 1991, Introduction to the theory of neural computing (Addison Wesley).
- Hinton, G., J. McClelland and D. Rumelhart, 1986, Distributed representations, in: D. Rumelhart and J. McClelland, Parallel distributed processing: exploration in the cognition (MIT Press, Cambridge, MA).
- Karels, G.V. and A. Prakash, 1987, Multivariate normality and forecasting of business bankruptcy, Journal of Business, Finance and Accounting (Winter) 573–593.
- Kryzanowski, L., M. Galler and D. Wright, 1989, Using artificial neural networks to pick stocks, Financial Analysts Journal (July/Aug.) 21-27.
- Kryzanowski, L. and M. Galler, 1994, Analysis of small business financial statements using neural nets, Journal of Accounting, Auditing and Finance, Forthcoming.
- Liang, T., J. Chandler, I. Han and J. Roan, 1992, An empirical investigation of some data effects on the classification accuracy of probit, ID3 and neural networks, in: Cont. Acc. Res., Fall.
- Odom, M. and R. Sharda, 1990. A neural network model for bankruptcy prediction, Proceedings of the IEEE International Conference on Neural Networks (San Diego, CA) 163–168.
- Pau, L. and C. Gianotti, 1990, Economic and financial knowledge-based processing (Springer, Berlin).
- Raghupathi, W., L. Schleade and B. Raju, 1991, A neural network approach to bankruptcy prediction, Proceedings of the IEEE 24th International Conference on System Sciences (Hawaii); Reprinted in Trippi and Turban (1992).
  Rahimian, E., S. Singh, T. Thammachofe and R. Virmani, 1992, Bankruptcy prediction by neural
- Rahimian, E., S. Singh, T. Thammachofe and R. Virmani, 1992, Bankruptcy prediction by neural network, in: R. Trippi and E. Turban, eds. Neural networks in finance and investing (Probus).
- Rumelhart, D. and J. McClelland, 1986, Parallel distributed processing: exploration in the cognition (MIT Press, Cambridge, MA).
- Rumelhart, D., G. Hinton and R. Williams, 1986, Learning internal representations by error propagation, in parallel distributed processing (Cambridge, MA) 318-362.
- Swales, G. and Y. Yoon, 1992, Applying artificial networks to investment analysis, Financial Analysts Journal (Sept./Oct.).
- Theodossiou, P., 1993, Predicting shifts in the mean of a multivariate time series process: an application in predicting business failures, Journal of the American Statistical Association 88(422) 441–449.
- Trippi, R. and D. De Sieno, 1992, Trading equity index futures with a neural network, Journal of Portfolio Management (Fall).
- Trippi, R. and E. Turban, eds., Neural networks in finance and investing (Probus).

- Varetto, F. and G. Marco, 1993, Diagnosi delle insolvenze e reti neurali: esperimenti e confronti con l'Analisi discriminante lineare, W.P. Centrale dei Bilanci, Sept. 1993 and forthcoming, Economia Aziendale.
- Varetto, F., 1990, Il sistema di diagnosi dei rischi di insolvenza della Centrale dei Bilanci, Bancaria ed., (Rome).
- Wong, F., P. Wang, T. Goh and B. Quek, 1992, Fuzzy neural systems for stock selection, Financial Analysts Journal (Jan./Feb.).