the Autonomous Management School of
Ghent University and Katholieke Universiteit Leuven

Vlerick Leuven Gent Working Paper Series 2004/16

ALTERNATIVE METHODOLOGIES IN STUDIES ON BUSINESS FAILURE: DO THEY PRODUCE BETTER RESULTS THAN THE CLASSIC STATISTICAL METHODS?

SOFIE BALCAEN

HUBERT OOGHE

Hubert.Ooghe@vlerick.be

D/2004/6482/17

ALTERNATIVE METHODOLOGIES IN STUDIES ON BUSINESS FAILURE: DO THEY PRODUCE BETTER RESULTS THAN THE CLASSIC STATISTICAL METHODS?

SOFIE BALCAEN[1]

Department of Accountancy and Corporate Finance,

Ghent University

HUBERT OOGHE[2]

Vlerick Leuven Gent Management School

Contact:

Hubert Ooghe

Vlerick Leuven Gent Management School

Tel: +32 09 210 97 86

Fax: +32 09 210 97 00

Email: Hubert.Ooghe@vlerick.be

---

[1] Research Assistant, Department of Accountancy and Corporate Finance, Ghent University, Belgium

ABSTRACT

Over the last 35 years, the topic of company failure prediction has developed to a major research domain in corporate finance. Academic researchers from all over the world have been developing a gigantic number of corporate failure prediction models, based on various types of modelling techniques. Besides the classic cross-sectional statistical methods, which have produced numerous failure prediction models, researchers have also been using several alternative methods for analysing and predicting business failure.

To date, a clear overview and discussion of the application of alternative methods in corporate failure prediction is still lacking. Moreover, frequently, different designations or names are used for one method. Therefore, this study aims to provide a clear overview of the alternative research methods, attributing each of them a fixed designation. More in particular, this paper extensively elaborates on the most popular methods of survival analysis, machine learning decision trees and neural networks. Furthermore, it discusses several other alternative methods, which can be considered to have a certain value added in the empirical literature on business failure: the fuzzy rules-based classification model, the multi-logit model, the CUSUM model, dynamic event history analysis, the catastrophe theory and chaos theory model, multidimensional scaling, linear goal programming, the multi-criteria decision aid approach, rough set analysis, expert systems and self-organizing maps. This paper discusses the main features of these methods and their specific assumptions, advantages and disadvantages and it gives an overview of a number of academically developed corporate failure prediction models. Several issues viewed in isolation by earlier studies are here considered together, which is of major importance for gaining a clear insight into the possible alternative methods of corporate failure modelling and their corresponding features.

A second aim of this paper is to find an answer to the question whether the more sophisticated, alternative modelling methods produce better performing failure prediction models than the rather simple classic statistical methods. The analysis of the conclusions of a large number of empirical studies comparing the classification results and/or the prediction abilities of failure prediction models based on different techniques seems to indicate that we may question the benefits to be gained from using the more sophisticated alternative methods.

---

[2] Professor, Vlerick Leuven Gent Management School, Belgium, Department of Accountancy and Corporate Finance, Ghent University, Belgium; Ernst & Young Chair of Growth Management

1 INTRODUCTION

Because of the large number of parties involved in corporate failure or 'business failure'[3] and the large costs of company failure, the avoidance of failure has always been an important issue in the field of corporate finance. Moreover, as a result of the negative spiral in the general economic environment, the increased availability of data and statistical techniques, the extended academic research on the impact of market imperfections and on information asymmetry and the introduction of the New Basel Capital, the topic of company failure prediction has become a major research domain in corporate finance. Over the last 35 years, academic researchers from all over the world have been dedicated to the search for the best corporate failure prediction model, which classifies companies according to their (financial) health or failure risk (Altman, 1984; Dimitras et al., 1996; Altman & Narayanan, 1997). They have been using numerous types of modelling techniques and estimation procedures, with different underlying assumptions and computational complexities.

The classic cross-sectional statistical methods can be considered the most popular methods for the development of corporate failure prediction models. A gigantic number of 'single-period' classification models or 'static' models have been developed, especially multivariate discriminant analysis or 'MDA' models and logit models (Zavgren, 1983; Van Wymeersch & Wolfs, 1996; Atiya, 2001). Besides these classic statistical methods, academic researchers have also been using several ***alternative methods*** for analysing and predicting business failure. These methods are the result of the strong progress in computational possibilities and in artificial intelligence (AI).

To date, literature does not seem to provide a clear overview and discussion of the application of alternative methods in corporate failure prediction. Moreover, frequently, different designations or names are used for one method. For this reason, this study aims to give a ***clear overview*** of the alternative research methods, attributing each of them a fixed designation. First of all, it will extensively elaborate on the ***most popular alternative methods*** of survival analysis, machine learning decision trees and neural networks. These three methods have produced a considerable number of business failure prediction models. It will explain each of these methods (how corporate failure is modelled), it will dilate upon their specific assumptions, advantages and disadvantages and it will give an overview of a number

[3] In this paper, the terms 'corporate' and 'business' refer to all kinds of non-financial, 'private sector' companies. Public sector firms from industries such as utilities, education and health are not considered, because they mostly have different priorities and goals and are strongly regulated by the Government. In addition, financial companies from industries such as banking, insurance and management activities of holdings are not analysed in this paper.

of academically developed corporate failure prediction models. Furthermore, this paper will discuss several ***other alternative methods***, which we can be considered to have a significant value added in the empirical literature on business failure. The following methods will be discussed: the fuzzy rules-based classification model, the multi-logit model, the CUSUM model, dynamic event history analysis, the catastrophe theory or chaos theory model, multidimensional scaling, linear goal programming, the multi-criteria decision aid approach, rough set analysis, expert systems and self-organizing maps. It will shortly discuss the main features of these methods, highlight their most important drawbacks and problems and state a number of academic studies on corporate failure prediction models. In this paper, several issues viewed in isolation by earlier studies are considered together. Therefore, this paper contributes towards a clear insight into the different popular alternative methods of corporate failure prediction modelling and their corresponding features.

An interesting question to be posed here is whether the alternative modelling methods, which are more complex, produce better performing failure prediction models than the classic cross-sectional statistical methods of MDA, logit, probit and linear probability models. The question of which modelling method produces the best performing failure prediction model(s) has already been raised in several papers. There are lot of empirical studies comparing the classification results and/or the prediction abilities of failure prediction models based on different techniques. Unfortunately, there is no study systematically comparing all possible modelling methods. So as to find an answer on the question being raised, this paper will summarize the conclusions of a large range of comparative studies on different kinds of corporate failure prediction models.

This paper is structured as follows. Section two extensively elaborates on the most popular alternative methods of survival analysis, machine learning decision trees and neural networks. The other popular alternative methods are enlarged upon in section three. Finally, section four tries to find an answer to the question whether the more sophisticated alternative methods are better than the classic cross-sectional statistical methods and gives an overview of the studies comparing different modelling methods.

2 MOST POPULAR ALTERNATIVE METHODS

This section extensively elaborates on the most popular alternative methods of survival analysis, machine learning decision trees and neural networks, which have produced a considerable number of business failure prediction models. It discusses the features of each of these methods (how corporate failure is modelled) and it dilates upon their specific assumptions, advantages and disadvantages. Finally, a summarizing table gives a clear overview of the characteristics of these methods and reports a number of academically developed corporate failure prediction models.

2.1 Survival analysis

Corporate failure can be modelled by statistical techniques drawn from *'survival analysis'*. This results in a *'hazard model'*. Besides Lane et al. (1986) and Luoma & Laitinen (1991), survival analysis has also been applied by, for example, Kauffman & Wang (2001). Survival analysis is based on the stringent assumption that the failing and non-failing firms belong to the same population of firms, with the non-failing firms being 'right-censored' cases for which the event of failure has not yet occurred[4]. In contrast with the classic statistical models, a survival analysis model does not assume a dichotomous dependent variable (Shumway, 1999). The dependent variable in a hazard model is the time that a firm spends in the group of non-failing companies or the 'survival time'. Just like in most other failure prediction studies, researchers start from a wide range of possible independent variables or 'failure predictors' and select the variables to be included in the hazard model by means of backward and forward elimination (Lane et al., 1986).

The basic concept of the survival analysis method is the *hazard rate* of a company. This is the conditional probability of failure in the next period given the survival of the firm to that period (Lane et al., 1986; Crapp & Stevenson, 1987). In a 'continuous time' hazard model, such as the 'Cox proportional hazards model', the hazard rate is the instantaneous failure risk at time t, given survival up to time t-1. In a 'discrete time' hazard model, such as the 'discrete time proportional odds model of Cox', the hazard rate is the failure risk in year t, given survival up to the previous year.

---

[4] 'Censored' means 'no longer observed'. 'Right censored' refers to the fact that these firms are no longer observed in the future and that their outcome with respect to their survival status is not known. The survival time of these non-failing firms is at least as long as the total time period from the start of the research period to the end. For these right censored firms, only the time to censoring or the 'censoring time' is known (Lane et al., 1986; Shumway, 1999).

Survival analysis estimates the *survival function S(t)*, which models the probability that a firm will live longer than t (i.e. the survival time is longer than t). The distribution function F(t), which gives the distribution of the time to failure, can be derived by 1-S(t). The *hazard function h(t)* is the most important function in survival analysis, because it models the hazard rate, which is the basic concept of survival analysis. The hazard function models the probability of failure in the next period given that the firm was alive at time t. It is modelled as follows:

$$h(t) = \lim_{\Delta t \rightarrow 0+} [P (t < T < t+\Delta t \mid T<t) / \Delta t] \qquad (1)$$

A hazard model which is perfectly suited for the analysis of corporate failure is the '*Cox proportional hazards method*' (Luoma & laitinen, 1991; Lane et al., 1986). The 'semi-parametric' Cox model assumes that the hazard function is based on an arbitrary baseline hazard function $h_0(t)$, for which no distributional assumptions are required (Lane et al., 1986; Luoma & Laitinen, 1991). As in practice, distributional assumptions are often violated, this method results in more efficient and robust models. This method is called a 'proportional' hazard method, because the hazard of each company is assumed to be a proportion of the hazard of another firm. In other words, the hazard functions of firms are assumed to be strictly parallel. A specific feature of the Cox proportional hazards method is that it allows to include time-dependent variables, variables showing the same value for each company on a certain moment, such as macro-economical variables, and lagged variables (Shumway, 1999).

It should be emphasized here that the focus of survival analysis is on determining the effect of the factors – the independent variables – upon the hazard rate and not on determining the actual hazard rate (Yang & Temple, 2000). Also, it should be emphasized that the values of the hazard function can not strictly be interpreted as failure probabilities: they may be greater than one (Laitinen & Kankaanpää, 1999).

Hazard models have several ***advantages*** over other statistical methods (Lane et al., 1986; Luoma & Laitinen, 1991; Shumway, 1999). Firstly, they resolve the problems of static models by explicitly accounting for the time dimension of company failure. A hazard model does not assume that failure is a steady-sate. It acknowledges that a firm's failure risk changes trough time and a hazard model gives additional information regarding the 'likely time to failure'. Hazard models also control for each firm's period at risk and they allow to incorporate explanatory variables that change with time (i.e. annual observations for each firm) and macro-economic variables that are the same for all firms at a given point in time. As survival analysis incorporates the time dimension of failure, it is also possible to simultaneously study firms that are in different phases of the failure process. Secondly, hazard models require no specification of the distributions of the data set and therefore overcome many of the problems concerning sample biases. Thirdly, survival analysis models allow to use more data than the classic statistical failure models. For example, they use the data from ‘right-censored’ cases for which the event of failure has not yet occurred. As a result, they provide more efficient out-of-sample failure predictions. In addition, they allow to use incomplete or ‘random censored’ data, in which the time to failure is not known. When a firm leaves the non-failing group for a reason other than failure, the classic statistical failure models wrongly consider this firm as non-failing, while the hazard model considers this firm to be ‘censored’[5]. Furthermore, the results of a hazard model can be interpreted easily. A positive coefficient indicates that an increase in the corresponding indicator leads to a decrease in survival probability or an increase in failure risk. Finally, survival analysis allows to analyse a wide variety of variables. For example, Crapp & Stevenson (1987) included an exhaustive set of variables in their hazard model. They did not only use the usual financial ratios, but also variables related to other areas, such as macro-economic conditions, management, firm growth and asset quality. Similarly, Shumway (1999) used a combination of accounting ratios and market-driven variables in his hazard model[6]. Luoma & Laitinen (1991) summarize the advantages of survival analysis as follows: “survival analysis as a failure prediction model is […] (1) more natural, (2) more exact and flexible, and (3) uses more information leading, therefore, to better predictions provided that it is correctly modelled and that the number of firms is sufficient (p. 678)”

Besides the numerous advantages, hazard models also have some ***disadvantages***. First of all, hazard models are not designed to be used for failure classification or failure prediction. If one aims to use a hazard model for classification, a specific procedure is required. A first example of such a classification procedure is the one applied by Lane et al. (1986).

[5] Survival analysis only allows for ‘randomly censored’ cases when the random censoring is non-informative. This means that the censoring times do not depend on the failure risk. At moment t, randomly censored firms need to be representative for all firms that have survived up to the moment t with similar values for the independent variables. The failure risk of these ‘randomly censored’ firms are required to be similar to the failure risk of the other, surviving companies.

[6] It should be remarked that the hazard model of Shumway (1999) is not really a hazard model in the strict sense. Shumway states that “ a multi-period logit model is equivalent to a discrete time hazard model” and he uses a multi-period logistic model in which each firm-year observation is considered as an independent observation. Each year in which the firm survives is included in the multi-period logit model as a firm that did not fail, while each failing firm contributes for only one observation to the model.

With a view to develop a short term failure classification model, he suggests to calculate the probability that a firm will live longer than t time units (for example, t = 12 to determine the survival probability 12 months) and to classify the firm as failing or non-failing on the basis of a comparison of its survival probability with a certain cut-off value. A second example is the classification procedure used by Luoma & Laitinen (1991). This simple method involves the division of the sample of firms into two groups - a failing and a non-failing group - on the basis of the firms' hazard rates estimated by the model. It is based on the assumption that the two groups of firms consist of an equal number of companies. A second disadvantage of hazard models is that the calculation of survival times is quite arbitrary, because the closing date of the annual account is implicitly considered as the natural starting point of the failure process (Luoma & Laitinen,1991). Thirdly, there is evidence that the number of failing and non-failing companies in the estimation sample may affect the hazard rates and hence may lead to sample-specific results of the survival analysis models (Luoma & Laitinen,1991)[7]. Fourthly, the efficiency of survival analysis in failure prediction is to a great extent determined by the diversity of the failure processes found in the estimation sample. If, for example, the sample contains a lot of 'acute' failing firms and a small number of 'chronic' failing firms, the model will result in a higher hazard rate for many companies with good financial ratios, while other firms with poor financial ratios will show lower hazard rates. It is clear that survival analysis will have the best results when it is used on samples where the time length of the failure process is homogenous (Luoma & Laitinen, 1991). A final disadvantage of survival analysis is that, the Cox proportional hazards model is subject to the problem of multicollinearity. Strong correlations between the independent variables must be avoided (Lane et al., 1986).

2.2 Machine learning decision trees

In the mid 1980s, *'machine learning'* was introduced in failure modelling research as a non-parametric technique for company classification. Machine learning involves pattern recognition and, based on a 'learning' process, it derives a set of rules. There are several approaches to learning a set of rules: the covering approach, the decision tree approach and the genetic algorithm approach. We will focus on the 'decision trees' approach, as it is most frequently used in corporate failure prediction studies. We would like to refer to Back et al. (1997) for a short explanation of these other approaches.

[7] This problem may, however, be avoided by using a sample of matched pairs of failing and non-failing firms.

Decision trees are the result of a process of 'supervised' learning[8], based on a certain decision-tree-building algorithm. It is a collection of branches (paths from the root to the leafs), leafs (classes of objects) and nodes (containing decision rules or 'splitting rules'), which classifies some 'objects' according to their attributes (Quinlan, 1986; Joos et al., 1998b). When applied to the topic of corporate failure classification, the decision tree classifies companies, represented by a range of features or variables. On the basis of the decision-tree-building algorithm, a sample of firms – the 'training set' of objects – is sequentially divided into several subsets of firms. In a learning or 'induction' process, the decision-tree-building algorithm recursively partitions regions of the attribute space into two sub-regions – failing and non-failing – until both sub-regions are sufficiently defined (Joos et al., 1998b). An attribute can be used numerous times in one decision tree. It is obvious that it is possible to find a complex tree that perfectly matches the data (100% correct classifications). However, as this perfectly matching tree is subject to the problem of over-fitting, it needs to be simplified. The decision-tree-building algorithm will offer the suitable approach to reducing the complexity of the tree. It should be emphasized that the method of decision trees does not provide a continuous scoring system, indicating a company's health. It is a discrete scoring system, indicating some kind of region of financial health or categories of failure risk (Frydman et al., 1985; Daubie et al., 2002).

The decision-tree-building algorithm determines the most important aspects of the decision tree: (1) the way to find the attribute that best discriminates between the two classes of firms (failing and non-failing) and (2) the approach to reducing the size of the tree. Consequently, the choice of this algorithm is extremely important. In corporate failure prediction literature, different algorithms have been used. Frydman et al. (1985), for example, used the ***recursive partitioning algorithm*** and called their decision tree method the '*recursive partitioning analysis*' or RPA[9]. The classification rules or (univariate) splitting rules in their binary classification three are derived in such a way that expected misclassification or 'resubstitution' risk of the three is minimized. In a first phase, a preliminary classification tree with a very small resubstitution risk is constructed. As this tree usually is very complex and subject to over-fitting, the final classification tree with the correct complexity is selected by minimizing the cross-validation risk and by a trade-off between complexity and resubstitution

[8] 'Supervised' learning means that the discretization of continuous-valued variables is done according to some concept: the discretization algorithm is concept-sensitive (Joos et al., 1998). In other words, a certain algorithm is used to 'guide' the process of learning.

[9] The RPA model of Frydman et al. (1985) has the characteristics of both the univariate and multivariate methods: it is based on a sequence of 'splitting rules' concerning one single variable (univariate splitting rules). For a comprehensive exposition

risk. In RPA, prior probabilities and misclassification costs[10] are taken into account. Another possible decision-tree-building algorithm is the ***entropy algorithm*** used by Joos et al. (1998b). This algorithm is based on the concept of 'information entropy'. Here, the attributes in the decision tree are selected in a way that information gains are maximized and hence entropy is minimized. Examples of entropy algorithms are CART, AQ, ID3, C4.5 and AC2 (Joos et al., 1997).

The decision trees method has a number of considerable ***advantages***. Firstly, as machine learning is a non-parametric method, there are no strong statistical requirements concerning the data in the training sample. A second advantage is that decision trees can handle incomplete and qualitative data (Joos et al., 1998b). In addition, according to a study of Quinlan (1986), decision trees can deal with noise or non-systematic errors in the values of attributes or class information[11]. Thirdly, it is possible that in some nodes of the decision tree, one can choose between different decision rules as different combinations of attributes, if these combinations result in the same splitting of these nodes (Daubie et al., 2002). Decision trees can also be considered as user-friendly (Joos et al., 1998b). There is no need to determine a cut-off point for the classification of firms and there is a clear 'failure/non-failure' output. Furthermore, the method of decision trees is very appealing because of its simplicity. The procedure is quite simple: looking for the best attribute for each subset of examples and building the tree. Finally, the graphical presentation form of the decision tree is appealing and allows for an easy distinction of the most significant attributes: they are in the root of the tree (Daubie et al., 2002).

On the other hand, the method of machine learning decision trees also has some ***drawbacks*** (Frydman et al., 1985; Dimitras et al., 1996; Joos et al., 1998b). First, just like classic statistical method of MDA, the decision trees require the specification of prior probabilities and misclassification costs. Both of these factors need to be incorporated in the 'induction' or learning process. Moreover, the decision trees method is more sensitive for changes in misclassification costs and prior probabilities than MDA. Secondly, decision trees are based on the assumption that the failing and non-failing groups of firms are discrete, non-overlapping and identifiable. A third disadvantage of the decision tree method is that the

on the RPA method, Frydman et al. (1985) refer to the book of Breiman et al. (1984) "Classification and regression trees", Belmont, CA: Wadsworth.

[10] The 'prior probabilities' are the probabilities of belonging to the failing and the non-failing group in the total population. The 'misclassification costs' are the costs of a type I and a type II error. In practice, defining these misclassification costs seems to be a very subjective decision. The costs of the consequences relied to both types of errors are mainly intangible and immeasurable and depend on the risk behaviour of the decision-maker and his or her attitude towards the proportion of the cost factors.

relative importance of the variables or attributes in the model can not be easily interpreted. The contributions of the various variables are ambiguous. There is no direct link between the variables and the output of the decision tree. Fourthly, as a decision tree is a discrete scoring systems, which classifies firms into risk categories, it can not be used to compare firms classified within the same ‘risk category’. Finally, a decision tree can not be ‘applied’ to new cases so as to assess the failure risk these firms. In each validation study, the decision tree induction process is repeated from the start, resulting in different trees with different variables. There is no way to determine a certain ‘optimal tree’ with ‘optimal variables’.

2.3 Artificial neural networks (NN)

In 1990, the *artificial neural networks (NN)* technique has entered the field of business failure prediction and, ever since, it has grown to a very popular technique (Charitou et al., 2004). Odom & Sharda (1990) were the first to apply NNs for the prediction of company failure. Other studies using the NN technique are: Cadden (1991), Coats & Fant (1991), Coats & Fant (1993), Fletcher & Goss (1993), Udo (1993), Weymaere & Martens (1993), Wilson & Sharda (1994), Altman et al. (1994), Boritz et al. (1995), Back et al. (1996a), Bardos & Zhu (1997), Yang et al. (1999), Atiya (2001) and Charitou et al. (2004). According to a study of Daubie & Meskens (2002), the NN technique dominates the literature on business failure in the second half of the 1990s. Currently, NNs are still frequently applied in corporate failure prediction.

Neural networks are computer systems that copy human learning processes and human intuition (Hawley et al., 1990). The networks consist of a number of highly interconnected processing elements, called ‘neurons’. In NNs, the independent variables offered to the network are called ‘inputs’, the dependent variables are known as ‘training values’ and the estimated values are called ‘output values’ (Shachmurove, 2002). A NN has a certain architecture or structure. An example of a frequently used architecture is the feed-forward[12] layered network, in which the neurons are divided into different subsets, called ‘layers’. This kind of layered network contains (1) an input layer of neurons containing the input information, (2) internal or ‘hidden’ layers of a number of neurons and (3) an output layer of one neuron (Coast & Fant, 1993). As the number of hidden neurons increases, the network becomes more complex.

[11] Quinlan (1986) points out that it is counter-productive to eliminate noise in the training sample, if the same attributes will also show noise when the decision tree will be put into practice.

When applied to the analysis of company failure, a NN allows to assess the failure risk of companies based on a vector of input information and an output vector. A NN does not use any kind of 'pre-programmed knowledge base' (Hawley et al., 1990). The neurons of the network allow to recognize meaningful patterns in the data. They process and transform the input – a vector of variables – by a vector of weights into one single output signal. The output signal of a neuron, in turn, is sent as an input signal to many other neurons and is possibly sent back to itself. As the signals are passed through the network via weighted interconnections between the neurons, the 'network knowledge' is stored (Hawley et al., 1990; Coats & Fant, 1993). The method of neural networks is based on *'supervised' learning*. The network is 'learned' or 'trained' on a 'training sample' of input-output pairs of data (and possibly a 'validation sample') and the appropriate, best possible sets of weights are determined on the basis of a training algorithm. This process of working towards an appropriate mapping is also called 'convergence' (Coast & Fant, 1993). Once a stable equilibrium configuration or mapping with acceptable error levels has been found, the learning phase (i.e. the weight adaptation mechanism) takes an end and the weightings are locked.

With a view to construct a NN for corporate failure prediction, the researcher needs to select a certain training algorithm, which will be used for pattern recognition. Several training algorithms are possible. The most popular algorithm is the back-propagation algorithm. This algorithm is based on the principle of continuous error feedback[13]. In the learning phase of a back-propagation NN, the training sample is used for fitting parameters, while the validation sample is used to check for over-fitting. In a second phase, the 'recall phase', a 'test sample' of new data is used as input and the NN predicts the output in the form of binary group membership. In this way, the network can be tested for its classification abilities or reliability. (Weymaere & Martens, 1993; Altman et al., 1994 ; Zain, 1994; Back et al., 1996a; Joos et al., 1998a; Yang et al., 1999; Cybinski, 2000). Despite its popularity, the back-propagation algorithm has some clear disadvantages. Yang et al. (1990) point out that (1) it has a high computational intensity, (2) it is unable to explain conclusions, (3) it lacks a formal theory which imposes a need for expertise on the user and (4) it requires a validation dataset[14]. An

[12] 'Feed-forward' means that the information between the layers is flowing only in one direction (Laitinen & Kankaanpää, 1999).

[13] The back-propagation algorithm works as follows. The values resulting from each neuron process through the network towards a final output layer and, once the values reach this output layer, the output of the network is compared to the desired output. Any error is then used to adjust the connection weights, working backwards through the network. This process is repeated until the network has learned the relationship between the inputs and the output. Details on this back-propagation algorithm can be found in Hertz et al. (1991) and in (Tucker, 1996).

[14] For these reasons, Yang et al. (1990) suggest to use a 'Probabilistic NN', which exhibits feed-forward and fast learning.

other popular training algorithm is cascade-correlation (Joos et al., 1997). The cascade-correlation learning algorithm has been used by, for example, Coats & Fant (1993). They suggest that the cascade-correlation NN has some significant advantages over back-propagation NN: "(1) It learns very quickly, (2) it determines its own size and design, (3) it retains the structure it has built even if the training sample changes and (4) it requires no […] error signals though the connections of the network (p. 144)".

It should be stressed that it is of vital importance that the training sample contains information about some carefully selected and appropriate independent variables. Several procedures can be used for 'pre-processing' the data. For example, a principal component analysis (Weymaere & Martens, 1993) or a procedure based on individual prediction accuracy of the variables, correlation analysis and cross-validation (Atiya, 2001) may be used to find an appropriate set of input variables. A few years ago, the genetic algorithm has been introduced as a very promising method for pre-processing the data and finding the best set of indicators for neural networks (Back et al., 1995; Back et al., 1996b; Back et al., 1997; Hekanaho et al., 1998).

In comparison with other methods, the NN technique has several advantages. First of all, NNs are able to analyse complex patterns quickly and with a high accuracy level (Shachmurove, 2002) and they are able to learn from examples, without any pre-programmed knowledge (Back et al., 1996b). Secondly, they are not subject to the restrictive statistical assumptions of MDA. More in particular, no distributional assumptions are imposed and the input data do not need to conform to linearity (Coats & Fant, 1993; Zain, 1994; Tucker, 1996; Cybinski, 2000; Shachmurove, 2002). As it can be argued that the relationship between failure risk and financial ratios shows saturation effects and that the effects of financial ratios on failure risk are multiplicative, the non-linear approach is an important advantage (Atiya, 2001). Thirdly, non-numeric data can easily be included in a NN, because of the absence of the linearity constraint (Coats & Fant, 1993). A fourth advantage is that a NN is perfectly suited for pattern recognition and classification in unstructured environments with 'noisy data', which are incomplete or inconsistent (Hawley et al., 1990; Tucker, 1996; Shachmurove, 2002). The network tolerates data errors and missing values by making use of the context and 'filling in the gaps'. Consequently, a NN is able to work with annual account data, which are often inconsistent and incomplete. In addition, a NN can overcome the problem of autocorrelation, which frequently arises in time series data (Hawley et al., 1990; Cybinski, 2000, 2001). Fifthly, the NN technique can be considered as user-friendly as it offers a clear 'failure/non-failure' output. Finally, when predicting company failure, neural networks generally seem to be more robust – especially when sample sizes are small – and more flexible than other methods (Cybinski, 2000).

Although the technique of NNs (based on back-propagation) mostly seems to deliver excellent performance results in corporate failure prediction studies and appears to have clear advantages over other methods, it also shows some serious drawbacks (Hawley et al., 1990; Coats & Fant, 1993; Fletcher & Goss, 1993; Zain, 1994; Altman et al., 1994; Tucker, 1996; Trigueiros & Taffler, 1996; Joos et al., 1998a; Tan & Dihardjo, 1999; Charitou et al. 2004; Cybinski, 2000, 2001; Shachmurove, 2002; Baesens et al., 2003). The most important problem related to the use of NNs is the 'black box' problem: a NN does not reveal the significance of each of the variables in the final classification and the derived weights can not be interpreted. We have completely no understanding or knowledge concerning how the network classifies companies into the failing and non-failing group. It is impossible to understand how the relations in the layer-structure are estimated and hence, the network can not be 'applied' as such in order to classify new cases. In other words, the black-box problem makes it impossible to use the NN in practice, in a decision context. A second major drawback of NNs is that they are very sensitive to the 'garbage in – garbage out' problem. Consequently, one has to carefully select the variables that are included in the training samples and assure the quality of the data. The appropriate selection of economically significant variables from an extensive set of available variables can be very time consuming. Thirdly, as a NN can be made to fit the data 'like a glove', it runs the risk of over-parametrization or over-fitting. This results in a sample-specific model with a low generalizing ability. There are some possible solutions to this over-fitting problem, but there is no uniform guideline. Trigueiros & Taffler (1996), for example, suggest to limit the number of estimated coefficients to maximum 10 percent of the number of cases in the sample so as to reduce the over-fitting problem. A fourth problem is that the physical architecture of the NN has to be defined by the researcher by means of trial and error. For example, in a network with multiple layers, the number of layers and the number of nodes in each layer has to be decided upon (arbitrarily) by the researcher. When defining the architecture of the network, one has to take account of the fact that a higher number of layers leads to a more complex NN with a higher internal validity, but also causes a higher degree of over-fitting and a lower external validity. Fifthly, a NN requires a long processing time before the training phase is finished[15] and runs the risk of not finding a stable, optimal configuration. In addition, the training phase may result in illogical weightings of the variables (i.e. illogical network behaviour) due to different variations of the input values. A final drawback of the NN technique is that a NN requires a large training sample of input-output values in order to sufficiently train the network. A small training sample seems to lead to an over-fitted network.

2.4 Overview

Table 1 gives an overview of the main advantages and drawbacks of the most popular alternative methods and reports a number of academically developed corporate failure prediction models. These alternative methods are clearly more sophisticated and computationally more complex than the classic statistical methods, such as MDA and logit analysis.

Insert Table 1 About Here

---

3. OTHER ALTERNATIVE METHODS

Besides the popular alternative methods mentioned in the previous section, academic researchers have also been using a number of other alternative methods, which can be considered to have a significant value added in the empirical literature on corporate failure. This section enlarges uopn the fuzzy rules-based classification model, the multi-logit model, the CUSUM model, dynamic event history analysis, the catastrophe theory and chaos theory model, multidimensional scaling, linear goal programming, the multicriteria decision aid approach, rough set analysis, expert systems and self-organizing maps. It shortly discusses the main features of these methods and their most important drawbacks and problems. Finally, it also gives an overview of a number of academic studies on corporate failure prediction models applying these methods.

It should be emphasized that this study does not aim to present an exhaustive overview of all possible alternative methods and models. It focuses on those methods and models that are frequently cited in literature and those we consider to have a significant added value in the

[15] In order to speed up the training phase, one could transform the data presented to the network. For example, Weymaere & Martens (1993) and Zain (1994) propose to normalize the ratios to variables with zero means and a variance of one before presenting them to the network.

empirical literature on corporate failure. For example, the methods of case based forecasting (Jo & Han, 1996), multi-factor analysis and multiplicative hazards models (Daubie & Meskens, 2002) fall outside this study.

For a fuller discussion or for a more extensive overview of all possible failure prediction methods and models, we would like to refer to Jones (1987), Keasey & Watson (1991), Ooghe et al. (1995), Dimitras et al. (1996), Altman & Narayanan (1997) and Altman & Saunders (1998). Jones (1987) and Keasey & Watson (1991) offer a comprehensive literature review. They focus on, respectively, the techniques used for failure prediction and the limitations and usefulness of several methods. Dimitras et al. (1996) is another important review study on failure prediction methods and models. Altman & Narayanan (1997) survey the studies on business failure classification models in 21 different countries, while Altman & Saunders (1998) elaborate on the development of credit risk models of all types, including credit scoring models, over the last 20 years, especially in the USA. Ooghe et al. (1995) give a detailed overview of the literature on failure models in Belgium.

3.1 The fuzzy rules-based classification model

A very special method to measure failure risk is the '*fuzzy knowledge' based decision aiding method*, used by Spanos et al. (1999). This method starts from a number of if-then rules, which are based on already existing, qualitative knowledge on corporate failure and are determined by the decision maker. These if-then rules make a link between a number of conditions concerning pre-defined variables and the failure status. Next, the relevance of each if-then rule is tested on an estimation data set. Each rule is attributed a rating index between zero and one, which denotes the probability of correctness of a rule. In this rating index, a higher rating corresponds to a better rule. Finally, according to the decision maker's preferences, a certain set of fuzzy rules – the 'fuzzy rules set' – is exported to a *fuzzy rules-based classification model* and, according to this model, firms are classified as failing or non-failing.

The most important ***advantage*** of the fuzzy rule model is its intuitive basis. A ***negative feature*** of the fuzzy rule model is that it strongly depends on the arbitrarily determined if-then rules, based on the knowledge of the decision maker. e

3.2 The multi-logit model

With a view to offer a solution to the fact that classic statistical models of failure prediction only consider data from one specific year, Peel & Peel (1988) introduced the *'multi-logit' methodology*. The multi-logit model allows to simultaneously use data from several years before failure and to simultaneously discriminate between failing and non-failing firms for several reporting periods prior to failure. The multi-logit method is based on the stringent assumption of 'signal consistency', which implies that for each firm, the data for the consecutive years prior to failure give consistent signals about the status of the firm.

The most important ***advantage*** of the multi-logit model is that, in contrast to the classic statistical failure models, it predicts corporate failure considering information from several consecutive years. On the other hand, a ***disadvantage*** of the multi-logit method is that it entirely depends on the assumption of 'signal consistency'. As this assumption is very likely to be violated in practice, the practical usefulness of the multi-logit method may be rather limited.

3.3. The CUSUM model

Theodossiou (1993) and Kahya & Theodossiou (1996) used the *'CUSUM model'*[16] in order to predict, respectively, business failure and corporate financial distress on the basis of financial variables. The CUSUM model is a dynamic extension of MDA. It analyses time-series behaviour of financial variable vectors. This method has the ability to distinguish between transitory changes in the financial variables that result from a serial correlation[17] and non-transitory changes that result from permanent shifts in the mean structure due to financial problems. CUSUM is a sequential procedure, that allows to detect the starting point at which a firm's financial variables shift from a 'good performance' joint distribution[18] to a 'bad performance' joint distribution. A shift in the joint distribution of a firm's financial variables is considered as a signal that the firm tends towards failure. In this way, the CUSUM models predicts a firm's tendency towards failure. The sequential procedure is based on the sequential probability ratio test and the theory of 'optimal stopping rules' (Kahya & Theodossiou, 1996). The CUSUM model involves solving an optimization problem concerning the CUSUM

[16] CUSUM stands for 'cumulative sum'.
[17] Serial correlation is a situation in which positive deviations from the mean structure of the variables are followed by positive deviations in subsequent periods, while negative deviations follow negative deviations.
[18] For a healthy firm, the sequence of attribute vectors exhibit a 'good performance joint distribution'. Evidence shows that for non-failing firms the mean of the vector of attributes is stationary over time. In other words, the means of all variables are stable over time (Theodossiou, 1993).

parameters, which determine the 'sensitivity' of the model to distributional changes (i.e. the time between the occurrence of change and its detection in the distribution of a firm's financial variables). In the optimization problem, the expected error costs of type I and type II errors are minimized (Theodossiou, 1993).

The CUSUM method has several ***advantages*** (Theodossiou, 1993). First, it analyses a firm's financial health, based on information about present and past performances of the firm. A second positive feature of the CUSUM model is that it has a very short memory with respect to a firm's good performances over the years, while it has a very long memory regarding bad performances.

3.4 Dynamic event history analysis (DEHA)

The *'dynamic event history analysis' (DEHA)* method was applied by Hill et al. (1996) in order to distinguish between financially distressed firms that survive and those that eventually become bankrupt. This DEHA method sees company failure as a process. It looks at the transitions to and from stable and financially distressed states and from the latter to the bankrupt state, using longitudinal data. A transition or a change in a firm's financial status (for example, stable, financially distressed or bankrupt) is measured by means of a 'transition rate' or a 'conditional probability'. Event history analysis considers the independent variables over time and their impact on the dependent variable of interest. '*Conditional probability*' is the key feature of DEHA: the likelihood that a firm will become bankrupt or financial distressed in the future (i.e. outside the observation period) is considered to be conditional upon whether a firm is financially distressed or bankrupt at a particular point in time.

DEHA has several ***advantages***. A first advantage of this method is that it moves away from the snapshot focus of the classic statistical models. It recognizes that corporate failure is a dynamic process that starts with some initial conditions and involves changes in these conditions over time. Secondly, it allows for time-varying independent variables (which may vary over the observation period) and for censored cases. Furthermore, the 'conditional probability' feature of the DEHA method is very appealing, because it closely corresponds to reality: the failure probability of a firm in the future strongly depends upon the current financial status of the firm.

3.5 The ‘catastrophe theory’ or ‘chaos theory’ model

Scapens et al. (1981) were the first researchers who considered company failure as a catastrophic event[19] and who used *‘catastrophe theory’* to explain corporate failure. Similarly, Lindsay & Campbell (1996) used *‘chaos theory’* in order to develop a corporate failure prediction model. A chaos theory model regards companies as chaotic systems which show chaotic behaviour. Consequently, it implicitly assumes that firms are deterministic and predictable, but only over short periods of time, due to extreme sensitivity to the initial conditions. A second assumption of the chaos theory model is that healthy or non-failing firms show more chaos than unhealthy or failing firms. This assumption stems from the application of the chaos theory statement that “healthy systems exhibit more chaos than unhealthy systems”. In other words, the returns of a firm approaching failure are assumed to be less chaotic than the returns of the same firm in an earlier time period. Lindsay & Campbell (1996) measured the amount of chaos of each firm for different time periods and then classified the firms as failing or non-failing on the basis of a (univariate or multivariate) decision rule, which includes information on the change in the amount of chaos.

It is clear that a chaos theory model requires a suitable measure of chaos. Lindsay & Campbell (1996) measured the amount of chaos of a firm by means of the ‘Lyapunov exponent’: the larger this exponent, the sooner the company becomes unpredictable.

An important ***advantage*** of the chaos theory method over other methods is that it involves a dynamic analysis of a firm’s financial health. It considers the amount of chaos in different time periods. On the other hand, a clear ***drawback*** of the chaos theory or catastrophe theory model is that its validity depends on the strong assumption that healthy firms exhibit more chaos than failing firms. In practice, this assumption may be violated and, hence, the chaos model may be applied inappropriately and have no validity.

3.6 Multidimensional scaling (MDS)

Mar-Molinero & Ezzamel (1991) introduced the technique of *‘Multidimensional Scaling’ (MDS)* into the domain of corporate failure. They explored the relationship between a sample of financial ratios that can be used to describe the financial health of a firm. Neophytou & Mar-Molinero (2001) applied MDS to the prediction of corporate failure and the examination of which attributes explain failure. MDS offers a totally different way of looking at the problem of company failure prediction. In contrast with other methods of modelling corporate failure, it is a kind of graphical ‘clustering’ method.

MDS starts from a dataset of companies and attributes (e.g. financial ratios) and takes the companies as variables and the attributes as cases. Based on a table of distances – the ‘distance matrix’ or the ‘dissimilarity matrix’[20] – it produces a graphical representation of the structure of the dataset in the form of a map (Mar-Molinero & Serrano-Cinca, 2001; Neophytou & Mar-Molinero 2001). First, each company is represented as a point in an X-dimensional space, of which the position is determined by a set of coordinates. The number of dimensions (X) in which the map is drawn, is decided by the researcher[21]. Next, MDS determines which dimensions provide an adequate representation of the most important features in the data. In view of determining these dimensions (Y), one may conduct a logit analysis using the firm’s coordinates in the X-dimensional space as explanatory variables and a dichotomous dependent variable expressing the failure status. The result is an *MDS map* with Y dimensions in which each failing or non-failing company is represented as a point.

Finally, the data are interpreted by means of a ‘ProFit analysis’[22]. This is a regression-based method that attempts to explain how each of the particular attributes (e.g. financial ratios) is associated with the position of a firm in the Y-dimensional space. It consists of a set of regressions – one regression for each of the attributes – in which the value of the attribute is used as the dependent variable and the firm’s coordinates in Y-dimensional space as the independent variables. If the R^2 of a regression falls below a certain cut-off value, the corresponding attribute is considered to be not relevant to the failure classification problem. The presentation of the result of the ProFit analysis in the Y-dimensional map clearly indicates how the various attributes relate to the different dimensions (Mar-Molinero & Serrano-Cinca, 2001; Neophytou & Mar-Molinero 2001).

It should be emphasized that, in contrast with the classic statistical modelling techniques, MDS has no particular demands for the data. For example, the information about similarity does not have to be limited to Euclidian distance. The only requirements are (1) that there is ‘a message’ in the data and (2) that the ratios are standardized, when they are initially

[19] In this respect, Cadden (1991) points out that “One can identify a series of mistakes on the part of management rather than a catastrophic error. This means that financial decline is generally gradual, although the actual event of bankruptcy is relatively sudden and abrupt.”
[20] This matrix is a square matrix that contains information about the proximity or similarity between firms. The number of rows and columns in this matrix is equal to the number of companies in the sample.
[21] This is an important decision, as the number of dimensions influences the quality of the map, its interpretability, its ease of use and its stability.
[22] ProFit stands for ‘property fitting’ (Neophytou & Mar-Molinero, 2001).

measured in different units (Mar-Molinero & Serrano-Cinca, 2001; Neophytou & Mar-Molinero 2001).

The MDS method has several advantages (Mar-Molinero & Serrano-Cinca, 2001; Neophytou & Mar-Molinero 2001). A first advantage is that the end result of MDS is a statistical map, which has a very intuitive interpretation. In contrast with many other statistical methods, the results can be interpreted without a deep understanding of the underlying statistical principles of the method. Consequently, MDS is a flexible and powerful tool. A second advantage of MDS is the robustness in the presence of outliers (discordant observations)23 and the ability to cope with highly correlated data. Thirdly, as already mentioned, MDS does not make any assumptions about the distribution of the data. Fourthly, MDS can cope with redundant information: there is no need to conduct an initial analysis in data reduction. All possible variables or firm attributes can be included into the analysis. Finally, an interesting feature of MDS is that it is able to explain the possible 'causes' of failure.

Nevertheless, MDS also has some ***drawbacks***. First, if a MDS failure model is based on annual account information, such as financial ratios, it is limited to the use of only one annual account (e.g. the most recent annual account). However, as mentioned in a paper of Balcaen & Ooghe (2004), several problems and critics arise when company failure is predicted on the basis of only one single annual account. Secondly, as the MDS model is not explicitly meant to be used in a predictive context – for failure prediction of new cases – specific procedures need to be applied (Mar-Molinero & Serrano-Cinca, 2001; Neophytou & Mar-Molinero, 2001). A first option is to repeat the study with the new firm added to the dataset, but this requires technical skills to conduct the study and is subject to the limitations of the computer packages. A second option is to conduct a 'reverse' ProFit analysis: one can derive auxiliary scales from the ProFit analysis in order to approximately locate the new firm in the space.

3.7 Linear goal programming (LGP)

Gupta et al. (1990), for example, applied 'linear goal programming' (LGP) to the bankruptcy classification problem. LGP is one of the various techniques derived from mathematical programming. A LGP model formulates intra-group and inter-group differences

[23] In MDS, outliers are considered as possibly influencing the results. Therefore, they are tracked and analysed. Nevertheless, the maps are visually more attractive if extreme observations (outliers) are excluded from the analysis (Neophytou & Mar-Molinero, 2001).

between failing and non-failing firms and, on the basis of these differences, it calculates a score for each firm and a boundary (cut-off point) for group discrimination. It generates a hyper-plane, which is used to distinguish between the failing and the non-failing group of firm observations. This hyper-plane is constructed in a way that the observations which lie within the boundary are as far within the boundary as possible, leading to sharpening the differentiation between the two groups. The cut-off point or boundary is determined by (1) maximizing the weighted sum of distances between the observations and the adjusted boundary and (2) minimizing the weighted sum of boundary violations.

The LGP method has several ***advantages***. First, it does not require the restrictive statistical assumptions of MDA. Furthermore, it is a flexible tool, which is easy to understand.

3.8 The multi-criteria decision aid approach (MCDA)

Zopoudinis (1987), Zopoudinis & Dimitras (1998) and Doumpos & Zopoudinis (1999) all applied ‘multi-criteria decision aid’ (MCDA) methods to the prediction of corporate failure. MCDA allows an assessment of the level of firm risk, based on both financial ratios and qualitative information concerning the firm. Zopoudinis (1987), for example, analysed several strategic criteria, such as the quality of management, the research and development level, the market trend and the market/niche position and ranked firms according to their ‘risk level’.

The ‘multi-group hierarchical discrimination’ (MGHDIS) method is a further extension of MCDA (Doumpos & Zopoudinis, 1999), which can be used as a corporate failure classification model. This method derives a set of two additive ‘utility functions’, one for the failing and one for the healthy firms and these functions are used to classify firms as failing or non-failing. In the classification procedure, it is assumed that the decision maker’s preferences concerning the classification of a firm into the failing or the non-failing class are monotonic functions of the attributes. Two different classification procedures are possible. According to the first procedure, a firm is classified as a failing one, if the utility of the firm corresponding with the utility function for the failing group is higher than the utility corresponding non-failing utility function. A second classification procedure involves determining an optimal classification rule. This is the optimal cut-off value for the difference between the utility corresponding with the non-failing utility function and the utility corresponding with the failing utility function (i.e. the ‘utility difference’), leading to the best classification results for the MGHDIS model. According to the optimal classification rule, a firm is classified as failing if its utility difference is lower than the optimal cut-off value.

The most important ***advantage*** of the MCDA and MGHDIS methods is that it allows to assess corporate failure risk based on quantitative (financial and other) and qualitative information.

3.9 Rough set analysis

Slowinski & Zopoudinis (1995) were the first to apply the methods of 'rough set' analysis in the evaluation of corporate failure risk. This study was followed by other – mainly Greek – studies. The rough set method considers the evaluation of failure risk as a multi-attribute sorting problem[24]. It attempts to describe a set of firms by a set of multi-valued attributes, being financial ratios and other variables. The rough set method accepts both quantitative and qualitative variables. From a set of examples – called the 'training sample' – which represents 'knowledge' and comprises a set of firms identified by a set of characteristics or attributes, it derives a number of decision rules (deterministic and non-deterministic sorting rules) or 'if… then rules'. First, a range of 'minimal subsets' of independent attributes is constructed. A subset of attributes is called a minimal subset, if this subset has the same sorting quality as the whole set of attributes. Then, the 'core of attributes' is defined as the intersection of all minimal subsets. Next, a reduced decision table is constructed, in which the redundant attributes are eliminated. Finally, on the basis of this decision table, the set of sorting rules – the 'sorting algorithm' – is derived and firms are classified by matching their description to the set of sorting rules. (Slowinski & Zopoudinis, 1995 ; Daubie et al., 2002).

The rough set approach has many ***advantages***. Firstly, a rough set analysis may include qualitative variables and hence company failure can be evaluated in a qualitative way (Slowinski & Zopoudinis, 1995). Secondly, the concept of rough sets is easy and, as it clearly explains the decision process of classification of a firm into the failing or the non-failing group, it can easily be applied to new firms by matching these firms to the sorting rules. Finally, rough set analysis is flexible in the application of the decision rules to new firms, in a sense that it can easily be adapted if a particular attribute seems to be difficult to obtain. As the rough set method proposes a large number of minimal subsets, the minimal subsets

[24] A comprehensive explanation of this method can be found in Pawlak (1982).

including this particular attribute may simply be excluded from the analysis, without reducing the quality of sorting (Daubie et al., 2002).

A disadvantage of the rough set analysis is that quantitative variables first need to be recoded into qualitative terms (discretized) before they can be used (Slowinski & Zopoudinis, 1995).

3.10 Expert systems

Expert systems with inductive learning algorithms are the result of the strong evolution of artificial intelligence. Messier & Hansen (1988), for example, applied the method of expert systems to the prediction of bankruptcy. An expert system depends on the representation of expert knowledge on company failure as a series of if-then rules - the *'expert base'* – which first need to be programmed into the expert system. In other words, it is based on a 'predefined knowledge base' (Messier & Hansen, 1988; Hawley et al., 1990). Also a search heuristic needs to be specified. When the expert base and the heuristics are specified, the if-then rules are tested on a number of representative examples of failing and non-failing firms (Messier & Hansen, 1988), which may be called the 'training sample'. Finally, the expert system provides a number of if-then rules and firms can be classified as failing or non-failing.

A first ***advantage*** of using an expert system for the analysis of corporate failure is that it allows for different kinds of variables (including qualitative variables). Secondly, expert systems have no specific demands concerning the statistical distribution of the data. Finally, as expert systems result in a number of if-then rules, they can easily be applied in practice to new cases by using the if-then rules in a classification decision. In this way, expert systems are regarded as user-friendly.

There are also a number of ***disadvantages*** related to the use of expert systems in corporate failure prediction (Messier & Hansen, 1988; Hawley et al., 1990). First, it is often difficult to program the intuition or 'knowledge base' of the expert and to determine which heuristics must be used. Second, the process of extracting the knowledge into rules is very time consuming and expensive and, consequently, expert systems are not beneficial for on-time decision problems. A third drawback is that expert systems are not considered to be flexible, because they are unable to use inductive learning to adapt the if-then rules to changing situations (i.e. changes in the knowledge base). Finally, expert systems are not capable to work with incomplete, noisy data or input information with errors.

3.11 Self organizing maps (SOM)

Besides using neural network technique in a supervised way – with a supervised learning algorithm, an input vector and a desired output or target vector – the neural network method can also be used in an 'unsupervised' way. In this case, NNs are used as a clustering method, which involves an unsupervised learning phase without any target vector concerning the failure status. Here, only an input vector concerning the independent variables is offered to the network. In the unsupervised learning phase, the network self-organises the input information, discovers the basic features of the data and associates these basic features with a certain outcome. In other words, the data are first mapped and then, they are analysed by using a type of visualization method. Because of its self-organizing capabilities, this method is called the 'self organizing map' (SOM). It was first used in 1993 by Martin-del-Brio and Serrano-Cinca (Daubie & Meskens, 2002) in order to visualize financial annual account data from one single year or two consecutive years. However, when applying the SOM to the topic of company failure prediction, there is a desired output or a target vector concerning the failure status. Consequently, when classifying companies as failing or non-failing, the SOM technique has to be used in a 'semi-supervised' way.

Kiviluoto & Bergius (1998) applied the SOM technique to bankruptcy prediction in a dynamic way. Their 'two-level' SOM allowed to analyse financial data from several consecutive years and resulted in three types of corporate behaviour, which is typical for failing companies. The two-level SOM proceeds in two phases. In a first phase, the single-year state of a firm is visualized in a first-level SOM: the firm is presented as a point on a map and hence is identified by a position vector. In a second phase, the change in the state of the firm is analysed by a second-level SOM. This second-level SOM is trained with vectors obtained by concatenating the position vectors (on the first SOM) from several consecutive years. So, each point on the second-level SOM corresponds to a trajectory on the first-level SOM, which allows rather simple interpretations.

The technique of self-organizing maps has some considerable ***advantages*** over other methods. First, it can be used to detect regions of increased failure risk or to view the evolution of the condition of a company over time. Moreover, the two-level SOM offers some possibilities to explore typical 'failure paths' (Kiviluoto & Bergius, 1998).

On the other hand, a clear ***drawback*** of the SOM method is that the first-level SOM requires the user to pre-select a small set of independent variables.

3.12 Overview

Table 2 gives an overview of the main advantages and drawbacks of the alternative methods discussed in this section, and it reports a number of academically developed corporate failure prediction models.

Insert Table 2 About Here

4 ARE THE ALTERNATIVE METHODS BETTER?

A closer look at the features of the alternative modelling methods, reveals that the alternative methods are computationally much more complex and advanced that the rather simple classic cross-sectional statistical methods of MDA, logit, probit and linear probability models. An interesting question to be posed here is whether the alternative methods produce better performing failure prediction models than the classic cross-sectional statistical methods.

The question of "which modelling method produces the best performing failure prediction model(s)?" has already been raised in several papers. There are lot of empirical studies comparing the classification results (ex post) and/or the prediction abilities (ex ante)[25] of failure prediction models based on different techniques[26]. Unfortunately, there is no study systematically comparing the performances of all these different methods. So as to find an answer on the question being raised, we discuss the conclusions of a large range of comparative studies on different corporate failure prediction modelling methods. Table 3 gives a summary of these conclusions. For each study, the methods that are compared, are indicated with an X. This symbol X remains, if the corresponding method shows no superior results, when compared to the other methods. An '$\mathbf{X}^{+}$' indicates that the corresponding method appears to be slightly better than the other methods, while an '$\mathbf{X}^{+\circ}$' means that, only in some situations, the method is slightly better than the other methods. A method that is generally or clearly the best is marked with an '**X***', while a method that is clearly the best method in certain situations, is indicated with an '**X*°**'.

[25] Most comparative studies limit their analysis to the comparison of classification results.

Insert Table 3 About Here

Some studies conclude that the different modelling techniques lead to *similar results*. In this context, Platt & Platt (1990) point out that the within-sample classification results of corporate failure models one year prior to failure seem to be fairly invariant with respect to methodology, which can be partly explained by the close mathematical relationships between the various methods (Neophytou & Mar-Molinero, 2001). Laitinen & Kankaanpaa (1999) conduct a very comprehensive comparative study, involving MDA, logit analysis, RPA, survival analysis and NNs and conclude that, generally, there is no superior method of classification. Dependent on whether the ex post classification results or the ex ante prediction results are used, the ranking of the methods differs. Moreover, the differences in prediction accuracy that appear at first sight seem to be not statistically significant. The only difference in predictive performances that is found to be significant, is the difference between the logit model and survival analysis, one year prior to failure. Here, the logit method seems to be better than the survival analysis model. Similarly, Charitou et al. (2004) compare the performances of NNs and logit models and find that NNs and logit models are both reliable alternatives for company failure prediction.

On the other hand, the large majority of comparative studies show varying conclusions, which point in different directions. Doumpos & Zopoudinis (1999) conduct a comparative study on the techniques of linear MDA, logit analysis and multi-group hierarchical discrimination (MGHDIS), using the same variables for the three methods (i.e. an equal comparative basis). They find (1) an inferior performance of the MDA model, when compared to MGHDIS, and (2) an equal performance of the logit model and the MGHDIS model. Theodossiou (1993) provides evidence for the CUSUM model being clearly superior in performance, when compared to an MDA model. In a study of Joos et al. (1998b), logit models seem to have an overall better accuracy than machine learning decision trees. Nevertheless, the decision trees method appears superior for qualitative variables and companies with an abbreviated form of the annual accounts. Shumway (1999) compares two pairs of models. Firstly, the MDA model of Altman (1968) is compared with a hazard model containing the same variables and, secondly, the logit model of Zmijwski (1984) is compared with a similarly constructed hazard model. The survival analysis technique reports equal or

[26] It is important to bear in mind that, even if the performances of the compared models are based on one specific data set, an

better out-of-sample predictions than MDA. Luoma & Laitinen (1991), on the contrary, find that MDA and logit analysis outperform survival analysis. Frydman et al. (1985) compare the technique of RPA to MDA and conclude that, dependent on the choice of misclassification costs, the RPA model does not always outperform the MDA model. Spanos (1999) compares the accuracy of a fuzzy rule-based classification model with the accuracy of MDA, logit analysis and probit analysis and concludes that that the fuzzy rule-based classification model overall provides the best results.

Especially in the studies comparing the NN method to the classic statistical techniques and other methods, the conclusions are widely divergent. In most studies, the NNs have proven their superior performance[27]. More in particular, the studies of Odom & Sharda (1990), Cadden (1991), Coats & Fant (1992) and Coats & Fant (1993) suggest that the NN technique is clearly superior to the MDA method. Coats & Fant (1993) find that NNs are particularly more effective for "early detection" of financial distress and hence for minimizing the type I errors. Wilson & Sharda (1994) also compare MDA with NNs and show that NNs, using the back-propagation algorithm, perform slightly better than MDA in predicting failure one year prior to failure. Fletcher & Goss (1993) and Udo (1993) show that NNs are better than logit models in extracting information from attributes for forecasting bankruptcy. Chung & Tam (1993) perform a comparative analysis of a back-propagation NN and two types of decision tree algorithms. They conclude that the NN method is the best, both in a 1 and a 2 year prior to failure predictive context. In a study of Back et al. (1996b) a genetic algorithm-based NN is compared with the techniques of MDA and logit analysis. One and three years prior to failure, the NN seems to perform better than the other methods, while two years prior to failure, MDA shows the best results. Based on a sample of Belgian data, Weymaere & Martens (1993) find that NNs generally perform slightly better than MDA and logit analysis, especially for the medium term (3 years) analysis. Only for short term failure prediction, the logistic regression model seems to perform equally. In addition, when the data are first subjected to a principal component analysis, the NN technique significantly outperforms the traditional models. Bardos & Zhu (1997) compare both logit analysis and NNs with the linear MDA method and conclude that the logit model does not perform better than the MDA model, while a simple NN with eight input variables performs slightly better than MDA, especially for the non-failing firms. Zain (1994) compares a NN with a logit model and finds that, 3 years prior to failure, the NN has the best classification results, while, one year prior to failure, the logit

---

accurate comparison of performance results may still be impossible, because the models in scope include different variables.

[27] I also would like to refer to the overview of studies in the paper of Atiya (2001).

model performs best. When comparing the stability of the models, the NN technique appears to be the best method.

In contradiction to the above mentioned studies attributing superior qualities to the NN technique, Trigueiros a Taffler (1996) point out that there is little evidence that the artificial intelligence NN approach dominates the conventional multivariate models, particularly in the case of out-of-sample prediction. They argue that the relative performance of NNs, as compared to traditional statistical models, depends on the sample size used. NNs have an increased prediction accuracy when small samples are used. In this respect, they stress that the superiority of the NNs found in many studies may be the result of the use of a small sample as a basis for comparison. In the case of a small sample, the NNs are likely to show a very high number of fitted coefficients, in some cases even higher than the number of cases in the model fitting, and hence this over-fitting may result in an overstated accuracy for the NN in comparison to the other techniques. In this context, Altman et al. (1994), using a large sample, find little or no differences in classification performance between neural networks and conventional multivariate statistical techniques. They find that complex NNs perform equally or better than the MDA in the original estimation sample, while their performance in a validation context is even poorer than MDA, due to illogical weights and over-fitting. Similarly, in a study of Pompe & Bilderbeek (2000), it is shown that in the case of large samples, NNs and linear MDA models perform equally, while in the case of small estimation and training samples, NNs deliver better results than linear MDA. Finally, when comparing different modelling techniques on a large sample of 570 companies, also Hekanaho (1998) finds that the differences between the various induction methods, especially between rule-based learning and NNs, are quite small and inconsistent. Depending on the sample size, NNs and rule-based learning may perform better than logit analysis. On the contrary, Back et al. (1997) show that NNs (and machine learning models) perform better than MDA and logit models when the sample size is large (400 cases), while there is no best performing method when the sample size is smaller (200 or 100).

Finally, some studies indicate that NNs have poorer performances than other methods. Boritz et al. (1995) find that NN have no superior classification abilities as compared to MDA, logit and probit analysis. Only for particular combinations of proportions of failing and non-failing firms in the training and validation samples and for particular misclassification costs for type I and type II errors, the NN shows superior results. Furthermore, the results of a comparative study of Yang et al. (1999) indicate that logit models perform better than models based on NNs. It should be noted here that the poor performances of NN models may be explained by the fact that the NN technique is a new technique in the topic of corporate failure prediction and its possibilities have not yet been fully explored. In this respect, Laitinen & Kankaanpää (1999) state that "…, neural networks, is in its present form as effective as discriminant analysis was as early as thirty years ago (p. 84)".

The above analysis of the comparative studies on different modelling methods indicates that, although the alternative methods are computationally more complex and sophisticated than the classic cross-sectional statistical methods, it is not clear whether they produce better performing corporate failure prediction models. In other words, we may question the benefits to be gained from using the more sophisticated alternative methods mentioned in this paper.

5 CONCLUSIONS

Over the last 35 years, academic researchers from all over the world have been dedicated to the search for the best corporate failure prediction model, which classifies companies according to their (financial) health or failure risk (Altman, 1984; Dimitras et al., 1996; Altman & Narayanan, 1997). They have been using numerous types of modelling techniques and estimation procedures, with different underlying assumptions and different computational complexities. The classic cross-sectional statistical methods can be considered the most popular methods: a gigantic number of 'single-period' classification models or 'static' models have been developed, especially multivariate discriminant analysis or 'MDA' models and logit models (Zavgren, 1983; Van Wymeersch & Wolfs, 1996; Atiya, 2001). Besides these classic cross-sectional statistical methods, academic researchers have also been using several ***alternative methods*** for analysing and predicting business failure. These methods are the result of the strong progress in computational possibilities and in artificial intelligence (AI).

This study gives a clear overview and discussion of the alternative research methods, attributing each of them a fixed designation. First of all, it extensively elaborates on the most popular alternative methods of survival analysis, machine learning decision trees and neural networks, which have produced a considerable number of business failure prediction models. It explains each of these methods, it dilates upon their specific assumptions, advantages and disadvantages and it gives an overview of a number of academically developed corporate failure prediction models. Furthermore, this paper discusses several other alternative methods, which we consider to have a significant value added in the empirical literature on business failure: the fuzzy rules-based classification model, the multi-logit model, the CUSUM model, dynamic event history analysis, the catastrophe theory and chaos theory model, multidimensional scaling, linear goal programming, the multicriteria decision aid approach, rough set analysis, expert systems and self-organizing maps. It shortly discusses the main features of these methods, highlights their most important drawbacks and problems and states a number of academic studies on corporate failure prediction models. Several issues viewed in isolation by earlier studies are considered together and, therefore, this paper contributes towards a clear insight into the different alternative methods of corporate failure prediction modelling and their corresponding features.

Finally, based on an extensive analysis of a large number of empirical studies comparing the classification results and/or the prediction abilities of corporate failure prediction models based on different modelling methods, it is indicated that, although the alternative methods are computationally more complex and more sophisticated than the classic cross-sectional statistical methods, it is not clear whether they produce better performing corporate failure prediction models. In other words, we may question the benefits to be gained from using the more sophisticated alternative methods.

However, finding a complete answer to the question whether the more sophisticated, alternative modelling methods produce better performing failure prediction models than the classic cross-sectional statistical methods, requires further research systematically comparing all possible methods.

REFERENCES

Altman E.I., 1984, The success of business failure prediction models – An international survey. Journal of Banking and Finance, Vol. 8, p. 171-198.

Altman E.I., Marco G., Varetto F., 1994, Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience). Journal of Banking and Finance, Vol. 18, p. 505-529.

Altman E.I., Narayanan P., 1997, An international survey of business failure classification models. Financial Markets, Institutions and Instruments, Vol. 6, nr. 2, p. 1-57.

Altman E.I., Saunders A., 1998, Credit risk measurement: Developments over the last 20 years. Journal of Banking and Finance, Vol. 21, nr. 11-12, p. 1721-1742.

Atiya A.F., 2001, Bankruptcy prediction for credit risk using neural networks: A survey and new results. IEEE Transactions on Neural Networks, Vol. 12, nr. 4, July 2001, p. 929-935.

Back B., Laitinen T., Sere K., 1996a, Neural networks and bankruptcy prediction: funds flows, accrual ratios and accounting data. Advances in Accounting, Vol. 14, P. 23-37.

Back B., Sere K., Van Wezel M.C., 1995, Choosing the best set of bankruptcy predictors. Proceedings of the First Workshop on Genetic Algorithms, Vaasa, Finland, January 1995, p. 1-20.

Back B., Laitinen T., Sere K., Van Wezel M., 1996b, Choosing bankruptcy predictors using discriminant analysis, logit analysis and genetic algorithms. Turku Centre for Computer Science Technical Report nr.40, September 1996, p. 1-18.

Back B., Laitinen T., Hekanaho J., Sere K., 1997, The effect of sample size on different failure prediction methods. Turku Centre for Computer Science Technical Report nr. 155, December 1997, p. 1- 23.

Baesens B., Setonio R., Mues Ch., Vanthienen J., 2003, Using neural network rule extraction and decision tables for credit-risk evaluation. Management Science, Vol.49, nr. 3, March, p. 312-329.

Bardos M., Zhu W., 1997, Comparison of discriminant analysis and neural networks: Application for the detection of company failures. In: Biometric approaches in management science, Kluwer Academic Publishers, 25 p.

Boritz J.E., Kennedy D.B., Albuquerque A., 1995, Predicting corporate failure using a neural network approach. Intelligent Systems in Accounting, Finance and Management, Vol. 4, nr. 2, p. 95-111.

Cadden D., 1991, Neural networks and the mathematics of chaos – An investigation of these methodologies as accurate predictions of corporate bankruptcy. Paper presented at the First International Conference on Artificial Intelligence Applications on Wall Street, New York.

Charitou A., Neophytou E., Charalambous C., 2004, Predicting corporate failure: empirical evidence for the UK. European Accounting Review, Vol. 13, nr. 3, September 2004, p. 465-497.

Chung H.M., Tam K.Y., 1993, A comparative analysis of inductive-learning algorithms. International Journal of Intelligent Systems in Accounting, Finance and Management, Vol. 2, nr. 1, January 1993, p. 3-18.

Coats P.K., Fant L.F., 1991, A neural network approach to forecasting financial distress. The journal of Business Forecasting, Vol. 10, nr. 4, Winter 1991-1992, p. 9-12.

Coats K.P., Fant L.F., 1993, Recognising financial distress patterns using a neural network tool. Financial Management, Vol. 22, nr. 3, p. 142-155.

Crapp H., Stevenson M., 1987, Development of a method to assess the relevant variables and the probability of financial distress. Australian Journal of Management, Vol. 12, nr. 2, p. 221-236.

Cybinski P., 2000, The path to failure : where are bankruptcy studies at now ? Journal of Business and Management, Vol. 7, nr. 1, p. 11-39.

Cybinski P., 2001, Description. Explanation, Prediction: The Evolution of Bankruptcy Studies? Managerial Finance, Vol. 27, nr. 4, p. 29-44.

Daubie M., Levecq Ph., Meskens N., 2002, A comparison of the rough sets and recursive partitioning induction approach: An application to commercial loans. International Transactions in Operational Research, Vol. 9, p. 681-694.

Daubie M., Meskens N., 2002, Business failure prediction : a review and analysis of the literature. Working Paper, Department of Productions and Operations Management, Catholic University of Mons, Belgium, p. 1-15.

Dimitras A., Zanakis S., Zopudinis C., 1996, A survey of business failures with an emphasis on failure prediction methods and industrial applications. European Journal of Operational Research, Vol. 90, nr. 3, p. 487-513.

Doumpos M., Zopoudinis C., 1999, A multicriteria discrimination method for the prediction of financial distress: the case of Greece. Multinational Finance Journal, Vol. 3, nr. 2, p. 71-101.

Fletcher D., Goss E., 1993, Forecasting with neural networks: An application using bankruptcy data. Information and Management, Vol. 24, nr. 3, p. 159-167.

Frydman H., Altman E.I., Kao D.L., 1985, Introducing recursive partitioning for financial classification: The case of financial distress. Journal of Finance, Vol. 40, nr. 1, p. 269-291.

Gupta Y., Rao R., Bagchi P., 1990, Linear goal programming as an alternative to multivariate discriminant analysis: A note. Journal of Business Finance and Accounting, Autumn 1990, p. 593-598.

Hawley D.D., Johnson J.D., Raina D., 1990, Artificial neural systems: A new tool for financial decision-making. Financial Analysts Journal, Vol. 46, nr. 6, November-December 1990, p. 63-72.

Hekanaho J., Back B., Sere K., Laitinen T., 1998, Analysing Bankruptcy Data with Multiple Methods, American Association for Artificial Intelligence, p. 1-7.

Hertz J., Krogh A., Palmer R.G., 1991, Introduction to the theory of neural computation. Addison-Wesley.

Jo H., Han I., 1997, Bankruptcy prediction using case-based reasoning, neural networks, and discriminant analysis. Expert Systems With Applications, Vol. 13, nr. 2, p. 97-108.

Jones F.L., 1987, Current techniques in bankruptcy prediction. Journal of Accounting Literature, Vol. 6, p. 131-164.

Joos Ph., Ooghe H., Sierens N., 1998a, Methodologie bij het opstellen en beoordelen van kredietclassificatiemodellen. Tijdschrift voor Economie en Management, Vol. 18, nr. 1, p. 1-48.

Joos Ph., Ooghe H., Sierens N., 1997, Methodologie bij het opstellen en beoordelen van kredietclassificatiemodellen. Working Paper, March 1997, Department of Corporate Finance, Ghent University, p. 1-41.

Joos P., Vanhoof K., Ooghe H., Sierens N., 1998b, Credit classification: A comparison of logit models and decision trees. Proceedings Notes of the Workshop on Application of Machine Learning and Data Mining in Finance, 10th European Conference on Machine Learning, April 24, Chemnitz (Germany), p. 59-72.

Kahya E., Theodossiou P., 1996, Predicting corporate financial distress: A time-series CUSUM methodology. Paper presented at the Third Annual Conference of the Multinational Finance Association, June 1996, p. 1-38.

Kaiser U., 2001, Moving in and out of financial distress: Evidence for newly founded service sector firms. Paper nr. 01-09, February 2001, Centre for European Economic Research Paper Series, University of Connecticut, p. 1-33.

Kauffman R.J., Wang B., 2001, The success and failure of Dotcoms: A multi-method survival analysis. Working Paper nr. 01-09, Management Information Systems Research Center, Carlson School of Management, University of Minnesota, Minneapolis, p. 1-7

Keasey K., Watson R., 1991, Financial distress models: a review of their usefulness. British journal of Management, Vol. 2, nr. 2, July 1991, p. 89-102.

Kiviluoto K., Bergius P., 1998, Exploring corporate bankruptcy with two-level self-organising map. Paper published in the Proceedings of the Fifth International Conference on Computational Finance, London Business School, Boston, Massachusetts, USA, Kluwer Academic Publishers, pp. 373-380.

Laitinen T., Kankaanpää M., 1999, Comparative analysis of failure prediction methods: the Finnish case. The European Accounting Review, Vol. 8, nr. 1, p. 67-92.

Lane W.R., Looney S.W., Wansley J.W., 1986, An application of the Cox proportional hazards model to bank failure. Journal of Banking and Finance, Vol. 10, p. 511-531.

Lindsay D.H., Campbell A., 1996, A chaos approach to bankruptcy prediction. Journal of Applied Business Research, Vol. 12, nr. 4, p. 1-9.

Luoma M., Laitinen E.K., 1991, Survival analysis as a tool for company failure prediction. Omega International Journal of Management Science, Vol. 19, nr. 6, p. 673-678.

Mar-Molinero C.M., Ezzamel M., 1991, Multidimensional scaling applied to corporate failure. OMEGA International Journal of Management Science, Vol. 19, nr. 4, p. 259-274.

Mar-Molinero C., Serrano-Cinca C., 2001, Bank failure: a multidimensional scaling approach. The European Journal of Finance, Vol. 7, p. 165-183.

Messier W.F., Hansen J.V., 1988, Including rules for expert system development: An example using default and bankruptcy data. Management Science, Vol. 34, nr. 2, December 1988, p. 1403-1415.

Neophytou E, Mar-Molinero C., 2001, Predicting corporate failure in the UK: a multidimensional scaling approach. Working Paper, School of Management, University of Southampton, February 2001, p. 1-23.

Odom M., Sharda R., 1990, A neural network model for bankruptcy prediction. Proceedings of the Second IEEE International Joint Conference on Neural Networks, San Diego, USA, Vol. II, p. 163-168. New York: IEEE, p. 63-68.

Ooghe H., Joos P., De Bourdeaudhuij C., 1995, Financial distress models in Belgium: The results of a decade of empirical research. International Journal of Accounting, Vol. 30, p. 245-274.

Pawlak Z., 1982, Rough sets. International Journal of Information and Computer Science, Vol. 11, nr. 5, p. 341-356.

Peel M.J., Peel D.A., 1988, A multilogit approach to predicting corporate failure: Some evidence for the UK corporate sector. Omega International Journal of Management Science, Vol. 16, nr.4, p. 309-218.

Quinlan J.R., 1986, Induction of decision trees. Machine Learning, Vol. 1, p. 81-106.

Scapens R.W., Ryan R.J., Flecher L., 1981, Explaining corporate failure: A catastrophe theory approach. Journal of Business Finance and Accounting, Vol. 8, nr. 1, p. 1-26.

Shachmurove Y., 2002, Applying artificial neural networks to business, economics and finance. Working Paper nr. 02-08, Center for Analytic Research in Economics and the Social Sciences (CARESS), University of Pennsylvania, USA, p. 1-47.

Shumway T., 1999, Forecasting bankruptcy more accurately: a simple hazard model. Working Paper, University of Michigan Business School, USA, July 1999, p. 1-31.

Shumway T., 2001, Forecasting bankruptcy more accurately: a simple hazard model. Journal of Business, Vol. 74, nr. 1, p. 101-124.

Slowinski R., Zopudinis C., 1995, Application of the rough set approach to evaluation of bankruptcy risk. Intelligent Systems in Accounting, Finance and Management, Vol. 4, p. 27-41.

Spanos M., Dounias G., Matsatsinis N., Zopoudinis C., 1999, A fuzzy knowledge-based decision aiding method for the assessment of financial risk: the case of corporate bankruptcy prediction. European Symposium on Intelligent Techniques (ESIT), June 1999, Orthodox Academy of Crete, Greece,p. 1-7.

Tan C.N.W., Dihardjo H., 1999, A study on using artificial neural networks to develop an early warning predictor for credit union financial distress with comparison to the probit model. Technical Report nr. 00-07, School of Information Technology, Bond University, Australia , p. 1-19.

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, Vol. 88, nr. 422, June 1993, p. 441-449.

Trigueiros D., Taffler R., 1996, Neural networks and empirical research in accounting. Accounting and Business Research, Vol. 26, nr. 4, Autumn 1996, p. 347-355.

Tucker J, 1996, Neural networks versus logistic regression in financial modelling: a methodological comparison. Paper published in Proceedings of the 1996 World First Online Workshop on Soft Computing (WSC1), Nagoya University, Japan, August 19-30, 1996.

Udo G., 1993, Neural network performance on the bankruptcy classification problem. Computers and Industrial Engineering, Vol. 25, September 1993, p. 377-380.

Weymaere N., Martens J.-P., 1993, Financial distress analysis with neural networks. Internal paper, Electronics Laboratory, Ghent University, p. 1-12.

Wilson R.L., Sharda R., 1994, Bankruptcy prediction using neural networks. Decision Support Systems, Vol. 11, nr. 5, p. 545-557.

Yang Q.G., Temple P., 2000, The hazard of Chinese enterprises under restructuring. Paper presented at CEPR/ESRC Transition Economics and Chinese Economy Conference, Centre for Economic Reform and Transformation (CERT), 24/25 August 2000, Heriot-Watt University, Edinburgh, p. 1-37.

Yang Z.R., Platt M.B., Platt H.D., 1999, Probabilistic neural networks in bankruptcy prediction. Journal of Business Research, Vol. 44, nr. 2, p. 67-74.

Zain S., 1994, Failure prediction: an Artificial Intelligence Approach. PhD Thesis, Department of Corporate Finance, Faculty of Economics and Business Administration, Ghent University, p. 1-121.

Zavgren C, 1983, The prediction of corporate failure: the state of the art. Journal of Accounting Literature, Vol. 2, p. 1-37.

Zopoudinis C., Dimitras A.I., 1998, Multicriteria Decision Aid Methods for the Prediction of Business Failure, Dordrecht: Kluwer Academic Publishers.

Zopoudinis C., 1987, A multicriteria decision-making methodology for the evaluation of the risk of failure and an application. Foundations of Control Engineering, Vol. 12, nr. 1, p. 45-67.

TABLE 1

Overview of the most popular alternative methods applied in corporate failure prediction

Method	**Main advantages**	**Main drawbacks**	**Failure prediction models**
Survival analysis	* accounts for time dimension of failure ! * gives likely time to failure * allows for time-varying independent variables * no assumption of dichotomous dependent variable * no distributional assumptions * uses more data * allows for random censoring * easy interpretation	* not designed for classification * assumption: failing and non-failing firms belong to the same population * sample construction may affect hazard rates * requires homogenous lengths of failure processes in sample * subject to multicollinearity	Lane et al. (1986) Luoma & Laitinen (1991) Kauffman & Wang (2001)
Decision trees	* no strong statistical data requirements * allows for qualitative data * can handle noisy and incomplete data * user friendly: clear output * simple procedure	* specification of prior probabilities and misclassification costs * assumption: dichotomous dependent variable * relative importance of variables unknown * discrete scoring system * can not be 'applied'	Joos et al. (1998b) Frydman et al. (1985)
Neural networks	* does not use pre-programmed knowledge base * suited to analyse complex patterns * no restrictive assumptions * allows for qualitative data * can handle noisy data * can overcome autocorrelation * user-friendly: clear output * robust and flexible	* black box problem * can not be 'applied' * requires high quality data * variables must be carefully selected a priori * risk of over-fitting * requires definition of architecture * long processing time * possibility of illogical network behaviour * large training sample required	Odom & Sharda (1990) Cadden (1991) Coats & Fant (1991) Coats & Fant (1993) Fletcher & Goss (1993) Udo (1993) Wilson & Sharda (1994) Altman et al. (1994) Boritz et al. (1995) Back et al. (1996a) Bardos & Zhu (1997) Yang et al. (1999) Atiya (2001) Charitou et al. (2004)

TABLE 2

Overview of the other alternative methods applied in corporate failure prediction

Method	Main advantages	Main drawbacks	Failure prediction models
Fuzzy rules based classification model	* intuitive basis	* dependence on arbitrarily if-then rules	Spanos et al. (1999)
Multi-logit model	* considers information from several years	* assumption of signal consistency	Peel & Peel (1988)
CUSUM model	* takes account of data from present and past * short memory concerning good performances - long memory concerning bad performances		Theodossiou (1993) Kahya & Theodossiou (1996)
DEHA	* sees failure as a process * allows for time-varying variables * allows for censored cases * conditional probability' feature		Hill et al. (1996)
Chaos theory model	* considers information from different times	* strong assumption: healthy firms are more chaotic	Scapens et al. (1981) Lindsay & Campbell (1996)
MDS	* statistical map with intuitive interpretation * robust, when outliers * deals with highly correlated data * no distributional requirements * no need for data reduction	* not dynamic * not designed for prediction * can not be 'applied'	Mar-Molinero & Ezzamel (1991) Neophytou & Mar-Molinero (2001)
LGP	* no distributional requirements * flexible	* complex	Gupta et al. (1990)
MCDA			Zopoudinis (1987) Zopoudinis & Dimitras (1998) Doumpos & Zopoudinis (1999)
Rough set analysis	* allows for qualitative variables * easy method * user-friendly: can easily be applied to new cases * flexible	* quantitative variables need to be recoded into discrete variables	Slowinski & Zopoudinis (1995)
Expert systems	* allows for qualitative variables * no statistical distribution requirements * user-friendly: can easily be applied to new cases	* 'predefined knowledge base' needs to be programmed * heuristics needs to be determined * time consuming, expensive * not flexible * sensitive to incomplete, noisy	Messier & Hansen (1988)

		data or input information with errors	
SOM	* allows to detect regions of increased failure risk or to view the evolution of the condition of a company *the two-level SOM offers some possibilities to explore typical ‘failure paths	* requires pre-selection of a small set of independent variables	Kiviluoto & Bergius (1998)

TABLE 3

Overview of the conclusions of comparative studies on modelling methods for corporate failure

Method Author	year	MDA	LA	PA	LPM	MGH DIS	CUSUM	Decision Tree	SA	RPA	Fuzzy rules	NN
Doumpos & Zopoudinis	1999	X	**X**$^{+}$			**X***						
Theodossiou	1993	X					**X***					
Joos et al.	1998b		**X***					**X**$^{+\circ}$				
Shumway	1999	X	X						**X***			
Luoma & Laitinen	1991	**X***	**X***						X			
Frydman et al	1985	X								**X**$^{+\circ}$		
Spanos	1999	X	X	X							**X***	
Odom & Sharda	1990	X										**X***
Cadden	1991	X										**X***
Coats & Fant	1991	X										**X***
Coats & Fant	1993	X										**X***
Wilson & Sharda	1994	X										**X**$^{+}$
Charitou et al.	2004		X									X
Fletcher & Goss	1993		X									**X***
Udo	1993		X									**X***
Chung & Tam	1993							X				**X***
Back et al.	1996b	**X***$^{\circ}$	X									**X***$^{\circ}$
Weymaere	1993	X	**X**$^{+\circ}$									**X**$^{+}$

& Martens												
Bardos & Zhu	1997	X	X									**X**$^{+}$
Pompe & Bilderbeek	2000	X										**X**$^{+\circ}$
Zain	1994		**X**$^{+\circ}$									**X**$^{+\circ}$
Boritz et al.	1995	X	X	X								X
Altman et al.	1994	X										X
Yang et al.	1999		**X***									X
Trigueiros a Taffler	1996											
Back et al.	1997	X	X					**X**$^{+\circ}$				**X**$^{+\circ}$
Hekanaho	1998		X							**X**$^{+\circ}$		**X**$^{+\circ}$
Laitinen & Kankaanpaa	1999	X	X						X	X		X

MDA = multivariate discriminant analysis

LA = logit analysis

PA = probit analysis

LPM = linear probability model

SA = survival analysis