**Vlerick Leuven Gent Working Paper Series 2004/15**

# 35 YEARS OF STUDIES ON BUSINESS FAILURE: AN OVERVIEW OF THE CLASSIC STATISTICAL METHODOLOGIES AND THEIR RELATED PROBLEMS

SOFIE BALCAEN

HUBERT OOGHE

Hubert.Ooghe@vlerick.be

D/2004/6482/16

# 35 YEARS OF STUDIES ON BUSINESS FAILURE: AN OVERVIEW OF THE CLASSIC STATISTICAL METHODOLOGIES AND THEIR RELATED PROBLEMS

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

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

**ABSTRACT**

Over the last 35 years, the topic of business failure prediction has developed to a major research domain in corporate finance. A gigantic number of academic researchers from all over the world have been developing corporate failure prediction models, based on various modelling techniques. The ‘classic cross-sectional statistical’ methods have appeared to be most popular. Numerous ‘single-period’ or ‘static’ models have been developed, especially multivariate discriminant models and logit models.

As to date, a clear overview and discussion of the application of the classic cross-sectional statistical methods in corporate failure prediction is still lacking, this paper extensively elaborates on the application of (1) univariate analysis, (2) risk index models, (3) multivariate discriminant analysis, and (4) conditional probability models, such as logit, probit and linear probability models. It discusses the main features of these methods and their specific assumptions, advantages and disadvantages and it gives an overview of a large number of academically developed corporate failure prediction models.

Despite the popularity of the classic statistical methods, there have appeared to be several problems related to the application of these methods to the topic of corporate failure prediction. However, in the existing literature there is no clear and comprehensive analysis of the diverse problems. Therefore, this paper brings together all criticisms and problems and extensively enlarges upon each of these issues. So as to give a clear overview, the diverse problems are categorized into a number of broad topics: problems related to (1) the dichotomous dependent variable, (2) the sampling method, (3) non-stationarity and data instability, (4) the use of annual account information, (5) the selection of the independent variables, and (6) the time dimension.

This paper contributes towards a thorough understanding of the features of the classic statistical business failure prediction models and their related problems.

# 1 INTRODUCTION

Over the past 35 years, the topic of 'business failure prediction' has developed to a major research domain in corporate finance. Many academic studies have been dedicated to the search for the best corporate failure prediction model, based on publicly available data and statistical techniques. Not only in developed but also in developing countries, researchers have been putting at lot of effort into building business failure prediction models (Altman, 1984; Dimitras et al., 1996; Altman & Narayanan, 1997). As a result, the academic research on business failure is extensive.

There are several reasons for the strong interest in the avoidance and the prediction of corporate failure or 'business failure'.

Firstly, business failure involves *many parties* and *large costs* and, therefore, research in the topic of corporate failure prediction has been stimulated both by private agents, who urge for an accurate failure prediction model so as to be able to take preventive or corrective actions (Charitou et al., 2004) in firms that are predicted to fail in the future, and by the Government, that aims to detect bad performing companies and to take corrective actions in order to prevent failure (Shumway, 1999)[3]. Several stakeholders rely on a firm's success. Evidence shows that the market value of a distressed firm declines substantially, which may severely affect different stakeholders of the firm (Zavgren, 1983; Bickerdyke et al., 1999; Charitou et al., 2004; Daubie & Meskens, 2002). Moreover, company failure may inflict negative shocks for each of the stakeholders and, therefore, the total (economic and social) cost of business failure may be large. Company failure generates various types of costs, not only for the direct (internal) stakeholders of the company – the entrepreneur, management and employees – but also for the direct environment of the firm – shareholders, equity and credit suppliers, clients and suppliers, the Government – and the economy as a whole. Due to 'contagion-effects', the costs of the failure of a firm with a large network of related companies may cause a downward spiral for the whole economy of a country (Doumpos & Zopoudinis, 1999; Bickerdyke, 1999) and, in his way, company failure may have important consequences with respect to the employment and the (regional) economic welfare (Van Caillie, Arnould,

[3] The Government In many countries, the "central bank" has been developing prediction models for company failure. For example, the National Bank of Belgium (Belgium), the Banque de France (France), the Deutsche Bundesbank (Germany), the Centrale dei Bilanci (Italy), the Oesterreichische Nationalbank (Austria) and the Bank of England (UK) all have a failure prediction model for the assessment of financial heath of firms. In Belgium, Flemish Government created the "Committee for the Supervision of Business Management" in 1985, which was to encourage firms to adopt a policy aimed at the prevention of difficulties that may threaten continuity (a preventive business management) and to give advisory services to firms in problems in order to reduce the failure probability.

2001). Consequently, prediction of company failure is important not only from the 'individual' point of view, but also for the 'society as a whole' (Amrhein, 1998).

Secondly, due to the *negative spiral in the general economic environment*, the measurement of company performance and research on causes of company failure (Tamari, 1966; Van Caillie & Dighaye, 2002) and failure prediction has been stimulated. Over the last 30 years, the general economic situation of the developed countries has changed at an enormous speed and companies have experienced a downward trend. The conditions in which companies operate have changed at lot. Companies have started operating in a global economy, competition has become much stronger and Government regulation has increased. In many countries (the UK, USA, Belgium, Israel,…), bankruptcy rates have risen spectacularly and a lot of companies have become increasingly vulnerable to failure (Tamari, 1966; Altman & Saunders, 1998; Doumpos & Zopoudinis, 1999; Blazy, 2000; Charitou et al., 2004; Daubie & Meskens, 2002).

Thirdly, the evolution in the *availability of data and statistical techniques* has offered increased possibilities for research concerning corporate failure prediction. First, the progress in quantitative sciences – mathematics, statistics, applications in informatics and artificial intelligence – has provided a large range of quantitative techniques, which can be used for the development of failure prediction models (Doumpos & Zopoudinis, 1999; Van Caillie & Dighaye, 2002). Second, many company data have become publicly available (Van Caillie & Dighaye, 2002), which allows researchers to use large computer databases of standardized company financial information. For example, Belgium shows an important evolution concerning the publication demands of annual accounts. Since the Royal Decree of 1976, prescriptions with respect to the layout of the annual accounts have changed a lot and many uniform data on the balance sheet and the income statements of Belgian companies have become available trough the 'Balanscentrale' of the National Bank of Belgium.

Fourthly, in line with the *extended academic research on the impact of market imperfections* and *information asymmetry*, research on credit rating and corporate failure prediction has been boosted.. In contradiction with Modigliani & Miller (1958), who assume that financial markets are perfect and that investment and financing decisions can be separated, it has become clear that financial markets are not perfect. The available funds are insufficient to fund all profitable or 'good' projects (i.e. projects with a positive net present value) and, consequently, some value-creating projects may be left without financing. For example, the banking market or loan market is capital-constrained and, also, investors may ration capital because of information-asymmetries. In this context, the provision of funds or capital depends on the ‘expected return’ and hence on the success probability of the project(s). Only those projects with the highest expected return will be provided with the necessary capital. In this context, risk assessment of the companies and their project(s) has appeared to be of major importance. Moreover, the use of failure prediction models may reduce the existing information asymmetry between funds providers and a firm’s management.

Fifthly, failure prediction models have been proven *necessary to obtain a more accurate assessment of a firm’s financial situation*. Although one could expect that independent auditors or other decision makers are able to make a correct assessment concerning the financial health of firms (‘going concern’ or ‘clean’ qualification), research has shown that, in practice, they do not perform as well as failure prediction models in classifying companies as failing (Altman & McGough, 1974; Deakin, 1977; Keasey & Watson, 1991).

Finally, research on corporate failure prediction has been stimulated as a result of the consultative papers of the *New Basel Capital Accord*, which is to replace the original Basel Capital Accord of 1988 and will become effective in 2005 in most industrialized countries. In the regulations of the original Basel Capital Accord regarding capital requirements, risk weights are based on certain fixed categories of risk, associated with some types of claims (Basel Committee on Banking Supervision, 1988). According to the New Basel Capital Accord, the risk-adequate equity coverage for corporate lending will become more risk-sensitive and more flexible. Another important feature of this New Basel Capital Accord is that banks are allowed to use their own internal rating systems in order to determine their risk-adequate equity coverage. In this context, the New Capital Accord creates a great incentive for banks to develop their own internal risk assessment models. In addition, it might also be expected that the use of information provided by credit rating agencies will become very important and that these rating agencies will pay a lot of attention to development of failure prediction models to determine the failure risk of companies (Odera et al., 2002; Becchetti & Sierra, 2003; Rime, 2003).

With a view to evaluating and predicting failure risk of companies and of finding an adequate model to classify companies according to their (financial) health, academic researchers from all over the world have been using numerous types of modelling techniques and estimation procedures, with different underlying assumptions and a different computational complexity. The most popular methods are the “***classic cross-sectional statistical methods***”. Numerous ‘single-period’ or ‘static’ failure prediction models have been developed, especially multivariate discriminant models and logit models (Zavgren, 1983; Van

Wymeersch & Wolfs, 1996; Atiya, 2001). However, despite the popularity of the classic statistical methods, there appear to be several problems related to the application these methods to the topic of corporate failure prediction. As a result, classic statistical failure prediction models have been the subject of many criticisms.

As to date, a clear ***overview and discussion*** of the classic cross-sectional statistical methods is still lacking, a first aim of this paper is to extensively elaborate on the application of (1) univariate analysis, (2) risk index models, (3) multivariate discriminant analysis, and (4) conditional probability models, such as logit, probit and linear probability models, in corporate failure prediction. This paper will discuss the main features of these methods and their specific assumptions, advantages and disadvantages and will give an overview of a large number of academically developed corporate failure prediction models.

Secondly, as in the existing literature there is no clear and comprehensive analysis of the ***diverse problems*** related to the application the classic statistical methods to the topic of corporate failure prediction, this paper aims to bring together all critics and problems. So as to give a clear overview, the diverse problems will be categorized into a number of broad topics and each of these problems will be extensively enlarged upon.

It is clear that this paper may contribute towards a thorough understanding of the features of the classic statistical failure prediction models and their related problems.

The remainder of this paper is structured as follows. Section two of this paper is entirely devoted to the overview and discussion of the classic statistical methods applied in corporate failure prediction modelling. Section three discusses the various problems that crop up when applying classic statistical methods to the issue of corporate failure prediction.

## 2 CLASSIC STATISTICAL FAILURE PREDICTION MODELS

Over the years, the classic cross-sectional statistical methods have been widely used for the development of corporate failure prediction models (Zavgren, 1983; Van Wymeersch & Wolfs, 1996; Atiya, 2001). These models are also called ‘single-period’ classification models or ‘static’ models (Shumway, 1999). They involve a certain classification procedure so as to classify firms into a failing group or a non-failing group of firms with a certain degree of accuracy or ‘misclassification rate’. The two different types of misclassification errors that can be made by applying business failure prediction models, are discussed in Appendix 1.

Multiple discriminant analysis is by far the dominant classic statistical method, followed by logit analysis (Altman & Saunders, 1998). Other classic methods are: univariate analysis, risk index models, probit analysis and linear probability models. This section elaborates on the different classic cross-sectional statistical methods. It explains the features of each of these methods and it discusses their specific assumptions, advantages and drawbacks. At the end of this section, a table gives an overview of the main advantages and drawbacks of the different methods and reports a number of academically developed corporate failure prediction models.

It should be emphasized that this paper does not aim to present an exhaustive overview of all classic statistical failure prediction models that have been developed until now. We focus on models developed by academic researchers, that (1) are published in (academic) literature, (2) are frequently cited in literature and (3) are considered to have a significant value added in the empirical literature on corporate failure. For a supplementary overview of failure prediction models, we would like to refer to the work of Zavgren (1983), Altman (1984), Taffler (1984), Jones (1987), Keasey & Watson (1991), Ooghe et al. (1995), Dimitras et al. (1996), Altman & Narayanan (1997) and Altman & Saunders (1998). Zavgren (1983), Altman (1984) and Taffler (1984) give an overview of the literature on failure models in the 1960s and 1970s. Zavgren (1983) surveyed different methods and failure prediction models developed on USA data, while Altman (1984) gives an overview of business failure prediction models developed in different countries. Jones (1987) and Keasey & Watson (1991) also 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

## 2.1 Univariate analysis

In 1966, *Beaver (1967a)* was the pioneer in building a corporate failure prediction model with financial ratios. He was the first researcher to apply a univariate model – a “univariate discriminant analysis model” – on a number of financial ratios of a paired sample of failing and non-failing companies in order to predict company failure. In view of selecting the financial ratios to be included in his univariate model, Beaver (1967a) applied a dichotomous classification test in order to identify those ratios that were the best in classifying the companies as failing or non-failing.

In a univariate failure prediction model, the emphasis is placed on individual signals of failure. A ***classification procedure*** is carried out separately for each measure or ratio in the model. When classifying a firm, the value for each measure or ratio is analysed separately and, according to the corresponding 'optimal cut-off point' of the measure – the point at which the percentage of misclassifications is minimized – the firm is classified as failing or non-failing. Generally, if a firm's ratio value is below the cut-off point, it is classified as failing and, if the firm's ratio is above the cut-off point, it is classified as non-failing. For those ratios where a higher value indicates a poorer financial health, the opposite classification rule is to be applied. In this kind of classification, the ***classification accuracy*** can be measured by the total misclassification rate and the percentage of the type I and type II errors.

An important ***advantage*** of the univariate failure prediction model is its simplicity. The application of a univariate model does not require any statistical knowledge: for each ratio, one simply compares the ratio value for the firm with a cut-off point and decides on the classification accordingly.

On the other hand, it should be stressed that the univariate analysis is based on the stringent ***assumption*** that the functional form of the relationship between a measure or ratio and the failure status is linear. It is obvious that this assumption is often violated in practice, where many ratios show a non-linear relationship with the failure status (Keasey & Watson, 1991). As a result, the univariate modelling technique is often applied in an inappropriate way and conclusions may be questionable.

Although the simplicity of the univariate model is appealing, this method also shows some important ***disadvantages***. Firstly, firm classification can only occur for one ratio at a time, which may give inconsistent and confusing classifications results for different ratios on the same firm (Altman, 1968; Zavgren, 1983). This problem is called the *'inconsistency problem'*. Secondly, when using financial accounting ratios in a univariate model, it is difficult to assess the importance of any of the ratios in isolation, because most variables are highly correlated (Cybinski, 1998). In the same context, the univariate model contradicts with reality in that the financial status of a company is a complex, multidimensional concept, which can not be analysed by one single ratio. Finally, the optimal cut-off points for the variables are chosen by ‘trial and error’ and on an ‘ex post’ basis, which means that the actual failure status of the companies in the sample is known (Bilderbeek, 1973). Consequently, the cut-off points may be sample specific and it is possible that the classification accuracy of the univariate model is (much) lower when the model is used in a predictive context (i.e. ‘ex ante’).

## 2.2 Risk index models

In response to Beaver, *Tamari (1966)* realized that the assessment of the financial health of a company cannot rely on one variable alone. Furthermore, he pointed out that, due to the inconsistency problem (see above) it is difficult to get a clear picture of a companies’ financial health. These are the reasons why he introduced a ‘risk index’. It is a simple ‘point system’, which includes different ratios, generally accepted as measures of financial health. Each firm is attributed a certain number of points, between 0 and 100, according to the values of the ratios for the firm. A higher total of points indicates a better financial situation. The risk index takes account of the fact that some ratios are more important than others. Points are allocated in a way that the most important ratios have higher weights (i.e. correspond to a higher maximum of points).

M*oses & Liao (1987)* presented another interesting type of risk index. This type of risk index first requires a univariate analysis, which allows to determine an optimal cut-off point for each of the financial ratios. Next, for each of the ratios, a dichotomous variable is created and these variables are assigned a score of one if a firm’s ratio value exceeds the optimal cut-off point and a score of zero if the value is lower. Finally, a risk index is created by simply adding the dichotomous variables. Similar to the risk index of Tamari, this risk index associates a high score to a healthy financial situation.

The risk index model has the ***advantage*** that it is very intuitive and simple in its application. On the other hand, this immediately points at a major ***drawback*** of the risk index: it is rather subjective in nature. For example, the weights of the ratios in the model of Tamari are determined subjectively.

## 2.3 Multiple discriminant analysis

In 1968, *Altman (1968)* introduced the statistical multivariate analysis technique into the problem of company failure prediction and estimated a model called the ‘Z-score model’. The method he used is called ‘multiple discriminant analysis’, which is “a statistical technique used to classify an observation into one of several a priori groups dependent upon the observation’s individual characteristics… [it] attempts to derive a linear [or quadratic] combination of these characteristics which ‘best’ discriminates between the groups (Altman, 1968, p. 592)”. Over the years, there has been an enormous volume of studies based on Altman’s Z-score model. In 1977, *Altman et al. (1977)* adjusted the original Z-score model – in order to take into account the new financial reporting standards – into a new, better performing model, known as ‘Zeta analysis’. Until the 1980s, the technique of MDA dominated the literature on corporate failure models. After the 1980s, its use has decreased (Dimitras et al., 1996), but the MDA method is frequently used as a ‘baseline’ method for comparative studies (Altman & Narayanan, 1997). In other words, MDA seems to be the generally accepted ‘standard method’. Most of the MDA studies used the linear MDA model, but also the quadratic MDA model was introduced, in order to (statistically) overcome the problem of unequal dispersion matrices in the data (see further).

An ***MDA model*** consists of a linear combination of variables, which provides the best distinction between the group of failing and the group of non-failing firms. For example, Altman’s Z-score model is a linear combination of the following ratios: working capital / total assets, retained earnings / total assets, earnings before interest and taxes / total assets, market capitalization / total debts and sales / total assets (Altman, 1968). The ***linear discriminant function*** is the following (Lachenbruch, 1975):

$$D_i = D_0 + D_1X_{i1} + D_2X_{i2} + \ldots + D_nX_{in} \qquad (1)$$

with $D_i$ = discriminant score for firm i (between $-\infty$ and $+\infty$),
$X_{ij}$ = value of the attribute $X_j$ (with j = 1, …, n) for firm i,
$D_j$ = linear discriminant coefficients with j = 0, 1,…, n.

In an MDA model, several (mostly financial) characteristics or ‘attributes’ of a company are combined into one single ***multivariate discriminant score*** $D_i$. This discriminant score is a one-dimensional measure which has a value between $-\infty$ and $+\infty$ and gives an indication of the financial health of the firm. This is why MDA is called a ‘continuous scoring system'. In most studies, a low discriminant score indicates a poor financial health[4]. The integration of several variables into one single performance measure or discriminant score is based on the principle of 'the whole being worth more than the sum of the parts' (Taffler & Agarwal, 2003). It is possible that seemingly insignificant variables on a univariate basis do supply significant information in a multivariate context (Altman, 1968) or that some coefficients have an unexpected, counter-intuitive sign, caused by the multivariate character of MDA (Ooghe & Verbaere, 1985).

In a ***classification context***, the essence of the MDA method is to assign a firm to the failing or the non-failing group based on its discriminant score. The firm will be assigned to the group it most closely 'resembles'. According to a certain optimal cut-off point for the MDA model, classification is achieved as follows: a firm is classified into the failing group if its discriminant score ($D_i$) is less than the cut-off point and it is classified into the non-failing group if its score $D_i$ exceeds or equals the cut-off point[5]. In the strict sense, a classification by an MDA model can not be considered as a prediction, but, in practice, when a firm is classified as failing because it most resembles the firms failing in the next year (i.e. the attributes of the sample of failing firms are measured in year t+1), this classification is treated as a prediction that the firm will fail in year t+1 (Blum, 1974).

The ***classification accuracy*** or 'performance' of an MDA model is mostly assessed on the basis of the type I and type II error rates. Furthermore, the percentage of correct classifications or the unweighted error rate[6] are frequently used. Besides these error measures, which all require the specification of a certain cut-off point, the performance of an MDA model can also be evaluated on grounds of stochastic dominance. In this respect, the 'Receiver Operating Curve (ROC)' (Steele, 2002) and the 'trade-off function' (Ooghe et al., 2003; Ooghe & Balcaen, 2002c) give a clear graphical presentation of the performance of a model (see Appendix 2) and do not require the specification of a cut-off point. The larger the area under the ROC or the closer the trade-off function to both axes, the better the model's performance. In the same context, the Gini-coefficient is a very suitable criterion for model performance. The Gini-coefficient is an aggregated performance measure that reflects the difference between the trade-off function of the model and the trade-off function of the non-discriminating model (see Appendix 3): the higher the Gini-coefficient, the better the

---

[4] In some studies - for example, in Ooghe et al. (1994a) - the MDA model is defined in the opposite direction. Here, a high discriminant score $D_i$ indicates a poor financial health and, hence, the score is seen as a risk measure.

[5] Ooghe et al. (1994a) use two cut-off points instead of one. The firms that have a score less than the lower threshold X are classified as non-failing (here, a lower score indicates a lower failure risk) and the firms with a score above higher threshold Y are classified as failing. Companies with a score between X and Y are "grey zone" companies, of which the financial health situation is unclear.

discrimination between failing and non-failing firms (Ooghe et al., 2003; Ooghe & Balcaen, 2002c). Other possible performance measures are $R^2$-type measures and measures based on entropy (Joos et al., 1998a). $R^2$-type measures indicate the percentage of the variance that is explained by the model. For example, the count $R^2$ measure reports the number of correctly and falsely classified firms.. Measures based on entropy are used as performance measures in failure prediction research by, for example, Zavgren (1985). These entropy measures only evaluate the discriminating ability of the model and do not allow taking misclassification costs and population proportions into account a posteriori.

Although Eisenbeis (1977) assumes that ***multicollinearity*** is an irrelevant concern in MDA models and Altman & Eisenbeis (1978) point out that multicollinearity among the independent variables does not pose any problem in MDA models, most authors agree that, when the correlation among the independent variables is severe, collinearity possibly inflicts some problems. It leads to unstable and difficult-to-explain parameter estimates and may influence the accuracy of the classification results. (Joy & Tollefson, 1975; Joy & Tollefson, 1978; Ooghe et al., 1994a; Back et al., 1996b; Doumpos & Zopoudinis, 1999). Therefore, one should perform correlation analyses and multicollinearity tests and avoid the inclusion of highly correlated variables in the MDA model. Edmister (1972) points out that "low levels of inter-correlation present few problems, but as the data set becomes increasingly multicollinear, the problem becomes increasingly severe (p. 1482)". In this respect, Lussier & Corman (1994) tested the variables for the presence of 'problematic' multicollinearity.

The technique of MDA starts from several assumptions (Edmister, 1972; Eisenbeis, 1977; Zavgren, 1983; Karels & Prakash, 1987; Joos et al., 1998a). First of all, MDA assumes that that the ***dataset is dichotomous***: groups are discrete, non-overlapping and identifiable. Problems related to this issue are discussed in point 2.2.1. Secondly, the use of MDA is also based on ***three restrictive assumptions***: (1) the independent variables included in the model are multivariate normally distributed, (2) the group dispersion matrices or 'variance-covariance matrices' are equal across the failing and the non-failing group and (3) the prior probability of failure and the misclassification costs are specified. Although some authors have stressed the importance of the first two restrictive assumptions and their potential biases, most corporate failure studies do not attempt to analyse whether the data satisfy these assumptions. As, in practice, the data rarely satisfies the three statistical assumptions, the MDA modelling technique is often applied in an inappropriate way and, consequently,

---

[6] The unweighted error rate is the unweighted sum of the two types of errors.

conclusions and generalizations are questionable (Joy & Tollefson, 1975; Eisenbeis, 1977; Richardson & Davidson, 1984; Zavgren, 1985).

In practice, it seems that the first assumption of *multivariate normality* is often violated (Deakin, 1976; Taffler, 1977; Barnes, 1987), which may result in a bias in the significance tests and in the estimated error rates (Eisenbeis, 1977; Richardson & Davidson, 1984; McLeay & Omar, 2000). It should be mentioned here that a multivariate normally distribution a priori requires univariate normality (Karels & Prakash, 1987). For this reason, some researchers test for univariate normality of the variables and implicitly neglect testing for multivariate normality. It should be noted that there is ample evidence that financial ratio variables, which are mostly used in MDA models, generally exhibit non-normal distributions[7] (Barnes, 1982; Ooghe & Verbaere, 1985; McLeay & Omar, 2000). Some researchers correct for univariate non-normality and try to approximate univariate normality by transforming the variables prior to the estimation of their model. Deakin (1976), Taffler (1983) and Altman et al. (1977), for example, forced their non-normal variables into a normal distribution by means of a normalizing transformation. In the literature, there are no general guidelines concerning the appropriate transformation in order to approximate normality. For example, Taffler (1983) transformed the variables by means of reciprocal or logarithmic transformations. Altman et al. (1977) used a log-transformation and Deakin (1976) provided evidence that using a square root or log-normal transformation of the financial ratios may result in a normal distribution. Other researchers approximate univariate normality by 'trimming' the outliers prior to the estimation of their model. Trimming may involve 'outlier deletion', which involves segregating outliers by reference to the normal distribution, or 'windsorising', which concerns changing an outlier's value into the value of the closest non-outlier so that, finally, the distribution fits the normal distribution (Taffler, 1983; Barnes, 1987; Ooghe et al., 1995; McLeay & Omar, 2000). Although transforming the variables may result in normally distributed variables, (1) the assumption of multivariate normality is still violated and (2) the transformation may change the interrelations among the variables (Eisenbeis, 1977) and hence may distort the MDA model. By way of conclusion, it is clear that this issue needs to be treated with care.

[7] Deakin (1976) found that financial ratios might be more normally distributed within a specific industry group. Hence, the violation of the normality assumption might be weaker when building industry-specific failure prediction models.

A second assumption which needs to be tested prior to the development of the MDA model is the assumption of *equal dispersion matrices*[8]. If this assumption is violated, the significance tests for differences in variable means between the failing and non-failing group of firms will be affected. Furthermore, in case of unequal dispersion matrices, a quadratic classification rule – a quadratic MDA model – needs to be used (Joy & Tollefson, 1975; Eisenbeis, 1977; Zavgren, 1983). In practice, however, researchers avoid working with quadratic MDA models, because these models are very complex and only seem to outperform linear MDA models in the case of (1) large samples, (2) a small number of independent variables relative to the sample and (3) very substantial differences in the dispersion matrices[9]. Therefore, they simply try to transform the data in a way that the dispersion matrices are not too different and apply linear MDA (Taffler, 1982).

The third assumption states that, in the selection of the optimal cut-off score of the estimated model, the *prior probabilities* of belonging to the failing or non-failing group (i.e. population) and the *costs of a type I and a type II error* should be considered (Edmister, 1972; Eisenbeis, 1977; Deakin, 1977; Zavgren, 1983; Hsieh, 1993; Steele, 1995). If this restrictive assumption is violated, the reported accuracy of the MDA model will be biased and will not indicate the accuracy of the model when applied to the total population. In this respect, Deakin (1977) points out that the specification of prior probabilities and misclassification costs is required in order to get an accurate image of the frequency of errors likely to be obtained in a 'real world' application of the model. The optimal cut-off point should result from the minimization of a 'total loss function', which includes the error rates and both the corresponding population proportions and misclassification costs. In practice, however, the specification of the error 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 depends on the risk behaviour of the decision-maker and his or her attitude towards the proportion of the cost factors. In addition, the specification of population proportions seems to be very difficult and subjective, as a certain reference period needs to be chosen[10]. This is why Steele (1995) calls it a 'subjective factor'. Due to these practical problems, most researchers

[8] It should be noted here that the equality of the dispersion matrices can only be tested in a correct way by Box's M criterion if the assumption of multivariate normality is met (Taffler, 1982). However, as the variables used in most MDA models are typically non-normally distributed, the results of this Box's M test are often misleading.

[9] An overview of studies on the use of quadratic classification models and equality of dispersion matrices is given in Eisenbeis (1977).

[10] When a prediction model 3 years prior to failure is constructed (i.e. a model based on a dataset of information on the annual accounts 3 years prior to the moment of failure of the failing firms), fluctuating failure rates from time to time make it difficult to determine which year the failure rate (and hence the population proportions) have to be taken from (Eisenbeis, 1977) or to determine how the 'average' failure rate has to be calculated. This problem will be even greater when a 'pooled sample' of observations from different time periods is used (see further).

applying MDA simply try to minimize the total error rate instead of the total loss function. Unlike Altman et al. (1977) and Taffler (1982), which are two of the small number of studies that do take note of the ratio of both error costs, most researchers neglect specifying error costs or/and population proportions. They implicitly assume that (1) the misclassification costs are equal and that (2) the sample proportions are equal to the population proportions[11]. It is obvious that neglecting these factors has some important implications. As, in practice, the cost of misclassification of a failing firm (type I error) often is much larger than the costs of misclassifying a non-failing firm (type II error), neglecting misclassification costs generally leads to relatively high type I errors. On the contrary, as the population frequency of non-failing firms is much larger than the population frequency of failing firms, neglecting population frequencies implies a too strong focus on reducing type I errors, which results in a relatively low type I error rate and a relatively high type II error. The latter aspect is the reason why El-Zayaty (1987) generally finds high type II error rates in many failure prediction studies.

There are, however, some possible solutions to the problems related to the definition of the optimal cut-off point. A first solution is to report the classification results (type I and type II errors) of the model for different cut-off values. The study of Pompe & Bilderbeek (2000) and Ooghe & Verbaere (1985) are one of the few studies that choose this option. Another option is the 'black-grey-white' method, as mentioned by Edmister (1972). This method specifies a lower cut-off score as the score where the model has a 0% type II error and a higher cut-off score as the score with 0% type I error. The area between these two scores is the grey area. When applying this "black-grey-white" method, the predictive power of the model is assessed by the percentage of firms that is classified into the grey area.

Although MDA is the most frequently used modelling technique in failure prediction, it has some serious ***disadvantages***, additional to the problems related to the violation of the basic assumptions. Firstly, MDA requires that the *classification rule* is linear, which means that a discriminant score above or below a certain cut-off point automatically signals a good or a poor financial health. In the same respect, the MDA classification rule intuitively contradicts with the fact that some variables do not show a linear relationship to financial health: some variables indicate financial problems both when they have a very low and a very high value.

[11] In case of matched samples, they implicitly assume that group membership in the population is equally likely.

Secondly, we should bear in mind that the discriminant scores are only *ordinal* measures, which allow for a relative (ordinal) ranking between firms. MDA can also generate failure probabilities, but this requires a subjective and possibly inaccurate assessment of the probabilities associated with particular discriminant scores (Zavgren, 1985). Thirdly, although MDA is very similar to the technique of multiple regression analysis, it is computationally not equivalent. The estimation method of least-squares is not suitable when estimating a linear relation with a binary dependent variable (Bilderbeek, 1978; Bilderbeek, 1979). Consequently, in MDA models, the standardized coefficients can not be interpreted like the β-coefficients of a regression and hence do not indicate the *relative importance* of the different variables (Altman, 1968; Blum, 1974; Joy & Tollefson, 1975; Eisenbeis, 1977; Taffler, 1983). The MDA coefficients are not unique – only the variables are – and they do not take into account the inter-correlations between the variables in the model. In addition, as Zavgren (1985) points out, we have to keep in mind that the attempt to assess the meaning of the individual coefficients is inappropriate in view of the purpose of the technique of MDA. In contrast, Scott (1978) argues that, if the requisite assumptions of the MDA model concerning collinearity are met, the standardized coefficients can be used to evaluate the importance of the individual variables. Also Blum (1974) makes conclusions about the relative importance of the variables by comparing the rankings of variables by relative size of the standardized coefficients. He suggests that these rankings "may yield an approximation of relative importance (p. 10)". Eisenbeis (1977) and Joy & Tollefson (1975) mention some possible methods proposed in the literature which attempt to assess the relative importance of the independent variables.

### 2.4 Conditional probability models

After the period in which MDA were clearly dominant, this method has been replaced by less demanding statistical techniques such as ***logit analysis*** (LA), ***probit analysis*** (PA) and ***linear probability modelling*** (LPM). These methods result in 'conditional probability models' (Zavgren, 1983; Zavgren, 1985; Doumpos & Zopoudinis, 1999), consisting of a combination of variables, which distinguish best between the group of failing and the group of non-failing firms. *Ohlson (1980)* pioneered in using logit analysis on financial ratios in order to predict company failure, while *Zmijewski (1984)* was the pioneer in applying probit analysis (PA). Until now, LA has appeared to be a very popular method in failure prediction. The number of studies using PA is much smaller, probably because this technique requires more computations (Gloubos & Grammatikos, 1988; Dimitras et al., 1996).

***Conditional probability models*** allow to estimate the probability of company failure conditional on a range of firm characteristics by a non-linear maximum likelihood estimation. The models are based on a certain assumption concerning the probability distribution. The logit models assume a logistic distribution (Maddala, 1977; Hosmer & Lemeshow, 1989), while the probit models assume a cumulative normal distribution (Theil, 1971). In the linear probability models, the relationship between the variables and the failure probability is assumed to be linear (Altman et al., 1981; Gloubos & Grammatikos, 1988). As logit analysis clearly is the most popular conditional probability method in corporate failure prediction literature, we will focus on this particular conditional probability technique and we will not further elaborate on LPM and PA.

In LA, a non-linear maximum likelihood estimation procedure[12] is used to obtain the estimates of the parameters of the following ***logit model*** (based on Hosmer & Lemeshow, 1989, p. 25 and Gujarati, 2003, p. 595-615):

$$P_1(\mathbf{X}_i) = 1 / [\, 1 + \exp -(B_o + B_1X_{i1} + B_2X_{i2} + \ldots + B_nX_{in})] = 1 / [\, 1 + \exp -(D_i)\,]$$

where $P_1 (\mathbf{X}_i)$ = probability of failure given the vector of attributes $\mathbf{X}$i;
$B_j$ = coefficient of attribute j with j = 1, …, n and $B_0$ = intercept
$X_{ij}$ = value of the attribute j (with j = 1, …, n) for firm i,
$D_i$ = the "logit" for firm i.

The LA model combines several characteristics or 'attributes' into a (multivariate) ***probability score*** for each company, which indicates the 'failure probability' or the 'vulnerability to failure'. The logistic function implies that the logit score (i.e. the probability of failure) $P_1$ has a value in [0,1] interval and is increasing in $D_i$. If $D_i$ approaches minus infinity, $P_1$ will be zero and if $D_i$ approaches plus infinity, $P_1$ will be one. In LA, failure probability $P_1$ follows the logistic distribution (see Laitinen & Kankaanpää, 1999, p. 70).

When the failed status is coded as one (zero), a high (low) logit score indicates a high failure probability and hence a poor financial health. In a ***classification context***, the essence of the LA model is to assign firms to the failing or the non-failing group based on their 'logit score' and a certain cut-off score for the model. In the case where failure is coded as one and a high logit score indicates a high failure probability, a firm is classified into the failing group if its logit score exceeds the cut-off point and it is classified into the non-failing group if its score is lower than or equals the cut-off point. Similar to MDA, the LA model is based on the 'resemblance' principle: firms are assigned to the group they most closely resemble.

Just like the MDA model, the ***classification accuracy*** or 'performance' of a LA model can be assessed on the basis of the type I and type II error rates, the percentage of correct classifications, the unweighted error rate, the Receiver Operating Curve (Steele, 2002), the trade-off function (Ooghe et al., 2003; Ooghe & Balcaen, 2002c), the Gini-coefficient (Ooghe et al., 2003; Ooghe & Balcaen, 2002c), $R^2$-type measures and measures based on entropy (Joos et al., 1998a).

When applying LA, no ***assumptions*** are made regarding the distribution of the independent variables – LA does not require multivariate normal distributed variables or equal dispersion matrices – nor concerning the prior probabilities of failure (Ohlson, 1980; Zavgren, 1983; Joos et al., 1998a). As LA does not require the restrictive assumptions of MDA and allows to work with disproportional samples, the LA method is commonly considered as 'less demanding' than MDA. Nevertheless, LA is based on two assumptions. First, the LA method requires the dependent variable to be *dichotomous*, with the groups being discrete, non-overlapping and identifiable. Problems related to this assumption are discussed in section two. Second, the *cost of type I and type II error rates* should be considered in the selection of the optimal cut-off probability. However, due to the subjectivity of the choice of these misclassification costs in practice (Steele, 1995), most researchers minimize the total error rate and, hence, implicitly assume equal costs of type I and type II errors (Ohlson, 1980; Zavgren, 1985; Koh, 1992; Hsieh, 1993). Ohlson (1980) is one of the few researchers who explicitly acknowledged the impact of the choice of the error costs on the corresponding error rates. He reports the error rates of his model for different cut-off points associated with different error costs. Similarly, Ooghe et al. (1993) report a table of several possible cut-off points (for any error cost ratio) and the corresponding performance results of the model and Ooghe et al. (1994a) report percentile tables with possible cut-off points for failing and non-failing companies. This allows the external user of the model to assess the performance of the model for any combination of error costs. On the other hand Koh (1992) showed that in his LA model, the choice of the optimal cut-off point is rather insensitive and hence robust to different specifications of misclassification costs.

---

[12] See McFadden (1973) for a comprehensive analysis of the logistic regression model.

In contrast with the former authors, he concludes that failure models should be applicable in a wide range of situations and the non-consideration of error costs should not be a serious problem.

Besides the fact that logit analysis has no assumptions concerning the distribution of the independent variables and the prior probabilities of failure, there are some other important ***advantages*** of LA. First, the output of the LA model, the logit score, is a score between zero and one, which immediately gives the '*failure probability*' of the company (Ohlson, 1980; Ooghe et al., 1993). Second, the estimated *coefficients* in a LA model can be interpreted separately as the importance or significance of each of the independent variables in the explanation of the estimated failure probability (Ohlson, 1980; Mensah, 1984; Zavgren, 1985), provided that there is no multi-collinearity among the variables. Third, LA models allow for *qualitative variables* with categories rather than continuous data. In this case, dummies are used (Ohlson, 1980; Keasey & Watson, 1987; Joos et al, 1998a). Finally, the *non-linear* shape of the logit function is appealing. The underlying logistic function implies that, compared to a firm that has an average health, an extremely healthy (or weak) company must experience a proportionally larger deterioration (or amelioration) in their variables in order to deteriorate (or ameliorate) its financial health score (Laitinen & Kankaanpaa, 1999) .

Nevertheless, LA also has several serious ***drawbacks.*** Firstly, LA models are extremely sensitive to the problem of *multi-collinearity*. The inclusion of highly correlated variables must be avoided (Ooghe et al., 1993; Ooghe et al., 1994a; Joos et al., 1998a; Doumpos & Zopoudinis, 1999). However, as most LA models are (mainly) based on financial ratios, which in se are highly correlated because they often share the same numerator or denominator, the multi-collinearity problem may be severe (Tucker, 1996). Secondly, LA models are very sensitive to *outliers* (i.e. discordant observations) and *missing values*. Therefore, the dataset first needs to be corrected for possible outliers and missing values (Joos et al., 1998b). Finally, although logit models do not require the variables to be normally distributed, there is evidence that they do remain sensitive to *extreme non-normality* (McLeay & Omar, 2000). Therefore, prior to the estimation of the LA model, the data first need to be transformed or deleted (outlier deletion) in order to approximate or improve normality.

## 2.5 Overview

Table 1 gives an overview of the main advantages and drawbacks of the different classic statistical methods and reports a number of academically developed corporate failure prediction models.

---

Insert Table 1

---

# 3 PROBLEMS CONCERNING CLASSIC STATISTICAL METHODS

Although the classic statistical methods of MDA and LA are widely used in corporate failure prediction studies, there are a number of ***common problems*** related to the application of these techniques to the topic of corporate failure prediction. This section gives an extensive overview and discussion of the problems related to (1) the assumption of the dichotomous dependent variable, (2) the sampling method, (3) the stationarity assumption and data instability, (4) the selection of the independent variables, (5) the use of annual account information and (6) the time dimension.

## 3.1 Dichotomous dependent variable

### 3.1.1 The arbitrary separation of populations

As mentioned in the previous section, the classic statistical techniques of MDA and LA assume the dependent variable to be dichotomous (Cybinski, 2001). Consequently, if these methods are applied to the topic of corporate failure prediction, the populations of failing and non-failing firms are assumed to be well defined and clearly separated from each other. However, in reality, corporate failure is **not a well-defined dichotomy**. There is no clear external criterion for the class labels. The populations of failing and non-failing firms do not seem to be clearly separated. Therefore, the use of a dichotomous dependent variable is in contrast with reality. Some researchers argue that it is only possible to construct a population of clearly failing companies and clearly non-failing firms and a population of 'grey zone' companies, for which the situation is unclear. Another possibility is to construct populations associated with multiple outcomes, such as failure, acquisition and non-failure, as in the study of Astebro & Winter (2001). In this context, the use of the classic statistical models does not seem to be suited for corporate failure prediction.

However, when constructing failure prediction models with a classic statistical technique, researchers **arbitrarily separate companies into a failing and a non-failing population** (arbitrary class definition). A first arbitrary factor concerns the definition of failure and a second arbitrary factor concerns the way in which this definition of failure is applied in order to separate the total population of companies into a failing and a non-failing population. As a result, the basic assumption of the dichotomous dependent variable is violated and the modelling techniques are applied inappropriately. The remainder of this point extensively comments on the problems related to the arbitrary construction of the two populations.

In corporate failure prediction studies, the population of failing and non-failing firms entirely depends on the researcher's choice of the ***definition of a 'failing company'***. The terms of 'bankruptcy', 'failure', '(cash) insolvency', 'liquidation' and '(loan) default' are commonly used and sometimes refer to the same failure concept. An overview of the meaning of these terms can be found in Altman (1993) and in Argenti (1976).

It is clear that most corporate failure studies are based on the *'legal definition'* of failure (a.o. Dirickx & Van Landeghem, 1994; Ward & Foster, 1997; Van Caillie, 1999; Charitou et al., 2004; Daubie & Meskens, 2002). In this context, a company is considered to be failing if it is characterized by a certain 'failing' legal situation. In most cases, the legal situation of 'bankruptcy' is used (Ooghe & Van Wymeersch, 2003). The popularity of the legal definition of failure (i.e. 'legal business failure') can be explained by the fact that it offers some important advantages. First, the moment of failure can be objectively dated. In the great majority of the studies, the change in the juridical situation is taken as the moment of failure. In addition, the legal definition of failure provides an objective criterion that allows the researcher to easily separate all companies into two populations (Ooghe & Joos, 1990; Ooghe et al, 1993; Ooghe et al., 1995; Dirickx & Van Landeghem, 1994; Charitou et al., 2004). It should however be noted that the legal definition of failure depends on the country in which the failure prediction model is developed and the corresponding legislation concerning company failure (see for example, Franks et al., 1996).

When comparing the legal definitions of failure in studies from different countries, it is clear that each country has its specific 'common' definition, according to the legal framework on company failure[13]. Karels & Prakash (1987) give an extensive overview of several legal definitions of failure used by researchers in empirical 'bankruptcy' studies.

Although the legal definition of failure seems to be widely accepted, it may also cause some problems. A first problem is that the moment of legal failure often does not reflect the 'real' failure event. It is possible that there is a great time leap between the real moment of failure and the moment of change in legal situation (for example, the declaration of bankruptcy) (Ooghe et al., 1995; Pompe & Bilderbeek, 2000). For example, there is often a great time leap between the deposit of the last published annual account, which can be considered as the "real" moment of failure, and the moment of bankruptcy. Theodossiou (1993) states that "in practice […] firms [in the US] stop reporting about two years before filing for bankruptcy (p. 442)". In this respect, the legal decision concerning failure can be seen as a 'subjective' decision. Secondly, it is possible a firm, showing many of the characteristics of a failing company, does not show a change in its legal situation (for example, the firm is not declared bankrupt). Finally, instead of showing a 'failing' legal situation (such as bankruptcy), troubled companies may also merge with an other firm or a reorganize (Daubie & Meskens, 2002).

Due to these problems related to the use of the legal failure definition, some researchers have indicated that the legal definition of failure is too narrow and have suggested to study '*financial distress*' (Keasey & Watson, 1991; Hill et al., 1996; Kahya & Theodossiou, 1996; Doumpos & Zopoudinis, 1999; Platt & Platt, 2002). They argue that it is more interesting to study financial distress or 'economic business failure', because (1) many financially distressed firms never file for a bankruptcy and because (2) stable and financially healthy (non-distressed) firms may file for bankruptcy for specific (strategic) reasons, that are not related to financial distress. If the former two situations occur frequently, using a legal definition of failure could result in contaminated samples of failing and non-failing firms[14] and, finally, in poor forecasting abilities of the developed failure prediction model. However, the use of a financial distress definition also suffers from a serious drawback: there is a lack of

[13] For example, in most UK studies, a company is considered as failing if it is characterized by a Liquidation, an Administration or a Receivership (Charitou et al., 2004). In USA studies, a company is considered as failing if it filed for bankruptcy under 'Chapter 7' (a liquidation bankruptcy) or 'Chapter 11' (a reorganization). In Belgium, the failing population mostly consists of companies that are declared bankrupt or filed for a judicial composition. In France, a company is attributed to the failing population if it shows a failing juridical situation of 'défaillance' or 'procédure collective'.

[14] Taffler (1982, 1983) recognizes the problem of 'contaminated' non-failing samples and argues that the very unhealthy firms (on the basis of financial analysis techniques) should be excluded from the non-failing sample: the sample of non-failing firms should only include non-distressed firms.

a consistent definition and the criterion of financial distress has to be chosen arbitrarily. Platt & Platt (2002) define a financially distressed firm as a firm which reports some of the following indicators: (1) several years of negative net operating income, (2) suspension of dividend payments or (3) major restructuring or layoffs. Other possibilities of defining financial distress are: making losses and selling shares to private investors, entering into a capital restructuring or a reorganization and experiencing a couple of years of negative shareholder's funds or accumulated losses (McLeay & Omar, 2000). Keasey & Watson (1991) mention that the criterion of financial distress is incomplete and arbitrary in nature and they conclude that "there may be a need to develop specific models for different types of financial distress (p. 93)"[15].

Besides bankruptcy and financial distress, several other *economical definitions of failure* are used in corporate failure prediction studies. A first example is 'cash insolvency', which means that the firm is unable to pay its financial obligations when the payments become due (Laitinen, 1994). This definition of failure is closely related to the Finnish juridical process of 'liquidity bankruptcy' (Laitinen & Kankaanpää, 1999). Another example is 'loan default'. Ward & Foster (1997) argue that loan default is a better way of defining failure, because it is an economically defined event, as opposed to bankruptcy, which is a legally defined event. They suggest that the loan default definition is more consistent with the economic reality. However, it is clear that this failure definition implicitly limits the failure prediction study to a context of credit or loan problems. Other failure definitions are based on events such as capital reconstructions, major closures or forced disposals of large parts of the firm, informal Government support and loan covenant renegotiations for solvency reasons with bankers (Taffler & Agarwal, 2003). Finally, in the light of the new framework of Basel II, some *default events* are explicitly defined, such as credit loss associated with any delay in payment of more than 90 days or with a distressed restructuring involving the forgiveness or postponement of principal amounts or interests by financial institutions (Hayden, 2003). It is clear that these default events occur more frequently than bankruptcy. However, these definitions of failure are purely credit-oriented and hence are not appropriate when analysing failure in a business context (i.e. corporate failure) instead of in a credit context.

It is clear that the definition of corporate failure should be carefully chosen, because there is no point in building a very accurate model to predict classes which are different from those of real interest.

[15] These kind of multi-state prediction models may contribute towards the satisfaction of 'user needs'.

When applying the chosen definition of failure in order to create a failing and a non-failing sample of companies, a second arbitrary factor enters the construction of the two populations. It is clear that when the failure definition is applied to a certain ***arbitrarily chosen year or time period,*** the separation of companies into a failing and a non-failing population is artificial. The two populations will only be mutually exclusive within the chosen time period (Ooghe & Verbaere, 1985; Ooghe & Joos, 1990; Ooghe et al., 1995; Altman & Narayanan, 1997). Moreover, the application of the failure definition to an arbitrarily chosen year or time period involves a certain 'selection bias' (Shumway, 1999) and may result in 'contaminated' populations. For example, a company that meets the legal definition of failure one year after the considered time frame will be included in the population of non-failing companies, although it shows many of the characteristics of the group of failing companies. This problem of selection bias may be solved by using an 'extended time frame' for the construction of the populations. For instance, one could ensure that the non-failing population only consists of companies with non-failing characteristics by only including those companies that do not meet the definition of failure up to five years after the considered time period. For example, Ooghe et al. (1993) use this 'extended time frame' technique in order to account for the selection bias. On the contrary, Back et al. (1997) recognize the existence of a selection bias, but they argue that it is better to included all types of non-failing companies, even those with many failing characteristics.

### 3.1.2 Other comments on the construction of populations

It should be noted here that the way in which the two populations (failing and non-failing) are separated is of major importance for the further development of the failure prediction model. If the selection of discriminating variables to be included in a failure prediction model is done empirically (i.e. from a large set of variables used as initial input of the model), the definition of failure will ***influence the selection of variables***. For example, if a legal failure definition is used, it may be expected that the role of solvency and liquidity variables will be important, bacause these two dimensions are explicitly integrated in most legislations concerning bankruptcy[16] (Blazy, 2000; Van Caillie & Dighaye, 2002). Similarly, if the criteria that are used in the financial distress definition or the economical failure definition are correlated with some of the variables in the initial battery of variables offered to the model, it is very likely that these variables will be selected as most discriminative (Ooghe et al., 1995). However, contrary to expectations, Hayden (2003) found that three different models developed for three different definitions of failure (bankruptcy, delay in payment and loan restructuring) have very similar structures regarding the selected variables.

A final important remark concerns the fact that most corporate failure prediction models are estimated on *two clearly separated populations* of companies: a group of risky, failing firms and a group of non-risky, surviving firms. As they all try to distinguish between two groups of firms that are already well separated in multidimensional space (Cybinski, 2000; Cybinski, 2001), we may argue that their good classification performances are not surprising. In this respect, Wood and Piesse (1987) point out that the ex post discrimination between risky companies that have failed and non-risky firms that have not failed is "not surprising" (p. 29). Moreover, we may argue that the reported 'predictive' accuracy or *'reliability'* of these models is *misleading*. Therefore, the real informative value of many failure prediction models should be questioned. In addition, the accuracy of the models should be tested on a sample consisting of different kinds of firms, especially 'grey zone' firms, which are not clearly failing or clearly non-failing. If the objective is to identify likely failures from a pool of problem companies, the existing failure prediction models, which are estimated on a sample of clearly failing and non-failing firms, may perform poorly.

### 3.2 Sampling method

It is clear that, if a failure prediction model is eventually to be used in a predictive context, the samples of failing and non-failing firms used for estimation of the model should be representative for the whole population of firms (Ooghe & Joos, 1990). There is no point in building a highly accurate model for an available sample that is not representative. Moreover, the classical paradigm[17] is based on the assumption that a random sampling design is used. The firms in the estimation sample and new, future samples of cases, for which a failure prediction is to be made, are assumed to come from the same distribution. Nevertheless, in the great majority of the classic statistical failure prediction models, the estimation of the models is based on **non-random samples**, whose compositions are different from the population's composition. Examples of the endless list of studies based on non-random samples are: Altman (1968), Deakin (1972), Blum (1974), Altman et al. (1977), Taffler & Tisshaw (1977), Van Frederikslust (1978), Dambolena & Khoury (1980), Ohlson (1980), Zavgren (1982),

[16] In the USA, on the contrary, the legislation on bankruptcy does not mention any requirements concerning insolvency (Warren & Westbrook, 1999).

Chalos (1985), Gentry et al. (1985a), Keasey & Watson (1987), Aziz et al. (1988), Gloubos & Grammitokos (1988), Keasey & Mc Guinness (1990), Mossman et al. (1998) and Altman et al. (1995). If the estimation samples are non-random, it might be expected that the parameter estimates of the models and the estimated failure probabilities are biased (Zmijewski, 1984) and that the overall classification results are affected. Piesse & Wood (1992) point out that, when a failure prediction model is based on non-random samples, the accuracy results of the model can not be generalized. They stress that the reported classification accuracy of the model is the 'ex-post' accuracy, which may be very different from the 'ex-ante' performance of the model in a predictive context. In other words: the reported accuracy of the model may be misleading. Zmijewski (1984), on the contrary, finds that the use of non-random samples does not significantly affect the overall accuracy rates. Moreover, they conclude that the statistical inferences on the impact of the independent variables are not affected. Only the individual group classification (type I and type II errors) and estimated probabilities seem to be influenced by the use of non-random samples.

If a failure prediction model is estimated on samples that are non-random with respect to certain general characteristics, such as *industry, size class and age* – certain industries, size classes or firm ages are under-represented – the model may be inefficient in a predictive context, when used on those types of companies that are under-represented in the estimation samples. This is probably one of the reasons why Pompe & Bilderbeek (2000) find that their model has a poor performance when tested on starting companies and large companies. It has to be noted here that this drawback of using non-random samples may be strongly reduced by building industry specific, size class specific and age specific models[18].

Non-random samples may be the result of (1) over-sampling the failing companies, in case of 'state-based' sampling, of (2) applying a 'complete data' sample selection criterion or of (3) using matched pairs of failing and non-failing firms. Zmijewski (1984) conducted a study concerning the existence the first two types of biases caused by using non-random sampling and found significant evidence.

First of all, because of the low frequency rate of failing companies in the economy, most researchers draw a 'state-based' sample (i.e. the selection of firms in the sample is based on the known survival outcome of firms) and thereby ***'over-sample' the failing companies***. However, as most estimation techniques are based on the assumption of random sampling,

[17] The classic paradigm is that given a set of firms with known descriptor variables and known outcome class membership, a rule is constructed which allows other companies to be assigned to an outcome class on the basis of their descriptor variables.
[18] When developing these kind of specific models, one needs to be sure that the estimation sample corresponds to the 'user needs'.

over-sampling the failing companies may result in a *'choice-based' sample bias* (Zmijewski, 1984; Platt & Platt, 2002). The over-sampled group of failing companies will show understated (low) misclassification error rates, because the model will pay more attention to accurately classifying the failing companies, at the expense of a higher probability of misclassifying non-failing firms[19]. Finally, over-sampling the failing companies will result in an overstatement of the (ex-post) accuracy of the model. In an ex-ante context, many failure signals will be given to survivor companies (Zavgren, 1983; Zmijewski, 1984; Piesse & Wood, 1992; Platt & Platt, 2002). Platt & Platt (2002) found empirical evidence for the choice-based sample bias and warns that, in many studies, the reported percentages of correct classifications of failing firms are misleading.

Secondly, as missing data often appear, many researchers (for example, Taffler (1982), Ooghe & Verbaere (1985) and Declerc et al. (1991)) use a ***'complete data' sample selection criterion***. This selection criterion may lead to a *'sample selection' bias*. If, for example, the failing companies are more likely to have incomplete data, estimating the model based on a 'complete data' sample selection criterion will cause the probability of failure to be understated (Zmijewski, 1984) and will lead to large misclassification errors for the failing firms. As, in practice, this supposition is very reasonable – failing firms tend to be younger and smaller compared to the population of non-failing firms and hence more likely to have incomplete data – we might expect that many studies show a 'sample selection' bias.

Thirdly, in the majority of studies, researchers use ***matched samples*** of failed and non-failed companies. For each company in the failed sample, a similar paired non-failed company is selected or some multiple (Ohlson, 1980; Platt & Platt, 2002). In this respect, Scott (1981) states that "most bankruptcy models are derived using a paired-sample technique. Part of the sample contains data from firms that eventually failed; the other part contains contemporaneous data from firms that did not fail (p. 320)." Examples of matched sampling can be found in the studies of Altman (1968), Blum (1974), Taffler & Tisshaw (1977), van Frederikslust (1978), Bildereek (1979), Zavgren (1983), Zavgren (1985), Keasey & McGuinness (1990), Platt & Platt (1990), Mossman et al. (1998), Charitou et al. (2004) and Charitou & Trigeorgis (2002). Usually, pairing is done on criteria of size, industry and age. Matching is a common practice, because it enables the researcher to control for some variables which are believed to have some predictive power but which are not included in the set of prediction variables (Zavgren, 1983; Ooghe & Verbaere, 1985; Keasey & Watson,

[19] It should be remarked that this may be considered as beneficial as the cost of a type I error (misclassification of a failing firm) is usually considered to be much larger than the cost of a type II error (misclassification of a non-failing firm).

1991; Dirickx & Van Landeghem, 1994). Jones (1987) explains the purpose of matching as follows: “Bankrupt firms are often disproportionally small and concentrated in certain failing industries. If non-bankrupt firms were drawn at random, there would probably be substantial differences between the two groups in terms of size and industry. The result is that the model attempting to discriminate between failing and healthy firms may actually be distinguishing between large and small firms or between different industries.” However, using matched pairs also has some serious drawbacks. If the sample of non-failing firms is constructed on the basis of characteristics of the failed sample, it is most likely that this sample will *not* be *representative* for the whole population of non-failing companies (Ooghe & Verbaere, 1985; Ooghe et al., 1993; Ooghe et al., 1995). It is very likely that some characteristics are over-represented or under-represented in the matched samples and this may result in a sample specific failure prediction model (Zavgren, 1983). Zmijewski (1984) shows that, under certain circumstances, the use of matched samples that differ significantly from the population proportions leads to biased coefficients in logit models[20] and Keasey & Watson (1991) warn that the matched samples may cause misleading indications of the model’s predictive accuracy. An additional problem rises with respect to the choice of matching criteria (Ooghe et al., 1993; Ooghe & Verbaere, 1985; Ohlson, 1980; Peel & Peel, 1987). Matching criteria are often chosen ad hoc (i.e. arbitrarily) and, if these criteria show any link with the failure probability, this may lead to a *selection bias*. The size criterion, for example, may incur some problems, as the size variable could itself be a significant discriminating variable (smaller firms are often more prone to failure than larger firms). As the predictive power of the matching variables are eliminated, this will result in a restricted (instead of a general) model of company failure (Taffler, 1982). This is why Eisenbeis (1977) states that “every effort should be made to avoid arbitrary grouping (p. 889)”. A final drawback of matched sampling is that, in practice, multivariate matching on the basis of several criteria is *difficult to perform* (Ooghe & Verbaere, 1985).

## 3.3 Non-stationarity and data instability

### 3.3.1 Non-stationarity problems and the predictive context

Using an MDA model or a conditional probability model in a **predictive context** requires that the relationships among the variables are stable over time and that the relationships in future samples of companies, which are to be classified by the model, are the

[20] On the other hand, Zmijewski (1984) concludes that the overall classification accuracy rates are not affected.

same as in the estimation samples of the model. First of all, this implies that the relationship between the independent variables and the dependent variable (the model score) are stable over time (Edmister, 1972; Zavgren, 1983). This is called the *‘stationarity assumption’* (Mensah, 1984; Jones, 1987). Secondly, if the independent variables in an MDA model are correlated, the inter-correlations between the independent variables should be stable and should be likely to be repeated in other samples (Edmister, 1972; Zavgren, 1983). It is clear that the problem of *‘non-stationarity’* is closely related to *‘data instability’*, which means that values of the independent variables – the mean structure – differ markedly between the estimation period and the forecast period (Mensah, 1984; Wood & Piesse, 1987). The classic paradigm assumes that the distributions of the variables do not change over time.

In the literature, ample evidence of ***data instability*** (also called ‘population drift) or ***non-stationarity*** can be found. For example, Barnes (1987) suggests that there is evidence that the relationships between financial ratios are unstable over time. For example, financial ratios may be sensitive to the use of alternative accounting methods over time. Also, Richardson & Davidson (1984) conclude that accounting ratios are unstable over time. In this respect, Mensah (1984) suggests that data instability may be due to changes in inflation, interest rates and/or phases of the business cycle. Similarly, Wood & Piesse (1987) point out that data instability may be attributable to, for example, phases of the business cycle, changes in the competitive nature of the market, changes in corporate strategy and technological changes. Dambolena & Khoury (1980) suggest that the data instability problem of financial ratios is the greatest for the firms that are about to fail. Consequently, many classic statistical models suffer from so-called ‘stationarity problems’ (Moyer, 1977; Mensah, 1984; Charitou et al., 2004).

In corporate failure prediction models, non-stationarity and data instability may have severe consequences.

Firstly, it may result in *poor predictive abilities* of the model: the model can not be used in samples from subsequent periods in time, unless with a great loss of performance. In other words, when the relations among the variables are not stable over time, a model may show poor predictive out-of-sample results when it is applied to a new sample of firms (forecast period), even if the model has good (ex post) classification results on the estimation sample (estimation period) (Mensah, 1984). In this context, Moyer (1977) and Joy and Tollefson (1975) suggested that ‘the proof is in the eating’, which means that, before one can really have confidence in the predictive abilities of a failure prediction model, the model needs to be tested on data subsequent to its construction. Also, Taffler (1983, 1984) stressed the importance of testing the efficiency of failure models on a new, future-dated sample. For example, Moyer (1977) analysed the classification performance of Altman's model on a sample of company data from the period 1965-1975 and found that the accuracy is much lower than the accuracy reported in the original Altman study, which is based on data from the period 1946-1965. Pompe & Bilderbeek (2000) point out that these lower 'predictive' performance results particularly show up in periods of a downward evolution in the economy, when the overall failure rate in the economy is higher[21].

Secondly, data instability implies that the models are fundamentally *unstable* or *not robust over time*. When classic statistical failure models are re-estimated on new, more recent data, the estimated coefficients generally appear to change. Consequently, corporate failure prediction models are fundamentally unstable and may need redevelopment from time to time (Joy & Tollefson, 1975; Taffler, 1982; Mensah, 1984). Therefore, Keasey & Watson (1991) argue that 'old' models, which are estimated in a time period far before the period over which predictions are to be made, are not useful. In this context, it seems appropriate to monitor and to test the performance of failure prediction models at regular time intervals (i.e. inter-temporal validation) and to update the models, if necessary (Dirickx & Van Landeghem, 1994; Ooghe et al., 1994a). Updating a model may involve the estimation of new coefficients, but it may also be limited to the calculation of new cut-off points.

In view of overcoming the problems related to data instability, a couple of researchers have searched for *remedies*. Dambolena & Khoury (1980) and Betts & Belhoul (1987) measured the stability of the financial ratios by their variation and included these (in)stability measures in their MDA model. In order to attenuate the data instability problem, Platt & Platt (1990) used industry-relative ratios in their failure prediction model. They contended that the use of industry-relative ratios leads to more stable financial ratios across estimation and forecast time periods. They find that the use of industry-relative ratios can help to stabilize forecasts based upon a multi-industry corporate failure prediction model. Platt & Platt (1991), on the other hand, studied the use of industry-relative ratios[22] and found that there is no significant difference between industry-relative models and unadjusted models with respect to the stability of the estimated coefficients (both have stable coefficients). Another type of remedy is investigated by Mensah (1983). He indicated that the use of deflated data (i.e. the

---

[21] This conclusion is in line with the above suggested relationship between data instability and changes in the business cycle.

[22] An industry-relative ratio is created by relating the value of the ratio for a company at time t to the value of the ratio for the average firm in its industry at time t. Industry-relative ratio for firm i in industry j = ratio for firm i / [ (mean ratio in industry j) * 100 ] (Platt & Platt, 1990).

‘real values’ of financial ratios) instead of unadjusted data may increase the predictive performances of failure prediction models.

### 3.3.2 Non-stationarity problems and the use of pooled samples

Besides the use of failure prediction models in a predictive context, also the general practice of **‘pooling’ of data across different years** requires that the relationships among the variables are stable over time. In the great majority of studies on corporate failure classification models, the estimation sample of failing companies is a *‘pooled sample’* and consists of companies that are failing in different years. Although it is obvious that the pooled data are subject to distinct macro-economic conditions, the resulting failure models do not consider the underlying economic events (Zmijewski, 1984; Mensah, 1984). They implicitly assume that the relations between the variables in the model are stable over time (Altman & Eisenbeis, 1978; Zmijewski, 1984). In other words: they assume *stationarity and data stability*. For example, studying a period of 20 years – as in the Altman study of 1968 – one implicitly assumes stationarity and data stability and neglects the fact that the average value of the financial ratios may be changing over time. However, as mentioned before, the stationarity assumption is very likely to be violated in practice (Zmijewski, 1984) and there is a lot of evidence for data instability. As a consequence, failure prediction models estimated on pooled data may be based on ‘temporarily distorted’ data and this may result in inconsistent coefficient estimates (Platt et al., 1994) and a low accuracy level (Back et al., 1997). Mensah (1984) examined the construction of a failure prediction model based on a pooled sample and models based on several smaller samples with homogeneous economic conditions and concluded that (1) the models are very unstable and that (2) the accuracy of the models differs across the economic environments.

Despite the stringent assumptions concerning pooled samples, it might be argued that pooling of data across different years is necessary to increase the ‘representativeness’ of a failure prediction model. If a model would be based on data from a very limited time period[23], the model may show poor predictive results when used on data from other time periods (Ooghe & Joos, 1990).

[23] One example of the small number of models that are based on a very short selection period is the Ooghe-Verbaere model (Ooghe & Verbaere, 1982).

## 3.4 Annual account information

### 3.4.1 The use of financial ratios

The majority of the classic cross-sectional models only use annual account information in the form of **financial (accrual and cash flow) ratios** in order to predict failure (Dimitras et al., 1996). The reasons for using financial ratios are that (1) they are 'hard', objective measures and (2) they are based on publicly available information (Micha, 1984; Laitinen, 1992; Dirickx & Van Landeghem, 1994). Nevertheless, financial ratios have been the subject of many criticisms and corporate failure prediction models that are limited to annual account information have proven to suffer from some serious drawbacks. All possible disadvantages related to the use of financial ratios are discussed below. We would like to emphasize here that, despite the criticisms, the importance of financial ratios and their corresponding meaning may not be neglected!

First of all, as the obligation to deposit and/or publish annual accounts mostly depends on certain criteria on ***firm type and/or firm size,*** failure prediction models that use financial ratios are restricted to data from those companies that meet the criteria. In many countries (for example, USA, UK, Germany,…) only 'large' firms are obliged to publish their annual accounts and, as a result, many studies on failure prediction models have been restricted to large companies, meeting certain criteria concerning the asset size and/or the sales level and/or the number of employees.

Secondly, when constructing failure predictions on the basis of financial ratios, researchers implicitly assume that the annual accounts give a ***fair and true view*** of the financial situation of companies. However, it is clear that many annual accounts are unreliable and do not give a fair and true view.

There is much anecdotal and academic evidence on the fact that firms generally have incentives to manage their earnings and to manipulate their annual accounts (Ooghe & Joos, 1990; Ooghe et al. 1995). High levels of intentional *earnings manipulation* may be referred to as 'fraud'. Firms generally try to maintain positive earnings and avoid reporting earnings decreases (Degeorge et al. 1999). For example, Burgstahler & Dichev (1997) and Degeorge et al. (1999) found strong evidence that firms generally manage earnings (by means of cash flow from operations and, to a lesser extent, changes in working capital) to avoid small income decreases and small losses. We may expect that, especially in failing firms, the annual accounts do not give a fair and true view of the companies' financial situation and are subject of 'creative accounting practices'. It is generally believed that failing firms manage their earnings upwards and give a more positive presentation of the financial situation, especially when the moment of failure is very near (Argenti, 1976; Ooghe & Joos, 1990; Ooghe et al. 1995; Rosner, 2003). In this context, creative accounting is used as a 'defence mechanism'. Charitou

& Lambertides (2003) state: “Extant theories on troubled firms with persistent earnings problems predict that managers’ accounting choices are expected to be income increasing (p. 1)” and Rosner (2003) points out that “prior literature […] suggests that failing firms […] may be motivated to engage in fraudulent financial reporting (“fraud”) to conceal their distress (p. 366)”. Sweeney (1994) and DeFond & Jiambalvo (1994) and Rosner (2003) found empirical evidence for income-increasing earnings management in failing firms. On the contrary, some authors – Charitou & Lambertides (2003) and De Angelo et al. (1994) – have provided evidence for income-decreasing earnings management in failing firms. Using unreliable financial ratios from manipulated annual accounts possibly leads to a significant ‘*inconsistency problem’* for many financial ratios used in failure prediction models (Tucker, 1996). For example, Joos & Ooghe (1993) point out that Belgian financial statement information is often inconsistent and incomparable. The study of Theunissen (1999) is the only study that extensively analysed the effect of financial decisions concerning the revaluation of fixed assets, the booking of provisions, the activation of establishment costs, the valuation method of stocks and the depreciation method on the results of some corporate scoring systems. Theunissen concluded that (1) ‘normal levels’ of accounting cosmetics do not have an enormous effect on the scoring of the firm by classic statistical models, (2) creative accounting involving several measures may have larger effects and (3) the effects depend on the statistical scoring model that is used.

Besides annual account manipulation, the *lack of an internal control system* is an other possible source of unreliability. The annual accounts of smaller companies generally are unreliable because of a lack of an internal *control* system (Keasey & Watson, 1986; 1987).

In addition, the annual account most close to the moment of failure is very likely to be unreliable, because, if this last annual account has become published only after the firm filed for bankruptcy, it may contain *adjustments made by the auditor* in the light of bankruptcy filing. The use of financial ratios from such an ‘accommodated’ annual account in the estimation process of a model for failure prediction one year prior to failure may result in a distorted model (Charitou & Lambertides, 2003). This problem may, however, be solved by using the previous annual account instead of the ‘accommodated’ last annual account.

A third problem related to the use of annual account data is the occurrence of ***missing values***. In Belgium, for example, the data for ‘overdue short-term priority debts’ often take the value of zero (Ooghe & Joos, 1990). In order to overcome the problem of missing values, one may delete the cases with missing values or fill in the missing values with mean or random values (Tucker, 1996).

Fourthly, corporate failure prediction models based on annual account information may be biased by ***extreme ratio values***. When using the original values of financial ratios, the model may be ‘contaminated’ by extreme values (Moses & Liao, 1987) and the presence of extreme values for some ratios may bias the coefficients for these ratios in the model. Still, the problem of extreme ratio values may be partly reduced by trimming the ratios at certain percentiles (for example, the $10^{th}$ and the $90^{th}$ percentile).

Fifthly, as the annual accounts may contain ***errors***, failure models based on financial ratios may be the result of 'erroneous' information. In Belgium, several studies on the formal quality of the Belgian annual accounts (see Dirickx & Van Landeghem, 1994, p. 453) have shown that the quality of many annual accounts – especially the abbreviated form annual accounts deposited by small firms – is rather poor (Ooghe & Joos, 1990). This urges researchers to carefully check the quality of the annual accounts before using the data in the development of a corporate failure prediction model. Ooghe et al. (1993) for example, have detected and corrected the errors in their samples of annual accounts before using the data. It should be noted here that, as some errors may result in extreme financial ratio values, trimming the ratios may partly reduce the effect of errors.

Furthermore, if researchers only include financial ratios into their failure prediction model, they implicitly assume that all relevant failure or success indicators – both internal and external – are reflected in the annual accounts. However, it is clear that ***not all relevant information*** is reflected in the balance sheet and the income statement. In this context, Argenti (1976) states that "while these [financial] ratios may show that there is something wrong … I doubt whether one would dare to predict collapse or failure on the evidence of these ratios alone (p. 138)" and Zavgren (1985) point out that "any econometric model containing only financial statement information will not predict with certainty the failure or nonfailure of a firm (p. 22-23)" Also, Maltz et al. (2003) mention that the use of financial measures as sole indicators of organizational performance is limited[24].

For this reason, some authors have advised to include *non-accounting* or *qualitative failure indicators* in failure prediction models (Ohlson, 1980; Zavgren, 1983; Keasey & Watson, 1987; Lussier, 1994; Sheppard, 1994; Slowinski & Zopoudinis, 1995; Lussier, 1995; Lehmann, 2003; Zopoudinis & Doumpos, 1999; Daubie & Meskens, 2002; Becchetti & Sierra, 2003). We assert that non-financial and qualitative information might be particularly appropriate when studying failure of small companies[25], as it might be expected that the

[24] In their study, Maltz et al. (2003) managed to incorporate many different kinds of information into their new performance measurement model, the Dynamic Multi-dimensional Performance framework (DMP). In this DMP model, five success dimensions (financial performance dimension, market/customer dimension, process dimension, people development dimension and future dimension) are related to some baseline success measures and some additional success measures are suggested, dependent on the firm's size, its technology type, its strategy and the particular industry and environment in which the firm operates.

[25] In this respect, I would like to refer to the study of Hall (1994), who analysed the impact of several non-accounting variables concerning the background of the company, the motivations of the owner, characteristics of the owners, strategy, financial management, relationships with banks, pricing policy, marketing and quality of workforce on a small firm's ability to survive.

annual account information of these kind of firms is less reliable. Examples of possible non-accounting and qualitative indicators are: staffing, which involves the attraction and retention of quality employees (Lussier, 1994; Hall, 1994), management experience (Lussier, 1994; Lehmann, 2003), the education of the owner/manager (Lussier, 1994; Lehmann, 2003), the age of the owner/manager (Lussier, 1994), the motivation of the owner (Hall, 1994), social skills and leadership quality (Lehmann, 2003), the quality of management information systems which allow for timely information about financial and operational risks (Lehmann, 2003), the number of partners (Lussier, 1994), the existence of a plausible long-term business strategy for the company (Lehmann, 2003), the productive efficiency (Becchetti & Sierra, 2003), customer concentration (Lehmann, 2003; Becchetti & Sierra, 2003), dependence on one or a few large suppliers (Lehmann, 2003), subcontracting status (Becchetti & Sierra, 2003), export status (Becchetti & Sierra, 2003), the presence of large competitors in the same region (Becchetti & Sierra, 2003), the relationship with banks (Hall, 1994) and strategic variables, such as the level of diversification, the profitability of the industry, the industry growth rate, the market share and the number of joint ventures in which the firm is involved (Sheppard, 1994; Hall, 1994). Furthermore, characteristics (composition and structure) of the board of directors – for example, whether the CEO is the president of the board, the number of director interlocks and the percentage of insiders and outsiders – may explain why businesses fail (Elloumi & Gueyié, 2002; Sheppard, 1994). Other possible indicators can be found in a study of Lussier (1995), who built a failure prediction model with quantitative and qualitative managerial factors in order to predict failure of young businesses. In addition, as mentioned by Ooghe & Joos (1990), group-relations may be an important explanatory factor for business failure. More in particular, group structures may prevent a company from failure, although it shows weak financial ratios, or, in contrast, they may be the principal cause of company failure. This is the reason why Ooghe et al. (1993) included a group-relations variable into their long-term failure prediction model. Furthermore, as it is obvious that a firm's financial health at a certain point in time is influenced by the risk of failure of the previous period (Cybinski, 2001, 2000), a failure prediction model should also include information on past failure risk. Furthermore, information on the stock value may add significant information on the financial health of a firm and may be an indicator of impending failure[26]. Finally, certain external failure events reflecting management (corrective) actions, such as reduction in

[26] Several studies already have analysed the predictive power of market data (for example, the market rate of return or the market value), but the conclusions are unclear. Blum (1974) for example, finds that the stock market can not anticipate the timing of failure (market variables do not discriminate between failing and non-failing companies), while other authors suggest that it is likely that market data do provide additional information in a failure prediction context.

dividends, going concern qualified audit opinions, troubled debt restructurings, and violations of debt agreements may be good bankruptcy indicators (Flagg et al., 1991).

Besides non-accounting and qualitative variables, also general firm characteristics concerning *industry type* and the *size* have proven to be very important variables in failure prediction. For example, in a study of Laitinen (1992), the size variable appeared to be an important factor when predicting failure of newly founded firms. In addition, a lot of studies have shown that size has an influence on the likelihood of firm exit: small firms are more likely to exit than large firms (Bickerdyke, 1999). Consequently, small firms may be expected to be more likely to fail. Large firms are expected to have a lower failure probability because they (1) are more likely to benefit from scale-effects, (2) have more power in negotiations with their financial and social partners and (3) are more likely to benefit from their experience or 'learning effects' (Blazy, 2000). Hill et al. (1996), on the other hand, provided evidence that industry effects are important and have to be included in models that try to identify the impact of several variables on company failure. Furthermore, in a study of Ooghe et al. (2003), it is shown that the performance results of the Ooghe-Joos-De Vos failure prediction model clearly depends on the industry type and the size class of the companies they are used for. In order to capture the effect of industry and size-class, several authors have included industry information – in the form of industry variables, industry-dummies or industry-relative ratios – and variables concerning firm size into their failure prediction models (Astebro & Winter, 2001; Daubie & Meskens, 2002). On the other hand, a few authors try to account for the effects of size and/or industry by building size specific and/or industry specific models. For example, Taffler (1983), Mensah (1984) and Taffler (1984) advised to construct industry specific models and Bilderbeek (1978) suggested to make a distinction between large and small companies and to construct size-specific models. Also, Ooghe et al. (2003) advise to build industry specific and size specific models. Besides industry and firm size, also *company age* may play an important role. There is extensive empirical evidence of the influence of age on exit in general (Bickerdyke et al., 1999; Thornhill & Amit, 2002) and consequently, the impact of age on company failure has become a 'stilized fact': newly-founded firms are more likely to fail than older firms[27]. Therefore, it might be interesting to build age-specific models or to accounting for the impact of age by adding an age variable to a range of other variables in a failure prediction model. It should be noted here that Laitinen

[27] Thornhill & Amit (2002), on the contrary, argue that "despite the strong correlative evidence that age is a strong predictor of failure, age needs to be seen as a proxy for internal organizational processes that evolve over time". In this respect, newly-founded firms are more likely to fail than older firms, because they are more likely to face deficiencies in resources and capabilities.

(1992) does not share this opinion. He showed that a failure model for older firms may also be appropriate for failure prediction of newly founded firms, provided that the cut-off points are adjusted. However, as his study concerns only small companies, we may argue that his conclusion may not be generalized.

In addition, because a firm is never standing alone, a failure prediction model should consider information about the external environment. A first element is the *macro-economical situation*. As the empirical literature provides ample evidence on the relation between the business cycle and ‘exit rates’ (exit rates seem to increase during an economic downturn), the macro-economic environment may be a significant explanatory factor for company failure. Moreover, changes in the macro-economic environment may amplify the impact of other firm specific factors on failure risk. Factors that may affect a firm’s financial health are, for example, a rise in the interest rates, a recession and the availability of credit (Zavgren, 1983). It is true that the effects of external (macro-economic) variables are reflected in the annual accounts of companies, but we should keep in mind that there may be a significant lag (Zavgren, 1983). In this respect, Johnson (1970) points out that financial ratios “do not contain information about […] the intervening economic conditions (p. 1166)” and that “the riskiness of a given value for [a] ratio changes with the business cycle (p. 1167)”. Similarly, Richardson et al. (1998) assert that the accounting-based failure models generally do not control for changes in the information content of accounting data that may occur due to a recession. Therefore, the classification and prediction power of failure prediction models can be improved by adding information on the occurrence of a recession and hence by controlling for the knowledge that the accounting data represent company operations during a recession. (Richardson et al., 1998).Several studies have analysed the effect of macro-economic variables on corporate failure. For example, Swanson & Tybout (1981) examined the impact of different macro-economic variables on business failure and concluded that interest rates and the occurrence of credit shocks is the most important one. An other example is Bhattacharjee et al. (2002), who studied the effect of macro-economic instability on bankruptcy (as opposed to acquisition) and concluded that there is a higher bankruptcy risk when the economy enters a downturn (i.e. when there is a macro-economic instability). On the other hand, they find that the business cycle itself has no significant direct impact on bankruptcy risk. A second group of variables giving information on the external environment of a firm relates to the *prospects of the industry* in which the firm operates. In this respect, it may be advised to include information on the potential of the market, the industry profitability and the competition into the failure prediction model (Lehmann, 2003).

Finally, besides non-accounting and qualitative variables, information concerning size class, industry and age and information about the external environment, also *socio-scientific factors* (i.e. sociological, psychological and ethical aspects) can be taken into account. (Bijnen & Wijn, 1994).

An additional problem associated with the use of financial ratios in failure prediction models relates to the fact that financial ratios are constructed from ***different components***, each of them reflecting other information on the financial health of the firm. It is possible that failing and non-failing firms show no differences for certain financial ratios, while the components of these ratios clearly differ (Beaver, 1967b). For this reason, it might be interesting to analyse the components of financial ratios instead of the financial ratios themselves. A similar remark goes for the 'overall ratios', which are composed of ***different 'detailed ratios***. If the positive influence of one detailed ratio is compensated by a negative influence of an other detailed ratio, it is possible that the overall ratio does not reflect any problematic situation in the company. For example, a very high 'rotation of total assets' may be compensated by an extremely low 'profit margin of sales', resulting in a normal level of 'profitability of total assets'. Therefore, in order to timely detect problems, the analysis of the detailed ratios seems necessary (Bilderbeek, 1978).

Finally, there seems to be no consensus considering ***which type of financial ratios*** are the best failure indicators. Although many studies have compared the predictive abilities of *accrual-based* financial ratios and *cash flow-based* ratios, there seems to be no consensus concerning which ratios lead to the most accurate failure predictions. Gentry et al. (1985a) found that cash flow-based funds flow components offer a viable alternative to accrual-based financial ratios for classifying failing and non-failing firms. Gentry et al. (1987) and Aziz & Lawson (1989)[28] stated that corporate failure prediction models based on cash-based funds flow components have better failure prediction abilities than models based on accrual-based financial ratios. Similarly, Gentry et al. (1985b) found that adding cash flow ratios to ratio-based models increases the accuracy and Gombola & Ketz (1983) concluded that: "when cash flow is measured as cash revenues from operations less cash expenses for operations […] cash-flow ratios may contain some information not found in profitability measures (p. 113)". In addition, Declerc et al. (1990) concluded that their cash flow-based model performed slightly better than a model with accrual-based financial ratios. Recently, Sharma & Iselin (2003) investigated the relevance of cash flow and accrual information and confirmed that cash flow information seems to be very useful in solvency assessment. On the contrary, Casey & Bartczak (1984) are strong opponents of the use of cash flow-based models. Firstly, they asserted that operating cash flow is a poor predictor of corporate failure. This finding has been confirmed by Gentry et al. (1985a), who found that cash flow from operations does not improve the classification results. Casey & Bartczak (1984) also found that,

[28] Aziz & Lawson (1989) compared the accuracy of their cash flow model with the accuracy of Altman's Z score model (1968) and the Zeta-model (1977) and found that the cash flow model is superior to the Z-score model and gives better early warning signs than the Zeta model.

when compared to accrual-based models, cash flow-based models (in fact, they mean 'working capital funds flow' models) misclassify non-bankrupt firm at a higher rate. They concluded that cash flow variables fail to even marginally improve the predictive results, when used in combination with accrual-based ratios. However, it should be emphasised here that, Casey & Bartczak excluded distressed companies from the non-failing sample and this may have affected their conclusions. Besides Casey & Bartczak, also Gombola et al. (1987) indicated that cash flow from operations provides no additional information in bankruptcy prediction. Finally, Aziz et al. (1988), who compared a cash flow model with two accrual-based ratio models, concluded that these two types of models show similar performances.

In this respect, it should be mentioned that, in failure prediction research, *value added* ratios generally have been neglected: only a few studies have included value added ratios. Nevertheless, we assert that value added ratios do have discriminatory power. For example, in a study of Declerc et al. (1991) it is shown that, when incorporated into a failure prediction model along with other financial ratios, value added ratios do increase the model's classification results.

### 3.4.2 The use of one single annual account

In addition to the above mentioned problems related to the use of annual account information in the form of financial ratios, most classic cross-sectional models are subject to a number of problems related to the use of only **one single annual account** for each company in the estimation samples.

First of all, as an annual account only gives a ***snapshot*** of a company's financial situation, classic statistical failure prediction models assessing the failure risk of companies based on information from one annual account, may be misleading. For example, a model may classify a healthy company suffering from a temporary adverse situation (characterised by a temporary negative value for profitability or a temporary high value for 'overdue short-term priority debts') as failing.

A second problem is that classic 'static' failure prediction models ***do not account for time-series*** behaviour of variables and hence ignore important past information regarding corporate performance (Theodossiou, 1993; Dirickx & Van Landeghem, 1994; Kahya & Theodossiou, 1996). It can be argued that the prediction of company failure should not only depend on one single annual account, but on more than one annual account or the change in financial health (Bilderbeek, 1973; Bilderbeek 1978; Shumway, 1999). In 1966, Tamari (1966) already indicated the importance of the trend analysis in corporate failure studies. For the same reason, Edmister (1972) included the trend of the financial ratios as failure indicators in his failure prediction model. Trends were included in the form of up-trend and down-trend dummies and three-years averages. Similarly, Dambolena & Khoury (1980) and Betts & Belhoul (1987) found that the inclusion of (in)stability measures in a failure prediction model improves the classification results and Chalos (1985) pointed out that the use of trend data, which depict the average change over several years, could reveal more information from the annual accounts and are better able to capture 'creative accounting' practices. In this respect, Taffler (2003) stressed that the analysis of the evolution of annual account ratios is crucial in the assessment of company health.

Thirdly, some problems stem from the fact that, when building a failure prediction model based on one single annual account, one arbitrarily has to decide ***which annual account*** of each firm is used. First, the choice of when to observe the firm's data introduces a *'selection bias'* in the resulting model (Mensah, 1984; Shumway, 1999). Second, the applicability of a model is determined by the annual accounts that are used when estimating the model: the accounts one, two or three years prior to the event of failure. In other words, from an 'ex ante' viewpoint, dependent on which annual account is used, the model provides a classification statement concerning the failure/survival status respectively in year t+1, t+2 or t+3[29] (Deakin, 1972). For example, if a model is estimated on annual account data three years prior to failure, the model will show poor predictive results or become totally unreliable (it has little or no meaning) when it is used to predict failure in a shorter term, because the model gives an indication of the failure probability in year t+3 (Lane et al., 1986). Also, when the model is re-estimated on annual account information of other years prior to failure, the coefficients are very likely to be inconsistent: dependent on which year prior to failure the annual accounts are taken from, the coefficients will be different. Similarly, when the model is re-developed on annual account information of other years prior to failure, the selection of the variables in the models is very likely to be inconsistent.

Fourthly, as a great majority of researchers observe the failing firms' data one year prior to failure so as to construct a 'short term' bankruptcy prediction models, many models are subject to some specific problems related to the use of ***annual account data one year prior to failure***. A first problem rises in the estimation phase of these short term models, because, for many firms, the annual account one year prior to the moment of failure (bankruptcy) is not available. Firm's often stop publishing annual accounts one or two years prior to failure. In these cases, researchers generally consider the last published annual account as the annual account one year prior to failure and hence implicitly consider the moment of ‘real’ economic failure instead of the moment of ‘legal’ bankruptcy. A second problem appears when the short term model is to be applied in practice to predict failure of firms in year t+1. In most cases, failing firms delay the deposit of their annual account when they approach failure (bankruptcy). In this context, Deakin (1977) remarks that “… in many cases that [annual] report is delayed for failing companies and may not be available until the failure event (p. 75)”. In these cases, one is unable to apply the model in the appropriate way and, as a result, the model becomes ‘useless’.

Finally, it should be borne in mind that classic statistical models based on one single annual account implicitly assume that consecutive annual accounts are independent observations. However, it is clear that consecutive annual accounts are not independent of each other and can not be simply interpreted as repeated measurements (Dirickx & Van Landeghem, 1994). As a result, when a failure prediction model is applied several times to several annual accounts of one firm, this may result in a whole list of predictions, which are potentially conflicting (Keasey et al., 1990; Dirickx & van Landeghem, 1994) and hence inconsistency problems may rise. This is often referred to as ***‘signal inconsistency’*** (Luoma & Laitinen, 1991). Keasey et al. (1990) suggest that signal inconsistency occurs frequently.

### 3.5 Selection of independent variables

The great majority of corporate failure prediction studies starts from a large **initial battery** of variables, often ***arbitrarily chosen*** on the basis of their popularity in literature and their predictive success in previous research. The reason for the arbitrary choice of variables is that the theoretical basis for the selection of variables has always been too limited in order to allow a better selection (Karels & Prakash, 1987; Dirickx & Van Landeghem, 1994). Nevertheless, it should be stressed that selecting variables on the basis of popularity may be problematic, because popular ratios are more likely to be subject to ‘window dressing’ and hence to be unreliable (Beaver, 1967b). Apart from the arbitrary selection, some researchers have composed their initial set of variables on the basis of statistical considerations, (2) a theoretical model or (3) a combination of an empirical method and a theoretical model.

From the initial battery of variables, a **final set** of variables is selected so as to construct a failure prediction model. This final set of variables can be based on (1) statistical

[29] Most researchers argument that it is necessary to develop separate models for each time frame of failure prediction, being one, two or more years prior to failure. Altman (1978), on the contrary, is a strong opponent of using separate models. He cautions that this may confusing, in the case when the different models give contradictory predictions for one firm.

(empirical) considerations, (2) a theoretical model, (3) a combination of an empirical method and a theoretical model or (4) no specific considerations.

In most studies, the final set of variables is selected on the basis of ***statistical considerations***. As there is no (financial) theory indicating the (financial) variables that are the best predictors (Scott, 1981), researchers generally select those variables that lead to the best failure prediction model for the used sample and/or that satisfy some distributional requirements. Many failure prediction models are the result of a statistical search trough a number of plausible financial indicators in order to empirically find some characteristics that distinguish between failing and non-failing firms. Researchers select the variables on the basis of, for example, the statistical significance of the estimated parameters, individual discriminating ability of each of the variables (in a univariate analysis), the sign of the variables' coefficients, principal components analysis, factor analysis, the classification results of different combinations of ratios or stepwise methods (such as the forward selection and backward elimination methods). In this context, we can refer to the term "brute empiricism". The statistical characteristics of the variables are stressed, while the economic importance of the variables is ignored (Moses & Liao, 1987; Back et al., 1996b). It is obvious that this way of variable selection has some negative consequences and hence is subject to serious criticisms. In this context, Keasey & Watson (1991) extensively elaborate on the usefulness of these data driven models and their related problems.

A first consequence of the empirical selection of variables is that there is *little agreement* concerning which variables are the best in distinguishing between failing and non-failing companies. In the empirical literature, no definite group of good failure predictor variables can be found. There is a wide range of corporate failure models with good classification results, each consisting of different variables and a different number of variables (Edmister, 1972; Back et al., 1996b; Altman & Narayanan, 1997 ; Altman & Saunders, 1998; Mossman et al., 1998; Becchetti & Sierra, 2003). Dimitras et al. (1996) and Daubie & Meskens (2002) give an extensive overview of financial ratios included in corporate failure prediction models. According to Daubie & Meskens (2002), the most frequently used financial ratios are: current assets / current liabilities, working capital / total assets, EBIT / total assets, quick assets / current liabilities and net income / total assets. These ratios also appear in the study of Dimitras et al. (1996).

Secondly, if the variables are selected empirically, the choice of variables will strongly depend on the sample that is used and the resulting (empirically founded) failure prediction model is very likely to be *sample specific* and *unstable* (Edmister, 1972; Zavgren, 1983;

Zavgren, 1985). The empirical findings may therefore not be suited for generalization (Edmister, 1972; Gentry et al., 1987) In this context, Blum (1974) points out that, if there is no general theory on the symptoms of failure and the financial ratios are chosen statistically, one can not expect that the correlation between the independent variables and the failure status to be predicted will remain the same in any sample (see the 'non-stationarity problems' in point 3.3.1.). It is possible that the failure model will not be appropriate for predicting corporate failure in a different economic or temporal setting. This is the problem of '*statistical over-fitting*'. In this respect, Karels & Prakash (1987) warn that a careful selection of predictor variables is needed in order to improve the predictive performance of failure prediction models.

A third consequence of the empirical selection of the variables is that the failure prediction model may be *not diversified* and even show *counter-intuitive signs* for some coefficients. The models of Bilderbeek (1979), Zavgren (1985), Gloubos & Grammatikos (1988) and Keasey & Mc Guinness (1990) are some of the large number of models with unexpected coefficients[30]. These models use a combination of variables that leads to the best classification of the firms in their estimation samples. They neglect the economical meaning of the variables and, the possibly high correlation among the individual ratios may cause the statistical estimation procedure to assign counter-intuitive signs to some coefficients (Moses & Liao, 1987; Ooghe & Balcaen, 2002c; Hayden, 2003). The fact that many corporate failure models are not diversified and show counter-intuitive signs contradicts to the general viewpoint that a good model should include some carefully chosen variables from the whole spectrum of financial analysis - liquidity, indebtedness, profitability and activity[31] - and should use these variables in the intuitively right sense.

There are, however, some researchers who have not entirely motivated the selection of variables by their empirical performance, but rather by a certain (limited) theory. These ***theoretical frameworks*** reduce the scope for statistical over-fitting (Scott, 1981). Most of these theoretically founded studies are based on a certain *cash flow theory*. Beaver (1967a) was the first to apply a cash flow theory, in which the firm is viewed as "a reservoir of liquid assets which is supplied by inflows and drained by outflows (p. 79-80)". Similarly, a cash flow framework was the basis of the 'failing company model' of Blum (1974).

[30] See Ooghe & Balcaen (2002c) for a discussion of the counter-intuitive signs of some coefficients in these models.
[31] Dambolena & Khoury (1980) state that ratios that measure these four aspects "have been shown to have considerable merit in financial analysis and in the measurement of financial well being of corporate entities (p. 1021)".

The study of Aziz et al. (1988) was based on the cash flow identity of Lawson. Charitou et al. (2004) selected several operating cash flow related ratios, because they asserted that organizations can not survive without generating cash from their normal operations. Ooghe et al. (1993) selected their initial battery of variables on the basis of the 'operational cash flow table'[32]. Similarly, the initial set of variables of the model of Declerc et al. (1990) was based on this 'financing table'. Other examples of studies that have been based on a cash flow theory are: Gentry et al. (1985b), Aziz et al. (1988) and Aziz & Lawson (1989). Besides these cash flow theories, *other theoretical models* have been used. Wilcox (1971) motivated his variable selection by the 'gambler's ruin model'. In this model, the firm is viewed as a gambler who has an initial amount of money. This amount of money will eventually grow or fall to zero (this latter case represents bankruptcy) through a series of independent trials. In these trials, the firm may win a dollar (with probability p) or may lose a dollar (with probability 1-p). In this model, the firm meets losses by selling its assets. Charitou & Trigeorgis (2002) used the conceptual framework of option-pricing to select their variables. Other examples are Ooghe & Verbaere (1982), who used an integrated ratio model and other theoretical considerations in order select the initial set of variables, and van Frederikslust (1978), who attempted to give a theoretical discussion of his choice of variables. Keasey & Watson (1987) and Keasey & McGuiness (1990), on the other hand, used a theoretical a priori, but also considered the choices of previous empirical studies. Although many authors agree that constructing stable failure prediction models requires a theoretical framework (Charitou et al., 2004) – a sound theoretical foundation concerning the primary variables that are relevant in distinguishing between failing and non-failing firms – there is, until now, *no general theory* concerning company failure. Dimitras et al. (1996), for example, state that " a unifying theory of business failure has not been developed, in spite of a few notable efforts (p. 487)".

Finally, some corporate failure prediction studies have selected their predictor variables ***without any theoretical or empirical consideration***. The Ohlson model (1980), for example, is a typical result of this type of variable selection.

[32] The 'operational cash flow table' is: change in cash = cash flow from operations (i.e. operational gross result after taxes + changes in operational net working capital) +/- cash flow from investments in fixed assets +/- cash flow from financing (i.e. – financial flows of liabilities – financial flows of equity capital +/- external financing involving equity capital and/or liabilities) (Ooghe & Van Wymeersch, 2003, part 2, p. 54).

### 3.6 Time dimension

It is clear that the 'static' classic statistical failure prediction models ***ignore that companies change over time***. The output of such a model is a fixed score – a discriminant or logit score – for each company, which is independent of time. This is, however, in contradiction with general intuition and hence it seems that classic statistical failure models are not suited for corporate failure prediction. In this respect, Altman & Eisenbeis (1978) pointed out that "…the concept of prediction […] is not strictly applicable to the standard discriminant analysis which does not explicitly incorporate a time dimension (p. 186)". In fact, the principal aim of the a classic statistical model is to summarize information to determine whether a firm's profile most 'resembles' the profile of a failing or the profile of a non-failing firm. In this context, Taffler (1982; 1983) and Taffler & Agarwal (2003) stressed that an MDA model in fact analyses the following question: "Does the company have a profile more similar to the failing group of companies from which the model was developed or the non-failing group?" A Z-score is doing little more than reflecting information conveyed by the annual accounts in an ordinal scaled measure. A Z-score below a certain threshold highlights impending financial difficulties (i.e. the firm 'might' fail), but it does not indicate that the company 'will' fail. It has a ***retrospective character***. What is demonstrated is that failing and non-failing firms have dissimilar characteristics, not that the variables have predictive power[33] (Ooghe & Joos, 1990). Altman et al. (1981) referred to this as the concept of "***resemblance***". The retrospective character and the concept of resemblance causes the classic statistical failure prediction models to have a ***descriptive*** or ***'pattern recognition'*** nature. Keasey & Watson (1991), Taffler (1982; 1983) and Taffler & Agarwal (2003) emphasized this descriptive nature. It is clear that, when a model is descriptive in nature, it must not be seen as a prediction device, but rather as a communication device. In this context, MDA models, such as Altman's model (1968) and Taffler's model (1982), are generally robust: they seem to reveal continuing success in practical applications.

Due to the retrospective character and descriptive nature of a z-score, it seems appropriate to analyse the financial health of a company by examining the evolution of the Z-scores. Taffler (1983) argued that, instead of using the score as a failure probability, it is more interesting to analyse the history of Z-scores of each firm. Therefore, he introduced a type of 'risk index of Z-scores', the PAS index, reflecting the percentile in which the Z-score of the

[33] With a view to predict company failure, the crucial problem is to make inference in the opposite, prospective direction (from variables to failure). This requires a model that links certain variables' values to failure or non-failure. It must be demonstrated that certain variable values imply failure or non-failure (Johnson, 1970).

firm lies, when ranking the Z-scores of all companies for a particular year. He suggested to analyse a company's PAS indices for several years, because "the PAS trajectory of a company indicates its relative performance over time (p. 305)". Similarly, Ooghe & Joos (1990) also mentioned that the evolution of discriminant scores may add important information about the financial situation of a company.

A second comment related to the issue of the time dimension is that classic statistical failure prediction models do not explicitly give the expected ***time to failure***, which lessens the practical usefulness of these models (Lane et al., 1986; Luoma & Laitinen, 1991). These models determine the variables that predict no more than whether the firm might fail or not. No conclusions can be made with respect to the timing of failure. In this context, Cole & Gunther (1995) concluded that only a small number of the variables that are commonly used to predict the failure/survival status, actually are related to the timing of failure[34]. Consequently, it is of critical importance to avoid drawing conclusions on a firm's time to failure on the basis of its model score.

Thirdly, when determining the model score for a company – reflecting the company's failure risk – the classic statistical models do not take into account the period during which the company has been exposed to the risk of failure. In other words, failure probability is considered to be independent of ***firm age*** (Shumway, 1999). This is contradictory to the general expectation that younger, starting companies face higher failure probabilities than older, more established firms (Bickerdyke et al., 1999).

A fourth problem is that classic statistical failure prediction models do not treat company failure as a process. Failure is approached as a ***discrete event*** (Altman, 1984), which is reflected in the dichotomous dependent variable. In addition, the models are based on cross-sectional data and, therefore, they do not consider information on the progress of the failure process. They do not use the dynamics of the failure process in order to predict failure (Van Wymeersch & Wolfs, 1996). Instead, the static modelling techniques assume that failure is a ***steady-state***. The underlying failure process is assumed to be stable over time and no phases are distinguished (Luoma & Laitinen, 1991; Laitinen, 1993; Laitinen & Kaankaanpää, 1999). However, reality clearly indicates that failure is not a sudden event that happens unexpectedly (Luoma & Laitinen, 1991). In most cases, company failure is the result of a (long term) failure process or a "failure path", which gradually leads a company to the final moment of failure. The failure path may consist of several phases, each characterised by specific behaviour of

[34] With a view to distinguish between those variables determining company failure and those determining the timing of failure, Cole & Gunther (1995) applied a split-population survival-time method.

certain variables or specific symptoms of failure[35]. Consequently, the steady-state assumption of the classic 'static' failure prediction models may have serious consequences. In an empirically founded model, the relative importance of the different variables and the accuracy of the model will be implicitly determined by the *frequency of occurrence of the different phases* of the failure process in the estimation sample of failing companies. For example, a model estimated on data one year prior to failure, which mainly concern the final phase of the failure process, will perform poorly when it is used for failure prediction many years prior to failure, rather concerning an earlier phase of the failure process. In the same way, the sample construction with respect to the different phases of the failure process may explain the unstable coefficients when a model estimated on data one year prior to failure is re-estimated for a sample of annual accounts two and three years prior to failure (Laitinen, 1993).

Finally, besides considering failure as a steady-state instead of a process, the classic statistical models do not consider possible differences in failure paths. They assume that all companies follow a ***uniform failure process***. However, in practice, there seems to be a wide variety of failure processes or failure paths (Laitinen, 1991).  As in most classic statistical models the independent variables are chosen in an empirical way, the relevance of the different variables and the general efficiency of the models are implicitly determined by the *frequency of occurrence of the different failure paths* in the estimation sample of failing companies. For example, if the frequency of 'sudden failure' firms (with a short term failure path) in the estimation sample is high, the resulting model may be expected to perform badly if it is used for long term failure prediction several years prior to failure (Laitinen, 1991). Similarly, if the estimation sample mainly contains 'acute' failing firms (which shows good performance results until a rapid decline), this may explain why the resulting model performs poorly when applied to a sample with a high frequency of chronic failing firms (which have a very poor performance in the years prior to failure).

[35] As indicated by Daubie & Meskens (2002), the relative importance of the variables for the detection of failure is not constant over time.

# 4 CONCLUSION

The topic of corporate failure prediction has developed to a major research domain in corporate finance. The large number of parties involved in corporate failure, the large failure costs, the negative spiral in the general economic environment are only a few reasons for the strong interest in this topic. Other factors that have boosted the research in corporate failure prediction are: the increased availability of data and statistical techniques, the extended academic research on the impact of market imperfections and information asymmetry and the introduction of the New Basel Capital Accord. Over the last 35 years, many academic studies have been dedicated to the search for the best failure prediction model, which classifies companies according to their (financial) health or failure risk. Academic researchers from all over the world have been using numerous types of modelling techniques and estimation procedures for the development of corporate failure prediction models. The 'classic cross-sectional statistical methods' seem to be the most popular methods. A gigantic number of 'single-period' classification models or 'static' models have been developed.

Four general types of classic statistical methods have been applied in corporate failure prediction studies: (1) univariate analysis, (2) risk index models, (3) multivariate discriminant analysis and (4) conditional probability models. Each methods has its the specific assumptions, advantages and disadvantages. The large majority of academically developed classic corporate failure prediction models seem to be MDA models and logit models.

Although the classic statistical methods of MDA and conditional probability models have proven to be very popular in corporate failure prediction, there appear to be several problems related to the application of these methods.

Firstly, some problems relate to the use of a *dichotomous dependent variable*. As business failure is not a well defined dichotomy, the use of a dichotomous dependent variable is in contrast with reality. However, when estimating a classic statistical model, researchers arbitrarily or artificially separate companies into a failing and a non-failing population. The arbitrary choice of the definition of failure and the arbitrary way in which the chosen definition of failure is applied to the total population of companies (a certain year or time period is chosen) results in an inappropriate application of the classic statistical modelling techniques. One should also bear in mind that the chosen definition of failure may strongly influence the empirical selection of variables and that the use of two clearly separated populations of companies may cause a misleading model reliability.

Secondly, the *sampling method* seems to inflict some particular problems. When using non-random estimation samples, the classic statistical methods are applied inappropriately and the resulting model can not be generalized. Non-random samples may be the result of (1) over-sampling the failing firms, using state-based samples, (2) applying a complete data sample selection criterion or (3) using matched pairs of failing and non-failing firms.

Thirdly, the classic models can be criticized because of problems concerning *non-stationarity and data instability*. Using a classic statistical model in a predictive context, requires that the relationships among the variables in the model are stable over time. However, as there is ample evidence of non-stationarity and data instability, the model is likely to have poor predictive abilities and may be unstable or not robust over time. In addition, in the presence of non-stationarity and data instability, the use of pooled samples is inappropriate and leads to inaccurate and unstable models.

Fourthly, the use of annual account information is subject to a large number of remarks. The use of financial ratios can be criticized because of, for example, doubts about the annual accounts giving a fair and true view of the financial situation, the occurrence of missing values, extreme ratio values or errors and differing opinions on which type of financial ratios are the best failure indicators. Also, it is clear that not all relevant information is reflected in the balance sheet and income statement. Non-accounting, qualitative information, information concerning the industry, size and age, concerning the macro-economical and industry-specific situation and socio-scientific factors should be considered in a failure prediction model. Besides, there are a number of serious problems related to the use of only one single annual account, such as the snapshot focus, the ignorance of time-series behaviour, the selection bias and signal inconsistency.

Fifthly, the *selection of independent variables* seems to be problematic. Besides the fact that, in most studies, the initial battery of variables is arbitrarily chosen, the selection of the final set of variables is subject to criticism. An empirical selection of variables may lead to a sample specific and unstable model, which are over-fitted, and to a model that is not diversified and shows counter-intuitive signs for some coefficients. On the other hand, the variables may be selected on the basis of a theoretical framework, but, although many authors agree that constructing stable failure prediction models requires a theoretical framework, there is, until now, *no general theory* concerning company failure.

A final group of problems relate to the *time dimension*. A ‘static’ classic statistical model ignores that companies change over time. Its retrospective character and its dependence on the concept of ‘resemblance’, results in the descriptive or ‘pattern recognition’ nature of the model. In a classic model failure is approached as a discrete event. Failure is assumed to be a steady state, not a process, and each company is assumed to follow a uniform failure process. However, as reality clearly indicates that failure is a process, possibly with different phases and that different failure paths exist, these assumptions may have serious consequences for the model. In an empirically driven model, the relative importance of the variables and the overall efficiency or accuracy of the model will be determined by the characteristics of the failure processes that are present in the estimation sample of failing firms.

In this paper, several issues viewed in isolation by earlier studies are considered together and, in this view, the paper contributes towards a better understanding of the features of the classic statistical failure prediction models and their related problems.

The alternative methods for modelling business failure – such as multi-logit analysis, survival analysis, dynamic event history analysis, multidimensional scaling, decision trees, expert systems and neural networks – are beyond the scope of this study. However, as literature does not provide a clear overview of the application of alternative methods to the topic of business failure prediction, further research concerning these methods is necessary. Furthermore, it seems interesting to generate a literature overview of all studies comparing the predictive performances of different types of failure prediction models and to systematically compare the performances of the various methods.

## REFERENCES


Altman E.I., 1968, Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. The Journal of Finance, Vol. 23, nr. 4, September 1968, p. 589-609.

Altman E.I., 1978, Examining Moyer's re-examination of forecasting financial failure. Financial Management, Vol. 7, nr. 4, Winter 1978, p. 76-81.

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., Avery R.B., Eisenbeis R.A., Sinkey J.F., 1981, Application of classification techniques in business, banking and finance, Greenwich (Connecticut): JAI Press Inc., 249pp., ISBN: 0-89232-004-4.

Altman E.I., Eisenbeis R.A., 1978, Financial applications of discriminant analysis: A clarification. Journal of Financial and Quantitative Analysis, March 1978, p. 185-195.

Altman E.I., Eom Y.H., Kim D.W., 1995, Failure prediction: Evidence from Korea. Journal of International Financial Management and Accounting, Vol. 6, nr. 3, p. 230-249.

Altman E.I., McGough T.P., 1974. Evaluation of a Company as a Going Concern. Journal of Accountancy, Vol. 138, nr.6, December 1974, p. 50-57.

Altman E.I., Haldeman R.G., Narayanan P., 1977, ZETA analysis: A new model to identify bankruptcy risk of corporations. Journal of Banking and Finance, Vol. 1, nr. 1, p. 29-51.

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.

Amrhein D.G., Katz J.A., 1998, Cash flow as a factor in the mortality risk of a business. International Council for Small Business, 1998 Singapore Conference Proceedings, Paper nr. 18.

Argenti J. , 1976, Corporate Collapse: The Causes and Symptoms. Londen: Mc Graw-Hill, 193 pp., ISBN: 0-07-084469-0.

Astebro T., Winter J.K., 2001, More than a dummy: the probability of failure, survival and acquisition of firms in financial distress. Working Paper, September 2001, p. 1-39.

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.

Aziz A., Emanuel D.C., Lawson G.H., 1988, Bankruptcy prediction – An investigation of cash flow based models. Journal of Management Studies, Vol. 25, nr. 5, p. 419-437.

Aziz A., Lawson G., 1989, Cash flow reporting and financial distress models : testing of hypotheses, Financial Management, Vol. 18, nr. 1, Spring 1989, p. 55-63.

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.

Barnes P., 1987, The analysis and use of financial ratios: A review article. Journal of Business Finance and Accounting, Vol. 14, nr. 4, Winter 1987, p. 449-461.

Barnes P., 1982, Methodological implications of non-normality distributed financial ratios. Journal of Business Finance and Accounting, Vol. 9, nr. 1, Spring 1982, p. 51-62.

Beaver W., 1967a, Financial ratios predictors of failure. Empirical Research in Accounting: Selected Studies 1966, Journal of Accounting Research, Supplement to Vol. 4, p. 71-111.

Beaver W., 1967b, Alternative accounting measures as predictors of failure. American Accounting Association Manuscript Contest for 1967, p. 113-121.

Becchetti L., Sierra J., 2002, Bankruptcy risk and productive efficiency in manufacturing firms. Journal of Banking and Finance, Vol. 27, p. 2099-2120

Betts J., Belhoul D., 1987, The effectiveness of incorporating stability measures in company failure models. Journal of Business Finance and Accounting, Vol. 14, nr. 3, p. 323-334.

Bhattacharjee A., Higson C., Holly S., Kattuman P., 2002, Macro economic instability and business exit: Determinants of failures and acquisitions of large UK firms. Working Paper nr. 0206, Cambridge Working Papers in Economics, Department of Applied Economics, University of Cambridge March 2002.

Bickerdyke I., Lattimore R., Madge A., 2000, Business failure and change: an Australian perspective. Productivity Commission Staff Research Paper, Ausinfo, Canberra, ISBN 1 74037 029 5, p. 1-192.

Bijnen E.J., Wijn M.F., 1994, Corporate prediction models: ratios or regression analysis? Research memoranda nr. 658, May 1994, Faculty of Economics and Business Administration, Tilburg University, The Netherlands, p. 1-23.

Bilderbeek J., 1979, An empirical study of the predictive ability of financial ratios in the Netherlands. Zeitschrift Für Betriebswirtschaft, May 1979, p. 388-407.

Bilderbeek J., 1973, Financieel-economische indices ten behoeve van de bedrijfsbeoordeling. Economisch en Sociaal Tijdschrift, Jaargang 27, nr. 2, April 1973, p. 141-155.

Bilderbeek J., 1978, Het voorspellen van falingen. Financiële kengetallen als thermometer voor de ondernemingsdoorlichting. Economisch en Sociaal Tijdschrift, Jaargang 32, nr. 1, p. 5-25.

Blazy R., 2000, La faillite, éléments d'analyse économique. Préface de Pierre Morin, Economica, Paris, France.

Blum M., 1974, Failing company discriminant analysis. Journal of Accounting Research, Vol. 12, nr. 1, p. 1-25.

Burgstahler D., Dichev I., 1997, Earnings management to avoid earnings decreases and losses. Journal of Accounting and Economics, Vol. 24, nr. 1, December, p. 99-126.

Casey C., Bartczak N., 1984, Cash flow: it's not the bottom line. Harvard Business Review, Vol. 4, p. 60-66.

Chalos P., 1985, Financial distress : a comparative study of individual, model and committee assessments. Journal of Accounting Research, Vol. 23, nr. 2, p. 527-543.

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.

Charitou A., Trigeorgis L., 2000, Option-based bankruptcy prediction. Paper presented at 6th Annual Real Options Confernce, Paphos, Cyprus, 4-6 July 2002, p. 1-25.

Charitou A., Lambertides N., 2003, Earnings management prior to bankruptcy. Working Paper University of Cyprus, Cyprus, p. 1-46.

Cole R.A., Gunther J.W., 1995, Separating the likelihood and timing of bank failure. Journal of Banking and Finance, Vol. 19, p. 1073-1089.

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.

Cybinski P.J., 1998, The dynamics of the firm's path to failure: towards a new methodology for modeling financial distress. Unpublished Ph.D. Thesis, Griffith University, Brisbane.

Dambolena I, Khoury S., 1980, Ratio stability and corporate failure. Journal of Finance, Vol. 33, nr. 4, p. 1017-1026.

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.

Deakin E., 1977, Business failure prediction: an empirical analysis, p. 72-98. In: Altman, Sametz, 1977, Financial crisis: institutions and markets in a fragile environment.

Deakin E., 1972, A discriminant analysis of predictors of business failure. Journal of Accounting Research, Vol. 10, nr. 1, Spring 1972, p. 167-179.

Deakin E., 1976, On the nature of the distribution of financial accounting ratios: some empirical evidence. The Accounting Review, Vol. 51, nr. 1, January 1976, p. 90-97.

Declerc M., Heins B., Van Wymeersch Ch., 1991, The use of value added ratios in statistical failure prediction models: Some evidence on Belgian annual accounts. Paper presented at the 1991 Annual Congress of the European Accounting Association, April 10-12 1991, Maastricht (The Netherlands), p. 1-24.

Declerc M., Heins B., Van Wymeersch Ch., 1990, Flux financiers et prévision de faillite: une analyse comportementale de l'entreprise. Colloque International « L'analyse financière par les flux : mode ou modèle? », Institut d' Administration des Entreprises, Nice, France, December 1990, p. 1-26.

DeFond M. L., Jiambalvo J., 1994, Debt covenant violation and manipulation of accruals. Journal of Accounting and Economics, Vol. 17, January, p. 145-176.

Degeorge F., Patel J., Zeckhauser R., 1999, Earnings management to exceed thresholds. Journal of Business, Vol. 72, nr. 1, p. 1-33.

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.

Dirickx Y., Van Landeghem G., 1994, Statistical failure prevision problems. Tijdschrift voor Economie en Management, Vol. 39, nr. 4, p. 429-462.

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.

Edmister R., 1972, An empirical test of financial ratio analysis for small business failure prediction. Journal of Financial and Quantitative Analysis, March 1972, p. 1477-1493.

Eisenbeis, 1977, Pitfalls in the application of discriminant analysis in business. Journal of Finance, Vol. 32, nr. 3, June 1977, p. 875-900.

Elloumi F., Gueyié J.P., 2002, Ernstige financiële problemen en vennootschappelijk bestuur, Finance & Control, Februar 2002, p. 56-66.

El-Zayaty A., 1987, Business failure prediction models: An empirical investigation of type II errors. Working paper, October 1987, Baruch College, The City University of New York.

Flagg J.C., Giroux G.A., Wiggins C.E., 1991, Predicting corporate bankruptcy using failing firms. Review of Financial Economics, Vol. 1, nr. 1, Fall 1991, p.67-79.

Franks J., Nyborg K., Torous W., 1996, A comparison of US, UK and German insolvency codes. Financial Management, Vol. 25, nr. 3, p. 86-101.

Gentry J.A., Newbold P., Whitford D.T., 1987, Funds flow components, financial ratios and bankruptcy. Journal of Business Finance & Accounting, Vol. 14, nr. 4, p. 595-606.

Gentry J.A., Newbold P., Whitford D.T., 1985a, Classifying bankrupt firms with funds flow components. Journal of Accounting Research, Vol. 23, nr. 1, Spring 1985, p. 146-160.

Gentry J.A., Newbold P., Whitford D.T., 1985b, Predicting bankruptcy: If cash flow’s not the bottom line, What is? Financial Analysts Journal, Vol. 41, nr. 5, September/October 1985, p. 47-56.

Gloubos G., Grammatikos T., 1988, The success of bankruptcy prediction models in Greece. Studies in Banking and Finance, Vol. 7, p. 37-46.

Gombola M.J., Ketz J., 1983, Note on cash flow and classification patterns of financial ratios. The Accounting Review, Vol. 58, nr. 1, January 1983, p. 105-114.

Gombola M., Haskins M., Ketz J., Williams D., 1987, Cash flow in bankruptcy prediction. Financial Management, Winter 1987, p. 55-65.

Gujarati, 2003, Basic Econometrics. Fourth Edition, 2003, London: McGraw-Hill.

Hall G., 1994, Factors distinguishing survivors from failures amongst small firms in the UK construction sector. Journal of Management Studies, Vol. 31, nr. 5, September 1994, p. 737-760.

Hayden E., 2003, Are credit scoring models sensitive with respect to default definitions? Evidence from the Austrian Market. Dissertation Paper, Department of Business Administration, University of Vienna, Austria, p. 1-43.

Hill N.T., Perry S.E., Andes S., 1996, Evaluating firms in financial distress: An event history analysis. Journal of Applied Business Research, Vol. 12, nr. 3, p. 60-71.

Hosmer D.W., Lemeshow S., 1989, Applied logistic regression. John Wiley & Sons: New York.

Hsieh S., 1993, A note on the optimal cutoff point in bankruptcy prediction models. Journal of Business Finance & Accounting, Vol. 20, nr. 3, April 1993, p. 457-464.

Johnson C.G., 1970, Ratio analysis and the prediction of firm failure. Journal of Finance, December 1970, p. 1166-1168.

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

Joos Ph., Ooghe H., 1993, The comparison between market determined and accounting determined measures of risk : an empirical assessment for the non-financial firms listed at the Brussels Stock Exchange. Working Paper, Department of Corporate Finance, Ghent University, Belgium, November 1993, p. 1-35.

Joos Ph., Ooghe H., Sierens N., 1997, Performance measures in credit classification. Working Paper, Department of Corporate Finance, Ghent University, Belgium, April 1997, p. 1-12.

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.

Joy O.M., Tollefson J.O., 1978, Some clarifying comments on discriminant analysis. Journal of Financial and Quantitative Analysis, March 1978, p. 197-200.

Joy O.M., Tollefson J.O., 1975, On the financial applications of discriminant analysis. Journal of Financial and Quantitative Analysis, Vol. 10, nr. 5, December 1975, p. 723-739.

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.

Karels G.V., Prakash A.J., 1987, Multivariate normality and forecasting of business bankruptcy. Journal of Business Finance & Accounting, Vol. 14, nr. 4, Winter 1987, p. 573-593.

Keasey K., McGuinness P., 1990, The failure of UK industrial firms for the period 1976-1984, logistic analysis and entropy measures. Journal of Business Finance & Accounting, Vol. 17, nr. 1, p. 119-135.

Keasey K., McGuinness P., Short H., 1990, Multilogit approach to predicting corporate failure: Further analysis and the issue of signal consistency. Omega International Journal of Management Science, Vol. 18, nr. 1, p. 85-94.

Keasey K., Watson R., 1986, The prediction of small company failure: some behavioural evidence for the UK. Accounting and Business Research, Vol. 17, nr. 65, p. 49-57.

Keasey K., Watson R., 1987, Non-financial symptoms and the prediction of small company failure: a test of Argenti's hypotheses. Journal of Business Finance & Accounting, Vol. 14, nr. 3, Autumn 1987, p. 335-354.

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.

Koh H.C., 1992, The sensitivity of optimal cutoff points to misclassification costs of Type I and Type II errors in the going-concern prediction context. Journal of Business Finance & Accounting, Vol. 19, nr. 2, January 1992, p. 187-197.

Lachenbruch P.A., 1975, Discriminant Analysis. New York: Hafner Press.

Laitinen E.K., 1992, Prediction of failure of a newly founded firm. Journal of Business Venturing, Vol. 7, p. 323-340.

Laitinen E.K., 1991, Financial ratios and different failure processes. Journal of Business Finance and Accounting, Vol. 18, nr. 5, p. 649-673.

Laitinen E.K., 1993, Financial predictors for different phases of the failure process. Omega The International Journal of Management Science, Vol. 21, nr. 2, p. 215-228.

Laitinen E.K., 1994, Traditional versus operating cash flow in failure prediction. Journal of Business Finance and Accounting, vol. 21, nr. 2, March 1994, p. 195-217.

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.

Lehmann B., 2003, Is it worth the while? The relevance of qualitative information in credit rating. Working Paper presented at the EFMA 2003 Meetings, Helsinki, p. 1-25.

Lízal L., 2002, Determinants of financial distress: what drives bankruptcy in a transition economy? The Czeck Republic case. William Davidson Working Paper nr. 451, January 2002, p. 1-45.

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.

Lussier R.N., 1995, A nonfinancial business success versus failure prediction model for young firms. Journal of Small Business Management, Vol. 33, nr. 1, p. 8-20.

Lussier R.N., Corman J., 1994, A success vs. Failure prediction model of the manufacturing industry. Paper nr. 48 , Conference of the Small Business Institute Director's Association, San Antonio, Texas, February 1994, p. 1-5.

Maddala G.S.,177, Econometrics. New York: McGraw-Hill Book Company.

Maltz A.C., Shenhar A.J., Reilly R.R., 2003, Beyond the balanced scorecard: refining the search for organizational success measures. Long Range Planning, Vol. 36, p. 187-204.

Mc Fadden D., 1973, Conditional logit analysis of qualitative choice behavior. In: Frontiers in Econometrics, edited by Zarembka P., New York: Academic Press.

Mc Leay S., Omar A., 2000, The sensitivity of prediction models tot the non-normality of bounded an unbounded financial ratios. British Accounting Review, Vol. 32, p. 213-230.

Mensah Y.M., 1983, The differential bankruptcy predictive ability of specific price-level adjustments: some empirical evidence. The Accounting Review, Vol. 58, nr. 2, April 1983, p. 228-246.

Mensah Y.M., 1984, An examination of the stationarity of multivariate bankruptcy prediction models: A methodological study. Journal of Accounting Research, Vol. 22, nr. 1, p. 380-395.

Micha B., 1984, Analysis of business failures in France. Journal of Banking and Finance, Vol. 8, p. 281-291.

Moyer R.C., 1977, Forecasting financial failure: A re-examination. Financial Management, Vol. 6, nr. 1, Spring 1977, p. 11-17.

Mossman Ch.E., Bell G.G., Swartz L.M., and Turtle H., 1998, An empirical comparison of bankruptcy models. The Financial Review, Vol. 33, nr. 2, p. 35-54.

Moses D., Liao S.S., 1987, On developing models for failure prediction. Journal of Commercial Bank Lending, Vol. 69, p. 27-38.

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.

Odera G., Dacorogna M.M., Jung T., 2002, Credit risk models – Do they deliver their promises? A quantitative assessment. Working Paper, Economics Working Paper Archive at WUSTL, November 2002, p. 1-18.

Ohlson J., 1980, Financial ratios and the probabilistic prediction of bankruptcy. Journal of Accounting Research, Vol. 18, nr. 1, Spring 1980, p. 109-131.

Ooghe H., Balcaen S., 2002c, Are failure prediction models transferable from one country to another? An empirical study using Belgian financial statements. Proceedings of the 9th Annual Conference of the Multinational Finance Society, 30/06/02 – 03/07/02, Cyprus**.**

Ooghe H., Camerlynck J., Balcaen S., 2003, The Ooghe-Joos-De Vos failure prediction models: a cross-industry validation. Brussels Economic Review, Vol. 46, nr. 1, Spring 2003, p. 39-70.

Ooghe H., Joos P., 1990, Failure prediction, explanation of misclassifications and incorporation of other relevant variables: result of empirical research in Belgium.

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.

Ooghe H., Joos P. en De Vos D., 1993, Risico-indicator voor een onderneming aan de hand van falingspredictie-modellen. Accountancy en Bedrijfskunde Kwartaalschrift, Vol. 18, nr. 3, september 1993, p. 3-26.

Ooghe H., Joos P., De Vos D., De Bourdeaudhuij C., 1994a, Towards an improved method of evaluation of financial distress models and presentation of their results. Working Paper, January 1994, Department of Corporate Finance, Ghent University, Belgium, 22p.

Ooghe H., Van Wymeersch Ch., 2003, Handboek financiële analyse van de onderneming. 2 parts, ISBN: 90-5095-330-1. Antwerpen: Intersentia.

Ooghe H., Verbaere E., 1982, Determinanten van faling: Verklaring en predictie. Research Report, Department of Corporate Finance, Ghent University, Belgium, p. 1-166.

Ooghe H., Verbaere E., 1985, Predicting business failure on the basis of accounting data: The Belgian experience. The International Journal of Accounting, Vol. 9, nr. 2, p. 19-44.

Peel M.J., Peel D.A., 1987, Some further empirical evidence on predicting private company failure. Accounting and Business Research, Vol. 18, p. 57-66.

Piesse J., Wood D., 1992, Issues in assessing MDA models of corporate failure: A research note. British Accounting Review, Vol. 24, p. 33-42.

Platt H.D., Platt M.B., 1990, Development of a class of stable predictive variables: the case of bankruptcy prediction. Journal of Business Finance & Accounting, Vol. 17, nr. 1, Spring 1990, p. 31-51.

Platt H.D., Platt M.B., 1991, A note on the use of industry-relative ratios in bankruptcy prediction. Journal of Banking and Finance, Vol. 15, p. 1183-1194.

Platt H.D., Platt M.B., 2002, Predicting corporate financial distress: reflections on choice-based sample bias. Journal of Economics and Finance, Vol. 26, nr. 2, Summer 2002, p. 184-199.

Platt H.D., Platt M.B., Pedersen J.G., 1994, Bankruptcy discrimination with real variables. Journal of Business Finance & Accounting, Vol. 21, nr. 4, July 1994, p. 491-510.

Pompe P., Bilderbeek J., 2000, Faillissementspredictie: een vergelijking tussen lineaire discriminant analyse en neurale netwerken. Economisch Sociaal Tijdschrift, Vol. 2, p. 215-242.

Richardson F.M., Kane G.D., Lobingier P., 1998, The impact of recession on the prediction of corporate failure. Journal of Business Finance and Accounting, Vol. 25, nr. 1 & 2, January/March 1998, p. 167-186.

Richardson F.M., Davidson L.F., 1984, On linear discrimination with accounting ratios. Journal of Business Finance and Accounting, Vol. 11, nr. 4, Winter 1984, p. 511-525.

Rime B., 2003, The new Basel Accord : Implications of the co-existence between the standardized approach and the internal ratings approach. Working Paper nr. 03.05, Studienzentrum Gerzensee, Stiftung der Schweizerschen Nationalbank, p. 1-36.

Rosner R.L., 2003, Earnings manipulation in failing firms. Contemporary Accounting Research, Vol. 20, nr. 2, Summer 2003, p. 361-408.

Scott J., 1981, The probability of bankruptcy: a comparison of empirical predictions and theoretical models. Journal of Banking and Finance, Vol. 5, nr. 3, September 1981, p. 317-344.

Scott E., 1978, On the financial applications of discriminant analysis: comment. Journal of Financial and Quantitative Analysis, March 1978, p. 201-205.

Sheppard J.P., 1994, Strategy and bankruptcy: an exploration into organizational death. Journal of Management, Vol. 20, nr. 4, p. 795-833.

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

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.

Steele A., 1995, Going concern qualifications and bankruptcy prediction. Paper presented at the European Accounting Association, February 1995, .., p. 1-28.

Steele A., 2002, Receiver Operating Characteristics (ROC) curves and stochastic dominance in discriminant functions: an application in bankruptcy diagnosis. Working paper, August 2002, Warwick Business School, University of Warwick, p. 1-36.

Swanson E., Tybout J., 1988, Industrial bankruptcy determinants in Argentina. Studies in Banking and Finance, Vol. 7, p. 1-25

Sweeney A.P., 1994, Debt-covenant violations and managers' accounting responses. Journal of Accounting and Economics, Volume 17, nr. 3, May 1994, p. 281-308.

Taffler R.J., 1982, Forecasting company failure in the UK using discriminant analysis and financial ratio data. Journal of the Royal Statistical Society, Vol. 145, Part 3, p. 342-358.

Taffler R.J., 1983, The assessment of company solvency and performance using a statistical model. Accounting and Business Research, Vol. 15, nr. 52, Autumn 1983, p. 295-307.

Taffler R.J., 1984, Empirical models for the monitoring of UK corporations. Journal of Banking and Finance, Vol. 8, p. 199-227.

Taffler R.J., Agarwal V., 2003, Do statistical failure prediction models work ex ante or only ex post? Paper read in the Deloitte & Touche Lecture Series on credit risk, University of Antwerp, February 2003, Belgium.

Taffler R.J., Tisshaw H., 1977, Going, Going, Gone - Four Factors Which Predict. Accountancy, Vol. 88, March 1977, p. 50-54.

Tamari M., 1966, Financial ratios as a means of forecasting bankruptcy. Management International Review, Vol. 4, p. 15-21.

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.

Theunissen L., 1999, How sensitive are statistical prediction models to changes in accounting methods and other misleading financial techniques? Proceedings of the Third international Stockholm Seminar on Risk Behaviour and Risk Management, Stokholm, June 1999.

Thornhill S., Amit R., 2003 (forthcoming), Learning about failure: Bankruptcy, firm age and the resource-based view. Organization Science.

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.

Van Caillie D., 1999, Business failure prediction models: what is the theory looking for? Paper presented at the Second International Conference on Risk and Crisis Management, May 1999, Liège, Belgium, p. 1-14.

Van Caillie D., Arnould S., 2001, The follow-up of financial value creation indicators to prevent bankruptcy in Belgian SMEs: an empirical multivariate approach. Paper presented at the 2001 Babson College Research Conference on Entrepreneurship, Jonköping, Sweden, June 2001.

Van Caillie D., Dighaye A., 2002, The concept of « economic added result », a new tool to prevent bankruptcy? Paper presented at the European Accounting Association Congress, April 2002, Copenhagen, Denmark, p. 1-30.

van Frederikslust R.A.I., 1978, Predictability of corporate failure: models for prediction of corporate failure and for evaluation of corporate debt capacity. Martinus Nijhoff Social Science Division, Leiden/Boston, PhD thesis in Economic Sciences, Erasmus University, Rotterdam, The Netherlands.

Van Wymeersch Ch., Wolfs A., 1996, La "trajectoire de faillite" des entreprises: une analyse chronologique sur base des comptes annuels. Working Paper, Université Notre-Dame de la Paix, Faculté des Sciences Economiques et Sociales, nr. 172, p. 1-32.

Ward T.J., Foster B.P., 1997, A note on selecting a response measure for financial distress. Journal of Business Finance and Accounting, Vol. 24, nr. 6, p. 869-879.

Warren E., Westbrook J.L., 1999, Financial characteristics of businesses in bankruptcy. American Bankruptcy Law Journal, Vol. 73, p. 499-590.

Wilcox J., 1971, A simple theory of financial ratios as predictors of failure. Journal of Accounting Research, Vol. 9, nr. 2, Autumn 1971, p. 389-395.

Wood D., Piesse J., 1987, The information value of MDA based financial indicators. Journal of Business Finance and Accounting, Vol. 14, nr. 1, p. 27-38.

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

Zavgren C.V., 1985, Assessing the vulnerability to failure of American industrial firms: A logistic analysis. Journal of Business Finance and Accounting, Vol. 12, nr. 1, Spring 1985, p. 19-45.

Zmijewski M.E., 1984, Methodological issues related to the estimation of financial distress prediction models. Journal of Accounting Research, Supplement to Vol. 22, p. 59-86.

## APPENDIX 1

The purpose of business failure prediction models (or failure classification models) is to classify firms into a failing group or a non-failing group of firms. Here, two types of classification errors can be made: a "type I error" or a "type II error". The table below summarizes the possible classification results, based on the most frequent presentation form of the classification results, the so-called "classification matrix" or "confusion matrix". If a failing firm is misclassified as a non-failing firm by the model, a type I error is made, and, if a non-failing firm is wrongly assigned to the failing group, a Type II error is made. In practice, both types of errors may have serious consequences. For example, when used in a credit decision context for a financial institution, refusing to grant a credit to a financially healthy firm (type II error) may cause the loss of future profits. This is why the type II error is also called 'commercial risk'. Approving a credit to a financially weak firm (type I error) may, on the other hand, cause a total loss situation, including the loss of due interests, the loss of the principal amount of the loan, juridical costs and opportunity costs (Altman et al., 1977; Joos et al., 1998a). This is why the type I error is often called 'credit risk'.

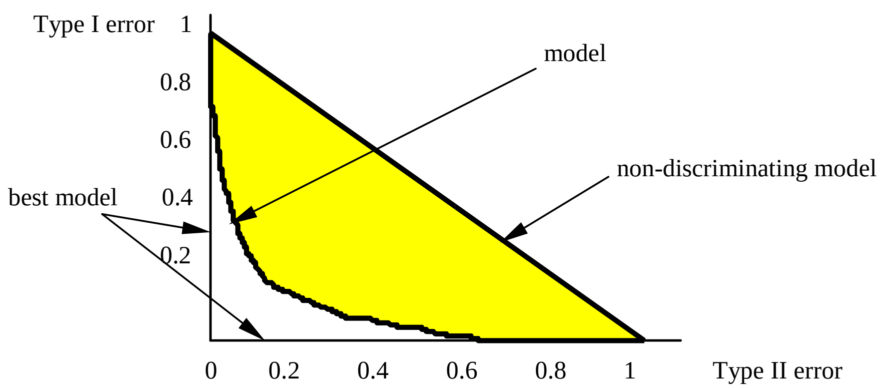
*Table : Classification results*

| | **Predicted group** | |
|---|---|---|
| **Actual group** | Failing | Non-failing |
| Failing | Correct classification (in %) | Type I error (in %) |
| Non-failing | Type II error (in %) | Correct classification (in %) |

## APPENDIX 2

The ROC represents the cumulative frequency distributions of the scores for the non-failing and the failing firms. The type II error rate (=$F_{non\text{-}failing}$ (y)) are on the X-axis and the type I error (=1-$F_{failing}$ (y)) are on the Y-axis (Steele, 1995), with $F_{failing}$(y) = cumulative distribution function of the scores of the failing firms and $F_{non\text{-}failing}$ (y) = cumulative distribution function of the scores of the non-failing firms. Each element of the trade-off function represents an optimal cut-off point for a given classification cost ($C_{Type\ I}$ and $C_{Type\ II}$) and population proportions ($\pi_{failing\ and}$ $\pi_{non\text{-}failing}$). It is clear that the best-performing (i.e., most discriminating) model has a trade-off function that coincides with the axes. By contrast, the non-discriminating model, which cannot distinguish between non-failing and failing firms, has a linear descending trade-off function from 100% type I error to 100% type II error. Comparing the location of the trade-off function of a failure prediction model with the location of the most discriminating and the non-discriminating models gives a clear indication of the performance of the model: a model has higher performance if its curve is located closer to the axes (Ooghe & Balcaen, 2002c).

***Figure: Trade-off function of a failure prediction model***



## APPENDIX 3

The Gini-coefficient is equal to the proportion of the area between the model and the non-discriminating model (i.e., the grey area in the figure in Appendix 1) and the area between the non-discriminating and the best model (i.e., the triangle with the axes as sides in the figure in Appendix 1) (Joos, Ooghe and Sierens, 1998). An empirical approximation of the Gini-coefficient is shown in the formula below (Joos, Ooghe and Sierens, 1998):

$$\hat{GINI} = \frac{\frac{x_{\max} y_{\max}}{2} - \sum_{i=1}^{n} (x_i - x_{i-1}) \frac{y_{i-1} + y_i}{2}}{\frac{x_{\max} y_{\max}}{2}}$$

$$= 1 - \sum_{i=1}^{n} (x_i - x_{i-1})(y_{i-1} + y_i)$$

with $x_i$ = type II error rate with cut-off point i;

$y_i$ = type I error rate with cut-off point i;

$x_{max}$ = maximum type II error rate, i.e., 100%;

$y_{max}$ = maximum type I error rate, i.e., 100%.

In a normal situation, this coefficient lies between zero and one. A high Gini-coefficient corresponds to a curve that is situated close to the axes, and hence, to a good performing model, while a low Gini-coefficient points out that the model performs badly. A negative Gini-coefficient implies that a model classifies most companies falsely.

# TABLE 1

**Overview of the classic statistical methods applied in corporate failure prediction**

| **Method** | **Main advantages** | **Main drawbacks** | **Failure prediction models** |
|---|---|---|---|
| Univariate analysis | simplicity | *assumes linearity<br>*'inconsistency problem'<br>*one-ratio model contradicts to multi-dimensional reality | Beaver (1967a) |
| Risk index models | *simple and intuitive<br>*multivariate model<br>*ratios are weighted | *subjective<br>*relative importance of ratios is unknown | Tamari (1966)<br>Moses and Liao (1987) |
| MDA models | *multivariate model<br>*continuous scoring model | *linearity assumption<br>*assumption on dichotomous dependent variable<br>*3 restrictive assumptions (multivariate normality; equal dispersion matrices; prior probabilities and misclassification costs)<br>*classification model (not prediction model!): ordinal scores<br>*relative importance of ratios is unknown<br>*not resistant to severe multi-collinearity | Altman (1968)<br>Altman et al. (1977)<br>Deakin (1972)<br>Edmister (1972)<br>Blum (1974)<br>Deakin (1977)<br>Taffler & Tisshaw (1977)<br>van Frederikslust (1978)<br>Bilderbeek (1979)<br>Dambolena & Khoury (1980)<br>Taffler (1982), model from 1974<br>Ooghe & Verbaere (1982)<br>Taffler (1983)<br>Micha (1984)<br>Betts & Belhoul (1987)<br>Gombola et al. (1987)<br>Gloubos & Grammatikos (1988)<br>Declerc et al. (1991)<br>Laitinen (1992)<br>Lussier (1994)<br>Altman et al. (1995) |
| Conditional probability models | *multivariate model<br>*continuous scoring model<br>*no distributional assumptions for independent variables – qualitative variables allowed<br>*no linearity assumption<br>*prior probabilities not required<br>*relative importance of ratios is known | *assumption on probability distribution<br>*assumption on dichotomous dependent variable<br>*classification model (not prediction !)<br>*extremely sensitive to multi-collinearity<br>*sensitive to extreme non-normality of independent variables<br>*sensitive to outliers and missing values | Ohlson (1980)<br>Zmijewski (1984)<br>Swanson & Tybout (1981)<br>Zavgren (1983)<br>Gentry et al. (1985a)<br>Zavgren (1985)<br>Keasey & Watson (1987)<br>Peel & Peel (1987)<br>Aziz et al. (1988)<br>Gloubos & Grammatikos (1988)<br>Keasey & McGuiness (1990)<br>Platt & Platt (1990)<br>Ooghe et al. (1993)<br>Sheppard (1994)<br>Lussier (1995)<br>Mossman et al. (1998)<br>Charitou et al. (2001)<br>Charitou & Trigeorgis (2002)<br>Lízal (2002)<br>Becchetti & Sierra (2003) |