Are Failure Prediction Models Widely Usable? An Empirical Study Using a Belgian Dataset

Hubert Ooghe

Vlerick Leuven Gent Management School and Ghent University, Belgium

Sofie Balcaen

Ghent University, Belgium

Faced with the question as to whether failure prediction models can easily be transferred and applied to a new data setting, this study examines the performance of seven models on a dataset of Belgian company failures after re-estimation of the coefficients. The validation results indicate that some models are widely usable: they are strongly predictive when applied to the new data set. The Gloubos-Grammatikos models and Keasey-McGuinness appear among the best performing models, and also Ooghe-Joos-De Vos and Zavgren seem to be widely usable, respectively for failure prediction 1 and 3 years prior to failure. At the same time, the Altman and Bilderbeek models show very poor results when applied to the Belgian dataset. The best performing models seem to combine the right variables in an intuitively right sense and it appears that the combination of some types of variables generally leads to good predictive results. On the contrary, the estimation technique, complexity and number of variables do not explain the predictive performances (JEL: G33,M49).

Keywords: failure prediction model, international comparison, validation, annual accounts, re-estimation.

I. Introduction

Over the years, failure prediction or financial distress models have been much discussed in accounting and credit management literature. From the late 1960s, when Altman (1968) and Beaver (1967) published their first failure prediction model, many studies have been devoted to the search for the most effective empirical method for failure prediction. In

many countries, not only in developed but also in developing countries, researchers have attempted to construct a good failure prediction model. Taffler and Tisshaw (1977) for the U.K., Altman et al. (1974) for France, Fernandez (1988) for Spain, Swanson and Tybout (1988) for Argentina are just some examples from the extensive list of studies on failure prediction models. A comprehensive overview of these studies can be found in Zavgren (1983), Altman (1984), Taffler (1984), Jones (1987), Altman and Narayanan (1997) and Altman and Saunders (1998). In Belgium, the first financial distress models were estimated in 1982 by Ooghe and Verbaere (1985).[1] In 1991, Ooghe, Joos and De Vos (1991) estimated a second generation of models. Ooghe, Joos and De Bourdeaudhuj (1995) give detailed overview of the literature on failure models in Belgium.

Recently, many papers comparing different modeling techniques (for example logit analysis, neural networks and decision trees) on the same dataset have been published, for example Bell, Ribar and Verchio. (1990), Altman, Marco and Varetto (1994), Curram and Mingers (1994), Joos, Ooghe and Sierens (1998), Kankaanpää and Laitinen (1999). In addition, some attention has been paid to the comparison of failure prediction models with different types of variables (Mossman et al., 1998).

When looking at the application of failure prediction models into practice, we find that many (international) financial information agencies apply failure prediction models on a totally different dataset of companies than the ones they are designed for. For example, the Altman model, which is originally designed to predict failure of large publicly traded manufacturing companies in the United States, is widely used to analyze the failure risk of different kinds of firms in several European countries. In this respect, one may well ask whether a given failure prediction model can easily be transferred and applied to totally new data settings. Consequently, we aim to compare the performance of a range of failure prediction models, on a dataset of Belgian company accounts over the sample period 1995 – 1999. We will identify those models and hence those combinations of variables that have the best predictive abilities in Belgium and put forward some possible explanations for the findings.

---

1. For an overview of financial distress studies in Belgium between 1982 and 1991, with an emphasis on those conducted at the Department of Corporate Finance of Ghent University, see Ooghe, Joos and De Bourdeaudhuij (1995).

It should be noted here that if we were to validate the models in their current form (i.e., with their original coefficients) on the Belgian data, we would not be able to determine whether the performance results of the models are a mere consequence of the choice of variables. Bad performance results could also be caused by coefficients that fail to capture the true relationships between the variables of the model and the failures of Belgian companies. In view of this, we will re-estimate the coefficients of all models over the Belgian data. As far as we know, there has been as yet no systematic comparative analysis of the predictive performances of failure prediction models on the same dataset using re-estimated coefficients.

This paper is divided into six parts. By way of introduction, section II gives a short explanation of the two modeling techniques that we focus on in this study: linear discriminant analysis and the logistic regression technique. In addition, the different performance measures that are used to examine the predictive abilities of the models are explained. Section III discusses the failure prediction models that are analyzed in this paper, with section IV focusing on the population and the sampling methodology. Section V reports the results of our empirical research and attempts to explain the findings. The final section will highlight the most important conclusions of this study.

II. Modeling Techniques and Performance Measures

A. Modeling Techniques

Modeling techniques for two-group classification in general, and failure prediction in particular, can generally be classified in four different groups: classical statistical techniques, recursive partitioning analysis (or tree classification), neural networks, and genetic algorithms.[2] The latter three classification methods may also be classified under the general heading of "inductive learning" (i.e., a learning process based on examples). It is more difficult to validate these kinds of models. As a result, this validation study only considers failure prediction models estimated with classical statistical techniques, such as linear discriminant analysis and logistic regression. A second reason why we

---

2. For a comprehensive summary of methodological issues on estimation and evaluation of credit scoring models, see Joos, Ooghe and Sierens (1998).

focus on linear discriminant analysis and logistic regression is because they are used in most failure prediction research, both in the earlier versions and in the most recent ones.[3]

Multiple discriminant analysis (MDA) compares the distribution of variables—which have a multivariate normal distribution—for different groups or populations (i.e., a failing and a non-failing group), which are known, identified and mutually exclusive (Altman et al., [1981]). It allows to estimate a model that consists of a linear combination of variables, which provides the best distinction between the group of failing and the group of non-failing firms. The essence of the MDA procedure is to assign a firm to the failing or the non-failing group based on its "(multivariate) discriminant score": it is assigned to the group it most closely resembles. The discriminant score, which has a value between $-\infty$ and $+\infty$, is the result of the combination of several financial characteristics of a company (variables) into one single score and gives an indication of the financial health of the firm. Classification is achieved as follows: a firm is classified into the failing group if its discriminant score is less than a certain cut-off point and it is classified into the non-failing group if its score exceeds or equals the cut-off point (Altman [1968], Lachenbruch [1975], Joos, Ooghe and Sierens [1998]).

The technique of logit analysis allows to estimate a conditional probability model. For each company, it estimates the probability of failure conditional on a range of characteristics (attributes) of the company, based on maximum likelihood estimation. This technique is based on the assumption that the failure probabilities follow the logistic distribution. The conditional probabilities have values between zero and one (on a sigmoidal curve) and are called "logit scores" (Hosmer and Lemeshow [1989]). On the basis of a logit score and a certain (optimal) cut-off point, a firm can be assigned to the failing or the non-failing group. Logit analysis is frequently used in classification studies because this method has some favorable qualities. For example, it is not necessary to adapt the method for disproportional samples (Zavgren [1983] and Joos, Ooghe and Sierens [1998]).[4]

---

3. In 1968, Altman started with his "Z-score" discriminant model, and the same risk analysis tool is still applied in the scoring models, developed by the Central Banks of, for example, Austria, France, Germany, Italy and the United Kingdom. The logistic regression technique was introduced at a later stage and is currently applied in both academic papers and in research from Central Banks.

4. In classification research, state-based samples (the probability of being selected

B. Performance Measures

The performance of a classification model indicates how well the model performs, and is called "goodness-of-fit" in the econometric literature. Two different kinds of performance measures will be discussed: measures based on a "classification rule", and measures based on the "inequality principle" (Joos, Ooghe and Sierens [1998]).[5] Because classification into a group of failing and non-failing companies is the main objective of a failure prediction model, it is obvious that performance measures based on a classification rule are frequently applied.

Measures Based on a Classification Rule

In this study, a high (logit or discriminant) score indicates a healthy financial situation, while a low score indicates a bad financial situation and hence a high failure probability. In this respect, a firm will be classified into the failing group if its score is lower than a certain cut-off point, while it will be classified into the non-failing group if its score is higher than the cut-off point. For a continuous score model, the classification rule can be formulated as follows:

$$y_i^* = \begin{cases} 1 \text{ if the logit or discriminant score } \hat{y}_i \text{ of firm } i > y^* \\ 0 \text{ if the logit or discriminant score } \hat{y}_i \text{ of firm } i \leq y^* \end{cases} \quad (1)$$

---

depends on the "state" of the firm, which is non-failing or failing) are often used instead of pure random samples. As, in the total population, the number of failing units is much smaller than the number of non-failing units, random sampling would lead to very small samples of failing firms and to inaccurate models.

5. Two performance criteria that are not used in this study are R^2-type measures and measures based on entropy (Joos, Ooghe and Sierens [1998]). Several R^2-type measures, which indicate the percentage of the variance that is explained by the model, are possible. The count R^2 measure, which reports the number of correctly and falsely classified firms, is the most suitable measure. However, these measures are not used, as the number of correctly and falsely classified firms is already measured by the Gini-coefficient. The concept of entropy originates from the information theory of Shannon (1948) and was originally introduced in econometrics by Theil (1971). Measures based on entropy were used as performance measures in failure prediction research by Zavgren (1985) and Keasey and McGuinness (1990). It has to be noted, though, that entropy measures only evaluate the discriminating ability of the model and do not allow taking misclassification costs and population proportions into account a posteriori. As a consequence, we do not use entropy measures in this study.

with y_i^* as the estimated class of firm I and y^* as the threshold or "cut-off point". Here, two types of misclassifications can be made: A type I error, denoted by E_f, represents a "credit risk" (a failing firm classified as a non-failing) and a type II error, denoted by E_{nf}, represents a "commercial risk" (non-failing firm classified as a failing).

In theory, the optimal cut-off point of a classification model is the point at which the "global cost function" is minimized. This global cost function, represented by equation 2, includes the misclassification error rates, the misclassification costs and the population proportions (Koh [1992]). The expected cost is:

$$EC = \pi_f C_f E_f + \pi_{nf} C_{nf} E_{nf} \ , \qquad (2)$$

where f is for failing, nf is for non-failing, π_f and π_{nf} are the population proportions of failing and non-failing firms, C_f and C_{nf} are the misclassification costs for the type I and type II errors, and E_f and E_{nf} are the type I and type II misclassification error rates (percentages).

However, the principal aim of this paper is not defining new optimal cut-off points for the models.[6] The scope of this paper is the comparison of predictive abilities and hence the comparison of error rates. In this respect, it is possible to assess the predictive performances of the models statistically, without taking into account the "subjective factors" (Steele [1995]) of misclassification costs and population frequencies. We examine the differences between the score distribution of the group of failing firms and score distribution of the non-failing group. This refers to "Dmax" of the Kolmogorov-Smirnov significance test for two samples (Siegel and Castellan [1988]). "Dmax" is the maximal difference between the cumulative distribution of the model scores of the failing group (i.e. F_f) and the cumulative distribution of the model scores of the non-failing group (i.e. F_{nf}). The score that corresponds to "Dmax" is reported as the "new optimal cut-off point" of the model. The misclassification errors for the group of failing firms (F_f) and the non-failing group ($1 - F_{nf}$), corresponding with this new, optimal cut-off point, are reported as the type I and the type II error rate. These are the key measures of performance used in this study.

As the maximization of Dmax (F_f–F_{nf}) is the same as the

---

6. Moreover, Ohlson (1980) states that logit analysis is not an econometric method which is designed to find an optimal frontier, trading off one type of error against another.

minimization of the unweighted sum of the misclassification errors ($F_f + 1 - F_{nf}$), the new optimal cut-off point is the point at which the (unweighted) sum of both misclassification error rates is minimal.[7] In section V, which reports the results, this unweighted sum of misclassification errors is reported as the "unweighted error rate" (UER). It should be stressed that the UER does not indicate the real percentage of the firms in the total population of failing and non-failing companies that is classified falsely by the models. The UER is only intended as a measure of accuracy, which is used to compare the predictive abilities of the models. In the comparative context of this study, abstraction is thus made of population proportions and misclassification costs.

Measures Based on the Inequality Principle

The performance of a model can also be demonstrated graphically with the construction of a trade-off function. Figure 1 shows an example. Here, the cumulative frequency distributions of the scores for the non-failing and the failing firms are located in a co-ordinate system, with the type II error rate ($= F_{nf}(y)$) on the X-axis and the type I error rate ($= 1 - F_f(y)$) on the Y-axis (Steele, 1995), where $F_f(y)$ is the cumulative distribution function of the scores of the failing firms and $F_{nf}(y)$ is the cumulative distribution function of the scores of the non-failing firms.

Each element of this trade-off function represents an optimal cut-off point for a given classification cost (C_f and C_{nf}) and population proportions (π_f and π_{nf}).

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.

The Gini-coefficient of a model is an aggregated performance measure that reflects the difference between the trade-off function of the

---

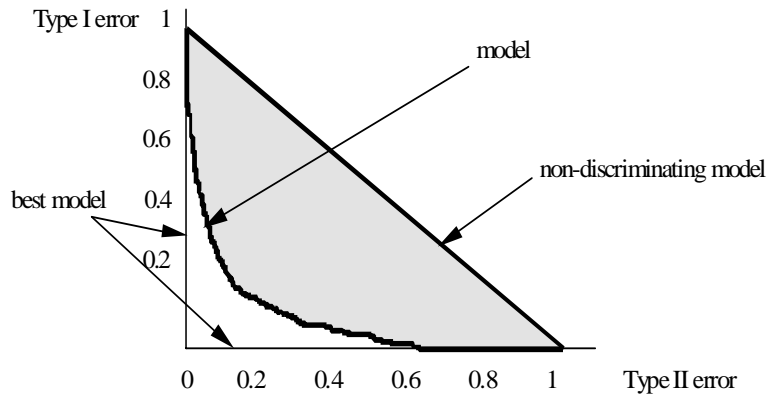7. Both terms are complementary: $F_{nf} - F_f + F_f + 1 - F_{nf} = 1$.

FIGURE 1.—Trade-off Functions of a Model

model and the trade-off function of the non-discriminating model. In a normal situation, this coefficient lies between zero and one and is equal to the proportion of the area between the model and the non-discriminating model (i.e., the grey area in figure 1) and the area between the non-discriminating and the best model (i.e., the triangle with the axes as sides). An empirical approximation of the Gini-coefficient is shown below (Joos, Ooghe and Sierens, [1998]):

$$GI\hat{N}I = \frac{\dfrac{x_{max}y_{max}}{2} - \displaystyle\sum_{i=1}^{n}(x_i - x_{i-1})\dfrac{y_{i-1} + y_i}{2}}{\dfrac{x_{max}y_{max}}{2}} \tag{3}$$

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

where x_i is the type II error rate with threshold I, y_i is the type I error rate with threshold I, x_{max} is the maximum type II error rate, i.e., 100%, y_{max} is the maximum type I error rate, i.e., 100%.

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.

Contrary to the discussion of the type I and type II errors separately, this measure gives a global judgement of performance: the Gini-coefficient is independent of changing cut-off points.

III. Failure Prediction Models

Because the aim of the study is to compare the validity of a range of failure prediction models, developed in different countries and time periods, we first select a number of models. In this selection process, several criteria are taken into account. Firstly, as already mentioned, this study only focuses on models estimated with linear discriminant analysis and logistic regression. Secondly, we restrict ourselves to the analysis of models that are frequently referred to in research papers. For example, on the basis of this criterion, the Altman (1968) model and the Zavgren (1983) model are included in this study. Thirdly, the availability of variable and coefficient information is an important criterion. As many recent models are licensed to commercial companies, they are not fully described in academic publications and therefore cannot be included in this study. The Taffler (1984) model is an example of a model that is excluded because of the unavailability of coefficients. Furthermore, the ease of use of the models with respect to the calculation of the variables is taken into consideration. Models that include non-financial data, such as gross national product, are left out of this study. For example, the Ohlson model (1980), which includes the GNP price index, is excluded from the analysis. In addition, this study only incorporates "developed country" models (Altman and Narayanan [1997]). Models from developing countries, where free market economies do not occur, fall outside the study, because we expect these models to show extremely large error rates when validated on Belgian annual accounts. In developing countries, it is difficult to detect company failure because of the degree of government protection. Finally, we opt for general models and hence exclude models investigating the probability of failure of, for example, new, high-tech or small firms. One example concerns the exclusion of the Laitinen model (1992), which was designed to predict failures of newly founded firms.

At the end of the selection procedure, eight models remain: Altman (1968) from the U.S.A., Bilderbeek (1979) from The Netherlands,

TABLE 1.   Samples of Failing and Non-failing Companies

Sample	Number of firms	Percentage
Failing firms	6,500	100.00
Failing in 1997	3,313	50.97
Failing in 1998	3,187	49.03
Non-failing firms	249,334	100.00
Termination of activity	122	0.05
Early dissolution- liquidation	5,795	2.32
Liquidation followed by a merger	44	0.02
Liquidation followed by an absorption	3,148	1.26
Closing of a liquidation	16,674	6.69
Without any particular legal status	223,552	89.66

Ooghe-Verbaere (1985) from Belgium, Zavgren (1985) from the U.S.A., Gloubos-Grammatikos (1988) from Greece, Keasey-McGuinness (1990) from the U.K., and Ooghe-Joos-De Vos (1991) from Belgium. In the appendix, table 1 summarizes the characteristics of each of these eight models, and in tables 2 to 8 the variables of the models are shown. In these tables, each different variable is attributed a unique name (X1, X2, …, X40). Detailed analysis of the variables reveals that the some variables appear in different models.

However, comparing the performances of the different models in their current form (i.e., with their original coefficients) will cause some difficulties. We should bear in mind that generally models perform badly out-of-sample when there is a large difference between the estimation sample and the validation sample, especially if the differences are the result of differences in the definitions of the dependent variables, exogenous factors (such as macroeconomic conditions and institutional and legal factors unique to the country of origin) and the sample period. It is clear that the calculation of the original coefficients of the various models is influenced by the correlations between the variables that are included in the original estimation samples and the variables that are not included in these samples, such as macroeconomic conditions, institutional factors and sample period. Consequently, if we would validate the models with their original coefficients on the Belgian data, we would be unable to determine whether bad performance of the models is the consequence of an inappropriate choice of the variables or of the use of coefficients that fail to capture the true relationship between the variables and

failures of Belgian companies. The re-estimation of the coefficients of all models over the Belgian data, will allow the models (i.e., combinations of variables) to take into account some factors specific to the Belgian validation dataset. As a result, we will be able to compare model performances more precisely, and to indicate whether the performance results can be explained by the choice of variables.

IV. Population and Samples

Before describing the population and the sampling method for re-estimation and validation, it seems appropriate to give some important definitions that are frequently used in this study.

A. Definitions of Failing and Non-failing Firms

A "failing" firm is a firm in the situation of bankruptcy, or with a request for a judicial composition (also called "moratorium on payments"), or with an official approval of a judicial composition. On the other hand, besides "normal" firms without any particular legal status, the group of "non-failing" firms also includes all firms characterized by the following juridical situations: termination of activity, early dissolution-liquidation followed by a merger with another company, liquidation followed by absorption by another company, closing of a liquidation, or without any particular legal status. In other words, all firms with associated doubts about the economic reasons for their juridical situation are included in the non-failing population. As it is our aim to validate failure prediction models, it is necessary to measure the performance of the models in a realistic situation, and hence consider these doubt-causing firms as non-failing ones. However, when re-estimating the coefficients of the failure prediction models, we want to reduce the influence of these doubt-causing firms. Therefore, in the re-estimation sample, which is discussed further, only firms without any particular legal status are included in the group of non-failing firms and all other firms are ruled out.

B. Population and Samples of Failing and Non-failing Companies

This study is based on Belgian accounting data from the period 1994 – 1999. It concerns published annual accounts of non-financial companies

subject to the legislation on the annual accounts of companies. The data were obtained from the CD-ROMs of Bureau Van Dijk and information supplier Graydon. It should be noted that, in Belgium, companies are required to deposit their annual accounts in a prescribed form, dependent on their size. A distinction is to be made between large firms that must prepare their annual accounts in a complete form, and small firms that are allowed to prepare their annual accounts in an abbreviated form. The group of large companies consists of firms with more than 100 employees and firms that meet at least two of the following three criteria: (1) more than 50 employees (yearly average), (2) turnover, V.A.T. excluded, of more than 200 million Belgian francs (yearly average), (3) total assets of more than 100 million Belgian francs. In Belgium, a major percentage of the companies have annual accounts in an abbreviated form: in 1999, only 6.2% of the total population of companies deposited a complete-form annual account.

The total population of companies consists of all firms having published at least one annual account in the period 1994 – 1999. Companies that are classified into the following activity classes, are excluded because of their special situations: financial intermediation, insurance and pension funding, management activities of holding companies and co-ordination centers, public administration, public defense, public services to the community, compulsory social security, education, health and social work, activities of membership organizations, private households with employed persons, and extra-territorial organizations. The total population comprises 268,465 companies, identified by their V.A.T. numbers. From this total population, two samples are taken: a sample of failing companies and a sample of non-failing firms.

The sample of failing companies consists of all firms that failed in 1997 or 1998. Only firms that failed in 1997 having annual accounts in 1994 or later, and firms that failed in 1998 having annual accounts in 1995 or later, are included. The failing sample comprises 6,500 companies. On the other hand, the sample of non-failing companies includes all firms that are non-failing on January 1, 1999, and that have annual accounts in 1994 or later. The non-failing sample involves 249,334 companies. Table 1 illustrates the number of failing and non-failing companies that are used in this study. It also reports the percentage of the non-failing sample that is made up of companies characterized by the judicial situations mentioned in the list in section A.

TABLE 2. Sampling Procedures

	Failing group		Non-failing group	
	Failing firms	Year annual accounts	Year annual accounts	Non-failing firms
1 *ypf*	Failing in 97	1996	1996	group C
	Failing in 98	1997	1997	group D
2 *ypf*	Failing in 97	1995	1995	group B
	Failing in 98	1996	1996	group C
3 *ypf*	Failing in 97	1994	1994	group A
	Failing in 98	1995	1995	group B

C. Samples of Failing and Non-failing Annual Accounts

The sampling procedure for the samples of failing annual accounts is rather simple. Because the aim of the study is to re-estimate and to validate the models one, two and three years prior to failure (i.e., 1 *ypf*, 2 *ypf* and 3 *ypf*), we select the annual accounts one, two and three years prior to failure (if available and if not concerning an extended fiscal year) for each company in the failing sample. However, not all companies deposit their annual accounts on December, 31. Consequently, the annual accounts 1, 2 and 3 *ypf* are defined as follows:

Account one year prior to failure: account with the closing date falling within the period [date of failure, date of failure – 365 days]
Account two years prior to failure: account with the closing date falling within the period [date of failure – 365 days, date of failure – (2 × 365 days)]
Account three years prior to failure: account with the closing date falling within the period [date of failure – (2 × 365 days), date of failure – (3 × 365 days)]

To select the samples of non-failing annual accounts, the group of non-failing companies is randomly divided into four equal groups: groups A, B, C and D. For each group of companies, the annual accounts of one specific year in the period 1994 – 1997 (if available and if not concerning an extended fiscal year) are taken. The following non-failing annual accounts are selected:

TABLE 3.   Number of Annual Accounts in the Samples of Failing and Non-failing Annual Accounts

	Failing annual accounts	Non-failing annual accounts
Sample 1 *ypf*	6,500	124,671
Sample 2 *ypf*	6,500	124,678
Sample 3 *ypf*	6,500	124,684

Non-failing firms in group A: annual accounts of 1994
Non-failing firms in group B: annual accounts of 1995
Non-failing firms in group C: annual accounts of 1996
Non-failing firms in group D: annual accounts of 1997

In this study, we link the failing annual accounts one, two and three years prior to failure to the non-failing annual accounts, bearing in mind that the annual accounts of the two different samples should refer to the same time frame. Accordingly, for each year prior to failure, the annual accounts of the two relevant years are taken together. This procedure is explained in table 2. The resulting numbers of annual accounts in the samples of failing and non-failing accounts are reported in table 3.

Because we want to re-estimate the coefficients of the eight models on the Belgian data before validation, we need validation and re-estimation samples of failing and non-failing annual accounts. Consequently, the samples of failing and non-failing annual accounts are randomly divided into separate re-estimation samples and validation samples. Within each sample of failing and non-failing annual accounts, 50% of the accounts are classified as a re-estimation sample, and 50% are included in a validation sample.

As already mentioned, in the re-estimation samples, the annual accounts of "doubt-causing" firms must be excluded from the group of non-failing annual accounts. Only companies without any particular legal status should be considered as non-failing. Furthermore, we are forced to reduce the large number of non-failing annual accounts in the re-estimation samples because of the practical limitations of the statistical program used. In this respect, about 20% of the non-failing annual accounts are selected randomly. In contrast, the number of failing annual accounts is not reduced. Finally, we eliminate all annual accounts that have not been deposited at the National Bank of Belgium

TABLE 4. **Number of Failing and Non-failing Annual Accounts in the Re-estimation Samples**

	Failing	Non-failing	Total
Sample 1 *ypf*	778	9,164	9,942
Sample 2 *ypf*	2,409	8,590	10,999
Sample 3 *ypf*	2,705	7,932	10,637
Total sample	5,892	25,686	31,578

and therefore are not available on the CD-ROMs of Bureau Van Dijk. This significantly reduces the original number of failing annual accounts in the sample, at 1 *ypf* in particular, as many failing companies cease to pay attention to financial reporting when they are close to failure. Also, the annual accounts of non-failing companies founded after 1 January, 1998 are excluded, because these companies do not have annual accounts in the period 1994 – 1997. Table 4 presents the number of failing and non-failing annual accounts in the 1 *ypf*, 2 *ypf* and 3 *ypf* estimation samples after the elimination of non-available annual accounts.

It is clear that some models have different variables and coefficients depending on the period within which they aim to predict failure. Ooghe-Verbaere, Keasey-McGuinness and Ooghe-Joos-De Vos contain different variables for failure predictions one, two and three years prior to failure. These models are re-estimated on the basis of the corresponding re-estimation samples, being the samples 1, 2 and 3 *ypf*. In addition, the Zavgren model, which uses the same variables but different coefficients for failure prediction one, two and three years prior to failure, is re-estimated on the basis of these three samples. On the other hand, the models that make no distinction between failure prediction one, two and three years prior to failure—being Altman, Bilderbeek, and Gloubos-Grammatikos discriminant and logit—use the same variables and coefficients independent of the year prior to failure. They only have "general" coefficients and hence are re-estimated on a "general" sample (i.e., "total sample" in table 4), which is composed of all annual accounts in the 1 *ypf*, 2 *ypf* and 3 *ypf* samples.

The validation samples of failing and non-failing annual accounts are taken in much the same way as the re-estimation samples. As with the re-estimation samples, the number of non-failing annual accounts is

**TABLE 5.   Number of Failing and Non-failing Annual Accounts in the Validation
            Samples**

	Failing	Non-failing	Total
Sample 1 *ypf*	738	9,943	10,681
Sample 2 *ypf*	2,445	9,391	11,836
Sample 3 *ypf*	2,653	8,987	11,640

much too large to be able to use the statistical program. Again, this
large number of non-failing annual accounts must be reduced: about
20% of the non-failing annual accounts are selected randomly. In
addition, all annual accounts that have not been deposited are excluded
from the validation samples. Again, this reduces the original number of
failing annual accounts in the sample 1 *ypf* significantly. Table 5 reports
the number of failing and non-failing annual accounts in the 1 *ypf*, 2 *ypf*
and 3 *ypf* validation samples after the elimination of non-available
annual accounts.

On the basis of each annual account in the re-estimation samples, we
calculate a range of variables or ratios (i.e., variables X1 to X40 referred
to in tables 2 to 8 in the appendix) to re-estimate the coefficients of the
models. On the other hand, on the basis of each annual account in the
validation samples, we compute a (logit or discriminant) score for each
model to determine the model performance. Here, it is important to
mention the influence of invalid observations, both in the re-estimation
and in the validation process.

Firstly, a detailed examination of the data concerning the variables
reveals a frequent occurrence of invalid variables, caused by zero values
in the denominators of the variables. This is particularly the case if the
denominator of a variable contains sales or inventories. According to
Belgian accounting law, approximately half of the small companies,
publishing their results in an abbreviated form, only state their gross
margin as they are not obliged to publish sales and operating costs.
Furthermore, some types of companies (for example service firms)
simply do not have inventories. As a result, when re-estimating each of
the models, a certain percentage of the annual accounts in the
re-estimation samples show invalid observations for some variables, and
hence cannot be used. Table 6 gives an overview of the percentage of
annual accounts that could be used for the re-estimation of each of the

TABLE 6. Number and Percentage of the Valid Annual Accounts in the Re-estimation Samples

Model	Year before failure	Number	Percentage*
Altman		31,235	98.9
Bilderbeek		14,292	45.3
Ooghe-Verbaere	1 *ypf*	9,755	98.1
	2 *ypf*	4,805	43.7
	3 *ypf*	10,192	95.8
Zavgren	1 *ypf*	2,620	26.4
	2 *ypf*	3,122	28.4
	3 *ypf*	3,112	29.3
Gloubos-Grammatikos	logit	31,504	99.8
	discriminant	31,132	98.6
Keasey-McGuinness	1 *ypf*	4,064	40.9
	2 *ypf*	4,966	45.1
	3 *ypf*	4,615	43.4
Ooghe-Joos-De Vos	1 *ypf*	9,703	97.6
	3 *ypf*	10,594	99.6

Note: *Percentage of the total number after elimination of the non-available annual accounts (see table 4).

models. For example, when re-estimating the Zavgren model, less than 30% of the annual accounts can be used, because the model uses ratios containing sales and inventories in their denominator. Furthermore, when re-estimating other models with ratios containing "sales" in their denominator (Bilderbeek, Ooghe-Verbaere 2 *ypf*, and Keasey-McGuinness), less than 50% of the annual accounts in the re-estimation samples can be used.

In this respect, we should bear in mind that, as the models use different variables, they are not re-estimated on the basis of the same samples of annual accounts. However, this causes no problems for the re-estimation process, because it is our aim to calculate the re-estimated coefficients for each of the models as precisely as possible. Consequently, we want to include as many annual accounts as possible for each individual model.

Secondly, detailed analysis of the data on the logit and discriminant scores also reveals the presence of invalid scores, caused by invalid variables. However, when validating the models it is important that the performance results are based on the same samples of annual accounts, as it is our aim to compare the results of the different models on an

equal basis. Consequently, all annual accounts that show an invalid score for at least one model are excluded from the validation samples. Only the annual accounts that have valid scores for each model are selected. Table 7 reports the number of annual accounts that are finally included in the validation sample.

V. Results and Interpretation

This section discusses the validation results of the different failure prediction models on the dataset of Belgian companies. Section A gives some preliminary remarks on the signs of the coefficients of the variables. As an illustration, table 9 in the appendix shows the validation results obtained by the authors of the models in their original studies. The performance results of the different models using the re-estimated coefficients are reported in section B. We discuss the type I, type II, and unweighted error rates corresponding to the optimal cut-off points of the models, and we calculate Gini-coefficients. We also give a graphical illustration of the results by means of trade-off functions. Finally, in section C, we give some possible explanations for the performance results.

A. Preliminary Remarks on the Signs of the Re-estimated Coefficients

The failure prediction models are re-estimated using the re-estimation samples and are attributed new coefficients and new optimal cut-off points. These new coefficients and cut-off points are reported in tables 2 to 8 in the appendix. A detailed analysis of the signs of the original and the new coefficients reveals that some of these signs do not correspond to expectations. In the following original models, some variables have counter-intuitive coefficients: variables X6 and X5 in the Bilderbeek model (see table 3); variables X19, X21, X6 and X22 in the Zavgren model (see table 5); variables X23 and X11 in the Gloubos-Grammatikos model (see table 6); variables X4 and X18 in the Keasey-McGuinness model (see table 7).

On the other hand, the following re-estimated models also have unexpected coefficients for certain variables: variables X1 and X5 in the Altman model (see table 2); variables X6, X7, X8 and X2 in the Bilderbeek model (see table 3); variables X9 and X14 in the 3 *ypf* Ooghe-Verbaere model (see table 4); variables X19 (2 *ypf* and 3 *ypf*),

TABLE 7. Number and Percentage of the Valid Annual Accounts in the Validation Samples

	Failing Number	Percentage*	Non-failing Number	Percentage*	Total Number	Percentage*
Sample 1 *ypf*	218	29.5	2,487	25.0	2,705	25.3
Sample 2 *ypf*	799	32.7	2,466	26.3	3,265	27.6
Sample 3 *ypf*	868	32.7	2,302	25.6	3,170	27.2

Note: *Percentage of the total number after elimination of the non-available annual accounts (see table 5).

TABLE 8. Performance Results of the Models with the Re-estimated Coefficients

	Error rate Type I*	Type II*	Unweighted error rate UER*	UER-based ranking	Gini Gini-coefficient*	Gini-based ranking
1 *ypf*						
Altman	60.55	28.47	44.51	7	9.32	7
Bilderbeek	97.71	0.56	49.13	8	–5.25	8
Ooghe-Verbaere	30.28	38.20	34.24	6	42.70	6
Zavgren	10.09	51.19	30.64	5	45.92	5
Gloubos-Grammatikos logit	36.24	22.11	29.18	4	50.44	4
Gloubos-Grammatikos discriminant	27.98	29.19	28.59	3	52.78	3
Keasey-McGuinness	20.18	28.07	24.12	1	64.77	1
Ooghe-Joos-De Vos	22.94	31.52	27.23	2	56.78	2
Average	38.24	28.66	33.45		39.68	
2 *ypf*						
Altman	55.19	38.93	47.06	6	5.46	6
Bilderbeek	2.00	97.32	49.66	7	–4.41	7
Ooghe-Verbaere	30.66	45.05	37.86	5	26.69	5
Zavgren	22.65	50.69	36.67	4	34.31	4

(*Continued*)

TABLE 8. (Continued)

	Error rate Type I*	Type II*	Unweighted error rate UER*	UER-based ranking	Gini Gini-coefficient*	Gini-based ranking
Gloubos-Grammatikos logit	30.16	38.89	34.53	3	41.91	3
Gloubos-Grammatikos discriminant	33.29	32.32	32.81	1	45.17	1
Keasey-McGuinness	24.03	43.80	33.91	2	42.94	2
Average	28.28	49.57	38.93		27.44	
3 ypf						
Altman	7.14	89.05	48.10	7	1.22	7
Bilderbeek	58.76	39.01	48.88	8	−0.81	8
Ooghe-Verbaere	37.90	43.53	40.72	6	24.45	5
Zavgren	17.63	56.47	37.05	1	31.75	2
Gloubos-Grammatikos logit	41.24	36.88	39.06	3	29.38	3
Gloubos-Grammatikos discriminant	50.69	25.41	38.05	2	32.51	1
Keasey-McGuinness	16.59	64.21	40.40	5	22.43	6
Ooghe-Joos-De Vos	32.95	45.92	39.43	4	25.27	4
Average	32.86	50.06	41.46		20.78	

Note: *Indicates percentages.

X21 (1 *ypf*), X6 (1 *ypf* and 2 *ypf*), X22 and X5 in the Zavgren model (see table 5); variables X26 (1 *ypf*), X18 and X5 in the Keasey-McGuinness model (see table 7); and, finally, variables X32 and X40 in the Ooghe-Joos-De Vos model (see table 8).

When making an overview of the variables with counter-intuitive coefficients, we find that the following models have at least two coefficients with unexpected signs and significant values of at least 0,001: Altman, Bilderbeek, Ooghe-Verbaere 3 *ypf*, Zavgren 1 *ypf*, and Ooghe-Joos-De Vos 3 *ypf*.[8] On the other hand, all coefficients in the Gloubos-Grammatikos discriminant and logit models and in Keasey-McGuinness 1 and 2 *ypf* have the expected, intuitive sign or have a value of 0,000.

The multivariate contexts in which the models are estimated allows for these unexpected signs. When we estimate the coefficients with the linear discriminant or logistic regression technique, we simply determine the combination that statistically offers the optimal classification of failing and non-failing firms. The coefficients can not be interpreted as the relative importance of the variables. As the variables may be highly correlated, the models are difficult to interpret. Moreover, due to the multivariate context, the (positive or negative) influence of some variables may counterbalanced by the (negative or positive) influence of other variables. Also, if the original estimation samples or the re-estimation samples in this study include any extreme ratio values for firms that are not representative for the firm"s true situation concerning failure/non-failure, the coefficients of those variables with extreme values may be biased.[9]

On the other hand, firm size and industry concentration may be a possible explanation to the unexpected signs. For example, the variable sales/total assets has an unexpected negative sign in the re-estimated Altman and Zavgren models. This negative sign may possibly be explained by the fact that the re-estimation sample of failing companies used in this study contains a very high concentration of (small)

---

8. In this study, only the coefficients with an absolute value of more than 0.001 are considered to be significant and to have a significant impact on the final model score.

9. In this respect, extreme ratio values may also influence the classification accuracy of the models. When trying to classify a firm with an extreme ratio value, the wrong signal provided by the non-representative ratio may counterbalance and even overwhelm the correct signals by the other ratios, which may result in an incorrect classification of the firm by the model (Moses and Liao [1987]).
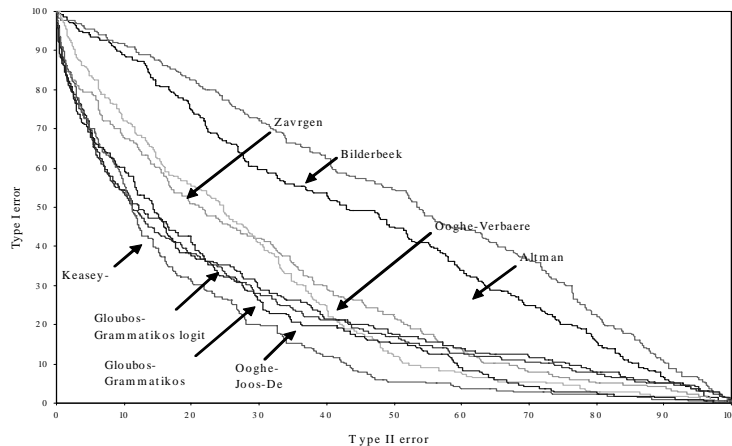
FIGURE 2.—Trade-off Functions of the models 1 Year Prior to Failure (re-estimated Coefficients).

companies from industries like restaurants and catering, wholesale and retail with traditionally high sales. The higher the sales/total assets ratio, the more likely the firm belongs to an industry with a high failure rate and hence, the worse the financial situation.

B. Performance Results

The validation results of the models with their re-estimated coefficients are shown in table 8. The best-performing models are indicated in bold letters, while the worst-performing models are printed in italic. Firstly, table 8 reports the type I, type II and unweighted error rates corresponding to the new optimal cut-off points of the models. Besides the unweighted error rate, the Gini-coefficient is also used to assess the fit of the models. The analysis of the unweighted error rates on one hand and the Gini-coefficients on the other hand generally leads to the same conclusions: we notice only small differences in the rankings of the performance results of the models. Finally, we plot the trade-off functions of the models 1, 2 and 3 *ypf* using the re-estimated coefficients in figures 2, 3 and 4.

The performance results of the models one year prior to failure indicate Keasey-McGuinness as the best performing model. In addition,
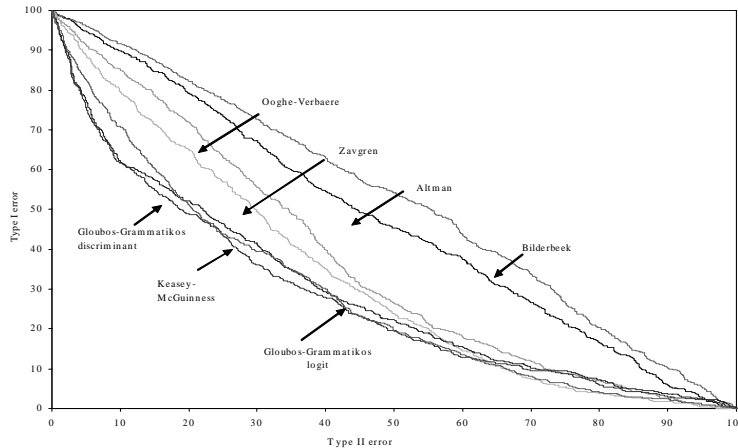
FIGURE 3.—Trade-off Functions of the Models 2 Years Prior to
Failure (Re-estimated Coefficients).

Ooghe-Joos-De Vos, Gloubos-Grammatikos discriminant, and Gloubos-
Gramatikos logit reveal very small unweighted error rates. Figure 2
concerning the short-term 1 *ypf* models also indicates these four models
as the ones with the best predictive abilities. On the contrary, the
Bilderbeek and Altman models seem to perform worst: they show the
highest error rates, the lowest Gini-coefficients and trade-off functions
that are located very far from the axes. The Bilderbeek model even has
a negative Gini-coefficient, which means that the model classifies most
companies falsely.

    With respect to the models two years prior to failure, Gloubos-
Grammatikos discriminant performs best, followed closely by Keasey-
McGuinness and Gloubos-Grammatikos logit. Similarly, figure 3
indicates Gloubos-Grammatikos discriminant and logit and Keasey-
McGuinness as the best-performing 2 *ypf* models. Furthermore, just as
in the short term case, Bilderbeek and Altman are the worst failure
prediction models.

    Examination of the performance results of the models three years
prior to failure and the trade-off functions in figure 4, reveals that the
Zavgren and the Gloubos-Grammatikos discriminant models are the
ones that perform best with respect to long term failure prediction.
Again, Bilderbeek and Altman clearly are the worst failure predictors.

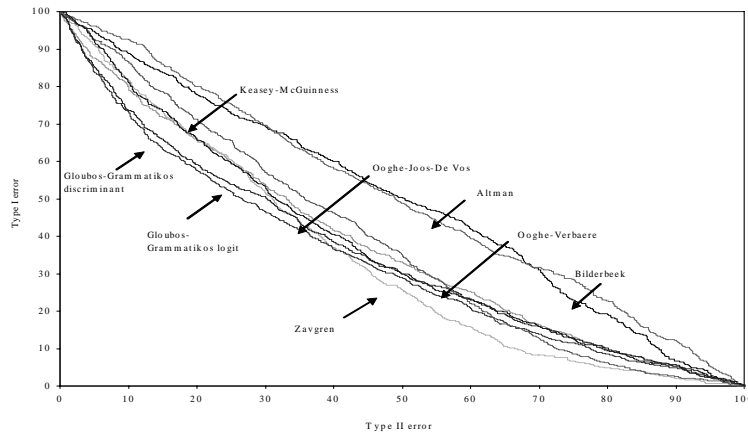   The validation results 1, 2 and 3 *ypf* indicate that some models are

FIGURE 4.—Trade-off Functions of the Models 2 Years Prior to
Failure (Re-estimated Coefficients).

widely usable: they can be applied to new data sets with a high
predictive performance. At the same time, the Altman and Bilderbeek
models show very poor results when applied to the new Belgian dataset.

When taking a closer look at the performance results in table 8 and
the trade-off functions, a couple of additional remarks can be made.

Firstly, it is noteworthy that if the term of the failure prediction is
longer, the average unweighted error rate of the models increases, and
the average Gini-coefficient decreases: the average UER of the short
term models is significantly lower than the average UER of the 2 *ypf*
and 3 *ypf* models. This finding is not surprising, as it is generally
believed that it is easier to predict failure, and hence discriminate
between failing and non-failing companies in the short term.[10] Specific
features of failing companies are less pronounced three years before
failure than they are one year before failure.

Secondly, when analyzing the general distribution of the trade-off
functions it is immediately apparent that the performance differences
depend on the term of failure prediction. The difference between the
trade-off functions of the best and the worst performing short term (1

---

10. The original Altman (1968) study and virtually every study since then has shown that
it is easier to predict failure as the failure data approaches.

ypf) models is extremely large. The performance differences for the 2 *ypf* models are significantly smaller and, finally, the long term (3 *ypf*) models reveal the smallest performance differences. Consequently, if the term of failure prediction is longer, it seems to be more difficult to make a distinction between good and bad performing models.

C. Possible Explanations

This section will analyze which factors explain why some models seem to be widely applicable, while others don't. As mentioned earlier, applying the re-estimated coefficients when validating the models allows us to eliminate the impact of the original coefficients on model performance. If we would apply the original coefficients, a bad performance result could simply be caused by model coefficients that fail to capture the true relationship between the independent variables and the failure of Belgian companies (even if the model includes the "right" variables and is based on the "right" modeling technique). Nevertheless, there are still several possible explanatory factors for the individual model performances. First of all, we will analyze the type of variables included in the models and try to indicate those common types of variables leading to the highest prediction accuracies. Secondly, we will take a closer look at some possible explanatory factors related to the model construction: (1) the estimation technique, (2) the complexity of the variables, and (3) the number of variables.

The conclusions reported in this section should, however, be treated with care. This study concerns only a limited number of models, which makes it impossible to reach any generally valid conclusions. Moreover, on the basis of this study, it is impossible to gain a clear insight into the real impact of these factors, because many different explanatory factors may interact and may influence each other. Clear statements concerning the real explanatory power of a certain factor requires a thorough investigation of many different models and systematic comparisons holding all other factors constant.

Type of Variables

First, it seems that both bad performing models—Altman and Bilderbeek—include the variables ("accumulated profits or losses" + retained earnings)/total assets' and "sales/total assets". Nevertheless, these variables do not seem to be an important explanatory factor for the poor predictive abilities of the models. The "sales/total assets" variable

is also present in the good performing Zavgren model and the good performing Ooghe-Joos-De Vos model also contains a variable very similar to the "accumulated profits and losses" variable.

On the other hand, when taking a closer look at the good performing models, there seem to be some common types of variables. On a term of three years, variables concerning the cash-liquidity situation of the company play an important role, besides variables concerning the debt situation and concerning gross return. On a term of two years, the models that are widely usable appear to include working capital variables, besides variables concerning the long term debt situation and gross and net return. Finally, good performing short term (1 *ypf*) failure models seem to include working capital and debt variables, besides gross return variables and variables concerning short term obligations (taxes, social security debts and short term financial debts).

When determining the common types of variables, we should keep in mind that the multivariate context in which the variables are used, probably is of great importance. The combination of variables may be the most important explanation for the model performances. In this respect, the overall good performance of the Gloubos-Grammatokos logit and discriminant models may be explained by the combination of the variables X1, X24 and X11, which are respectively a working capital variable, a long term debt variable and a gross return variable. On the other hand, the poor performance of the Altman model may be explained by the fact that the combination of some strong failure indicators (a working capital variable, a gross return variable and a debt situation variable) is counterbalanced by the impact of the poor indicator "sales/total assets", which is also included in the model.

Finally, when looking at the overview of the coefficients with counter-intuitive signs, there appears to be a strong association between the presence of intuitive signs and good performance results on the Belgian data set. All variables in the models that perform best (i.e., Gloubos-Grammatikos discriminant 1, 2 and 3 *ypf*; Gloubos-Grammatikos logit 1 *ypf* and 2 *ypf*; Keasey-McGuinness 1 *ypf* and 2*ypf*) seem to have a coefficient with the expected, intuitive sign (or with a value of 0.000). On the other hand, the models with the worst accuracy rate seem to have many coefficients with counter-intuitive signs. This finding makes a strong case for using the predictor variables "in the right sense", instead of statistically building models that seem to fit the data but neglect "wrong" coefficient signs.

In this respect, an important point to bear in mind is that the choice of the variables in the models is determined by the original estimation

samples that were used. There is little doubt that the industry concentration, the size class concentration, the definition of failure and the time period of the data in the original estimation samples and the social-cultural and legislative characteristics of the country of origin have an important impact on the chosen combination of certain types of variables and therefore could possibly explain the poor or good predictive performances of the models on the new dataset of Belgian companies. At first sight, the industry concentration appears to be an important explanatory factor. For example, the fact that both worst performing models were developed on a sample of (mainly) industrial companies might partly explain why these models perform badly on the Belgian sample, which is characterized by a high concentration of trade firms (i.e., wholesale and retail firms) and service companies. Besides the industry concentration, the time frame of the estimation sample seems to be related to model performance. Both worst models—Altman and Bilderbeek—are estimated much earlier than the other models and use much older data. On the contrary, the influence of size class seems to be rather limited. Not only the poorly performing Altman model, but also the good performing Zavgren and Keasey-McGuinness models are estimated on a sample of large companies. Likewise, an examination of the definitions of failure used for the development of the models, reveals no remarkable differences. Finally, although we expected that the performance results of the models would be related to the social-cultural and legislative characteristics of the country of origin, this study reveals no significant association. The best performing models are models from Greece, the U.K., Belgium and the U.S.A.

Construction of the Model

First, based on this study, we might argue that the estimation technique (i.e., discriminant or logistic regression) does not have a major influence on model performance. Although the worst-performing models—Altman and Bilderbeek—both are estimated using linear discriminant analysis, also the widely usable Gloubos-Gramatikos discriminant model is based on this technique. On the other hand, all models that can easily be used on the Belgian data set are based on logistic regression.

   Second, when analyzing the complexity of the variables included in the various models, it appears that Altman and Bilderbeek—the worst-performing models—only include very simple variables, which can be calculated with a small number of annual account items. On the

other hand, also good performing models, like the 1 *ypf* Keasey-McGuinness model and the 3 *ypf* Zavgren model, exclusively use very simple variables. Consequently, the complexity of the variables does not seem to be an explanatory factor for model performance.

Third, the number of variables does not seem to have an important impact on the model performance. The worst performing Altman and Bilderbeek models consist of five variables. If we take a look at the best performing models one year prior to failure, it seems that some of the models are composed of five variables or more (Gloubos-Grammatikos discriminant has five variables and Ooghe-Joos-De Vos contains eight items), while Keasey-McGuinness and Gloubos-Grammatikos logit contain only three variables. Also for failure prediction two and three years prior to failure, the number of variables does not appear to be of great importance.

VI. Summary and Conclusions

By examining the validity of a range of eight failure prediction models on a dataset of Belgian company accounts, the aim of this study was to answer the question whether failure prediction models can be widely transferred and applied to other data sets of company accounts. The failure prediction models first were re-estimated over the Belgian dataset, which allowed determining whether the performances are the result of the choice of variables and/or of the applied modeling technique. The predictive abilities of the models were assessed on the basis of several performance indicators: the type I, type II and unweighted error rates (UER) and the Gini-coefficients. Finally, the trade-off functions provided a graphical presentation of the results.

The validation results indicated that some models are widely usable: they can be applied to a data set totally different from the original estimation sample with a high predictive performance. More in particular, one year prior to failure, Keasey-McGuinness, Ooghe-Joos-De Vos, and both Gloubos-Grammatikos models (discriminant and logit) performed best. Two years prior to failure, Gloubos-Grammatikos discriminant performed best, followed closely by Keasey-McGuinness and Gloubos-Grammatikos logit. Finally, Zavgren and the Gloubos-Grammatikos discriminant model had the best predictive performances three years prior to failure. Overall, Gloubos-Grammatikos discriminant showed good predictive results on the Belgian dataset. At the same time, the Altman and Bilderbeek models presented very poor results: they are

not widely usable.

In search for possible explanations for these findings, we first examined the type of variables included in the models and tried to indicate those common types of variables leading to the highest prediction accuracy. On the basis of the study, it appeared that the combination of some types of variables generally leads to good predictive results. One year prior to failure, working capital and debt variables together with gross return variables and variables concerning short term obligations seemed to lead to accurate failure predictions. Good failure prediction models two years prior to failure appeared to include working capital variables along with variables on the long term debt situation and gross and net return variables. Finally, on a term of 3 years, variables concerning the cash-liquidity situation, gross return and the debt situation played an important role. This study also gave an indication that the way in which the variables are combined may explain whether a failure prediction model can be widely used or not. The combination of variables may be the most important explanation for model performances. Also, it seemed that models that are widely usable combine their variables in the "intuitively right sense".

Besides the choice of variables, we also took a closer look at some possible explanatory factors related to the model construction, but the estimation technique and the complexity and number of variables did not seem to explain the performance results.

The basic conclusion of this study is that business failure prediction models that are the result of a diversified combination of several types of variables, used in the intuitively right sense, seem to be widely usable on new data sets of firms. This makes a strong case for building "common sense" or "intuitive" failure prediction models, instead of statistically constructing models that have a limited use. A similar conclusion was reported by Moses and Liao (1987).

The conclusions concerning the impact of the possible explanatory factors should, however, be treated with care. Firstly, this study concerns only a limited number of models, which makes it impossible to reach any generally valid conclusions concerning the common predictor variables, the estimation technique and the complexity and number of variables in the models that can be widely used. Secondly, on the basis of this study, it is impossible to gain a clear insight into the real impact of these factors, because many different explanatory factors may interact and may influence each other. Further research involving a more extended list of models, systematic comparisons of explanatory factors and other new data sets (i.e,. from other countries) is required.

References and Appendix

Altman, E. I. 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance* 23(4): 589–609.

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

Altman, E. I. 1993. *Corporate Financial Distress and Bankruptcy*. New York, John Wiley & Sons.

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.

Altman, E. I.; Marco, G.; Varetto, F. 1994. Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks. *Journal of Banking and Finance* 18(3): 505–529.

Altman, E.I.; Margaine, M.; Schlosser, M.; Vernimmen, P. 1974. Statistical credit analysis in the textile industry: A French experience. *Journal of Financial and Quantitative Analysis* 9 (March).

Altman, E. I., Narayanan, P. 1997. An international survey of business classification models. *Financial Markets, Institutions & Instruments* 6(2): 1–57.

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

Beaver, W. 1967. Financial Ratios as predictors of failure. Empirical Research in Accounting, Selected Studies 1966. *Journal of Accounting Research* 4(Supplement): 71–111.

Bell, T.; Ribar, G.; Verchio, J. 1990. Neural nets versus logistic regression: A comparison of each model's ability to predict commercial bank failures. Paper presented at the International Cash Flow Accounting Conference, December 1990, Nice (France).

Bilderbeek, J. 1979. De continuïteitsfactor als beoordelingsinstrument van ondernemingen. *Accountancy en Bedrijfskunde Kwartaalschrift* 4(3): 58–61.

Curram, S., Mingers, J. 1994. Neural networks, decision tree induction and discriminant analysis: An empirical comparison. *Operational Research Society* 45(4): 440–450.

Fernandez, A. I. 1988. A Spanish model for credit risk classification. *Studies in Banking and Finance* 7: 115–25.

Gloubos, G., Grammatikos, T. 1988. Success of bankruptcy prediction models in Greece. *Studies in Banking and Finance* 7: 37–46.

Hosmer, D. W., Lemeshow, S. 1989. *Applied Logistic Regression*. New York, John Wiley & Sons.

Jones, F. L. 1987. Current techniques in bankruptcy prediction. *Journal of Accounting Literature* 6: 131–164.

Joos, Ph.; Ooghe, H.; Sierens, N. 1998. Methodologie bij het opstellen en beoordelen van kredietclassificatiemodellen. *Tijdschrift voor Economie en*

Management 43(1): 3–48.

Kankaanpää, M., Laitinen, T. 1999. Comparative analysis of failure prediction methods: The Finnish case. *European Accounting Review* 8(1): 67–92.

Keasey, K., McGuinness, P. 1990. The failure of U.K. industrial firms for the period 1976–1984: Logistic analysis and entropy measures. *Journal of Business Finance and Accounting* 17(1): 119–135.

Koh, H. C. 1992. The sensitivity of optimal cut-off points to misclassification costs of type I and type II errors in the going concern prediction context. *Journal of Business Finance and Accounting* 19(2): 187–197.

Lachenbruch, P.A. 1975. *Discriminant Analysis*. London, Hafner Press.

Laitinen, E.K. 1992. Prediction of failure of a newly founded firm. *Journal of Business Venturing* 7(4): 323–340.

Moses, D., Liao, S. S. 1987. On developing models for failure prediction. *Journal of Commercial Bank Lending* 69: 27–38.

Mossman, C. E.; Bell, G. G.; Turtle, H.; Swartz, L. M. 1998. An empirical comparison of bankruptcy models. *Financial Review* 33(2): 35–54.

Ohlson, J. A. 1980. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research* 18(1): 109–131.

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

Ooghe, H.; Joos, Ph.; De Vos, D. 1991. Failure prediction models. Working Paper, Department of Corporate Finance, Ghent University (Belgium).

Ooghe, H.; Joos, Ph.; De Bourdeaudhuj, C. 1995. Financial distress models in Belgium: The results of a decade of empirical research. *International Journal of Accounting* 30(3): 245–274.

Siegel, S., Castellan, N. J. 1988. *Nonparametric Statistics for the Behavioral Sciences*. New York, McGraw-Hill.

Steele, A. 1995. Going concern qualifications and bankruptcy prediction. Working Paper presented at the Doctoral Workshop 1995, March 1995, Leuven (Belgium).

Swanson, E., Tybout, J. 1988. Industrial bankruptcy determinants in Argentina. *Studies in Banking and Finance* 7: 1–25.

Taffler, R. J. 1984. Empirical models for the monitoring of U.K. corporations. *Journal of Banking and Finance* 8(2): 199–227.

Taffler, R. J., Tisshaw, H. J. 1977. Going, going, gone: Four factors which predict. *Accountancy* 88(March): 50–54.

Theil, H. 1971. *Principles of Econometrics*. Amsterdam, North-Holland Publishing Company.

Zavgren, C. 1983. The prediction of corporate failure: The state of the art. *Journal of Accounting Literature* 2: 1–33.

Zavgren, C. 1985. Assessing the vulnerability to failure of American industrial firms: A logistic analysis. *Journal of Business Finance and Accounting* 12(1): 19–45.

TABLE A1.   Characteristics of the Models

	Altman	Bilderbeek	Ooghe-Verbaere	Zavgren	Gloubos-Grammatikos	Keasey-Mc Guinness	Ooghe-Joos-De Vos
Country	United States	The Netherlands	Belgium	United States	Greece	United Kingdom	Belgium
Population	American industrial companies	Dutch industrial and trade companies	Belgian enterprises publishing their accounts in a complete form	American companies listed on the Stock Exchange with data available on Compustat	Greek enterprises	U.K. companies with data available on Datastream	Belgian enterprises publishing their results in a complete or abbreviated form
Period	1946-1965	1950-1975	1977-1980	1972-1978	1977-1985	1976-1984	1985-1990
Definition of failure	Declaration of bankruptcy by court	Declaration of bankruptcy by court	Declaration of bankruptcy by court. Request or approval of legal composition	Request of chapter 10 or 11 from the bankruptcy law	Declaration of bankruptcy by court. Failed companies sustained in operation by Greek government are excluded from the non-failed sample	Declaration of bankruptcy by court	Declaration of bankruptcy by court. Request or approval of legal composition
Sample:							
Non-failing	33 annual accounts	43 (original) 220 (validation)	753 (original) 347 (validation)	45 annual accounts	30 (original) 24 (validation)	43 (original) 15 (validation)	347 (original) 170 (validation)
Failing	33 annual accounts	40 (original) 127 (validation)	395 (original) 268 (original)	45 annual accounts	30 (original) 24 (validation)	43 (original) 15 (validation)	268 (validation) 218 (validation)
Method	Matched on industry and size of total assets	Matched on industry turnover. Size of total assets and number of employees	Random selection	Matched on industry and size of total assets	Matched on industry and size of total assets	Matched on industry and size of total assets	Systematic selection

(Continued)

TABLE A1.   (Continued)

	Altman	Bilderbeek	Ooghe-Verbaere	Zavgren	Gloubos-Grammatikos	Keasey-Mc Guinness	Ooghe-Joos-De Vos
Method	Matched on industry and size of total assets	Matched on industry turnover. Size of total assets and number of employees	Random selection	Matched on industry and size of total assets	Matched on industry and size of total assets	Matched on industry and size of total assets	Systematic selection
Estimation technique	Linear discriminant analysis	Linear discriminant analysis	Linear discriminant analysis	Logistic regression	Linear discriminant logit and probit analysis and linear probability models	Logistic regression	Logistic regression
Model	1 model applicable to data 1-5 *ypf*	1 model (and 2 derived classification functions) applicable to data 1-5*ypf*	3 models. 1-3 *ypf* and one general model. Each with different variables and coefficients	5 models. 1-5 *ypf*. Each with different variables and coefficients	1 model applicable to data 1-3 *ypf* (based on data 1*ypf*)	5 models. 1-5 *ypf* Each with different variables and coefficients	2 models. 1-3 *ypf*. Each with different variables and coefficients
No. of variables	5	5	5 for each model	7	5 (discriminant analysis)/3 (logit)	3/4/4	8/6
Classic. rule	Cut-off score	Cut-off score	Cut-off score	Cut-off score and entropy	Cut-off score	Cut-off score and entropy	Cut-off score

TABLE A2.   **Altman (1968), USA**

Variables	Original coefficients			New coefficients		
	1 ypf	2 ypf	3 ypf	1 ypf	2 ypf	3 ypf
Intercept	0.000	0.000	0.000	0.133	0.133	0.133
X1 Net working capital / Total assets	0.012	0.012	0.012	–0.016	–0.016	–0.016
X2 (Accumulated profits or losses + Retained earnings) / Total assets	0.014	0.014	0.014	0.014	0.014	0.014
X3 Earnings before interest and taxes (EBIT) / Total assets	0.033	0.033	0.033	0.068	0.068	0.068
X4 Equity / Liabilities	0.006	0.006	0.006	0.000	0.000	0.000
X5 Sales / Total assets	0.0099	0.0099	0.0099	–0.132	–0.132	–0.132
Cut–off point	0.0190	0.0221	0.0229	–0.1757	–0.1203	0.0574

TABLE A3.   Bilderbeek (1979), The Netherlands

Variables	Original coefficients			New coefficients		
	1 ypf	2 ypf	3 ypf	1 ypf	2 ypf	3 ypf
Intercept	–0.45	–0.45	–0.45	–0.218	–0.218	–0.218
X6 Net return on equity after taxes	–0.15	–0.15	–0.15	–0.008	–0.008	–0.008
X7 Accounts payable / Sales	–4.55	–4.55	–4.55	0.003	0.003	0.003
X5 Sales / Total assets	–0.17	–0.17	–0.17	0.107	0.107	0.107
X8 Gross added value / Total assets	1.57	1.57	1.57	–0.029	–0.029	–0.029
X2 (Accumulated profits or losses + Retained earnings) / Total assets	5.03	5.03	5.03	–0.006	–0.006	–0.006
Cut–off Point	–1.5495	–0.8523	–0.4955	–0.2145	0.4924	–0.0737

TABLE A4.　(Continued)

Variables	Original coefficients	New coefficients
3 years prior to failure		
Intercept	0.2153	–0.735
X9 Overdue taxes and social security charges / Short term liabilities	–18.3474	0.044
X16 (Accumulated profits or losses + Retained earnings) /		
(Equity + Liabilities)	3.3847	0.032
X13 Cash / Current assets	2.3601	4.137
X14 (Work in progress, finished goods and contracts in progress) /		
Current working assets	–1.9230	0.010
X17 Earnings before interests and taxes (EBIT) /		
(Equity + Long term liabilities)	0.0617	0.031
Cut–off Point	–0.2047	–0.4459

TABLE A5. Zavgren (1985), USA

Variables	Original coefficients			New coefficients		
	1 ypf	2 ypf	3 ypf	1 ypf	2 ypf	3 ypf
Intercept	−0.2388	−2.6106	−1.5115	1.596	0.656	0.462
X18 Inventories / Sales	0.0011	0.0419	0.0626	0.001	0.020	0.009
X19 Amounts receivable / Inventories	0.0158	0.0222	0.0083	−0.001	0.001	0.001
X20 (Cash + Short term investments) / Total assets	0.1078	0.1123	0.4248	1.528	4.974	3.065
X21 Quick ratio	−0.0307	−0.0269	−0.0155	−0.962	0.252	0.314
X6 Net return on equity after taxes	−0.0049	−0.0144	0.0052	−0.010	−0.001	0.029
X22 (Liabilities – Accrued charges and deferred income)/Equity	0.0435	0.0446	0.0182	0.001	0.000	0.001
X5 Sales / Total assets	−0.0011	0.0006	0.0000	−0.053	−0.079	−0.038
Cut–off Point	0.4340	0.0672	0.1829	0.915	0.737	0.723

TABLE A6.   Gloubos and Grammatikos (1988), Greece

Variables	Original coefficients			New coefficients		
Discriminant analysis	1 ypf	2 ypf	3 ypf	1 ypf	2 ypf	3 ypf
Intercept	4.423	4.423	4.423	−0.017	−0.017	−0.017
X23 Current ratio	−2.044	−2.044	−2.044	0.002	0.002	0.002
X1 Net working capital / Total assets	4.421	4.421	4.421	0.001	0.001	0.001
X24 Long term liabilities / Total assets	−4.404	−4.404	−4.404	−0.014	−0.014	−0.014
X11 Gross return on total assets before taxes	−2.778	−2.778	−2.778	0.203	0.203	0.203
X25 Earnings before interest, taxes, depreciation and amortization (EBITDA) / Short term liabilities	4.423	4.423	4.423	0.000	0.000	0.000
Cut–off point	2.2441	2.2070	2.5515	−0.0033	−0.0006	−0.0043
Logistic regression	1 ypf	2 ypf	3 ypf	1 ypf	2 ypf	3 ypf
Intercept	3.548	3.548	3.548	1.473	1.473	1.473
X1 Net working capital / Total assets	5.585	5.585	5.585	0.000	0.000	0.000
X24 Long term liabilities / Total assets	−8.504	−8.504	−8.504	−0.001	−0.001	−0.001
X11 Gross return on total assets before taxes	13.070	13.070	13.070	0.051	0.051	0.051
Cut–off point	0.9444	0.9848	0.9573	0.8138	0.8142	0.8142

TABLE A7.   Keasey and McGuinness (1990), U.K.

Variables	Original coefficients	New coefficients
1 year prior to failure		
Intercept	–0.0881	2.017
X4 Equity / Liabilities	–0.0316	1.318
X26 Purchases / Accounts payable	0.2710	–0.000
X27 Profit or losses before taxes / Sales	0.3227	0.035
Cut-off point	0.7939	0.8980
2 years prior to failure		
Intercept	–3.3612	1.280
X18 Inventories / Sales	–8.4286	0.007
X1 Net working capital / Total assets	2.7244	0.151
X28 Net return on operating assets before taxes	0.1081	0.064
X6 Net return on equity after taxes	0.01947	0.001
Cut-off point	0.0131	0.7853
3 years prior to failure		
Intercept	–6.4202	1.059
X21 Quick ratio	1.5599	0.074
X26 Purchases / Accounts payable	0.3010	0.001
X5 Sales / Total assets	0.8799	–0.067
X27 Profit or loss before taxes / Sales	0.4216	0.024
Cut-off point	0.1387	0.7400

TABLE A8. Ooghe, Joos and De Vos (1991), Belgium

Variables	Original coefficients*	New coefficients*
1 year prior to failure		
Intercept	+	+
X29 Direction of the financial leverage = Net return on total assets before taxes – Average interest rate of debt (1 if > 0, 0 if <0)	+	+
X30 (Accumulated profits or losses + Retained earnings)/ (Equity + Liabilities – Accrued charges and deferred income)	+	+
X31 Cash and short term investments / Total assets	+	+
X32 Overdue taxes and social security charges (1 if >0, 0 else)	–	+
X33 (Inventories + Accounts receivable – Accounts payable – Taxes, remuneration and social security debts – Advances received on contracts in progress) / Total assets	+	+
X34 Net return on operating assets before taxes	+	+
X35 Short term financial debt / Short term debt	–	–
X36 Debts guaranteed / Total debt	+	+
Cut-off Point	0.6588	0.9011

(*Continued*)

TABLE A8.   **(Continued)**

Variables	Original coefficients*	New coefficients*
3 years prior to failure		
Intercept	+	+
X30 (Accumulated profits or losses + Retained earnings)/ (Equity + Liabilities – Accrued charges and deferred income)	+	+
X37 Publication lag of the annual accounts	–	–
X32 Overdue taxes and social security charges (1 if >0, 0 else)	–	+
X38 (Earnings before interest, taxes, depreciation and amortization (EBITDA) – Capital investments) / Total assets	+	+
X39 Relationships with affiliated enterprises = (Amounts receivable from them + Commitments guaranteed on their behalf + Other financial commitments in their favor) / Total assets		+
X40 Total debt / (Equity + Liabilities – Accrued charges and deferred income)	–	+
Cut-off Point	0.8023	0.7631

Note: *Only the sign and not coefficients can be given because of a licence agreement with Graydon NV.

TABLE A9.   Classification Results in the Original Publications

	Cut-off	Type I error (percentage)		Type II error (percentage)		Unweighted error rate (percentage)	
		Original sample	Validation sample	Original sample	Validation sample	Original sample	Validation sample
1 year prior to failure							
Altman	2.6750	6.0	n/a	3.0	n/a	5.0	n/a
Bilderbeek	0.0250	n/a	n/a	n/a	n/a	32.0	n/a
Ooghe-Verbaere	3.1492	n/a	13.6	n/a	21.7	n/a	17.6
Zavgren	n/a	11.0	n/a	24.0	n/a	18.0	n/a
Gloubos-Grammatikos logit	0.5000	16.6	33.3	10.0	12.5	13.3	22.9
Gloubos-Grammatikos discriminant	0.0000	3.3	33.3	13.3	33.3	8.3	33.3
Keasey-McGuinness	n/a	14.0	44.0	14.0	30.0	14.0	37.0
Ooghe-Joos-De Vos	0.3117	n/a	14.7	n/a	22.4	n/a	18.5
2 years prior to failure							
Altman	2.6750	28.0	n/a	6.0	n/a	17.0	n/a
Bilderbeek	0.0250	n/a	n/a	n/a	n/a	27.0	n/a
Ooghe-Verbaere	0.1663	n/a	27.9	n/a	22.8	n/a	29.6
Zavgren	n/a	11.0	n/a	22.0	n/a	17.0	n/a
Gloubos-Grammatikos logit	0.5000	n/a	39.1	n/a	17.4	n/a	28.3
Gloubos-Grammatikos discriminant	0.0000	n/a	39.1	n/a	17.4	n/a	28.3
Keasey-McGuinness	n/a	16.0	22.0	21.0	29.0	18.5	25.5

TABLE A9.   (Continued)

	Cut-off	Type I error (percentage) Original sample	Type I error (percentage) Validation sample	Type II error (percentage) Original sample	Type II error (percentage) Validation sample	Unweighted error rate (percentage) Original sample	Unweighted error rate (percentage) Validation sample
3 years prior to failure							
Altman	2.6750	52.0	n/a	n/a	n/a	n/a	n/a
Bilderbeek	0.0250	n/a	n/a	n/a	n/a	29.0	n/a
Ooghe-Verbaere	0.3355	n/a	26.2	n/a	32.9	n/a	29.6
Zavgren	n/a	31.0	n/a	24.0	n/a	28.0	n/a
Gloubos-Grammatikos logit	0.5000	n/a	50.0	n/a	21.4	n/a	35.7
Gloubos-Grammatikos discriminant	0.0000	n/a	35.7	n/a	14.3	n/a	25.0
Keasey-McGuinness	n/a	28.0	27.0	19.0	44.0	23.5	35.5
Ooghe-Joos-De Vos	0.2137	n/a	18.3	n/a	34.1	n/a	27.7