INTELLIGENT SYSTEMS IN ACCOUNTING, FINANCE AND MANAGEMENT
Intell. Sys. Acc. Fin. Mgmt. **13**, 95–112 (2005)
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/isaf.259

BANKRUPTCY PREDICTION: THE INFLUENCE OF THE YEAR PRIOR TO FAILURE SELECTED FOR MODEL BUILDING AND THE EFFECTS IN A PERIOD OF ECONOMIC DECLINE

PAUL P. M. POMPE* AND JAN BILDERBEEK
Department of Finance and Accounting, University of Twente, Enschede, The Netherlands

SUMMARY
Using large amounts of data from small and medium-sized industrial firms, this study examines two aspects of bankruptcy prediction: the influence of the year prior to failure selected for model building and the effects in a period of economic decline. The results show that especially models generated from the final annual report published prior to bankruptcy were less successful in the timely prediction of failure. Furthermore, it was found that economic decline coincided with the deterioration of a model's performance. With respect to the methods used, we found that neural networks had a somewhat better overall performance than multiple discriminant analysis. Copyright © 2005 John Wiley & Sons, Ltd.

1. INTRODUCTION

The research described in this article focuses on the prediction of bankruptcy through the use of bankruptcy models. A model for predicting bankruptcy sets out to establish a relationship between failure and a number of financial ratios that can be calculated from a firm's annual report. Timely prediction of bankruptcy is important for all parties involved: shareholders, managers, workers, lenders, suppliers, clients, the community and the government (Dimitras *et al.*, 1996). Bankruptcy models are useful to those stakeholders that are able to take action to prevent failure. These actions include corporate restructuring or merger of the firm. Furthermore, these models are very useful in aiding investors in selecting firms to invest in, and in then monitoring them. These models can also be helpful for the pricing of loans. Many articles on failure prediction stress the importance of timely prediction; however, none of them define the term 'timely'. The prediction of failure only 1 year before the moment of failure is too late for undertaking actions such as the extraction of cash by an investor, or the execution of a turnaround plan. In order to take effective measures, the failure needs to be predicted a few years in advance. However, it is difficult to say exactly how many years in advance, since this depends on the specific situation. For example, it depends on the time needed to make and execute a turnaround plan. In this study, as a rough indication, we consider timely prediction to be the prediction of failure in year 4 prior to bankruptcy, or earlier. This article reports on the results of a study that is based on two research questions:

1. Which annual report (in terms of number of years prior to failure) is best used for building a model that timely predicts bankruptcy?

* Correspondence to: Paul P. M. Pompe, Department of Finance and Accounting, Faculty of Business, Public Administration and Technology, University of Twente, PO Box 217, 7500 AE Enschede, The Netherlands.
E-mail: paul.pompe@nl.abnamro.com

2. Does the performance of a model change during a period in which there is economic decline and a rapid increase in the number of bankruptcies?

With respect to the first research question, we have found that in many studies the model produced is based on the annual report from year 1 prior to failure. Presuming that there is usually a gradual decay leading up to the moment of failure, it is questionable whether this annual report is the most appropriate for building an early-warning model. There has already been some research on this subject using limited data from large listed firms (e.g. Altman *et al.*, 1977; El Hennawy and Morris, 1983). In our research we use large amounts of data from small and medium-sized firms. Regarding the second research question, the studies that have investigated the performance of models over time are relevant (e.g. Moyer, 1977; Mensah, 1984; Zavgren, 1985; Holmen, 1988; Altman, 1993). However, in the literature we could not find a study that was similar to ours. We focus on a short, interesting, period of 4 years, in which there was economic decline and a rapid increase in the number of bankruptcies. We are able to use enough annual reports to build a model for each calendar year.

Our data set contains 1356 bankrupt and 3600 nonbankrupt firms from Belgium. This is a large quantity of data in comparison with other studies on failure prediction. Furthermore, our data set is unusual in that virtually all the firms are small or medium-sized. In the data set, nearly all the firms employed less than 50 people. Other researchers have tended to focus on large, listed firms. In our study, both multiple discriminant analysis (MDA) and neural networks are used to generate models. In applying these methods, we follow a rigorous procedure that is not generally used in other studies on bankruptcy prediction. For example, we apply 10-fold cross-validation to the training set to determine a good parameter setting for each method. Results obtained using test sets are tested for significance.

Our results show that especially models generated from the most recent annual report published prior to bankruptcy were less successful in the timely prediction of failure. The final published annual report was not very representative of the years more remote from the moment of failure. Regarding the second research question, we found that the performance of a model can change considerably over a few years. Economic decline and a rapid increase in the number of bankruptcies accompanied the deterioration in a model's performance. With respect to the methods used, we found that neural networks had a somewhat better overall performance than MDA.

This article contains eight sections. The next section discusses the relevant literature, and the third section presents the data used. We describe the methods and procedures in Section 4. The results with respect to the first and second research questions are addressed in the fifth and sixth sections respectively. Section 7 discusses the performance of MDA and neural networks, and the final section presents some concluding remarks.

2. LITERATURE REVIEW

Researchers often state explicitly that they want to produce an early-warning model, but in other studies it seems also likely that a model that can predict failure as early as possible (for example, in year 4 or year 5 prior to failure) is aimed for. In many studies, the model produced is based on the annual report from year 1 prior to failure (e.g. Altman, 1968; Altman *et al.*, 1995; Richardson *et al.*, 1998; Lennox, 1999). However, it is possible that annual reports from earlier years are more suitable, if, over the years heading towards failure, firms experience a gradual decay. Starting from this premise then, the best possible prediction in a certain phase of this decay should be achieved by a model that is based on reports from the same phase. After all, these reports are the most representative of the phase concerned. Thus, if the goal is to generate a model that best achieves a late signalling of failure, then it is sensible to use only annual reports published just before the moment of failure. For a model to predict failure timely, only annual reports published a few years earlier should be used.

There are few articles on this subject in the literature. Altman *et al.* (1977) present the ZETA model, which is a model based on annual reports from year 1 prior to failure (i.e. a year-1 model). Models based on year 2, year 3, year 4 and year 5 were also considered. Of the five models, the year-1 model proved to have the best overall performance in the 5-year period prior to failure (the detailed results for all five models are not given in the article). El Hennawy and Morris (1983) also considered models ranging from year 1 through to year 5. Their results give some support to the view that an early-warning model should be based on annual reports from a few years prior to failure. The study by Altman *et al.* (1977) does not suggest this is the case. We do a similar analysis to Altman *et al.* (1977) and El Hennawy and Morris (1983) for two reasons. These researchers focused on large, listed firms, whereas in our research virtually all the firms are either small or medium-sized. Furthermore, since we are able to use data from many more firms our results are statistically more reliable. Both Altman *et al.* (1977) and El Hennawy and Morris (1983) had only limited data: they used annual reports from about 50 failed firms and 50 nonfailed firms.

Regarding the second research question, several studies have shown that the ratio values found in the population of annual reports alter over time. For example, Pinches *et al.* (1973) showed that the values of many ratios changed considerably during a period of 19 years (1951–1969); the average value of the ratio *equity*/*total assets*, for instance, decreased as a result of the greater use of debt capital. Suppose one wants to maximize the average success rate for the classes ‘failed’ and ‘nonfailed’ when predicting failure.[1] Then, changing ratio values in the population would seem to have two possible effects. We consider two years, 1990 and 2000.[2] The two effects are:[3]

I. The predictability of failure may alter. In other words, the predictability of failure in one period may be different from the predictability of failure in another period. It is conceivable that predicting failure in 1990 (using a model from 1990) is either more difficult or easier than predicting failure in 2000 (using a model from 2000).
II. The fitness of a model may alter. The fitness of a certain model in one period may differ from the fitness of the same model in another period. Possibly, in 1990, a model from 1990 will predict failure very well but, in 2000, this model is less able to predict failure than a model from 2000. Naturally, it is preferable if the fitness of a model does not deteriorate rapidly over time, so that the model does not have to be frequently replaced with a newer one.

[1] The success rate for a certain class is defined as the percentage of annual reports of this class that are classified correctly by a model. Reports of class ‘failed’ are from failed firms, and reports of class ‘nonfailed’ are from nonfailed firms.
[2] These two years are used as an example to make clear the meaning of effect I and effect II; they do not reflect the study period of this research.
[3] We define the predictability in period i as the highest possible average success rate for the two classes in period i. This result is achieved by the optimal division of the classes in period i, i.e. where each point x in the feature space is allocated to the class with the highest value for $p(x|\text{class})$. The fitness of a model in period i is defined as the difference between the predictability in period i and the average success rate for both classes in period i achieved by the model (the smaller the difference, the better the model fits). We assume that a model generated using data from period i fits very well for period i.

Inevitably, over time, a model's classification result for the population (defined as the average success rate for both classes) changes; for example, a model build in 1990 will almost certainly achieve different classification results in 1990, 1995 and 2000. This change in the classification result is caused by the two effects. Several studies have shown that, in the course of time, models have performed less well to a certain extent (e.g. Moyer, 1977; Zavgren, 1985; Holmen, 1988; Altman, 1993). For example, the ZETA model, built using data from the period 1962–1975, was better able to classify annual reports from the same period than annual reports from a later period (Altman, 1993). The studies mentioned usually provide insufficient information to estimate the importance of effects I and II. A study that somewhat resembles our own research is that by Mensah (1984), who divided the years 1972–1980 into two kinds of period: expansionary periods and recessionary periods. Annual reports from failed and nonfailed firms were drawn from the expansionary periods (set A), and from the recessionary periods (set B). In each set, the ratio between the numbers of reports from the two classes was 0.5:0.5. Two models were generated: model A was generated from set A, and model B was derived using set B. The classification result for set A, achieved by model A, was 86% (using leave-one-out), and the classification result for set B, achieved using model B, was 88% (leave-one-out). Thus, it would seem that the predictability of failure is about the same in expansionary and recessionary periods. Further, the classification result for set B, achieved using model A, was 73%; and the classification result for set A, obtained using model B, was 76%. Therefore, a distinct model for each type of period seems to be advisable: model A for expansionary periods, and model B for recessionary periods. In terms of the two effects, it would seem that effect II occurred and effect I did not.

What is new about our research is that we study both effects in a short, but interesting period (1988–1991), in which there was economic decline and a rapid increase in the number of bankruptcies. We have enough annual reports to generate a model for each calendar year, so we are able to observe both effects from year to year. In Lennox (1999) and Richardson *et al.* (1998), variables that indicated the general economic climate were entered in models, and these variables proved to be significant. However, adding such a variable is not appropriate in our research, since a model is not generated using annual reports from multiple calendar years (and hence different economic climates), but using annual reports from one calendar year only.

3. DATA

Data from Belgian firms are used because, in Belgium, virtually all firms are legally obliged to file their annual reports with the Belgian National Bank. In this research, we study the period 1986–1994 and, in order to obtain a homogeneous sample, we only select firms from the industrial sector. We distinguish two classes of annual reports: the class 'nonbankrupt' and the class 'bankrupt'. An annual report of the first class belongs to a firm that did not go bankrupt in the period studied. An annual report of the second class belongs to a firm that went bankrupt.[4] If the number of years between the calendar year to which the annual report refers and the year of bankruptcy equals i ($i = 1, 2, 3, \ldots$), we say that the annual report is from year i. There are seven sets of annual reports, with $i = 1, 2, \ldots, 7$. When defining the year of bankruptcy one can choose between two years. In

[4] At the time of failure, the legal status of this firm was 'Bankrupt'. Firms with this legal status have suspended payments to creditors and have lost all creditworthiness.

most studies on failure prediction the year of the formal legal timing of bankruptcy is chosen. However, there are studies in which the year following the calendar year of the last published annual report is considered to be the year of bankruptcy, e.g. see Bilderbeek (1979) and Lennox (1999). We decided to choose the latter definition, because the duration of the judicial procedure that leads to the legal status of bankruptcy differs for each firm. Therefore, the moment a firm stops publishing annual reports seems to us to be a more objective definition of the onset of failure. As a consequence of our choice, an annual report from year 1 is always the final annual report published prior to bankruptcy. The period between the closing date of the final published annual report and the legal moment of bankruptcy was between 0.5 and 2 years for 93% of the firms studied.

Most annual reports in our data set have not been audited, since most firms in our data set are not legally obliged to include an auditor's certificate. However, with most smaller firms, the annual report that is filed with the Belgian National Bank is virtually identical to the fiscal annual report. Therefore, we would not expect the figures that are reported to the Belgian National Bank to be an unreliable indicator of the financial performance and condition of the firm (although the reliability of annual reports with an auditor's certificate is probably higher).

All the available annual reports from industrial firms that went bankrupt in the period studied were obtained. We reduced this sample in three ways. The first reduction was by the removal of annual reports referring to a period other than a 1-year period. Then, annual reports containing obvious arithmetical errors were removed. An example of an arithmetical error is that the sum of the fixed assets and the current assets does not equal the total assets. Finally, we removed annual reports that did not have a value for each of the 39 ratios mentioned later in this section. The first row in Table I ('all firms') reports the final number of annual reports from each year. These annual reports belong to 1356 firms. In the first row, it is evident that the number of reports decreases considerably as the years increase. There are two reasons. First, in the set of annual reports from year 7 there are naturally no annual reports for firms that existed for too short a period to have an annual report in this set. Second, in the research base there is only an awareness of bankruptcies that took place during the period studied (1986–1994). Therefore, an annual report from year 7 can refer to fewer calendar years than an annual report from year 1. For example, there cannot be an annual report for year 7 that refers to 1993, since in the period studied the latest bankruptcy took place in 1994.

Every annual report of class 'bankrupt' refers to a calendar year between 1986 and 1993. Table II shows the calendar years of the reports from year 1. The number of annual reports for 1993 is very low, since the database we used was insufficiently recent to contain most annual reports from this year. Likewise, the number of reports from 1992 would have been higher if we had had a more recent database. As noted earlier, when considering the second research question, we use the years 1988, 1989, 1990 and 1991. Table II shows that this is an interesting period, because the real gross

Table I. The number of annual reports from the two classes

	Bankrupt							Nonbankrupt
	Yr 1	Yr 2	Yr 3	Yr 4	Yr 5	Yr 6	Yr 7	
All firms	1266	956	758	600	446	312	170	3600
Old firms	542	470	420	368	316	252	146	1800
Young firms	724	486	338	232	130	60	24	1800

Table II. The GDP growth and the number of annual reports from year 1 in the period 1986–1994

	1986	1987	1988	1989	1990	1991	1992	1993	1994	Total
GDP growth (%)	2.1	2.4	4.7	3.6	3.0	1.6	1.5	−1.5	2.6	
Annual reports from year 1	142	111	117	168	201	260	237	30	—	1266

domestic product (GDP) growth[5] was undergoing a strong decline, and the number of annual reports from year 1 (and so the number of bankruptcies) increased considerably. The low point of the recession was 1993. As might be expected, there is an inverse relationship between GDP growth and the growth of the number of bankruptcies in Belgium (VCPB, 1996).

For the class 'nonbankrupt', we use 3600 annual reports (see Table I); they are taken from 1988, 1989, 1990, and 1991 (900 reports from each year). Each annual report belongs to a different firm.[6] The 3600 reports remained after the same three reductions in the data took place as with the data for the class 'bankrupt'. In answering the first research question we use all annual reports of class 'bankrupt' in Table I (from the period 1986–1993), and all 3600 annual reports of class 'nonbankrupt' in Table I (from the period 1988–1991). In answering the second research question we only use the reports of class 'bankrupt' from the period 1988–1991, and again we use all 3600 reports of class 'nonbankrupt' for the same period.

In Table I, a distinction is made between two groups: annual reports of old firms and annual reports of young firms. For the class 'bankrupt', we place in the first group annual reports of firms that had a life of more than 8 years, and the second group contains reports of firms whose life span was 8 years or less. The life is calculated as the time between the start of the firm and the deposit date of the final annual report published before bankruptcy. For example, there are 1266 annual reports from the first year before bankruptcy, of which 542 belong to old firms and 724 to young firms. For the class 'nonbankrupt', the first group contains annual reports from firms that existed for more than 8 years at the deposit date of the annual report, and the second group consists of reports from firms that existed for 8 years or less at this date. Table III presents the percentiles (quartiles and median values) of the total assets, and of the number of persons employed, for six groups of annual reports taken from Table I. 1000 BEF is about €25.

Table III. The percentiles of the total assets and the number of persons employed

		Total assets (in million BEF)			Employees		
		25%	50%	75%	25%	50%	75%
All firms	Nonbankrupt	5	15	52	1	5	18
	Bankrupt yr 3	7	16	48	2	8	24
Old firms	Nonbankrupt	8	25	83	3	9	29
	Bankrupt yr 3	10	25	76	4	13	37
Young firms	Nonbankrupt	4	9	28	1	3	9
	Bankrupt yr 3	4	12	26	1	5	13

[5] GDP is at constant 1990 prices (OECD, 1999).
[6] For each of the 3600 firms, we ensured that the firm had the legal status 'Without any particular legal status' over the whole period studied.

For each annual report, 64 ratios were calculated.[7] These are ratios that are often mentioned in the literature, and that evaluate different dimensions of a firm's financial position: profitability ratios, activity ratios, liquidity ratios, solvency ratios. From these, only the 39 ratios with very few missing values in the annual reports were retained. The final selection of ratios to be included in all the models was created by applying stepwise selection to the group of 39 ratios. We used the statistical software package SPSS to carry out the stepwise selection (using Wilks's lambda as the class separability measure).[8] Stepwise selection was applied to a data set consisting of annual reports from the first row of Table I: 1190 annual reports of class 'bankrupt' (170 reports from each of the 7 years), and all 3600 annual reports of class 'nonbankrupt'. Before applying stepwise selection, we removed annual reports with extreme ratio values (using the procedure described in Footnote 12 with $i = 5$). The variable selection process resulted in the following nine ratios:[9] (1) (Profit before taxes + Debt charges + All non-cash expenses)/Total assets, (2) Profit before taxes/Equity, (3) Profit before taxes/(Operating income − Goods and services purchased), (4) (Operating income − Goods and services purchased)/Total assets, (5) Quick assets/Amounts payable within 1 year, (6) Cash/Current assets, (7) Stocks/Total assets, (8) Trade debts/Total assets, (9) Equity/Total assets.

The capital structure and sources of finance for small and medium-sized enterprises (SMEs) are likely to be fundamentally different from those of large firms (Altman, 1993). This might have an effect on which ratios are predictive. However, the ratios identified in this study can also be found in models for large firms that were generated in other studies on failure prediction. More research using data from SMEs is needed before it is possible to conclude that certain ratios are especially predictive with SMEs.

4. METHODS AND PROCEDURES

The data set is divided into training sets and test sets. A training set is used to generate a failure model, and a test set is used to estimate the predictive power of a model. A failure model is derived from a training set using a specified method. A frequently used method is MDA (e.g. Altman *et al.*, 1977; Bilderbeek, 1979; El Hennawy and Morris, 1983; Mensah, 1984; Laitinen, 1991). In the last decade, there has been interest in a new method known as neural networks (e.g. Altman *et al.*, 1994; Wilson and Sharda, 1994; Olmeda and Fernandez, 1997; Pompe and Feelders, 1997; Boritz *et al.*, 1995; O'Leary, 1998; Zhang *et al.*, 2003). In our research, both methods are applied. We do this for the following reason: if the results using one method show a certain pattern, and this pattern is also present in the results with the other method, then probably the pattern is not caused by the specific method used. Nearly all the analyses are completed using the software package S-Plus. We apply a feedforward neural network containing one hidden layer.

Before a method can be used, a number of parameters have to be assigned values. The setting of the parameters influences the form of the model produced. The parameters need to be assigned

[7] In Pompe and Bilderbeek (2005), the univariate importance of each of these ratios in the prediction of bankruptcies is determined.
[8] For F-to-enter and F-to-remove, we used the default values of 3.84 and 2.71, and for the maximum tolerance we used the value 0.7 to prevent high correlations in the selection.
[9] Operating income = Turnover + Increase/decrease in stocks of finished goods, and in work and contracts in progress + Own construction of fixed assets + Other operating income. Belgian firms are legally obliged to use this definition of 'operating income' in the annual report.

values such that the model achieves a good classification result for data that were not used in deriving the model. To determine a good parameter setting for each method, we follow a rigorous procedure that is not generally used in other studies on failure prediction. A preliminary trial, consisting of short analyses using some of the data, indicated which parameter values were sufficiently promising to be retained in the research. After the preliminary research, we concluded that, with the neural networks, several values should be tried for two parameters. The first parameter is the number of hidden units. We test five values: 0, 1, 2, 3 and 4 units. The second parameter amounts to two sets of starting values for the weights in the network being tried. Thus, there are 5 (first parameter) × 2 (second parameter) = 10 different combinations of parameter values possible. Each of these 10 settings will be tried.[10] During the preliminary research, we also experimented with several values for more than two parameters, leading to 252 parameter settings. Because the results achieved using 252 settings were almost equal to the results obtained using 10 settings, we retain only the 10 settings. This leads to a considerable saving of time.

With MDA there are also two parameters for which different values are tried. The first parameter is the combination of prior probabilities. During the preliminary research, we found that small changes in the combination (0.50, 0.50) sometimes led to better results. The following 11 combinations (prior probability of class 'nonbankrupt', prior probability of class 'bankrupt') are tested: (0.40, 0.60), (0.42, 0.58), (0.44, 0.56), . . . , (0.60, 0.40). Note that only this combination is varied, the actual ratio between the numbers of reports from the two classes in the training set (before removing reports with extreme ratio values) is always 0.5:0.5. The second parameter is the degree of removal of annual reports with extreme ratio values from the training set;[11] 12 degrees are tested.[12] Thus, for the MDA, there are 11 (first parameter) × 12 (second parameter) = 132 different combinations of parameter values possible. Each of them will be tried.

We apply 10-fold cross-validation to the training set to determine a good parameter setting for each method (see Stone (1974), Olmeda and Fernandez (1997) and Zapranis and Refenes (1999)). For example, with MDA we carry out 10-fold cross-validation 132 times,[13] each time using a different parameter setting.[14] The parameter setting leading to the highest cross-validation result is selected. Using this setting, an MDA model is derived from the training set, and then this model is used for classifying annual reports in the test sets.[15] We define the test result for a test set as (The percentage of annual reports of class 'nonbankrupt' in the test set that are classified correctly + The percentage of annual reports of class 'bankrupt' in the test set that are classified correctly)/2. To improve the reliability of the research results, each experiment is always carried out 10 times. The

[10] In this research, we use the function *nnet* in S-Plus (Venables and Ripley, 1994). We use the cross-entropy criterion, links from inputs to output, 0.2 as the degree of weight decay, and 300 as the maximum number of iterations for the optimizer. The parameters of the function that have not yet been given are assigned default values. The ratio values from all the annual reports are rescaled. After rescaling, most (90%) values of a ratio are in the interval [0, 1].

[11] The removal of extreme ratio values from the training set did not affect the results using neural networks.

[12] With respect to the removal of extreme ratio values, the procedure for each annual report in the training set is as follows. If at least one ratio value differs more than i standard deviations from the mean of the class concerned, the annual report is removed from the training set. For the degree of removal i, we try 11 values: 1, 1.5, 2, . . . , 6. In addition, we try not removing extreme values.

[13] Naturally, in the case of neural networks we carry out 10-fold cross-validation 10 times.

[14] With the 10 parts $D_1, D_2, \ldots, D_{10}$ used in 10-fold cross-validation, we ensure that in each part the numbers of annual reports of both classes are the same. Furthermore, we ensure that $D_1, D_2, \ldots, D_{10}$ always contain the same annual reports regardless of the specific method and the specific parameter setting used.

[15] Because we tried several parameter settings, the cross-validation result of the parameter setting selected will be optimistically biased; therefore, this result is less appropriate for using as a measure of the predictive power of the model. To obtain a better impression of the predictive ability we use test sets to obtain non-biased test results for the model.

important change in each round of an experiment, compared with the other nine rounds, is that again the total data set is randomly divided into training sets and test sets. A test result in a table is always the average result based on the 10 rounds of the experiment.

5. THE INFLUENCE OF THE YEAR PRIOR TO FAILURE SELECTED

Table IV details the number of annual reports that are used in this part of the research. First we carry out the analysis using reports from all the firms (see the upper panel of Table IV). Starting with the data in Table I, to achieve more even amounts of data from the different years, we reduced the number of data in year 1, year 2 and year 3 to the number of data in year 4. For example, for year 1, 600 annual reports were randomly selected from the total sample of 1266. In Table IV, we distinguish two elements: part I consists of annual reports for the training set and part II consists of annual reports for the test set. The 3600 annual reports of class 'nonbankrupt' are randomly divided into two sets of 1800 reports. Within the class 'bankrupt', several annual reports could be from the same firm, e.g. a firm could have an annual report from each of the 7 years. Therefore, we randomly divide the 1356 bankrupt firms into two groups of 678 firms. Then, in each year, reports of firms in the first group go to part I, and reports of firms in the second group go to part II. As a consequence, in Table IV, part I and part II are not exactly the same size in years 1, 2, . . . , 7. By following this procedure, the annual reports from any firm are all in part I, or all in part II, and so the undesirable situation is prevented that one annual report from a firm is used for training a model, and a second annual report is used for testing the model.

We derive seven models, a year-1 model, a year-2 model, a year-3 model, etc. The training set for each model contains equal numbers of reports of both classes; reports of class 'nonbankrupt' are randomly selected from part I of class 'nonbankrupt'. To take an example, the training set for the year-3 model consists of the 'bankrupt' reports in part I of year 3 plus the same number of reports randomly selected from part I of class 'nonbankrupt'. There are likewise seven test sets, one for each year. For example, the test set for year 5 consists of part II of year 5 'bankrupt' and part II of class 'nonbankrupt'. Each model is tested on each test set. In Table V, the test results using both MDA and neural networks (NN) are given (the separate test results for the two classes can be found in Appendix A). For example, a year-3 MDA model was derived and tested on the seven test sets in each of the 10 rounds of the experiment. The average of the 10 test results for the year-2 test set, for instance, was 71% as shown in the table. We also consider three models that are based on data from more than 1 year: a model for years 1, 2, 3, and 4, a model for years 3,

Table IV. The number of annual reports used in answering the first research question

		Bankrupt							Nonbankrupt
		Yr 1	Yr 2	Yr 3	Yr 4	Yr 5	Yr 6	Yr 7	
All firms	Part I	~300	~300	~300	~300	~223	~156	~85	1800
	Part II	~300	~300	~300	~300	~223	~156	~85	1800
	Total	600	600	600	600	446	312	170	3600
Old firms	Part I	~210	~210	~210	~184	~158	~126	~73	900
	Part II	~210	~210	~210	~184	~158	~126	~73	900
	Total	420	420	420	368	316	252	146	1800

Table V. The test results relevant to the first research question (all firms)

Model	Test set																				
	MDA							NN							MDA − NN						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7	1	2	3	4	5	6	7
1	77	69	65	62	61	58	57	80	70	68	65	63	59	58	−3	−1	−3	−3	−2	−1	−1
2	77	71	68	66	65	63	62	79	73	71	67	66	63	63	−2	−2	−3	−1	−1	0	−1
3	75	71	69	66	66	64	63	76	73	72	67	67	64	63	−1	−2	−3	−1	−1	0	0
4	74	70	68	66	67	65	65	74	72	70	68	67	66	63	0	−2	−2	−2	0	−1	2
5	73	70	68	66	66	65	64	73	71	70	68	68	66	65	0	−1	−2	−2	−2	−1	−1
6	71	69	67	65	65	64	65	71	69	68	67	67	66	64	0	0	−1	−2	−2	−2	1
7	68	67	65	64	64	63	63	71	69	68	66	67	66	65	−3	−2	−3	−2	−3	−3	−2
1–4	76	71	68	65	65	62	61	77	72	70	66	65	62	61	−1	−1	−2	−1	0	0	0
3–7	72	69	68	66	66	64	63	73	70	69	67	67	66	64	−1	−1	−1	−1	−1	−2	−1
1–7	75	72	68	66	66	65	63	74	71	70	67	67	65	65	1	1	−2	−1	−1	0	−2

4, 5, 6, and 7, and a model using all 7 years.[16] Again, the test results are presented in Table V. The differences between the test results using the two methods are given on the right-hand side of the table. We see that the neural networks have a somewhat better overall performance than MDA.

In Section 2 we indicated that we expected that the best possible prediction, in a certain phase before failure, would be achieved by a model based on reports from the same phase. In line with our expectation, Table V shows that, indeed, the best result for the test set of year *i* is usually achieved using the year-*i* model. Note that comparing the year-6 and the year-7 models with the other models is not completely fair: the predictive power of these two models is reduced by the availability of many fewer training data. The most striking finding is the poor performance of the year-1 model against the test sets for the earliest years (years 4, 5, 6, and 7), whereas the performance of the year-2 model for these sets is still good. This finding is strengthened by the results of a significance test (see Appendix B). It would seem that the last published annual report has an exceptional character and is not very representative of the years more remote from the moment of failure. Many of the annual reports used for generating the year-1 model are from very young firms. Possibly, these annual reports are in particular less suitable for building a model to predict failure timely. Therefore, we try repeating the analysis using only the annual reports from more established firms. The bottom section of Table IV indicates the number of annual reports used. In comparison with Table I, we have reduced the number of data in year 1 and year 2 to match the number in year 3. Each model is generated using reports from only 1 year. The test results are given in Table VI. Again, we find that the year-1 model performs badly with respect to the test sets from the earlier years. Returning to our first research question, we conclude that especially the final annual report published prior to bankruptcy does not appear to be very suitable for use in building an early-warning model.

[16] In the training set for each model, there are approximately 300 annual reports of class 'bankrupt'. The training set for the first model contains 75 (=300/4) annual reports from part I of year 1, 75 from part I of year 2, 75 from part I of year 3, and 75 from part I of year 4. The 75 annual reports are randomly selected from part I of the required year. With the second model there are 60 (=300/5) annual reports from part I of years 3, 4, 5, 6, and 7 in the training set. With the third model, the training set contains 43 (≈300/7) annual reports from part I of years 1 to 7. In terms of the cross-validation, the following addition to Footnote 14 applies: 'In addition, we ensure that all annual reports of the same firm (that are from different years) are always in the same part D_i'.

Table VI. The test results relevant to the first research question (old firms)

Model	Test set																				
	MDA							NN							MDA − NN						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7	1	2	3	4	5	6	7
1	78	73	70	66	65	60	59	81	74	69	67	63	58	56	−3	−1	1	−1	2	2	3
2	77	74	72	68	67	64	63	79	77	72	69	68	63	62	−2	−3	0	−1	−1	1	1
3	75	74	71	68	67	65	64	77	76	73	70	68	64	63	−2	−2	−2	−2	−1	1	1
4	76	73	70	67	67	64	65	76	75	72	69	68	65	64	0	−2	−2	−2	−1	−1	1
5	73	72	70	67	67	66	67	74	74	71	70	69	67	66	−1	−2	−1	−3	−2	−1	1
6	69	68	68	65	65	64	64	71	70	69	68	67	66	65	−2	−2	−1	−3	−2	−2	−1
7	67	66	66	64	63	63	64	70	68	67	67	65	65	64	−3	−2	−1	−3	−2	−2	0

We use the same selection of nine ratios in all the models that are built. As noted in Section 3, this selection is based on data from all 7 years prior to bankruptcy. In a separate analysis, we took a different approach: both building the model and selecting the ratios that were included in the model were based on data from the same year. For example, for the year-1 model, stepwise selection was only applied to annual reports from year 1. Again the results showed that the last published annual report in particular was not very successful in generating an early-warning model.

6. THE EFFECTS IN A PERIOD OF ECONOMIC DECLINE

To investigate the effects in a period of economic decline we generate four sets, one for each of the four calendar years 1988–1991 (see Table VII). As can be seen in the table, we carry out a separate analysis for all firms, for old firms, and for young firms. We now focus on the analysis for all firms. As an example, the set for 1988 consists of 900 annual reports of class ‘nonbankrupt’ and 234 annual reports of class ‘bankrupt’. All the annual reports in one set come from the same calendar year, i.e. all 1134 annual reports in the set for 1988 are dated 1988. We randomly divide each set into a training set and a test set. In the training set for 1988, for instance, there are 125 annual reports of class ‘nonbankrupt’ and 125 annual reports of class ‘bankrupt’. The training set

Table VII. The number of annual reports used in answering the second research question

		1988		1989		1990		1991	
		Nonbankrupt	Bankrupt	Nonbankrupt	Bankrupt	Nonbankrupt	Bankrupt	Nonbankrupt	Bankrupt
All firms	Training set	125	125	125	125	125	125	125	125
	Test set	775	109	775	183	775	277	775	265
	Total	900	234	900	308	900	402	900	390
Old firms	Training set	70	70	70	70	70	70	70	70
	Test set	380	31	380	65	380	119	380	113
	Total	450	101	450	135	450	189	450	183
Young firms	Training set	70	70	70	70	70	70	70	70
	Test set	380	63	380	103	380	143	380	137
	Total	450	133	450	173	450	213	450	207

Table VIII. The test results relevant to the second research question

Model		Test set											
		MDA				NN				MDA − NN			
		1988	1989	1990	1991	1988	1989	1990	1991	1988	1989	1990	1991
All firms	1988	78	74	74	70	77	75	73	71	1	−1	1	−1
	1989	78	75	73	70	77	76	73	71	1	−1	0	−1
	1990	77	74	73	71	78	75	73	71	−1	−1	0	0
	1991	78	74	73	70	77	74	73	70	1	0	0	0
Old firms	1988	78	73	75	70	77	73	74	71	1	0	1	−1
	1989	78	75	75	71	78	76	75	72	0	−1	0	−1
	1990	78	75	76	72	80	76	75	72	−2	−1	1	0
	1991	78	75	75	71	78	75	74	70	0	0	1	1
Young firms	1988	76	73	71	69	76	73	72	69	0	0	−1	0
	1989	74	73	70	68	75	75	70	69	−1	−2	0	−1
	1990	76	74	71	70	76	74	71	70	0	0	0	0
	1991	75	73	71	69	76	73	71	68	−1	0	0	1

for each year contains the same number of reports in order to make the comparison of the models as fair as possible. For the class ‘bankrupt’, annual reports from year 1 and from year 2 are used.[17] Thus, an annual report of class ‘bankrupt’ in the set for year i belongs to a firm whose legal moment of bankruptcy was probably in year $i + 1$, year $i + 2$ or year $i + 3$. For example, for the 1990 set, the bankruptcy probably occurred in 1991, 1992 or 1993. Four models are generated, one using each training set. Then, each of the models is tested on all four test sets. The upper section of Table VIII shows the test results for all firms (the separate test results for the two classes can be found in Appendix A). For example, a 1990 MDA model was derived and tested on the four test sets in each of the 10 rounds of the experiment. The average of the 10 test results for the 1988 test set, for instance, was 77%, as shown in the table. As noted earlier, we carry out separate analyses for old firms and for young firms. Table VII indicates the number of annual reports used, and the test results are given in Table VIII.

Considering the results for all firms, we can observe that each of the four models shows the same trend: the test result decreases as the calendar year of the test set gets more recent. The difference between the result for the 1988 test set and the result for the 1991 test set is 6%, 7%, or 8% (the results of a significance test can be found in Appendix C). A similar trend can be found with the models for the old firms and the models for the young firms. The results suggest that over a few years the performance of a model can change considerably, at least during a period in which the number of bankruptcies changes rapidly. The large change in the performance of each model seems to be caused mainly by effect I (a changing predictability of failure) and not by effect II (a changing fitness of the model). Effect I would seem to occur, since classifying annual reports from 1988 (using the 1988 model) is easier than classifying annual reports from 1991 (using the 1991 model). To a

[17] We arrange it such that in each training set, and in each test set, the number of annual reports from year 1 equals the number of annual reports from year 2. The annual reports are taken from the 1266 (year 1) and 956 (year 2) annual reports in Table I. It is possible that two annual reports from the same firm are used: an annual report from year 2 in the set for year i, and an annual report from year 1 in the set for year $i + 1$. To prevent one of these annual reports being used for deriving a model and the other being used for testing the same model, we ensure that each report is in a training set or each one is in a test set.

large extent, effect II does not seem to occur, because the 1988 model can classify annual reports from 1991 about as well as the 1991 model. Likewise, the 1991 model can classify annual reports from 1988 about as well as the 1988 model.

An interesting finding in this research is that the deterioration of a model's performance coincided with economic decline and a rapid increase in the number of bankruptcies. The differences between a firm that goes bankrupt and a firm that survives appear to be clearer when the general economic climate is sound. It seems plausible that the number of sickly firms will be greater in a generally weak economic situation. These are the sorts of firm whose future is the most difficult to predict; they are clearly not healthy, but they need not be heading for bankruptcy. Furthermore, there is possibly a difference in the attitude of creditors. In an economic trough many firms have problems, and so a strict application of credit rules might lead to many bankruptcies and substantial losses for credit institutions and suppliers. Therefore, in a weak economic situation, these parties may apply their rules less strictly (and so more ambiguously) than they do when the economic climate is sound. This would lead to the bankruptcy of a firm becoming less predictable in a trough; the failure depends on the flexibility of credit institutions and suppliers with regard to a specific firm.

7. THE PERFORMANCE OF MDA AND NEURAL NETWORKS

A disadvantage of MDA is that it only produces an optimal model if the data have certain statistical distribution characteristics, namely a multivariate normal distribution for each class and equal covariance matrices. Financial ratios generally do not meet these requirements well (e.g. Beaver, 1966; Bilderbeek, 1979; El Hennawy and Morris, 1983). With neural networks, the underlying belief is that this method might produce better failure models than MDA, since neural networks do not require a specific distribution of the data. The results from the studies mentioned in Section 4, however, give an ambiguous picture. Neural networks did not achieve better results in all situations. It is difficult to explain why the method sometimes outperforms MDA and in other situations it fails to achieve an improvement. A problem with some studies is the limited data: some studies use data from less than 100 firms for both training and testing the models. In our study, with a large quantity of data, we found that the neural networks did have a somewhat better overall performance than MDA.

In the preliminary research, we found that not removing annual reports with extreme ratio values from the training set clearly negatively affected the MDA results, especially when annual reports from year 1 prior to failure were in the training set (the removal of extreme ratio values from the training set did not affect the results using neural networks). Surprisingly, hardly any of the articles that compare MDA and neural networks in the field of failure prediction record that annual reports with extreme ratio values were removed from the training set, as one might reasonably expect.

Overfitting the training data is a well-known problem, especially with nonparametric methods such as neural networks. Cross-validation is an effective way of preventing overfitting, and so we applied 10-fold cross-validation to the training set and selected the parameter setting leading to the highest cross-validation result (see Section 4). As could be expected, this cross-validation result was usually worse than the in-sample result, and the differences between the cross-validation results and the in-sample results were larger with neural networks than with MDA. Furthermore, the differences were larger with smaller training sets (overfitting is easier if there are fewer data). The use of a separate test set to obtain a non-biased test result (as noted in Footnote 15) was also justified by the results: the cross-validation result with the parameter setting selected was usually better than the result with the test set from the same year as the training set.

In our analysis, we ensured that the number of unique reports of class 'bankrupt' in the training set always equalled the number of unique reports of class 'nonbankrupt'. With the first research question, we also experimented with training sets that contained more unique reports of class 'nonbankrupt' than unique reports of class 'bankrupt'. To obtain equal numbers of reports of both classes, we duplicated the annual reports of class 'bankrupt'. The function *nnet* in S-Plus, which we applied for training neural networks in this study, uses the batch mode of training. We knew that redundancy in the training set (i.e. the data set contains several copies of exactly the same pattern) causes the back-propagation learning in batch mode to be computationally slow (see Bishop (1995) and Haykin (1999)). However, redundancy can apparently also lead to worse classification results; in our study we saw a worse classification of annual reports from years (prior to failure) that were not represented in the training set. For example, with redundancy in the training set for the year-7 model (with 21 copies of each unique report of class 'bankrupt') the result of this model with the test set for year 1 was only 64% compared with the 71% given in Table V. The result of this model with the test set for year 7 was 64%, approximately the same as the 65% presented in Table V.

8. CONCLUSIONS

In particular, models generated from the final annual report published before bankruptcy had difficulty in the timely prediction of failure. Since great importance is attached to early prediction, it appears wise not to use this annual report when generating a model. This is an important conclusion, since, in the literature, many models are based on the annual report from year 1 prior to failure. Returning to the two studies discussed in Section 2, both Altman *et al.* (1977) and El Hennawy and Morris (1983) do not comment that they found any specific problems with the year-1 model. Perhaps the character of annual reports from year 1 belonging to large listed firms is less exceptional. Another reason could be that the numbers of annual reports used in the two studies were too small to uncover a problem with the year-1 model.

Regarding the second research question, we found that economic decline, and a rapid increase in the number of bankruptcies, accompanied the deterioration in a model's performance. The deterioration seemed to be caused mainly by a worsening predictability of failure (effect I), and not by a changing fitness of the model (effect II). As discussed in Section 2, it would seem that, in Mensah's (1984) study, effect II occurred and effect I did not. This is exactly the opposite of our finding. However, it is not that easy to compare Mensah's results directly with ours. Mensah (1984) had many fewer data (about 75 firms of each class), drawn from a longer period, and only from large listed firms. As a consequence, Mensah (1984) was able to make a crude distinction between expansionary and recessionary periods. We were able to examine a period of only 4 years (a period of economic decline), but we had enough annual reports to generate a model for each calendar year.

In the last decade there have been several studies to see whether the newer method of neural networks can generate better failure prediction models than the classical method using MDA. In those studies, neural networks have not achieved better results in all situations. It is difficult to explain why the method sometimes outperforms MDA, and in other situations it fails to show any improvement. In our study, using a large quantity of data and following a rigorous procedure, we did find that neural networks performed somewhat better overall than MDA.

Comparing other studies on failure prediction with our research, we must conclude that our results are fairly poor: classification results in other studies are often better. It seems that predicting the failure of large firms, which is the focus of most other studies, is easier than predicting the failure of small and medium-sized firms. One factor that could play a role here is the reliability of the data. As noted in Section 3, the reliability of our annual reports is probably somewhat less than the reliability of annual reports of large firms, since most annual reports in our data set have not been audited.

APPENDIX A

The test results for the two classes for all firms are shown in Table A.I (first research question) and in Table A.II (second research question).

Table A.I. The test results for the two classes: first research question, all firms (Nb: nonbankrupt)

Model	Test set															
	MDA								NN							
	Nb	1	2	3	4	5	6	7	Nb	1	2	3	4	5	6	7
1	80	73	58	51	43	42	35	33	77	82	63	60	52	48	40	38
2	71	83	72	65	61	59	55	53	70	87	77	73	64	63	56	56
3	62	88	80	75	70	69	66	63	65	88	81	78	69	69	63	61
4	59	88	82	77	74	74	71	70	64	85	80	77	73	70	68	63
5	62	83	77	74	70	70	67	66	64	83	78	77	72	73	69	66
6	59	83	78	75	72	71	68	70	62	79	75	74	72	71	71	66
7	49	86	84	82	80	80	78	77	62	80	76	74	71	73	70	68
1–4	67	85	74	69	63	63	57	55	68	86	76	72	65	61	56	55
3–7	60	84	79	75	72	72	69	66	61	84	79	78	73	74	72	66
1–7	64	86	79	72	68	67	65	61	64	85	78	76	70	69	66	65

Table A.II. The test results for the two classes: second research question, all firms (nonbankrupt, bankrupt)

Model	Test set							
	MDA				NN			
	1988	1989	1990	1991	1988	1989	1990	1991
1988	75, 80	71, 77	71, 76	67, 74	77, 78	73, 76	73, 74	69, 72
1989	74, 81	73, 78	71, 74	66, 75	76, 78	75, 77	73, 73	69, 73
1990	71, 83	67, 81	68, 79	64, 77	74, 82	72, 79	71, 75	68, 74
1991	70, 85	67, 81	67, 79	63, 78	77, 78	74, 75	74, 72	70, 69

APPENDIX B

In Table A.III, the values in the left half relate to MDA, and the values in the right half relate to neural networks. The following discussion applies to each table space (*year-i model*, *test set for year*

Table A.III. The number of rounds with a significant difference between the test results

Model	Test set													
	MDA							NN						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
1			7	8	8	8					8	7	8	
2														
3								8						
4	8							10						
5	9							10						
6	10	7						10	8	7				
7	10	9	7					10	8	8	7			
1–4								8						
3–7	9							10	7	7				
1–7								10						

j)[18] in these panels. The difference between the result of the year-*i* model with the test set for year *j* and the result of the year-*j* model with the test set for year *j* was tested for significance using the McNemar test. For example, in eight out of the ten rounds of the MDA analysis, the difference between the result of the year-1 model with the test set for year 6 and the result of the year-6 model with the same test set was significant. We considered a *p*-value smaller than 0.1 (one-sided test) to be significant. In Table A.III, the number of rounds with a significant difference is only given where this number was at least seven out of ten. To ensure a fair procedure, it is necessary that a test set contains equal numbers of annual reports from both classes. Therefore, we removed a large number of annual reports of class 'nonbankrupt' from the seven test sets (this was only done during the testing for significance, so the results in Table V are based on test sets that contain 1800 annual reports of class 'nonbankrupt'). As an example, the results of the significance test are based on a test set for year 1 that contains about 300 bankrupt annual reports from year 1 and the same number of annual reports of class 'nonbankrupt'. Table A.III shows that, for the year-1 model, there are many rounds with a significant difference, whereas with the year-2 model this is not the case.

APPENDIX C

With respect to the second research question, the difference between the result with the test set for 1988 and the result with the test set for 1991 was tested for significance. Only the models for all firms from 1988 and 1991 were considered. With respect to the 1988 MDA model, the *p*-value was less than 0.1 (one-sided test) in nine out of the ten rounds of the experiment. For the 1991 MDA model, the *p*-value was smaller than 0.1 (one-sided test) in eight out of the ten rounds. For the neural network models from 1988 and 1991, the corresponding values are seven out of ten and eight out of ten respectively. We tested for significance by means of the function *prop.test* in S-Plus. To ensure a fair procedure, it is necessary that a test set contains equal numbers of annual reports

[18] Where $i \in \{1, 2, 3, 4, 5, 6, 7, 1\text{–}4, 3\text{–}7, 1\text{–}7\}$ and $j \in \{1, 2, 3, 4, 5, 6, 7\}$.

from both classes. Therefore, we randomly removed a large number of annual reports of class 'nonbankrupt' from each test set (this was only done during the testing for significance, so the results in Table VIII are based on test sets that contain 775 annual reports of class 'nonbankrupt'). The results of the significance test are based on a 1988 test set that consists of 218 annual reports (109 reports from each class) and a 1991 test set that consists of 530 annual reports (265 from each class).

REFERENCES

Altman EI. 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance* **23**: 589–609.

Altman EI. 1993. *Corporate Financial Distress and Bankruptcy: A Complete Guide to Predicting & Avoiding Distress and Profiting from Bankruptcy*. John Wiley & Sons: New York.

Altman EI, Haldeman RG, Narayanan P. 1977. ZETA analysis: a new model to identify bankruptcy risk of corporations. *Journal of Banking and Finance* **1**: 29–54.

Altman EI, Marco G, Varetto F. 1994. Corporate distress diagnosis: comparisons using linear discriminant analysis and neural networks (the Italian experience). *Journal of Banking and Finance* **18**: 505–529.

Altman EI, Eom YH, Kim DW. 1995. Failure prediction: evidence from Korea. *Journal of International Financial Management and Accounting* **6**: 230–249.

Beaver WH. 1966. Financial ratios as predictors of failure. *Empirical Research in Accounting: Selected Studies* (supplement to *Journal of Accounting Research*) **4**: 71–111.

Bilderbeek J. 1979. An empirical study of the predictive ability of financial ratios in the Netherlands. *Zeitschrift für Betriebswirtschaft* **5**: 388–407.

Bishop CM. 1995. *Neural Networks for Pattern Recognition*. Oxford University Press: Oxford.

Boritz JE, Kennedy DB, Albuquerque A. 1995. Predicting corporate failure using a neural network approach. *International Journal of Intelligent Systems in Accounting, Finance and Management* **4**: 95–111.

Dimitras AI, Zanakis SH, Zopounidis C. 1996. A survey of business failures with an emphasis on prediction methods and industrial applications. *European Journal of Operational Research* **90**: 487–513.

El Hennawy RHA, Morris RC. 1983. The significance of base year in developing failure prediction models. *Journal of Business Finance & Accounting* **10**: 209–223.

Haykin S. 1999. *Neural Networks, a Comprehensive Foundation*. Prentice-Hall International: London.

Holmen JS. 1988. Using financial ratios to predict bankruptcy: an evaluation of classic models using recent evidence. *Akron Business and Economic Review* **19**: 52–63.

Laitinen EK. 1991. Financial ratios and different failure processes. *Journal of Business Finance & Accounting* **18**: 649–673.

Lennox CS. 1999. The accuracy and incremental information content of audit reports in predicting bankruptcy. *Journal of Business Finance & Accounting* **26**: 757–778.

Mensah YM. 1984. An examination of the stationarity of multivariate bankruptcy prediction models: a methodological study. *Journal of Accounting Research* **22**: 380–395.

Moyer RC. 1977. Forecasting financial failure: a re-examination. *Financial Management* (Spring): 11–17.

OECD. 1999. *OECD Economic Surveys 1998–1999: Belgium/Luxembourg*. Organisation for Economic Co-operation and Development: Paris.

O'Leary DE. 1998. Using neural networks to predict corporate failure. *International Journal of Intelligent Systems in Accounting, Finance and Management* **7**: 187–197.

Olmeda I, Fernandez E. 1997. Hybrid classifiers for financial multicriteria decision making: the case of bankruptcy prediction. *Computational Economics* **10**: 317–335.

Pinches GE, Mingo KA, Caruthers JK. 1973. The stability of financial patterns in industrial organizations. *Journal of Finance* **28**: 389–396.

Pompe PPM, Feelders AJ. 1997. Using machine learning, neural networks, and statistics to predict corporate bankruptcy. *Microcomputers in Civil Engineering* **12**: 267–276.

Pompe PPM, Bilderbeek J. 2005. The prediction of bankruptcy of small and medium-sized industrial firms. *Journal of Business Venturing*, forthcoming.

Richardson FM, Kane GD, Lobingier P. 1998. The impact of recession on the prediction of corporate failure. *Journal of Business Finance & Accounting* **25**: 167–186.
Stone M. 1974. Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society B* **2**: 111–147.
VCPB. 1996. *Jaarverslag (1995/1996)*. Vlaamse Commissie voor Preventief Bedrijfsbeleid. Ministerie van de Vlaamse Gemeenschap: Brussels.
Venables WN, Ripley BD. 1994. *Modern Applied Statistics with S-Plus*. Springer-Verlag: New York.
Wilson RL, Sharda R. 1994. Bankruptcy prediction using neural networks. *Decision Support Systems* **11**: 545–557.
Zapranis AD, Refenes AN. 1999. *Principles of Neural Model Identification, Selection and Adequacy: with Applications to Financial Econometrics*. Springer-Verlag: London.
Zavgren CV. 1985. Assessing the vulnerability to failure of American industrial firms: a logistic analysis. *Journal of Business Finance & Accounting* **12**: 19–45.
Zhang L, Altman EI, Chen S. 2003. Financial distress warning models based on MDA and ANN techniques—China's experience (unpublished manuscript).

