The Effect of Sample Size on Different Failure Prediction Methods

Barbro Back
Turku School of Economics and Business Administration,
Information Systems Science,
P.O. Box 110, FIN-20521 Turku, Finland

Teija Laitinen
University of Vaasa, Department of Accounting and Finance,
P.O.Box, 65101 Vaasa, Finland

Jukkapekka Hekanaho
Åbo Akademi University, Department of Computer Science,
Lemminkäisenkatu 14, FIN-20520 Turku, Finland

Kaisa Sere
Åbo Akademi University, Department of Computer Science,
Lemminkäisenkatu 14, FIN-20520 Turku, Finland

Turku Centre for Computer Science
TUCS Technical Report No 155
December 1997

ISBN 952-12-0134-7
ISSN 1239-1891

Abstract

Neural networks and machine learning methods have proved in many ways and in a number of publications to be real challengers to statistical methods - especially to logit and discriminant analysis - in predicting failures. However, most of the studies have used a rather small data set, very often close to only one hundred observations. Therefore, it has been difficult to say whether there are any significant differences between the methods tested. In this study, we compare neural networks, a machine learning method, discriminant analysis and logit analysis using a large data set consisting of 570 companies. We investigate the effects of the prediction capabilities of the methods when using different sample sizes for estimation and testing, i.e., 400-170, 200-90 and 100-50. Our study shows that neural networks and the machine learning method perform better than discriminant analysis and logit analysis when the sample size is 400 while there is no best performer when the sample size is decreased to 200 and 100.

Keywords: bankruptcy predictions, neural networks, machine learning, backpropagation, discriminant analysis, logit analysis, genetic algorithms

TUCS Research Group
Computational Intelligence in Business

1. Introduction

Neural networks and machine learning methods have proved in many ways and in a number of publications to be real challengers to statistical methods - especially to discriminant analysis and logit analysis - in predicting failures. During the last years almost in all of these publications the conclusion has been that the prediction accuracy can be increased by using neural networks instead of traditional linear discriminant analysis or logit analysis.

However, Trigueiros and Taffler (1996) criticise the way the methods have been applied in these studies. They are above all concerned with the problems connected with the number of fitted coefficients compared to the number of observations. They review a number of studies and find that this problem is obvious. Most of the studies have used a rather small data set, very often less than one hundred observations. Despite of this, the number of fitted coefficients has been very high, in some cases even higher than the number of cases in the model fitting. To avoid the overfitting problem they suggest an old rule-of-thumb to restrict the number of coefficients to a maximum of 10% of the number of cases. To our knowledge, the cause of overfitting within the feed forward neural networks is not fully understood and no exact rules can be given on how to avoid the problem. Different authors give different guidelines and no uniform approach has been found.

Furthermore, Trigueiros & Taffler (1996) state that the results achieved by studies with a large sample such as the one by Altman et al (1994) are contradictory to findings in neural network studies with small samples. Altman et al found little or no difference in classification performance between neural networks and conventional multivariate statistical techniques. In contrast to this, the studies e.g. by Coats and Fant (1991), Salchenberger, Cinar, and Lash (1992), Tam and Kiang (1992), Fletcher and Goss (1993), Wilson and Sharda (1994), and Back, Laitinen and Sere (1996) using small data samples generate evidence that the prediction accuracy in failure prediction can be increased by using neural networks instead of traditional statistical methods.

The sample size is often restricted, because the traditional failure prediction research is based on matched pairs. In our study, we are not paying attention to matching failed and nonfailed companies. Instead of this, we vary the number of observations in estimation and training between 100 and 400 firms. The aim is twofold. First, we avoid the problems connected with statistical issues in most of the previous studies using data with insufficient size. Second, a large sample size can show if neural networks can outperform logit analysis when the data sample is big and provide generalisable results instead of those of experiment types in previous studies. Observe that we do not compare our results directly with those obtained using data with matched pairs.

The paper is organized as follows. Section 2 briefly presents the central features of the methods. In Section 3 the model is specified. Section 4 presents the data. Section 5 reports the used variables and the results when changing the sample size. The study ends with discussion and conclusions in Section 6.

2. Methods

2.1 Discriminant analysis

Discriminant analysis tries to derive a linear combination of two or more independent variables trying to discriminate the best group among a priori defined groups, which in our case are failed and nonfailed companies. This is achieved by the statistical decision rule of maximizing the between-group variance relative to the within-group variance. This relationship is expressed as the ratio of the between-group to the within-group variance. The discriminant analysis derives the linear combinations from an equation that takes the following form:

$$Z = w_1x_1 + w_2x_2 + \ldots + w_nx_n$$

where

Z	= discriminant score
w_i (i=1, 2, ... ,n)	= discriminant weights
x_i (i=1, 2, ... ,n)	= independent variables, the financial ratios

Thus, each firm receives a single composite discriminant score which is then compared to a cut-off value, which determines to which group the company belongs to.

Discriminant analysis does very well provided that the variables in every group follow a multivariate normal distribution and the covariance matrices for every group are equal. However, empirical experiments have shown that especially failed firms violate the normality condition. In addition, the equal group variances condition is also violated. Moreover, multicollinearity among independent variables is often a serious problem, especially when stepwise procedures are employed (Hair et al., 1992). However, empirical studies have proved that the problems connected with normality assumptions do not weaken its classification capability, but its prediction ability.

The two most frequently used methods in deriving the discriminant models have been the *simultaneous (direct) method* and the *stepwise method.* The former is based on model construction by e.g. theoretical grounds, so that the model is ex ante defined and then used in discriminant analysis. When the stepwise method is applied, the procedure selects a subset of variables to produce a good discrimination model using forward selection, backward elimination, or stepwise selection. The stepwise method that we used is a built in function in the SAS-program.

The stepwise selection begins with no variables in the model. At each step, if the variable that contributes least to the discriminatory power of the model measured by Wilks` lambda fails to meet the criterion to stay, then it will be removed. The variable not in the model that contributes most to the discriminatory power of the model is entered. When all the variables in the model meet the criterion to stay and none of the other variables meet the criterion to enter, the stepwise selection process stops (SAS 1988:910).

There are also other modeling strategies, like the one suggested by Hosmer and Lemeshow (1989). In some contexts it has been stated that the disadvantage of the stepwise procedure is that the economic importance of variables is ignored and the statistical grounds are stressed. The modeling strategy by Hosmer and Lemeshow is less mechanical in that it allows for the analyst's judgment concerning variables to be included in the model.

2.2 Logit analysis

Logistic regression analysis has also been used to investigate the relationship between binary or ordinal response probability and explanatory variables. The method fits linear logistic regression model for binary or ordinal response data by the method of maximum likelihood. Among the first users of logit analysis in the context of financial distress was Ohlson (1980). Like discriminant analysis, this technique weights the independent variables and assigns a Z score in a form of failure probability to each company in a sample. The advantage of this method is that it does not assume multivariate normality and equal covariance matrices as discriminant analysis does. Logit analysis incorporates nonlinear effects, and uses the logistical cumulative function in predicting a bankruptcy, i.e.,

$$probability\ of\ failure = \frac{1}{1+e^{(-Z)}} = \frac{1}{1+e^{-(w_0+w_1x_1+...+w_nx_n)}}$$

Logistic analysis applies the same variable selection methods as discriminant analysis presented above. For model construction we selected, as in the case of discriminant analysis, the stepwise method that is a built in function in the SAS-program. The procedure starts by estimating parameters for variables forced into the model, i.e. intercept and the first possible explanatory variables. Next, the procedure computes the adjusted chi-squared statistic for all the variables not in the model and examines the largest of these statistics. If it is significant at the specified level, in our study 0.05, the variable is entered into the model. Each selection step is followed by one or more elimination step, i.e. the variables already selected into the model do not necessarily stay. The stepwise selection process terminates if no further variable can be added to the model, or if the variable just entered into the model is the only variable removed in the subsequent elimination (see SAS 1990).

2.3 Neural networks

An (artificial) *neural network* consists of a large number of processing elements, *neurons*, and connections between them. It basicly implements some *function* f that maps a set of given input values x to some output values y: $y = f(x)$. A neural network tries to find the best possible approximation of the function f as a complex combination of elementary nonlinear functions. This approximation is coded in the neurons of the network using *weights* that are associated with each neuron.

A formal *neuron* is the basic element of any neural network. A neuron is a simple processing element that as inputs takes an n-dimensional vector $[x_1, \ldots, x_n]^T$, extended with a constant component $x_0 = 1$. The neuron forms the weighted sum

$$w^T x = w_0 + \sum_1^n w_i x_i,$$

where $x = [1, x_1, ..., x_n]^T$ and where $w = [w_0, ..., w_n]^T$ is the *weight vector* which is stored in the neuron. In the simplest case the output of a neuron is the sign of this expression, $y = sgn(w^T x)$. Such a neuron can classify n-dimensional vectors into two different classes when the weights are determined so that $y = 1$ for class 1 vectors and $y = -1$ for class 2 vectors.

The weights of a neural network are *learned* using an iterative procedure during which examples of correct input-output associations are shown to the network and the weights are modified so that the network starts to mimic the desirable input-output behavior. Learning in a neural network then means finding an appropriate set of weights. This ability to learn from examples and based on this learning the ability to generalize to new situations is the most attractive feature of the neural network paradigm. For a more thorough description of neural networks we refer to Hecht-Nielsen (1991), and Hertz et al. (1991).

The variables for the input vectors can be chosen by an exhaustive search from the available variables, but this becomes very time consuming when the choice is to be done among several variables. Another method to choose the variables for the networks is to use a *genetic algorithm*. A genetic algorithm simulates Darwinian evolution. It maintains *a population* of *chromosomes*, where a chromosome is a candidate-solution to the problem we want to solve. Chromosomes are often called *strings* in a genetic algorithm context. A string in its turn, consists of a number of *genes*, which may take some number of values, called *alleles.* The genetic algorithm terms for genes and alleles are *features* and *values*. Associated with each string is a *fitness value*, which determines how 'good' the string is. The fitness value is determined by a *fitness function*, which we can think of as some measure of profit or goodness that we want to maximize. Basically, there are three operators that lead to good results in a genetic algorithm, namely *reproduction*, *crossover*, and *mutation*. For a more thorough description of genetic algorithms we refer to Goldberg (1989).

2.4 Rule learners

Symbolic machine learning aims at analysing data in order to induce intelligible descriptions that are comrehensible to humans. In the case of rule learners the prediction models consist of sets of rules, where each rule is of form

if $condition_1$ and $condition_2$ and ... and $condition_n$ **then** $class = cl_1$.

In other words each rule specifies some conditions on the independent variables under which we may predict the value of the dependent variable. Each condition expresses possible values of a variable. We distinquish between two different kinds of conditions; nominal expressions and disjunctive expressions. The nominal expression ($v = s$) simply relates a variable v to a value s. The expression is true for an example firm if the value of variable v in firm is s. A disjunctive expression ($v = s1$) or ($v = s2$) or ... or ($v = sm$) is a disjunction of nominal expressions ($v = si$), all refering to the same variable v. A disjunctive expression is true for a firm if at least one of the nominal expressions is true. If all the conditions in a rule are true for a firm, we say that the rule covers the firm. The final model is a set of rules and a firm belongs to a certain class if it is covered by at least one of the rules for the class.

The most common way of learning a set of rules, is to learn one rule at a time using a covering approach, see e.g. Michalski (1983). In this method a rule is first learned by generalising examples, so that the rule covers a large number of the examples in the target class. After this the examples covered by the rule are removed and the process is repeated. An alternative approach to rule learning is to use decision trees, see e.g. Quinlan (1993). In this method a decision tree is first learned from the data. In the second phase the tree is converted to a set of rules. Here we concentrate on a third approach to rule learning based on genetic algorithms, where the whole set of rules in learned simultaneously.

The genetic algorithm used is somewhat different from the one described in the previous section. Here the genetic algorithm is used directly to learn the rules. Like in the previous section the population consists of a set of chromosomes. In this case however, each chromosome encodes a rule. The purpose of the genetic algorithm is to learn a model consisting of a set of rules that characterises the target class.

The search for a good rule-based model is guided by a fitness function. We use a combination of the minimal description length principle (Rissanen, 1982) and information gain criterion (Quinlan, 1990). The minimal description length principle is a stochastic measure of model complexity and model fit. The principle states that the best model m for some data D is the one that minimises the total complexity of the model and the exceptions of the model. Both the model complexity and the exceptions are measured as the number of bits needed to encode them

$$bits(m_i, D) = lm(m_i) + le(D \mid m_i),$$

where *lm* and *le* are the lengths of the encoding of the model and the exceptions, respectively. One of the benefits of the minimal description length principle is that it quantifies model error and complexity within a single information-theoretic framework, thus allowing comparisions between models of different complexity. One difficulty however, is the need to find a suitable encoding function for models and exceptions. The encoding we use is described in Hekanaho (1997) and is similar to the encoding defined by Quinlan (1995).

Information gain quantifies how much information is gained by a rule, i.e., how useful the rule is in separating between examples of different classes. The criterion is derived from Shannons Information Theory, Shannon and Weaver (1949).

3. Model specification

Among the studies focusing on failure prediction there are some papers that have been especially contributive. Among the first studies that identify the problemacy with selecting predictors and prediction methods were the studies of Ramser and Foster (1931), Fitzpatrick (1932), Winakor and Smith (1935), and Merwin (1942). These studies settled the fundamentals for failure prediction research. A change in the research tradition took place when Beaver (1966) presented the univariate analysis approach. Soon after Beaver, Altman (1968) pioneered with the use of multivariate approach in the context of bankruptcy models. After the Altman study this multivariate approach became dominant in these models. Actually the study of Beaver remained as the main study applying univariate analysis. The studies by Deakin (1972), Edminster (1972), Blum (1974), Altman et al. (1977) and El Hennawy and Morris (1983) are representative examples of studies that used multiple discriminant analysis technique.

There is no theoretical basis for which predictors should be used for different models. In our study we have selected 33 ratios that previous research has found useful in bankruptcy predictions as a starting point. The ratios are presented in Table 1. The final predictors for linear discriminant analysis and logit analysis are selected by stepwise regression and for neural networks and machine learning by genetic algorithms and are presented in Section 5.

	Ratio	**Study**	**Compustat Item**
1	Cash/Current Liabilities	E, D	1 / 5
2	Cash Flow/Current Liabilities	E	(18+14) / 5
3	Cash Flow/Total Assets	E-M	(18+14) / 6
4	Cash Flow/Total Debt	Bl, B, D	(18+14) / 181
5	Cash/Net Sales	D	1 / 12
6	Cash/Total Assets	D	1 / 6
7	Current Assets/Current Liabilities	M, B, D, A-H-N	4 / 5
8	Current Assets/Net Sales	D	4 / 12
9	Current Assets/Total Assets	D,E-M	4 / 6
10	Current Liabilities/Equity	E	5 / 60
11	Ebit / Total Interest Payments	A-H-N	178 / 15
12	Equity/Net Sales	R-F, E	60 / 12
13	Inventory/Net Sales	E	3 / 12

14	Long Term Debt/Equity		9 / 60
15	Long Term Debt/Net Capital	E-M	9 / (6-70-71)
16	MV of Equity/Book Value of Debt	A, A-H-N	(24*25) / 181
17	Net Income/Total Assets	B, D	172 / 6
18	Net Quick Assets/Inventory	Bl	(1+2-5) / 3
19	Net Sales/Total Assets	R-F, A	12 / 6
20	Net worth / Total Liabilities	Bl	(6-181) / 181
21	Operating Income/Total Assets	A, T, A-H-N	13 / 6
22	Quick Assets/Current Liabilities	D, E-M	(1+2) / 5
23	Quick Assets/Net Sales	D	(1+2) / 12
24	Quick Assets/Total Assets	D, T, E-M	(1+2) / 6
25	Rate of Return to Common Stock	Bl	(237 / 60)*100
26	Retained Earnings/Total Assets	A, A-H-N	(36 / 6)*100
27	Total debt / Equity	S	181 / 60
28	Total Debt/Total Assets	B, D	(5+9) / 6
29	Total Liabilities/ Net Capital Emp	T	181 / (6-70-71)
30	Working Capital/Net Sales	E, D	179 / 12
31	Working Capital/Net Worth	T	179 / (6-181)
32	Working Capital/Total Assets	W-S,M,B,A,D	179 / 6
33	Ebit/Total assets		100*(178/6)

Legend:

A	Altman 1968	E-M	El Hennawy and Morris 1983
A-H-N	Altman, Haldeman, and Narayanan 1977	F	Fitzpatrick 1932
B	Beaver 1966,	M	Mervin 1942
Bl	Blum	R-F	Ramser and Foster 1931
D	Deakin 1972,	W-S	Winakor and Smith 1935
E	Edminster 1972		

Table 1. Financial ratios found to be useful in previous bankruptcy prediction studies.

4. Data

4.1 Bankruptcy year

The data was collected from the Compustat database which covers 16,775 American companies. In the first phase we selected all the firms that were reported as failed in Compustat between 1985 and 1993, 380 in all. Table 2 presents the numbers of failed companies per year, 1985-1993.

Year	Number of failed companies
1985	27
1986	36
1987	37
1988	54
1989	59
1990	68
1991	49
1992	38
1993	12
Total	*380*

Table 2. Yearly frequence of companies coded "bankruptcy" in Compustat Annual Industrial File.

In the following phase we rejected the firms that had all the variables missing for some year or had some evident mistakes in their financial statements (i.e. negative net sales). Furthermore, we rejected firms that did not have financial statements for the whole research period, one to three years prior to failure. The data therefore included firms that failed in 1987 or after so that the financial statements for the whole three years research period could be used. Finally, the number of failed companies was 285. The same number of nonfailed companies was also then included, i.e., the total data set included 570 companies.

After sampling the data it was divided into two parts: training or estimation set and validation set. About one third of the data, i.e., 85 failed and 85 nonfailed companies, were put aside as test data ($test_1$). Thus, in the first phase we included 400 companies in the training set. In the second phase the estimation data was divided into two subsamples of 200 companies each. One of these subsamples was then used in estimating the DA- and logit functions and learning the neural networks to examine the correct results. Similarly, the test data was divided into two subsamples of 90 firms per sample ($test_2$, $test_3$).

Finally, the smallest estimation and learning sample comprised only half of the previous data, i.e., 50 failed and 50 nonfailed companies. As test data we used the original test sample with 170 observations as well as smaller data sets of 50 firms ($test_4$, $test_5$). Tables 3a and 3b present yearly number of failed and nonfailed companies in every data set.

	Estimation or training sample	**Validation sample**
Sample size	**400**	**170**
Year		
1987	21	9
1988	31	16
1989	38	15
1990	42	19
1991	34	11
1992	25	12
1993	9	3
Total	200	85

Table 3a. Distribution of failed companies in the sample by year coded "bankruptcy" in Compustat Annual Industrial File. The original estimation or learning sample and validation sample $test_1$.

	Estimation or training sample	Validation sample		Estimation or training sample	Validation sample	
Sample size	200	90		100	50	
Year		$test_2$	$test_3$		$test_4$	$test_5$
1987	11	5	5	4	2	2
1988	17	9	8	8	5	4
1989	18	8	8	10	5	5
1990	22	9	11	9	5	6
1991	16	6	6	9	4	5
1992	12	7	5	8	4	2
1993	4	1	2	2	0	1
Total	100	45	45	50	25	25

Table 3b. Distribution of failed companies in the sample by year coded "bankruptcy" in Compustat Annual Industrial File. Reduced estimation or learning samples of 200 and 100 companies and reduced validation samples of 90 companies ($test_2$, $test_3$) and 50 companies ($test_4$, $test_5$).

4.2 Industry and size

The failed companies included in Compustat during the years 1985-1993 are presented in Table 4 by industry. The main industry among the failed companies was manufacturing industries which was represented by 26% of the companies. The next largest industry was retail and services, about 18%. The smallest proportions were represented by agriculture production and health, educational & related services, together less than 6%.

The largest industry among the nonfailed companies was also the manufacturing industry including 16% of the companies. Other industries were represented with equal proportions, except the same industries than those among failed companies, agriculture production and health, educational & related services, together about 7%.

ISIC - Codes		F	NF
100	Agriculture Production	5	2
1000	Mining,Oil and Gas Extraction, Construction	42	49
2000	Finished Products	46	45
3000	Manufacturing Industries	99	60
4000	Transportation, Communications, Electricity, Gas	21	47
5000	Retail and Services	67	45
6000	Depository and Credit Institutions, Investment Offices	39	45
7000	Lodging Places, Services Amusement & Recreation Services	44	47
8000	Health, Educational & Related Services	15	26
9000	Other	2	14
		380	380

Table 4. The industries of bankruptcy coded firms in Compustat during 1985-1993 and the industries in the sample of nonfailed companies.

The industries of failed and nonfailed companies in the subsamples are presented in Tables 5 and 6. The ISIC- codes are corresponding the ones in Table 4.

Sample Size	**400**		**200**		**100**	
ISIC - Codes	**F**	**NF**	**F**	**NF**	**F**	**NF**
100	2	2	2	0	2	0
1000	17	23	7	10	2	6
2000	19	24	11	14	6	9
3000	63	28	29	16	12	9
4000	14	27	8	14	3	5
5000	36	23	16	9	10	5
6000	21	27	10	14	6	5
7000	22	27	12	12	6	6
8000	6	10	5	5	3	3
9000	0	9	0	6	0	2
Total	200	200	100	100	50	50

Table 5. The industries of failed and nonfailed companies in the sample and subsamples. Estimation Samples

Sample Size	170 test1		90 test2		90 test3		50 test4		50 test5	
ISIC - Codes	F	NF	F	NF	F	NF	F	NF	F	NF
100	0	0	0	0	0	0	0	0	0	0
1000	8	13	6	12	3	4	4	4	2	2
2000	12	17	5	4	7	12	3	3	4	7
3000	21	15	11	8	11	7	6	5	6	4
4000	6	7	4	3	3	5	3	2	2	2
5000	14	11	7	7	8	4	3	5	3	2
6000	8	6	5	3	3	4	2	3	1	3
7000	9	8	5	4	5	5	2	1	3	3
8000	5	6	0	3	5	3	0	1	4	2
9000	2	2	2	1	0	1	2	1	0	0
Total	85	85	45	45	45	45	25	25	25	25

Table 6. The industries of failed and nonfailed companies in the sample and subsamples. Validation Samples

5. Predictors and prediction results selected

In our analysis we did not apply any preclassification system for the data, i.e., we did not use any technique to ensure that all nonfailed companies in our sample are sound. Thus, the nonfailed firms were selected randomly from the basic data. This means that in our sample of sound firms there are firms suffering from financial distress, maybe even firms that are going to fail within a year, i.e., in 1994. This was done on purpose because we wanted to have all types of nonfailed companies in our data. The absolute prediction accuracy may suffer from this starting point, but the comparison between neural network, machine learning, logistic and discriminant analysis is more realistic when the data is not manipulated ex ante.

5.1 Discriminant analysis

The variables for the discriminant models were selected using the SAS stepwise procedure. The procedure started with all the 33 financial ratios as variables. If all variables in the estimated model had intuitive signs, the model was accepted with no adjustments. If the model included variables with counter-intuitive signs, these variables were rejected and the model was re-estimated. This procedure was repeated until all the variables in the model had intuitive signs.

The models were estimated for one, two, and three years prior to failure for all three data sets, i.e. 400, 200, and 100 firms. The analysis was conducted using the SAS statistical package. The models were estimated for every year and every data set separately. This means that at the end we had three alternative models for every year, estimated from three alternative data sets. The variables that were selected into the models using discriminant analysis are presented in Table 7.

Sample Size n=400		n=200	n=100
Year			
1	Cash Flow/Current Liab.	Cash Flow/Current Liab.	Cash Flow/Total Ass.
	Cash/Total Assets	Cash/Total Assets	Equity/Net Sales
	Equity/Net Sales	Current Liab./Equity	Quick Ass./Total Ass.
	Net Worth/Total Liab.	Equity/Net Sales	Rate of Ret/ Comm. Stock
	Operating Inc./Total Ass.	Operating Inc./Total Ass.	
	Total Debt/Total Ass.	Total Debt/Total Ass.	
2	Cash Flow/Current Liab.	Cash Flow/Current Liab.	Cash Flow/Current Liab.
	Cash/Total Assets	Cash/Total Assets	Current Liab./ Equity
	Equity/Net Sales	Equity/Net Sales	Equity/Net Sales
	Quick Assets/ Total Ass.	Rate of Ret./Common Stock	Quick Ass./Current Liab.
	Rate of Ret./Common Stock	Ebit/Total Assets	
	Total Liab./Net Cap.Emp.		
	Ebit/Total Assets		
3	Net worth/Total Liab.	Cash Flow/Current Liab.	Cash/Net Sales
		Cash/Total Assets	Equity/Net Sales
		Quick Assets/Net sales	Operating Inc./Total Ass.
		Total Debt/Total Assets	
		Ebit/Total Assets	

Table 7. The variables in DA-models one, two, and three years prior to the failure.

In the first phase the sample of 400 firms, 200 failed and 200 nonfailed companies were used to estimate a discriminant model. After the model was estimated it was first crossvalidated and then tested with a separate holdout sample with 85 failed and 85 nonfailed companies. These results are presented in Table 8. The results indicate that the discriminant function classifies 85% of the nonfailed companies correct in the crossvalidation and even 87.1% in the holdout sample. Two years prior to failure the total prediction accuracy decreases about 10% in both the crossvalidation, the estimation, and in the holdout sample. Three years prior to failure a decrease of about 10% in the total prediction accuracy can again be found in the estimation and the crossvalidation but the results in the holdout sample remain quite stable. The prediction accuracy for nonfailed companies three years prior to failure is extremely low. Here the cutting point could have been changed.

The overall prediction accuracy is somewhat lower than in studies in which the data is collected from the same calendar year, i.e., the failure has occurred during the same year. As it was described earlier the data in this study is from Compustat and the failure has occurred during the years 1987-1993. When the time period is as long as in this study, e.g., changes in macroeconomics factors are affecting the values of the financial ratios. This makes the models more insensitive to the differences between the failed and nonfailed companies. Anyhow, when we are studying the effect of the sample size on the prediction accuracy of alternative methods and on the generalization of the results the length of time-series is a problem common to all three methods and should not affect the comparison of the results.

In the second phase the data was decreased to 100 failed and 100 nonfailed companies (see Table 9). After the DA-function had been estimated it was tested first in crossvalidation with the same data set of 200 companies. After that the model was tested in three alternative holdout samples. The first holdout sample, $test_1$, consisted of the same 170 firms as the original holdout sample that was used when testing the DA-model estimated from the data of 400 companies. Furthermore, the holdout sample $test_1$ was divided into two parts, $test_2$ and $test_3$. This was done to examine the results when the data set used in testing is reduced.

When the data set used in the estimation is reduced, the results in crossvalidation are quite stable. Opposite to this, testing the function indicates problems in the generalization of the results. In all test sets the prediction accuracy decreases when we move away from the first year prior to failure to the second year. Then suddenly three years prior to failure the prediction accuracy increases 1-7 percentage units, depending on which data set is used in the testing.

Finally, the DA-function was estimated from the data of 50 failed and 50 nonfailed companies. Again, it was tested with the data set $test_1$, and in addition to this, the previous data sets $test_4$ and $test_5$ were reduced randomly to 50 firms each. The results are presented in Table 10.

The results are becoming more difficult to interpret. In addition to the previous phenomena concerning the increase in prediction accuracy with the data sets $test_1$ and $test_2$, the results in $test_5$ indicate instability also one year prior to failure. Furthermore, the variation between the prediction results is quite high. One year prior to failure it is 10 percentage units, two years prior to failure 4 and three years prior to failure even 14. This indicates that the results are hardly generalisable if the estimation data is as small as 100 observations.

n	400						170		
	estimation			crossvalidation			$test_1$		
year	nf	f	t	nf	f	t	nf	f	t
1	85.0	68.5	**76.8**	85.0	67.0	**76.0**	87.1	62.4	**74.8**
2	71.0	66.5	**68.8**	68.5	65.5	**67.0**	62.4	62.4	**62.4**
3	27.5	88.5	**58.0**	27.5	88.5	**58.0**	32.9	89.4	**61.2**

Table 8. Prediction accuracy when the data set of 400 firms was used in estimation and 170 firms in testing.

n	200						170			90			90		
	estimation			crossvalidation			$test_1$			$test_2$			$test_3$		
year	nf	f	t	nf	f	t	nf	f	t	nf	f	t	nf	f	t
1	81.0	69.0	**75.0**	81.0	66.0	**73.5**	85.9	60.0	**73.0**	80.0	60.0	**70.0**	88.9	55.6	**72.3**
2	68.0	72.0	**70.0**	65.0	71.0	**68.0**	52.9	67.1	**60.0**	51.1	66.7	**58.9**	51.1	66.7	**58.9**
3	72.0	64.0	**68.0**	67.0	63.0	**65.0**	60.0	68.2	**64.1**	53.3	66.7	**60.0**	64.4	68.9	**66.7**

Table 9. Prediction accuracy when the data set of 200 firms was used in estimation and 170 ($test_1$), 90 ($test_2$), and 90 ($test_3$) firms in testing.

n	100						170			50			50		
	estimation			crossvalidation			$test_1$			$test_4$			$test_5$		
year	nf	f	t	nf	f	t	nf	f	t	nf	f	t	nf	f	t
1	66.0	80.0	**73.0**	64.0	78.0	**71.0**	54.1	65.9	**60.0**	60.0	68.0	**64.0**	44.0	64.0	**54.0**
2	62.0	76.0	**69.0**	62.0	76.0	**67.0**	57.7	60.0	**58.9**	68.0	56.0	**62.0**	56.0	60.0	**58.0**
3	58.0	80.0	**69.0**	54.0	76.0	**65.0**	41.2	77.7	**59.5**	52.0	84.0	**68.0**	32.0	76.0	**54.0**

Table 10. Prediction accuracy when the data set of 100 firms was used in estimation and 170 ($test_2$), 50 ($test_2$), and 50 ($test_2$) firms in testing.

5.2 Logit analysis

We selected the variables for the logit models in the same way as we did for the discriminant analysis models. The selected models are presented in Table 11.

Sample Size	n=400	n=200	n=100
Year			
1	Cash Flow/Current Liab.	Cash Fl./Current Liab.	Cash Flow/Total Debt
	Cash/Total Assets	Cash/Total Assets	Cash/Net Sales
	Net Worth/Total Liab.	Equity/Net Sales	Cash/Total Assets
	Operating Inc./Total Ass.	Net Worth/Total Liab.	Equity/Net Sales
	Quick Assets/Net Sales	Operating Inc./Tot.Ass.	Rate of Ret./ Comm. Stock
	Total Debt/Total Assets	Quick Assets/Net Sales	
		Rate of Ret./Comm. Stock	
2	Cash Flow/Current Liab.	Cash Fl./Current Liab.	Cash Flow/Current Liab.
	Cash/Total Assets	Cash/Total Assets	Current Liab./Equity
	Ebit/Total Assets	Ebit/Total Int. Paym.	Ebit/Total Assets
	Quick Assets/Net Sales	Equity/Net Sales	Quick Assets/Net Sales
	Rate of Return/Comm. Stock	Quick Assets/Net Sales	Working Cap./Tot Ass.
	Tot. Liab./Net Cap. Emp.		
3	Cash Flow/Current Liab.	Cash Fl./Current Liab.	Operating Inc./Total Ass
	Cash/Total Assets	Cash/Total Assets	Quick Assets/Net Sales
	Total Debt/Total Assets	Ebit/Total Assets	Total Debt/Total Assets
		Equity/Net Sales	

Table 11. The logit-models one, two, and three years prior the failure.

The prediction results using logit analysis with the largest sample size, 400 companies in estimation and 170 companies in testing are presented in Table 12. In comparison with the results of discriminant analysis, the prediction accuracy is somewhat higher in all three years. Again, the results are stable so that the prediction accuracy increases when approaching the failure year.

Similarly as with DA, the results are affected immediately when the data is reduced by 50%. As Table 13 illustrates, one year prior to failure the prediction accuracy in the holdout sample varies between 63.4% and 74.5%. In the test data with the logistic function estimated from the sample of 400 firms the prediction accuracy was even higher, i.e., 75.9%. Two and three years prior to failure with the $test_1$ and $test_3$ data sets the prediction accuracy is again increased while the $test_2$ data indicates the decrease in the prediction accuracy (which should be the case).

As Table 14 indicates, decreasing the data to 100 firms affects also the results of the estimation. In addition to the crossvalidation results and to the two test data results the function in the estimation sample classifies the firms better three years prior to failure than it does two years prior to failure. The $test_1$ data leads to an opposite finding.

	n=400						n1=170		
	estimation			crossvalidation			$test_1$		
year	nf	f	t	nf	f	t	nf	f	t
1	81.5	77.5	**79.5**	81.0	76.5	**78.8**	80.0	71.8	**75.9**
2	70.5	67.5	**69.0**	69.0	66.5	**67.8**	71.8	60.0	**65.9**
3	63.0	64.0	**63.5**	62.0	63.0	**62.5**	61.2	67.1	**64.2**

Table 12. Prediction accuracy when the data set of 400 firms was used in estimation and 170 firms in testing

n	200						170			90			90		
	estimation			crossvalidation			test1			test2			test3		
year	nf	f	t	nf	f	t	nf	f	t	nf	f	t	nf	f	t
1	69.0	73.0	**71.0**	69.0	72.0	**70.5**	67.0	74.0	**70.5**	55.6	71.1	**63.4**	75.6	73.3	**74.5**
2	62.0	79.0	**70.5**	61.0	76.0	**68.5**	42.4	77.6	**60.0**	44.4	75.6	**60.0**	37.8	77.8	**57.8**
3	68.0	68.0	**68.0**	65.0	65.0	**65.0**	51.8	71.8	**61.8**	51.1	66.7	**58.9**	51.1	75.6	**63.4**

Table 13. Prediction accuracy when the data set of 200 firms was used in estimation and 170 ($test_1$), 90 ($test_2$), and 90 ($test_3$) firms in testing.

n	100						170			50			50		
	estimation			crossvalidation			test1			test4			test5		
year	nf	f	t	nf	f	t	nf	f	t	nf	f	t	nf	f	t
1	76.0	80.0	**78.0**	72.0	74.0	**73.0**	75.3	63.5	**69.4**	68.0	60.0	**64.0**	76.0	72.0	**74.0**
2	66.0	62.0	**64.0**	66.0	56.0	**61.0**	72.9	55.3	**64.1**	76.0	48.0	**62.0**	68.0	40.0	**54.0**
3	66.0	72.0	**69.0**	58.0	70.0	**64.0**	50.6	75.3	**63.0**	52.0	80.0	**66.0**	52.0	76.0	**64.0**

Table 14. Prediction accuracy when the data set of 100 firms was used in estimation and 170 ($test_1$), 50 ($test_4$), and 50 ($test_5$) firms in testing.

5.3 Neural networks

The neural networks were trained and tested extensively with different architectures focusing on the same aspects as in the previous sections using the backpropagation algorithm and the SNNS neural network package.

Year	Size n=400	Size n=200	Size n=100
1	2.Cash Flow/Current Liab.	2.Cash Flow/Current Liab.	1.Cash/Current Liab.
	5.Cash/Net Sales	3.Cash Flow/Total Ass.	2.Cash Flow/Current Liab.
	6.Cash/Total Ass.	4.Cash Flow/Total Debt	3.Cash Flow/Total Ass.
	7.Current Ass./Current Liab.	6.Cash/Total Ass.	5.Cash/Net Sales
	8.Current Ass./Net Sales	7.Current Ass./Current Liab.	6.Cash/Total Ass.
	9.Current Ass./Total Ass.	8.Current Ass./Net Sales	8.Current Ass./Net Sales
	10.Current Liab./Equity	13.Inventory/Net Sales	10.Current Liab./Equity
	13.Inventory/Net Sales	17.Net Income/Total Ass.	13.Inventory/Net Sales
	16.MV of Equity/BV of Debt	18.Net Quick Ass./Inventory	18.Net Quick Ass./Inventory
	17.Net Income/Total Ass.	19.Net Sales/Total Ass.	19.Net Sales/Total Ass.
	18.Net Quick Ass./Inventory	20.Net worth / Total Liab.	23.Quick Ass./Net Sales
	22.Quick Ass./Current Liab.	21.Operating Income/Total Ass.	28.Total Debt/Total Ass.
	24.Quick Ass./Total Ass.	28.Total Debt/Total Ass.	
	25.Rate of Return/Comm. Stock	30.Working Cap./Net Sales	
	28.Total Debt/Total Ass.	32.Working Cap./Total Ass.	
	29.Total Liab./Net Cap. Emp.		
	30.Working Cap./Net Sales		
	31.Working Cap./Net Worth		
	32.Working Cap./Total Ass.		
2	2.Cash Flow/Current Liab	1.Cash/Current Liab.	2.Cash Flow/Current Liab
	3.Cash Flow/Total Ass.	4.Cash Flow/Total Debt	5.Cash/Net Sales
	4.Cash Flow/Total Debt	5.Cash/Net Sales	6.Cash/Total Ass.
	8.Current Ass./Net Sales	6.Cash/Total Ass.	7.Current Ass./Current Liab.
	18.Net Quick Ass./Inventory	7.Current Ass./Current Liab.	8.Current Ass./Net Sales
	19.Net Sales/Total Ass.	9.Current Ass./Total Ass.	9.Current Ass./Total Ass.
	22.QuickAss./Current Liab.	12.Equity/Net Sales	15.Long Term Debt/Net Cap.
	24.Quick Ass./Total Ass.	17.Net Income/Total Ass.	20.Net worth / Total Liab.
	26.Retained Earn./Total Ass.	18.Net Quick Ass./Inventory	23.Quick Ass./Net Sales
	28.Total Debt/Total Ass.	19.Net Sales/Total Ass.	28.Total Debt/Total Ass.
	29.Total Liab./ Net. Cap. Emp.	22.Quick Ass./Current Liab.	31.Working Cap./Net Worth
	33.Ebit/Total Ass.	24.Quick Ass./Total Ass.	33.Ebit/Total Ass.
		32.Working Cap./Total Ass.	
3	1.Cash/Current Liab.	1.Cash/Current Liab.	2.Cash Flow/Current Liab.
	2.Cash Fl./Current Liab.	2.Cash Flow/Current Liab.	6.Cash/Total Ass.
	4.Cash Flow/Total Debt	7.Curr.Ass./Current Liab.	8.Current Ass./Net Sales
	7.Curr.Ass./Current Liab.	8.Current Ass./Net Sales	9.Current Ass./Total Ass.
	13.Inventory/Net Sales	9.Current Ass./Total Ass.	14Long Term Debt/Equity
	16.MV of Equi/BV of Debt	10.Current Liab./Equity	15.Long Term Debt/Net Cap.
	17.Net Income/Total Ass.	15.Long T,Debt/Net Cap.	16.MV of Equi/BV of Debt
	18.Net Quick Ass./ Inventory	23.Quick Ass./Net Sales	17.Net Income/Total Ass.
	19.Net Sales/Total Ass.	24.Quick Ass./Total Ass.	20.Net Worth / Total Liab.
	20.Net Wort/ Total Liab.	26.Ret. Earn./Total Ass.	22.Quick Ass./Curr. Liab.
	23.Quick Ass./Net Sales	28.Total Debt/Total Ass.	23.Quick Ass./Net Sales
	26.Ret. Earn./Total Ass.	29.Total Liab./NetCap.	24.Quick Ass./Total Ass.
	28.Total Debt/Total Ass.	32.Work. Cap./Total Ass.	25.Rate of Ret./ComStock
	29.Tot.Liab./ Net Cap. Emp		26.Ret. Earn./Total Ass.
	31.Work. Cap./Net Worth		28.Total Debt/Total Ass.
			29.Total Liab./ Net Cap. Emp
			31.Working Cap./Net Worth

Table 15a. The variables in the NN-models one and two years prior to failure.

Below we give the results of these tests. The table layouts differ slightly from the previous ones, as the estimation results are not relevant here. The variables in the different neural network models were selected using a genetic algorithm. The models per year are shown in Table 15.

In the final network architecture the learning rate was 0.30 and the momentum parameter was 0.20. The network had one hidden layer with four nodes. The results are presented in table 16-18.

The prediction rate for NN (80.0%, 75.0% and 70.0%) in Table 16 is higher than that for LA in Table 12 (78.8%, 67.8%, and 62.5%) and that for DA in Table 8 (76.0%, 67.0% and 58.0%) for the crossvalidation set for all three years. In the test set, NN has higher prediction accuracies for all three years prior to failure as well.

Table 17 shows that the results are affected when the data set is decreased from 400 to 200. The overall prediction rate decreases and the prediction accuracy three years prior to failure is better than two years prior to failure as was the case also for DA and for LA and which should not be the case. The prediction accuracy one year prior to failure in the hold-out samples varies between 77.8% and 84.4%. The variation range is bigger than in the DA- (70.0-73.0%) and about the same as in the LA experiments (63.4-74.5%). The accuracy in the holdout sample consisting of $test_3$ is better three years prior to failure than two years prior to failure which again should not be the case.

Table 18 shows that the prediction accuracy is 76.7% in the validation sample consisting of 100 observations. The prediction accuracy is again higher three years before failure than two years before failure in the crossvalidation set. There is a big drop in the prediction accuracy when moving from the crossvalidation set to the test sets, at most almost twenty percentage units (from 76.7% to 60.0%).

n	400			170		
	validation			$test_1$		
year	nf	f	t	nf	f	t
1	76.7	83.3	**80.0**	74.1	81.2	**77.7**
2	65.0	85.0	**75.0**	63.5	72.9	**68.2**
3	61.7	78.3	**70.0**	54.1	76.5	**65.3**

Table 16. Prediction accuracy when the data set of 400 firms was used in validation and 170 firms in testing.

n	200			170			90			90		
	validation			$test_1$			$test_2$			$test_3$		
year	nf	f	t	nf	f	t	nf	f	t	nf	f	t
1	76.7	76.7	**76.7**	75.3	81.2	**78.2**	86.7	82.2	**84.4**	77.8	77.8	**77.8**
2	66.7	70.0	**68.3**	60.0	61.2	**60.6**	66.7	62.2	**64.4**	60.0	55.6	**57.8**
3	80.0	63.3	**71.7**	60.0	60.0	**60.0**	64.4	62.2	**63.3**	60.0	60.0	**60.0**

Table 17. Prediction accuracy when the data set of 200 firms was used in estimation and 170 ($test_1$), 90 ($test_2$), and 90 ($test_3$) firms in testing.

n	100			170			50			50		
	validation			$test_1$			$test_4$			$test_5$		
year	nf	f	t	nf	f	t	nf	f	t	nf	f	t
1	73.3	80.0	**76.7**	64.7	72.9	**68.8**	64.0	56.0	**60.0**	68.0	76.0	**72.0**
2	53.3	80.0	**66.7**	65.9	69.4	**67.7**	64.0	60.0	**62.0**	56.0	64.0	**60.0**
3	46.7	93.3	**70.0**	65.9	82.4	**65.9**	40.0	84.0	**62.0**	44.0	76.0	**60.0**

Table 18. Prediction accuracy when the data set of 100 firms was used in validation and 170 ($test_1$), 50 ($test_4$), and 50 ($test_5$) firms in testing.

5.4. Symbolic rules

We use the genetic algorithm DOGMA by Hekanaho (1997), that can take advantage of both background domain knowledge and examples, to learn the rules. Since DOGMA is only able to work on discrete data, we have applied a discretiser that turns the continuous data into discrete intervals. We applied the Discretize utility in the MLC++ program, Kohavi et al (1994), with its default settings. This performs an entropy based discretisation of the examples. The result is a set of variables, and for each selected variable a set of intervals. Tables 19a-c show the variables selected by the discretiser.

Size=400	Size=200	Size=100
1.Cash/Current Liab.	1.Cash/Current Liab.	2.Cash Flow/Current Liab.
2.Cash Flow/Current Liab.	2.Cash Flow/Current Liab.	3.Cash Flow/Total Ass.
3.Cash Flow/Total Ass.	3.Cash Flow/Total Ass.	4.Cash Flow/Total Debt
4.Cash Flow/Total Debt	4.Cash Flow/Total Debt	8.Current Ass./Net Sales
5.Cash/Net Sales	5.Cash/Net Sales	12.Equity/Net Sales
6.Cash/Total Ass.	6.Cash/Total Ass.	16.MV of Equity/BV of
7.Current Ass./Current Liab.	7.Current Ass./Current Liab.	17.Net Income/Total Ass.
8.Current Ass./Net Sales	11.Ebit / Total Interest	19.Net Sales/Total
10.Current Liab./Equity	12.Equity/Net Sales	20.Net worth / Total Liab.
11.Ebit / Total Interest	16.MV of Equity/BV of Debt	23.Quick Ass./Net Sales
12.Equity/Net Sales	17.Net Income/Total Ass.	26.Retained Earn./Total
13.Inventory/Net Sales	18.Net Quick Ass./Inventory	28.Total Debt/Total Ass.
14Long Term Debt/Equity	19.Net Sales/Total Ass.	29.Total Liab./ NetCap.
16.MV of Equity/BV of	20.Net worth / Total Liab.	32.Working Cap./Total
17.Net Income/Total Ass.	21.Operating Income/Total	
18.Net Quick Ass./Inventory	22.Quick Ass./Current Liab.	
19.Net Sales/Total Ass.	23.Quick Ass./Net Sales	
20.Net worth / Total Liab.	26.Retained Earn./Total Ass.	
21.Operating Income/Total	27. Total debt / Equity	
22.Quick Ass./Current Liab.	28.Total Debt/Total Ass.	
23.Quick Ass./Net Sales	29.Total Liab./ NetCap. Emp.	
25.Rate of Ret./ComStock	32.Working Cap./Total Ass	
26.Retained Earn./Total Ass.	33.Ebit/Total Ass.	
27. Total debt / Equity		
28.Total Debt/Total Ass.		
29.Total Liab./ NetCap.		
30. Working Capital/Net		
31.Working Cap./Net Worth		
32.Working Cap./Total Ass		
33.Ebit/Total Ass.		

Table 19a. The variables in the DOGMA-models one year prior to failure.

Size=400	Size=200	Size=100
1.Cash/Current Liab.	1.Cash/Current Liab.	8.Current Ass./Net
2.Cash Flow/Current Liab.	2.Cash Flow/Current Liab.	12.Equity/Net Sales
3.Cash Flow/Total Ass.	3.Cash Flow/Total Ass.	19.Net Sales/Total Ass.
4.Cash Flow/Total Debt	4.Cash Flow/Total Debt	
6.Cash/Total Ass.	6.Cash/Total Ass.	
7.Curr.Ass./Current Liab.	7.Curr.Ass./Current Liab.	
10.Current Liab./Equity	11. Ebit / Total Interest	
11. Ebit / Total Interest	12.Equity/Net Sales	
12.Equity/Net Sales	13.Inventory/Net Sales	
13.Inventory/Net Sales	16.MV of Equi/BV of Debt	
16.MV of Equi/BV of Debt	17.Net Income/Total Ass.	
17.Net Income/Total Ass.	19.Net Sales/Total Ass.	
19.Net Sales/Total Ass.	20.Net Wort/ Total Liab.	
20.Net Wort/ Total Liab.	22.Quick Ass./Curr. Liab.	
22.Quick Ass./Curr. Liab.	26.Ret. Earn./Total Ass.	
25.Rate of Ret./ComStock	28.Total Debt/Total Ass.	
26.Ret. Earn./Total Ass.	29.Total Liab./ Net Cap. Emp	
27. Total debt / Equity		
28.Total Debt/Total Ass.		
29.Total Liab./ NetCap. Emp.		
33.Ebit/Total Ass.		

Table 19b. The variables in the DOGMA-models two years prior to failure.

Size=400	Size=200	Size=100
1.Cash/Current Liab.	2.Cash Flow/Current Liab.	13.Inventory/Net Sales
2.Cash Flow/Current Liab.	3.Cash Flow/Total Ass.	28.Total Debt/Total Ass.
3.Cash Flow/Total Ass.	11.Ebit / Total Interest	29.Total Liab./ Net Cap.
4.Cash Flow/Total Debt	12.Equity/Net Sales	
11.Ebit / Total Interest	13.Inventory/Net Sales	
12.Equity/Net Sales	16.MV of Equi/BV of Debt	
13.Inventory/Net Sales	20.Net Wort/ Total Liab.	
16.MV of Equity/BV of Debt	25.Rate of Ret./ComStock	
17.Net Income/Total Ass.	26.Ret. Earn./Total Ass.	
18.Net Quick Ass./Inventory	28.Total Debt/Total Ass.	
20.Net Wort/ Total Liab.	29.Total Liab./ NetCap.	
25.Rate of Ret./ComStock		
26.Ret. Earn./Total Ass.		
28.Total Debt/Total Ass.		
29.Total Liab./ NetCap. Emp.		

Table 19c. The variables in the DOGMA-models three years prior to failure.

The evaluation was performed in the following manner. First we discretised the data and run a well-known machine learning program C4.5Rules (Quinlan 1993) on it. Then, in the second step, we run DOGMA on the discretised data using the rules generated by C4.5Rules as background knowledge. Notice that we could also have run DOGMA directly on the discretised data to obtain similar results. The use of background rules is however motivated, because the rules are used to automatically set some parameters in the system.

The prediction results on the large dataset of 400 example companies are summarised in Table 20. The results for the second and third year are better than for the other methods, whereas the first year results are comparable to logit and discriminant analysis and slightly worse than the neural network model. One interesting observation conserning the model complexities is that the number of rules increases as we use the second and third year training sets. See column #r in Table 20. This means that from the first year training data DOGMA is able to learn quite a simple and strong model, consisting of only one rule in each class. However, for the second and third year data the models get more complicated and less accurate.

When we move on to the smaller sample of 200 companies we notice, in Table 21, a decrease of the second and third year accuracies. However, the second year performance is clearly better than for the other methods. The third year is similar to logit analysis, but worse than discriminant analysis.

As can be seen in Table 22, also the first year prediction accuracy drops when the smallest sample of 100 companies is used. For each year DOGMA learns simple models, consisting of 3-4 rules. But quite clearly the size of the sample is becoming too small to support accurate estimation of the underlying distribution.

Comparing the training accuracies and test accuracies of $test_1$ we notice that the training accuracies are consistently superior. The difference varies between 5.4% and 16.2% units. Thus even though the fitness function of DOGMA is built to avoid overfitting, some overfitting is occuring, indicating that some fitness parameters are not set optimally. A possible remedy to this would be to find more accurate values for the parameters by performing more elaborate parameter tuning runs on the training sets.

n	400				170		
	training				$test_1$		
year	nf	f	t	#r	nf	f	t
1	90.0	74.5	**82.3**	**1+1**	89.4	62.4	**75.9**
2	73.0	83.0	**78.0**	**4+1**	69.4	75.3	**72.4**
3	81.0	71.0	**76.0**	**5+3**	63.5	68.2	**65.9**

Table 20. Prediction accuracy and number of rules when the data set of 400 firms was used in estimation and 170 firms in testing. The #r column gives the number of rules in the learned model for each class.

n	200				170			90			90		
	training				$test_1$			$test_2$			$test_3$		
year	nf	f	t	#r	nf	f	t	nf	f	t	nf	f	t
1	86.0	85.0	**85.5**	**4+1**	89.4	64.7	**77.1**	86.7	66.7	**76.7**	91.1	60.0	**75.6**
2	78.0	85.0	**81.5**	**4+1**	60.0	70.6	**65.3**	53.3	75.6	**64.4**	64.4	64.4	**64.4**
3	85.0	69.0	**77.0**	**2+2**	68.2	55.3	**61.8**	64.4	55.6	**60.0**	66.7	55.6	**61.1**

Table 21. Prediction accuracy when the data set of 200 firms was used in estimation and 170 ($test_1$), 90 ($test_2$), and 90 ($test_3$) firms in testing.

n	100				170			50			50		
	training				$test_1$			$test_4$			$test_5$		
year	nf	f	t	#r	nf	f	t	nf	f	t	nf	f	t
1	80.0	88.0	**84.0**	**3+1**	81.2	62.4	**71.8**	84.0	64.0	**74.0**	68.0	56.0	**62.0**
2	36.0	96.0	**66.0**	**2+1**	30.6	84.7	**57.6**	24.0	80.0	**52.0**	24.0	92.0	**58.0**
3	80.0	76.0	**78.0**	**2+1**	72.9	49.4	**61.2**	64.0	48.0	**56.0**	76.0	36.0	**56.0**

Table 22. Prediction accuracy when the data set of 100 firms was used in estimation and 170 ($test_1$), 50 ($test_4$), and 50 ($test_5$) firms in testing.

6. Conclusions

In this study we have investigated whether changes in the sample size affect the prediction accuracy for logit analysis, discriminant analysis, neural networks and machine learning as to which method is best. We have used both crossvalidation for LA, DA and NN and hold-out samples for all four methods. The sample sizes have consisted of 400, 200 and 100 observations and the hold-out samples have consisted of 170, 90 and 50 observations.

Our study shows that in the crossvalidation set one year prior to failure and three years prior to failure NNs have the highest prediction accuracy for all three sample sizes. NNs have the highest prediction accuracy in the validation set with 400 observations also for two years prior to failure. However, this does not hold for the set with 200 observations,where LA has the best prediction accuracy and for the set of 100 observations, where DA is best. It is clear that NNs have the best overall outcome in the validation set independently whether the sample size is 400, 200 or 100. The differences in the accuracy are rather clear in the sample size of 400 observations two and three years prior to failure but less clear one year prior to failure while the differences are somewhat bigger in the sample size of 100 observations. However, we have not tested whether the differences are statistically significant.

Considering the holdout samples we found that NNs perform best one year prior to failure in the test sample with 170 observations and 90 observations but two years prior to failure the machine learning method performs best in the test sample with 170 observations as well as in the test sample with 90 observations. However, there is no best performer in the smallest hold-out sample. Three years prior to failure the machine learning is slightly better than NN in the test sample with 170 observations. The decrease in the prediction accuracy is bigger for NNs than for DA and LA. There is no clear best performer in the test samples with 90 and 50 observations.

The results show that neural networks and the machine learning method perform overall best where the sample size is 400 while there is no overall best performer when the sample size is decreased to 200 and 100 observations. Our results contradict the results achieved by Altman et al. (1994), who found little or no difference in classification performance between neural networks and discriminant analysis.

We do no agree with Triqueiros and Taffler (1996), who criticised NN studies for using too many fitted coefficients in the prediction models. There are no exact rules as to the number of fitted coefficients, only rules of thumbs. We have exceeded in this study the rule of thumb of restricting the number of coefficients to a maximum of 10% in our small and medium size samples. The results for the holdout samples give some evidence that there is a bigger decrease in the prediction accuracy for NNs in the small holdout sample than for DA and LA. This is not the case in the medium and big holdout samples.

Acknowledgments

An earlier version of this article has been published in the Proceedings of the Third Conference on AI in Accounting, Finance and Taxation. The work has been supported by the Academy of Finland and Liikesivistysrahasto. We like to thank Marko Grönroos who trained and tested the neural networks used in this work.

References

Altman, E. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance* 4 (September), 589-609.

Altman, E., R. Haldeman & P. Narayanan (1977). ZETA analysis. A new model to identify bankruptcy risk of corporations. *Journal of Banking and Finance* 1, 29-54.
Altman, E., G. Marco & F. Varetto (1994). Corporate Distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience). *Journal of Banking and Finance* 18, 505-529.

Back, B., T. Laitinen & K. Sere (1996). Neural networks and bankruptcy prediction: Funds flows, accrual ratios, and accounting data. *Advances in Accounting* 14, 23-37.

Beaver, W. (1966). Financial ratios as predictors of failure. Empirical research in accounting. Selected Studies, 71-111.

Blum, M. (1974). Failing company discriminant analysis. *Journal of Accounting Research* 12:1 (Spring), 1-25.

Coats, P. K. & L. F. Fant (1991). A neural network approach to forecasting distress. *The Journal of Business Forecasting* (Winter), 9-12.

Deakin, E. (1972). A discriminant analysis of predictors of business failure. Journal of Accounting Research 1 (Spring), 167-179.

Edminster, R. (1972). An empirical test of financial ratio analysis for small business failure prediction. *Journal of Financial and Quantitative Analysis* 2 (March), 1477-1493.

El Hennawy, R. H. A. & R. C. Morris (1983). The significance of base year in developing failure prediction models. *Journal of Business Finance & Accounting* 10:2 (Summer), 209-223.

Fitzpatrick, P.J. (1932). A comparison of ratios of successful industrial enterprises with those of failed firms. Certified Public Accountant. 12. October, November, December. 598-605, 656-662, 727-731.

Fletcher, D. & E. Goss (1993). Forecasting with neural networks. An application using bankruptcy data. Information & Management. 24. 159-167.

Goldberg, D. (1989). Genetic Algorithms in Search, Optimization, and Machine Learning. Addsion-Wesley, Reading, MA.

Hair, J., Anderson, R., Tatham, R. and Black,W. (1992). *Multivariate Data Analysis.* Macmillan Publishing Company. New York, NY.

Hecht-Nielsen, R. (1991). *Neurocomputing.* Addison-Wesley.Reading, MA.

Hekanaho J. (1997). GA-based rule enhancement in concept learning. In *Proceedings of the 3rd International Conceference on Knowledge Discovery and Data Mining*, Newport Beach, CA. AAAI Press, 183-186.

Hertz, J., Krogh, A. and Palmer, R.(1991). Introduction to the Theory of Neural Computing, Addison-Wesley.Reading, MA.

Kohavi, R., John, G., Long, R. and Manley, D. (1994). MLC++: A Machine Learning

Library in C++. In *Tools with Artificial Intelligence*, IEEE Computer Society Press. 740-743.

Merwin, C.L. (1942). Financing small corporations in five manufacturing industries 1926-36. New York. National Bureau of Economic Research.

Michalski, R.S. (1983). A theory and methodology of inductive learning. In Michalski, R.S., Carbonell, J. and Mitchell, T. (Eds.). *Machine Learning: An AI Approach*. Morgan Kaufmann.

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

Quinlan, J.R. (1990). Learning logical definitions from relations. *Machine Learning*, 5, 239-266.

Quinlan, J.R. (1993).*C4.5: Programs for machine learning*. Morgan Kaufmann. San Mateo.

Quinlan, J.R. (1995). MDL and categorical theories (continued). In *Proceedings of the 12th International Conference on Machine Learning*, Tahoe City, CA. Morgan Kaufmann, 464-470.

Ramser, J.R. & L.O. Foster (1931). A demonstration of ratio analysis. Bulletin No. 40. Urbana: University of Illinois. Bureau of Business Research.

Rissanen, J. (1989). *Stochastical complexity in statistical inquiry*. World Scientific. Singapore.

Salchenberger, L., E. Cinar & N. Lash (1992). Neural networks: A new tool for predicting thrift failures. *Decision Sciences*, Vol. 23, No. 4. 899-916.

SAS 1988:910).

SAS 1990).

Shannon, C. E. and Weaver, W. (1949). *The mathematical theory of communication*. The university of Illinois Press. Urbana.

Tam, K. & M. Kiang (1992). Managerial applications of neural networks: The case of bank failure predictions. *Management Science* 38:7 (July), 926-947.

Trigueiros, D. & R. Taffler (1996). Neural networks and empirical research in accounting. *Accounting and Business Research* 26:4, 347-355.

Wilson, R. L. & R. Sharda (1994). Bankruptcy prediction using neural networks. *Decision Support Systems* 11, 545-557.

Winakor, C.H. & R.F. Smith (1935). Changes in financial structure of unsuccessful industrial companies. Bulletin No. 51. Urbana: University of Illinois Press. Bureau of Economic Research.

Turku Centre for Computer Science
Lemminkäisenkatu 14
FIN-20520 Turku
Finland

http://www.tucs.abo.fi/

University of Turku
• Department of Mathematical Sciences

Åbo Akademi University
• Department of Computer Science
• Institute for Advanced Management Systems Research

Turku School of Economics and Business Administration
• Institute of Information Systems Science