
Arjana Brezigar-Masten, Igor Masten¹

Comparison of Parametric, Semi-Parametric and Non-Parametric Methods in Bankruptcy Prediction

Working Paper No. 6/2009, Vol. XVIII

Abstract: This paper compares parametric, semi-parametric and non-parametric methods in prediction of bankruptcy. Special care is devoted to the effect of choice-based sampling. The choice of the sampling and estimation method lead to a similar trade off. Using choice-based sampling and logit model leads to minimization of risk exposure. Samples unbalanced across groups and Klein and Spady (1993) semi-parametric method allow for better overall prediction accuracy and thus profit maximization. Both the choice of sampling method and the choice of estimation method should be thus made conditional on an explicit objective function of the financial institution in assessing credit risk.

Key words: bankruptcy prediction, semi-parametric methods, CART

The Working Paper Series is intended for the publication of the findings of research work still in progress, the analysis of data series, and the presentation of methodologies in particular research areas. The aim of the series is to encourage the exchange of ideas about economic and development issues and to publish findings quickly, even if they are not fully conclusive.

The opinions, findings, and conclusions expressed are entirely those of the authors and do not necessarily represent the views of the Institute of Macroeconomic Analysis and Development.

The contents of this publication may be reproduced in whole or in part provided that the source is acknowledged.

¹ University of Ljubljana, Faculty of Economics.

IMAD Working Paper Series

Publisher:

Institute of Macroeconomic Analysis and Development

Gregorčičeva 27

SI-1000 Ljubljana

Slovenia

Phone: (+386) 1 478 1012

Fax: (+386) 1 478 1070

E-mail: gp.umar@gov.si

Editor in Chief: Barbara Ferk, MSc (barbara.ferk@gov.si)

Working Paper: Comparison of Parametric, Semi-Parametric and Non-Parametric Methods in Bankruptcy Prediction

Authors: Arjana Brezigar-Masten, PhD (arjana.masten@gov.si); Igor Masten, PhD (igor.masten@ef.uni-lj.si)

Language edited.

Peer reviewed.

Ljubljana, September 2009

CIP - Kataložni zapis o publikaciji
Narodna in univerzitetna knjižnica, Ljubljana

336.279

330.4

BREZIGAR Masten, Arjana

Comparison of parametric, semi-parametric and non-parametric methods in bankruptcy prediction [Elektronski vir] / Arjana

Brezigar-Masten, Igor Masten. - El. knjiga. - Ljubljana : Urad RS za makroekonomske analize in razvoj, 2009. - (Zbirka Delovni zvezki UMAR ; letn. 18, št. 6)

Način dostopa (URL): http://www.umar.gov.si/fileadmin/user_upload/publikacije/dz/2009/dz06-09.pdf

ISBN 978-961-6031-91-2

1. Masten, Igor

247683328

CONTENTS

1 INTRODUCTION.....	1
2 DATA AND SAMPLE DESIGN.....	4
3 FORECASTING MODELS.....	5
3.1 Logit model.....	5
3.2 Klein and Spady semi-parametric estimator.....	6
3.3 CART.....	7
4 VARIABLE SELECTION	9
4.1 Three-stage approach.....	9
4.2 CART approach.....	10
5 RESULTS	11
6 CONCLUSION	16

List of tables and figures

Table 1: Estimates of the logit models 1 and 2	11
Table 2: Estimates of the Klein and Spady semi-parametric model 1.....	13
Table 3: Estimates of the Klein and Spady semi-parametric model 2.....	13
Table 4: In-sample classification accuracy	14
Table 5: Prediction accuracy	15
Figure 1: Estimated distribution function with the semi-parametric Klein and Spady model 1 (matched sample, without Zmijewski correction, $e=6$).....	12

Summary

Financial stability is of concern to employees, investors, bankers and government and regulatory authorities alike. Application of good methods of bankruptcy prediction in financial institutions can be seen as crucial in its procurement. Appropriate risk assessment is crucial for the allocation of resources and credit, which, in turn results in a positive growth effect and reduction of overall macroeconomic variability.

The aim of this paper is to contribute to recent bankruptcy prediction literature by investigating the merits of using some recently developed semi-parametric and non-parametric methods in such applications. We take the classic logit model as a benchmark for comparison.

As regards different estimation methods, we find non-parametric CART to be a very useful complementary method of variable selection. Augmenting classic models with variables selected by CART considerably improves forecasting accuracy. The choice between classic parametric method - logit - and the semi-parametric model of Klein and Spady (1993) interestingly induces the similar trade off as choice-based sampling. While logit appears to be more precise in detecting bad risks, it is also true that the semi-parametric model better captures the characteristics of healthy firms. A considerably larger share of the latter group in the population also implies better overall prediction accuracy. Both the choice of sampling method and the choice of estimation method should be thus made conditional on an explicit objective function of the financial institution in assessing credit risk.

Povzetek

Finančna stabilnost je ključna za dobro delovanje gospodarstva in je pomembna tako za zaposlene kot za investitorje, bankirje, državo in regulacijske organe. Uporaba učinkovitih metod napovedovanja stečajev v finančnih inštitucijah igra pomembno vlogo pri njenem zagotavljanju, saj je samo pravilno obvladovanje tveganj ključno za učinkovito alokacijo kreditov ter s tem povečuje gospodarsko rast in zmanjšuje makroekonomska tveganja.

Cilj tega delovnega zvezka je prispevek k literaturi napovedovanja stečajev z aplikacijo novih semi-parametričnih in ne-parametričnih ekonometričnih metod. Ti dve metodi sta nato primerjani z logit modelom, ki je standard v tovrstni literaturi napovedovanja.

Ugotovili smo, da je ne-parametričen CART zelo dobra komplementarna metoda izbire spremenljivk, poleg tega pa obogatitev modela s spremenljivkami, ki jih je izbral CART precej izboljša natančnost napovedi. Izbira med semi-parametrično cenilko Klein in Spady (1993) in logit cenilko je podobna izbiri med slučajnim in neslučajnim vzorčenjem. Medtem, ko je logit model bolj učinkovit pri identifikaciji podjetij z večjim tveganjem stečaja, semi-parametrični model bolje zajema karakteristike zdravih podjetij, ki predstavljajo glavnino podjetij v populaciji, tako da slednji model v celoti napoveduje boljše. Izbira metode vzorčenja kot izbira cenilke zavisi od specifičnega cilja, ki ga zasleduje finančna inštitucija pri obvladovanju kreditnih tveganj.

1 INTRODUCTION

Financial stability is of concern to employees, investors, bankers and government and regulatory authorities alike. Application of good methods of bankruptcy prediction in financial institutions can be seen as crucial in its procurement. Appropriate risk assessment is crucial for the allocation of resources and credit, which, in turn results in a positive growth effect and reduction of overall macroeconomic variability.

An additional reason for the growing interest in bankruptcy prediction is also the relevant impact of unsound credit on bank balances and, consequently, the minimum regulatory capital required by Basel Committee (2001). The new Basel Proposal and its latest revision in April 2003 are based on the three-pillar approach to capital adequacy: first, minimum capital requirements; second, a review of the supervisory process of internal bank assessments of capital; and third, the market disclosure involving the quality of information provided to the market. One of the most important innovations of the first pillar is the chance for banks to develop an internal rating system. The procedure to define an internal rating system can basically be divided into three steps (Moody's, 2000). First, the bank needs to make a choice about the classification model, which assigns to each borrower a posterior probability (or a score) of belonging to groups of sound or unsound borrowers. Second, starting from posterior probabilities definition of a "splitting rule", each borrower should be assigned to one of the several discrete classes in the rating system. Finally, the evaluation of the probability of default for each class, which is one of the input variables to work out the capital requirements, is applied.

The first step of development of an internal rating system thus faces an econometric challenge of choosing and evaluating the bankruptcy prediction model. This procedure also includes the selection of relevant explanatory variables and the choice of the cut-off point i.e. the value of posterior probability used to classify observations into classes of sound and unsound debtors. Circumstances faced by researchers in bankruptcy analysis have changed significantly in recent decades. This can be attributed to three factors. First, the availability of larger data sets with the median number of failing companies exceeding 1,000 (20 years ago the median was around 40 companies) allows for valid statistical inference where no conclusion could be reached before. Second, the spread of computer technologies and advances in statistical techniques allow for identification of more complex data structures. Basic methods may no longer be adequate for analyzing expanded data sets. Finally, there is an increased demand for advanced methods of controlling and measuring default risks due to the New Basel Capital Accord adoption. The Accord emphasizes the importance of risk management and encourages improvements in financial institutions' risk assessment capabilities.

At the beginning of the research period of failure prediction (see e.g. Fitzpatrick, 1932), there were no advanced statistical methods or computers available for the researchers. The values of financial ratios in failed and non-failed firms were compared with each other and it was found that they were poorer for failed firms. In 1966, the pioneering study of Beaver presented the univariate approach to discriminant analysis, while in 1968 Altman expanded this analysis to multivariate analysis. Until the 1980s, discriminant analysis was the dominant method in failure prediction. However, it suffered from assumptions that were violated very often. The assumption of the normality of the financial ratio distributions was problematic, particularly for the failing firms. During the 1980s, the method was replaced by logistic analysis, which has been until recent years the most used statistical method for failure prediction purposes. However, the assumption of logistic distribution of default probabilities may in many empirically relevant cases be violated. The potential heterogeneity of firms may be better captured by models that do not rely on overly restrictive distributional assumptions. This led to the application of fully non-parametric data mining methods in bankruptcy prediction.

During the 1990s artificial neural networks produced very promising results (Odom and Sharda, 1990, Tam and Kiang, 1991). However, no systematic way of identifying the predictive variables for the neural networks has been used in these studies. Drawing from good experience and success in several optimization problems in technical fields, genetic algorithms offer a new promising method for finding a suitable set of indicators for neural networks. Another class of data-mining methods with some desirable properties for predicting financial distress is classification and regression trees. The latter two methods may not only improve the selection of suitable predictors, but can also be used as independent forecasting tools. The second strand of literature that also does not rely on the overly restrictive assumptions of binary-choice models are semi-parametric models. Thus far, these have had few advances in bankruptcy prediction and consequently only offer an interesting research topic. One may consider them as the middle way. While they share the property of parametric models of offering a clear interpretation of modeled processes, they are much less rigid in structure and offer more flexibility in capturing the relevant information and complexity of data.

The aim of this paper is to contribute to recent bankruptcy prediction literature and investigate the merits of using some recently developed semi-parametric and non-parametric methods in such applications. We take the classic logit model as a benchmark for comparison for a number of reasons².

First, the logit model is widely used and taught. Second, it is relatively easy to understand and readily available in virtually all software packages. Finally, it has been proven to be a fairly robust and reliable tool for forecasting financial distress. Comparison of models is not confined only to out-of-sample forecasting precision. We also wish to examine whether relaxing the distributional assumptions underlying the logit model also yields insights that may help us better understand the determinants of financial distress.

Among alternative methods, we concentrate on two; the first is the semi-parametric estimator of binary choice models developed by Klein and Spady (1993), which we choose because of its superior theoretical properties among the available semi-parametric estimators³. The second method is based on classification and regression trees (CART hereafter). From the class of non-parametric methods, we chose it because of its simplicity and clarity of interpretation and, foremost, because it does not suffer from the "black-box deficiency" that is very often the main reason for criticism of artificial neural networks as the most prominent representative of other semi-parametric methods. In addition to bankruptcy prediction, CART is used also in the phase of variables selection.

We compare the performance of these methods on two different sample designs. Namely, the bankruptcy literature in all applications faces the problem of a low share of bankruptcy cases in the population and hence also data. A fairly common approach especially in the early studies was to use choice-based sampling of observations in order to obtain a more balanced sample of bankrupt and healthy firms (see Zmijewski 1983, for comparison). While such an approach may produce much better in-sample classification accuracy for bankrupt firms, it has a major deficiency: non-random sampling induces a bias in parameter estimates (Zmijewski, 1983; Maddala, 1983). As a consequence of the bias, it may be seriously questioned whether

² The second very popular parametric method is discriminant analysis (and its multivariate extension). This method was severely criticised in the literature, see Joy and Tollefson (1975), Eisenbenis (1977), Scott (1978), Altman and Eisenbeis (1978), Ohlson (1980) and Karson and Martell (1980) among others.

³ The second semi-parametric estimator of binary choice models with a single-index restriction is the estimator developed by Ichimura (1993).

balancing of observations in the sample is of value for practitioners in financial institutions. Balancing of the sample puts a disproportionate weight in the likelihood function on bankrupt firms. This may increase in-sample classification and out-of-sample prediction accuracy of bankrupt firms, but it also reduces the same types of accuracy for healthy firms. Because the share of healthy firms in the population is considerably larger, this usually results in reduction of overall in-sample classification and out-of-sample prediction accuracy. Similar reasoning led the authors in many recent applications to rely on random sampling. This is especially so in applications of non-parametric methods.

Hence, choice-based sampling may be fully acceptable only if the dominant objective of financial institutions and regulators is the minimization of risk exposure of financial portfolios. If important weight is also given to overall allocation of credit and profit maximization, one should not overlook that choice-based sampling leads to over-rejection of good and profitable lines of credit. For these reasons, we decided to analyze both approaches in sample design and compare the relative performance of methods in order to see whether some methods lead to a smaller trade-off.

The methods are tested on a sample of Slovenian firms. Note that many other applications disposed with data with only a limited coverage of industries. Our dataset, however, covers all industries and sizes of firms, which makes this analysis quite general. However, this also implies that the data contain various sources of real-life firm heterogeneity. These are also the circumstances that justify that use of methods that are at least theoretically better suited to account for these features of the data.

I find that choice-based sampling significantly affects prediction accuracy. Balancing group shares in the estimation sample in favor of bankrupt firms increases prediction accuracy for potentially bad risks. However, in real life financial institutions are faced with credit applications coming from a population with heavily unequal group shares. Using choice-based sampling thus leads to over rejection of potentially good risks. This implies that choosing to minimize risk exposure should be traded off with profit maximization. Because the share of healthy firms is considerably larger, this problem should not be neglected.

With regards different estimation methods, we find non-parametric CART to be a very useful complementary method of variable selection. Augmenting classic models with variables selected by CART considerably improves forecasting accuracy. The choice between classic parametric method (logit) and the semi-parametric model of Klein and Spady (1993) interestingly induces a similar trade off as choice-based sampling. While logit appears to be more precise in detecting bad risks it is also true that the semi-parametric model better captures the characteristics of healthy firms. A considerably larger share of the latter group in the population also implies better overall prediction accuracy. Both the choice of sampling method and the choice of estimation method should be thus made conditional on an explicit objective function of the financial institution in assessing credit risk.

The paper is structured as follows. Section 2 describes the data and the design of estimation samples and samples on which we test out-of-sample prediction accuracy. Section 3 describes different modeling approaches to bankruptcy prediction used in the paper. Section 4 contains a detailed description of procedures used in selecting predictor variables for the models. Section 5 discusses the main results, while Section 6 concludes and summarizes the findings.

2 DATA AND SAMPLE DESIGN

The data come from two databases of Slovenian companies. The first are data of annual financial statements for all Slovenian firms for the 1995-2001 period provided by Agency for Public Legal Records and Related Services (AJPES). From the initial database, we eliminated all observations for which, due to missing data, we could not calculate all the potential predictive variables (various financial ratios). This resulted in 39,005 observations on healthy firms in the sample. The second is the database of bankrupt firms for the same time period collected by I d.o.o., from which we are able to obtain 592 bankruptcy cases in the whole period. Industries in the sample mainly cover the manufacturing sector. We decided to omit financial institutions due to significant structural differences and different exposure to the risk of default provided by the regulatory framework.

As noted above, there are two approaches to sample design in the literature and we analyze the relative performance of both. The first approach, used less often in the literature, is to work with the sample of data as it is, i.e. with usually much larger share of healthy firms in the sample. Very similar to the majority of studies, the share of bankruptcy cases in our sample is rather small, roughly 1.5%.

The second approach uses choice-based sampling in order to obtain a more balanced share of bankrupt and healthy firms in the sample. We opted for equal shares and performed the selection in the following way. From the initial sample, we created ten sub-samples with 592 bankrupt firms and their 592 non-bankrupt mates. Matching is based on the following characteristics: size (measured by total asset), industry and year of bankruptcy. The last matching criterion ensures that financial statements of matched pairs are always of the same time period. Because matching is primarily used to obtain a balanced sample of bankrupt and healthy firms, the samples mainly consist of small and medium-sized companies, since the incidence of bankruptcy in the large-asset-size firm was quite rare.

In both approaches, 75 percent of observations were allocated to a sub sample on which the models were estimated, and 25 percent to a sub sample on which out-of-sample prediction accuracy was tested.

There is one important deviation from this general approach to sample design. In the application of the Klein and Spady (1993) semi-parametric model, the computational burden was excessive for estimation of the model on the complete dataset. For this reason, we considerably reduced the number of healthy firms entering the estimation sample. In particular, only 10% (or 3,900) of healthy firms were added to the 592 bankrupt firms. This implies that the sample contained 13.2% of bankrupt firms. In- and out-of-sample divisions use the same proportions as above.

From the balance-sheet and income statement data, we calculated 64 financial ratios as candidate predictors.⁴ Financial ratios can be broadly classified into four categories: liquidity, profitability, solvency and activity. The ratios are chosen on the basis of their popularity in the literature and their potential relevance to the study of financial distress. A dependent variable is a binary variable that takes on value one if the firm operates in time t , and zero if the firm filed for bankruptcy in time t . All independent variables are dated $t-1$.

⁴ Financial ratios, by their nature, have the effect of deflating statistics by size, implying that their potential predictive power is not contaminated by firm size (Altman, 2000).

3 FORECASTING MODELS

In terms of basic statistical characteristics, we use three different classes of methods. The first two methods engage the binary choice probability model with a single-index restriction as a basic structure, but differ in terms of distributional assumptions of the single index. The first method assumes a fully parametric and standard logit specification. The second used milder distributional assumptions and is estimated with the semi-parametric method developed by Klein and Spady (1993). Because of their relatively common basic structure, we treat their exposition in a similar way. The third method is based on classification and regression trees. This is a fully non-parametric method, whose main properties are described below.

3.1 Logit model

The logit model is, together with the probit model and discriminant analysis (DA), among the most common procedures in estimating bankruptcy. Unlike discriminant analysis that begins with the conditional distribution of X given y , logit and probit models specify the conditional distribution of y given X (the explanatory variables). Interestingly, if y is dichotomous, and X follows a multivariate normal distribution, the implied form for $P(y|X)$ is the same as that for the logit model (Maddala, 1983). However, logit analysis is valid under more general distributional assumptions about X than those implied by discriminant analysis. In contrast, Ohlson (1989) claimed that logit does not avoid all the problems discussed with respect to DA. If the explanatory variables are normally distributed, then DA should be used, since it is more efficient. However, if the explanatory variables are not normally distributed, then discriminant analysis gives inconsistent estimates, and one is better off using logit analysis in this case. These findings are supported in the literature by Maddala (1983), Amemiya (1981), Amemiya and Powell (1980), Kennedy (1991), Lo (1986) and Malhorta (1983).

As a starting point consider a single-index binary choice model:

$$y = \begin{cases} 1 & \text{if } \theta'x \geq \varepsilon \\ 0 & \text{otherwise} \end{cases}$$

where

$$P(y = 1|x) = h(\theta'x) \quad (1)$$

that links the probability that the binary dependent variable equals one given the covariates is equal to a probability transformation of the single index $\theta'x$. In principle, both the parameters of the single index θ and the probability transformation function h need to be estimated. Parametric methods assume a known form of h . In this class, the most widely used model found in the bankruptcy prediction literature is the logit model. In such a case, h is a logistic cumulative distribution function

$$h(\lambda) = e^\lambda / (1 + e^\lambda)$$

With this assumption, the parameter vector θ can be estimated consistently and efficiently by maximizing:

$$L = \sum [y_i \ln(P_i) + (1 - y_i) \ln(1 - P_i)] \quad (2)$$

Some of the first authors to applying logit methodology to the problem of bankruptcy were Santomero and Vinso (1977) and Martin (1977) who employed it to examine failures in the US banking sector. Ohlson (1980) applied it more generally to 105 bankrupt and 2,058 non-bankrupt firms. For recent examples, one can refer to Zmijewski (1984), and Wilson (1992). The accuracy of classification ranged from 76% in the work of Zmijewski (1984), where he employed probit and weighted exogenous sample likelihood models to investigate firms

listed on the American and New York stock exchanges from 1972 to 1978, to 96% in the study by Pantalone and Platt (1987), where the authors used logit analysis to determine the causes of banks bankruptcy in the US after deregulation.

The logit model has one appealing feature when matched or choice-based samples are used in the analysis. In such samples the probability of an observation entering the sample depends on the value of dependent variable, which violates the random sampling design assumption and in general causes both parameter and probability estimates to be asymptotically biased (Zmijewski, 1984). The logit model is more convenient in choice-based samples because it gives consistent results, without using any weighting procedures. The coefficients of the explanatory variables are not affected by the unequal sampling rates from the two groups. It is only the constant rate that is affected, and should be increased by $\log p_1 - \log p_2$, where p_1 and p_2 are the proportions sampled from the two groups (see Maddala, 1983 for a detailed discussion). Other coefficients are unaltered, and the standard errors also remain valid. Such correction of the constant was used in our application when the logit model was estimated on the choice-based sample.

3.2 Klein and Spady semi-parametric estimator

One important and potentially empirically relevant deficiency of the logit model is that it requires the validity of the assumption that the cumulative distribution of the error term is logistic. Consequently, it makes sense to investigate alternative specifications, which require less severe distributional assumptions. A good alternative offered by the literature in this respect are semi-parametric models.⁵ These models allow for simultaneous estimation of h and θ and, as such, provide a specification that is more flexible than a parametric model but retains many of the desirable features of parametric models (Horowitz, 2001). The single-index property is crucial for good properties of semi-parametric estimators because it allows avoidance of the curse of dimensionality. This is because the index $\theta'x$ aggregates the dimensions of x . Consequently, the difference between the estimator of h and the true function can be made to converge to zero at the same rate that would be achieved if $\theta'x$ were observable. Moreover, θ can be estimated with the same rate of convergence that is achieved in a parametric model. Thus, in terms of the rates of convergence of estimators, a semi-parametric single index model is as accurate as a parametric model for estimating θ and as accurate as a one-dimensional nonparametric model for estimating h . This dimension reduction feature of single index models gives them a considerable advantage over nonparametric methods in applications where X is multidimensional and the single index structure is plausible.

The main estimation challenge in single index models is estimating θ . Several estimators of θ are available in the literature. Ichimura (1993) developed a non-linear least squares estimator. Theoretically the semi-parametric maximum likelihood estimator of Klein and Spady (1993) is superior, which in addition to exhibiting $N^{-1/2}$ -consistency and asymptotic normality, also achieves the semi-parametric efficiency bound, assuming that the regressors and the errors are independent.

⁵ Manski (1985) proposed a semi-parametric estimator that does not rely on a single-index restriction. Subsequently, Horowitz (1992) developed it into the smoothed maximum score estimator. Although a smoothed maximum score requires very weak distributional assumptions it has some drawbacks. Its rate of convergence is lower than ordinary parametric estimators. Moreover, it only allows one to estimate the index, but not the probability transformation.

The estimate of θ is obtained by maximizing the quasi-log likelihood function given by:

$$\log L(\hat{\theta}) = n^{-1} \sum_{i=1}^n (\sigma_i / 2) [y_i \log(\hat{P}_i(\theta))^2 + (1-y_i) \log(1-\hat{P}_i(\theta))] \quad (3)$$

τ_i represents the trimming function as specified by Klein and Spady (1993) and is needed to weigh down the influence of observations with a very low probability and to ensure the usual convergence rate of the asymptotic distribution of the parameters. Probability $\hat{P}_i(\theta)$ is estimated using the fourth-order kernel with probability trimming. Klein and Spady (1993) showed that with these modifications the proposed estimator of θ is consistent, asymptotically normal and efficient. In addition, their Monte Carlo experiment indicates that there may be only modest efficiency losses relative to maximum likelihood estimation when the distribution of the disturbances is known, and the small sample behavior of the semi-parametric estimator in other cases is good.

Because choice-based sampling may lead to significantly biased results, we also considered a modification of the quasi-likelihood function in the spirit of Zmijewski (1994). In particular, we optimize:

$$\log L(\hat{\theta}) = n^{-1} \sum_{i=1}^n (\sigma_i / 2) \left[\frac{P_1}{P_2} y_i \log(\hat{P}_i(\theta))^2 + \left(\frac{1-P_1}{1-P_2} \right) (1-y_i) \log(1-\hat{P}_i(\theta)) \right] \quad (4)$$

where P_1 and P_2 are proportions of bankrupt firms in the population and estimation sample, respectively. The prediction accuracy of the coefficients obtained with the sampling correction are compared to prediction accuracy of the model without such correction to assess the influence of choice-based sampling on bankruptcy prediction accuracy.

3.3 CART

Datamining techniques offer a number of methods that can be successfully applied to predict bankruptcy. The most commonly used techniques in datamining are artificial neural networks, genetic algorithms (Sung, Chang and Lee, 1999) and decision trees. Among the latter, the most frequently used are Classification and Regression Trees (CART). An explicit comparison of datamining techniques is very difficult since each application has different goals and circumstances, which require different data mining techniques. Also, each data mining technique has its inherent limitations as well as assumptions that limit its application to specific actual cases.

Among non-parametric methods, we concentrated on the CART method.⁶ CART builds classification and regression trees for predicting continuous dependent variables (regression) and categorical predictor

⁶ Artificial neural networks were not considered in the analysis because this method suffers from the "black box problem," i.e. they cannot explain the results they obtain. In addition, the evidence of usefulness of applying ANN the literature is mixed. While some studies find it to be the preferred method relative to multivariate discriminant analysis (Salchenberger et al., 1992; Coats and Fant, 1993) other authors report less convincing evidence (Altman et al., 1994; Leshno and Spector, 1996). In some cases, decision tree algorithms proved to be better (Martinelli et al., 1999; McKee and Greenstein, 2000). The application of genetic algorithms that may also prove to be successful in bankruptcy prediction (Back, Laitine and Sere, 1999) was left for future research.

variables (classification). The classic CART algorithm was popularized by Breiman et al. (1984) (see also Ripley, 1996). The CART model is a flexible method for specifying the conditional distribution of a variable y , given a vector of predictor values X . Such models use a binary tree to recursively partition the predictor space into subsets where the distribution of y is successively more homogeneous. The terminal nodes of the tree correspond to the distinct regions of the partition, and the partition is determined by splitting rules associated with each of the internal nodes. By moving from the root node through to the terminal node of the tree, each observation is then assigned to a unique terminal node where the conditional distribution of y is determined. CART is nonparametric and can detect complex relationships between dependent variable and explanatory variables. Therefore, CART is particularly suited for discovering non-linear structure and variable interactions in datasets with a large number of potential explanatory variables.

In sum, the strengths of decision tree methods are: (1) ability to generate understandable rules; (2) performing classification without requiring much computation; (3) ease of calculation at classification time; (4) ability to handle both continuous and categorical variables; (5) providing a clear indication of which fields are the most important for prediction and classification, (6) enabling validation of a model using statistical tests, so the reliability of the model can be checked.

The two pioneering studies where the technique has been used for bankruptcy prediction are those of Frydman, Altman and Kao (1985), and Marais, Patell and Wolfson (1984) who employed it to assess loan classifications. The first mentioned study compared CART to the classificatory power of two discriminant models. Overall, the classification-tree models were found to perform best. In contrast, Marais et al. (1984) compared their recursive partitioning results against those of a multinomial probit model. Interestingly, they concluded that in estimating loan classifications there was very little difference between the two procedures.

4 VARIABLE SELECTION

Many bankruptcy prediction studies were centered on the search for individual or groups of predictors (financial ratios) that lead to the lowest misclassification rate. Despite some efforts to provide theoretical economic grounds in failure prediction context, no unified theory has been generally accepted as a basis for the ratio selection. Most of the previous studies used a brute empirical approach of initial choice of variables (also based on some economic criteria) followed by step-wise procedure to select the variables in the final logit or discriminant model. This procedure is not statistically rigorous and different sequencing or initial ordering of variables need not result in a unique selection. As an attempt to overcome this deficiency, some authors started using datamining techniques (Shirata, 1998). These are also better suited to capturing potential non-linearities in the relations between financial distress and predictor variables.

We decided to use two approaches with the aim of determining which could lead to better results for our dataset. The first approach is a more traditional three-stage approach, and the second uses CART analysis as one to the datamining techniques. Detailed descriptions of both are provided below.

4.1 Three-stage approach

For the first variable selection approach, we propose a three-stage strategy, which combines expert knowledge and evidence on most successful predictors found in the literature with statistical testing. In the first step, bivariate logistic regressions were run for each of the 64 ratios on each of the ten matched samples. Each of the ratios was screened for its classification precision. The ratios that classify correctly at least 60 percent of bankrupt firms and 60 percent of non-bankrupt firms on average were kept for further stages. This left a group of 27 financial ratios, 14 measuring profitability, 9 solvency and 4 liquidity of firms. There were nine ratios that classify neither bankrupt nor non-bankrupt firms at 60 percent accuracy. Seven describe firm activity and two are profitability measures. The remaining 28 ratios classified at the required precision either bankrupt or non-bankrupt firms, but not both. This also means that they were not considered in subsequent steps of variable selection.

In the second step, seven groups of highly correlated indicators were formed, using 0.5 as the correlation threshold. From each of the groups, we extracted one principal component. As a representative of each group, we then took the variable with the largest loading to the principal component. We prefer to proceed in this way of using the principal component in prediction models in order to avoid the efficiency problem due to generated regressors, and because principal components can be hardly given any direct economic interpretation.

In the last step, a logistic step-wise procedure was used to select the final variables. It starts by estimating parameters for variables forced into the model. 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 conventional levels, the variable enters into the model. One or more elimination steps follow each selection step, i.e. the variables already selected into the model do not necessarily stay. The step-wise 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.

After step three, I obtained with four financial ratios as the most suitable variables for bankruptcy prediction. Two of the ratios measure liquidity, one solvency and one profitability.

4.2 CART approach

In addition to using it as a non-parametric method for bankruptcy prediction, I used CART also as the second approach to variable selection. This approach is based on fitting a regression tree, specifying the default variable as the dependent variable and using all 65 financial ratios as independent variables. The aim of this approach is to identify the variables that resulted in being the most significant in the decision tree that partitions firms into bankrupt and healthy groups. In CART-based selection, one needs to avoid over-fitting because it may lead to bad out-of-sample prediction accuracy. Specifically, some of the lower branches in a tree may be strongly affected by outliers and other artifacts of the current data set. For this reason, it is preferable to find a simpler tree. The tree pruned to the best size was obtained with the process of cross validation (see Breiman et al., (1998) for details). On the matched sample, this resulted in a tree with four terminal nodes obtained on three variables. On the complete sample, the respective figures are five and four. Estimated regression trees were subsequently used as predictors of bankruptcy as discussed above.

Alternatively, I used the final nodes identified in CART analysis for each variable to create dummy variables that take a value of one if the values of the variable fall into the regions identified by CART threshold values and zero otherwise. These dummy variables are then added to the set of explanatory variables obtained in the three-step approach in logit and semi-parametric models. The motivation to do this is quite straightforward: CART is, by definition, better suited to identify potential non-linearities in the determinants of financial distress of firms. Including dummy variables that correspond to such non-linearities may in this respect be a useful way to augment standard models. In addition, using the variables obtained under two alternative search strategies in the same forecasting directly provides an insight about the relative merits of the two selection methods in bankruptcy prediction.

5 RESULTS

Table 1 presents the estimates of the logit model. There are two models presented. The first model uses four ratios obtained in the three-stage selection procedure as explanatory variables (model 1). The second model augments the set of explanatory variables with the dummy variables corresponding to final nodes of the estimated regression tree using CART (model 2). Note, however, that not all dummies are included due to perfect co-linearity with other explanatory variables. Both models are estimated on three different samples. The first is the matched sample with equal number of bankrupt and healthy firms in the sample. The sample labeled "Full" contains all available observations. This implies that the sample contains 592 bankrupt firms and 39,005 healthy firms. For completeness, the model is also estimated on a middle-sized (labeled "Larger") sample containing 592 bankrupt firms and 2,925 healthy firms. As explained in Section 2, construction of this sample was necessary to facilitate computational feasibility in optimization of the Klein&Spady semi-parametric model. For each of the samples, 75% of randomly selected observations are used for estimation, while the rest is retained for testing out-of-sample prediction accuracy.

Table 1: Estimates of the logit models 1 and 2

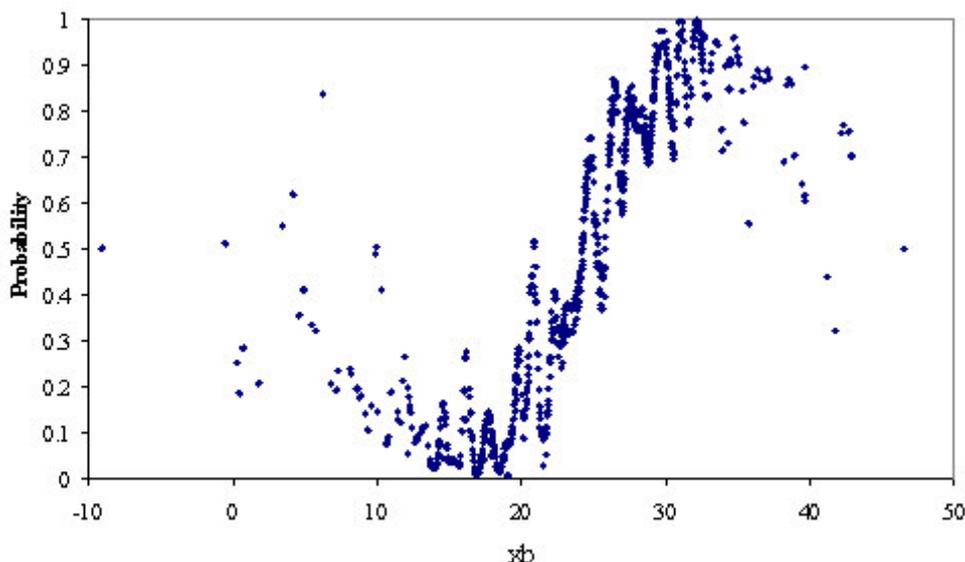
Sample	Matched		Larger		Full	
	Coefficient					
Constant	-6.98	-5.27	-5.14	-9.7	-0.87	-5.49
	-0.69	-0.77	-0.47	-1.24	-0.59	-0.55
Constant*	-11.17	-9.46	-7.16	-11.72		
Tfs	-3.86	-3.44	-3.12	-1.06	-3.42	-0.54
	-0.51	-0.66	-0.35	-0.46	-0.42	-0.32
Pppo	0.11	0.07	0.1	0.07	0.07	0.05
	-0.01	-0.01	-0.005	-0.01	-0.003	-0.003
Kol	0.6	0.5	0.85	0.81	1.1	0.78
	-0.19	-0.2	-0.15	-0.16	-0.25	-0.13
cf2d	2.54	1.28	3.01	3.1	3.5	4.22
	-0.74	-0.75	-0.7	-0.72	-1.48	-0.66
D1 _{CART}		2.2		4.04		3.15
		-0.33		-1.08		-0.33
D2 _{CART}		2.56		5.99		5.33
		-0.31		-1.05		-0.29

Notes: * Constant corrected by $\log p1 - \log p2$, where $p1$ and $p2$ are the proportions sampled from the two groups (see Maddala, 1983).

As seen in Table 1, all coefficients are significant and correctly signed. Since detailed discussion of the coefficients is not at the centre of our attention, it should suffice to say at this stage that sample design does not have a negligible effect on the estimated coefficient that, according to theory, remain consistent (with exception of the constant) regardless of the sample design. Below, we shall see how this affects the out-of-sample prediction accuracy.

The motivation for considering the semi-parametric model is clearly seen from Figure 1. It plots the distribution function of the estimated Klein and Spady model for one of the specifications (similar results emerge for any other specification), which is significantly different from the logistic distribution.

Figure 1: Estimated distribution function with the semi-parametric Klein and Spady model 1 (matched sample, without Zmijewski correction, $e=6$)



Tables 2 and 3 contain the estimation results for the Klein&Spady semi-parametric models 1 and 2, respectively. Each model was estimated with and without Zmijewski correction for shares of bankrupt and healthy firms that do not correspond to population shares (see Section 3.2). In addition, we report results for two different choices of trimming intensity in optimization of the quasi-likelihood.

Comparison of in-sample classification accuracy of the models is given in Table 4 (p. 14). In this respect, six comments are in order. First, it must be noted that the semi-parametric Klein&Spady model does not offer a better overall fit to the data than the logit model even though logit relies on distributional assumptions that are not fully supported by the data. This is a rather surprising finding, which clearly indicates a certain robustness and reliability of the logit model. Second, the fully non-parametric CART method offers the best fit on matched and larger sample, but also does not outperform logit on the full sample. Third, choice of sampling clearly demonstrates the trade-off faced by researchers. Choice-based sampling improves the classification accuracy of bankrupt firms, but the smaller precision for healthy firms results in an inferior overall fit of the model. Fourth, looking at the effect of corrections for choice-based sampling, we can observe that constant correction results in the smaller overall fit of the model. Fifth, the Zmijewski-type correction of the quasi-likelihood of the Klein&Spady semi-parametric model improves the fit only for model 2 estimated on a matched and completely balanced sample. In all other cases, it actually results in a deterioration of fit. This clearly indicates that such rather ad-hoc corrections of the likelihood in favor of the under-represented group of observations in the sample do not necessarily lead to improvement of classification accuracy of those observations. Finally, it can be noted that higher degree of both likelihood and probably trimming in the estimation of the Klein&Spady model that more intensively weighs down the influence of outlying observations in the sample improves the model's classification accuracy. The same also holds for out-of sample prediction accuracy (see below).

Table 2: Estimates of the Klein and Spady semi-parametric model 1

Sample								
Matched					Larger			
[Zmijewski correction, trimming intensity (e)]								
	[yes, 6]	[yes, 4.3]	[no, 6]	[no, 4.3]	[yes, 6]	[yes, 4.3]	[no, 6]	[no, 4.3]
tfs	1	1	1	1	1	1	1	1
pppo	-0.84 (-0.04)	-2.32 (-0.18)	-5.57 (-0.13)	-2.33 (-0.13)	-8.46 (-0.05)	-24.65 (-0.33)	-2.38 (-0.01)	-3.24 (-0.09)
kol	-13.41 (-0.55)	-3.51 (-0.25)	-4.34 (-0.1)	-2.94 (-0.16)	0.88 (-0.03)	-9.5 (-0.13)	-6.42 (-0.03)	-1.32 (-0.04)
cf2d	-7.19 (-0.3)	-2.63 (-0.2)	-2.26 (-0.1)	-2.81 (-0.24)	-13.54 (-0.09)	-10.34 (-0.15)	-6.58 (-0.02)	-1.34 (-0.05)

Notes: Standard errors in parentheses. A higher value of parameter implies less trimming and vice versa.

Table 3: Estimates of the Klein and Spady semi-parametric model 2

Sample								
Matched					Larger			
[Zmijewski correction, trimming intensity (e)]								
	[yes, 6]	[yes, 4.3]	[no, 6]	[no, 4.3]	[yes, 6]	[yes, 4.3]	[no, 6]	[no, 4.3]
tfs	1	1	1	1	1	1	1	1
pppo	-2.37 (-0.16)	-2.34 (-0.08)	-4.7 (-0.12)	-2.2 (-0.06)	-4.34 (-0.05)	-17.85 (-0.13)	-3.66 (-0.05)	-2.42 (-0.11)
kol	-1.85 (-0.16)	-1.6 (-0.07)	-4.37 (-0.1)	-1.32 (-0.06)	-2.73 (-0.03)	-6.28 (-0.05)	-1.57 (-0.06)	-1.08 (-0.05)
cf2d	-0.38 (-0.02)	-1.53 (-0.07)	-0.51 (-0.01)	-0.64 (-0.02)	-3.54 (-0.04)	-6.72 (-0.06)	-2.47 (-0.08)	-2.2 (-0.09)
D1 _{CART}	-1.11 (-0.15)	-0.99 (-0.08)	-1.04 (-0.1)	-0.01 (-0.13)	1.25 (-0.03)	5.8 (-0.09)	-0.32 (-0.18)	2.26 (-0.1)
D2 _{CART}	-4.57 (-0.27)	-4.71 (-0.19)	-4.65 (-0.16)	-1.82 (-0.11)	-2.38 (-0.14)	-7.01 (-0.08)	-3.35 (-0.09)	-5.7 (-0.32)

Notes: Standard errors in parentheses. A higher value of parameter implies less trimming and vice versa.

The constant is not reported because it cannot be identified within the semi-parametric model. For the same reason, one of the coefficients needs to be normalized to unity. Virtually all coefficient results are statistically significant and with signs similar to the logit model (note that the first coefficient is normalized to unity). What clearly emerges from the tables are significant differences in estimated parameters when compared to the logit model even after taking into account the normalization of the first coefficient. A second finding is that in the present context both Zmijewski's likelihood correction and trimming intensity importantly affect the estimated coefficients. As there is no theoretical proof of which choice would be better, I consider the effect of all these features on the prediction accuracy of the model.

Table 5 (p. 15) shows the central set of results of the paper, reporting out-of-sample prediction accuracy. Depending on the estimation sample, I consider different samples on which the models are tested for prediction accuracy. As explained in Section 2, all three different samples of data were divided so that 75 percent of observations were used for estimation and 25 percent for testing out-of-sample prediction accuracy. The label M→M denotes estimation on the matched sample and out-of-sample prediction also on a matched sample, i.e. with equal shares of bankrupt and healthy firms. The label M→P stands on estimation on matched sample, but prediction accuracy is tested on a sample with population group shares (roughly 1.5% of bankrupt firms). With the L→P I label, estimation is on a larger sample while prediction is made on a full sample with population group shares.

Table 4: In-sample classification accuracy

		Sample				
		Matched		Larger		Full
Model		Sample-size correction				
		yes	no	yes	no	
Logit model 1	Healthy	15.5	89	86.2	98.1	99.8
	Bankrupt	98.6	82.4	84.2	54.9	20.7
	Overall	57.5	85.7	86	92.4	98.6
Logit Model 2	Healthy	15.3	90.3	91.5	98.7	99.8
	Bankrupt	99.3	84	82.7	62.4	40.8
	Overall	14.9	87.2	90.4	93.9	99
K&S model 1 e=6	Healthy	77.9	81.7	96.1	96.6	
	Bankrupt	73.8	79.7	49.5	59.9	
	Overall	75.9	80.7	89.9	91.8	
K&S model 1 e=4.3	Healthy	80.4	83.6	96.4	97.1	
	Bankrupt	82	83.8	43.9	55.4	
	Overall	81.2	83.7	89.5	91.6	
K&S model 2 e=6	Healthy	89.6	87.2	97.1	97.7	
	Bankrupt	84.9	79.7	64	64.8	
	Overall	87.3	83.4	92.8	93.4	
K&S model 2 e=4.3	Healthy	89.2	84.8	96.6	97.7	
	Bankrupt	85.4	81.8	48.6	65.8	
	Overall	87.3	83.3	90.3	93.5	
CART	Healthy	97.7		98.9		99.9
	Bankrupt	84		69.8		39
	Overall	89.9		95.1		99

First we can note that the CART method, even though attaining comparably high levels of prediction accuracy, practically never yields the best results. Inclusion of dummy variables corresponding to CART terminal nodes, however, significantly improves the performance of both logit and Klein&Spady model. Second, correction for the bias induced by choice-based sampling does not yield any measurable benefits. Third, comparison on the full sample is possible only between the logit model and the CART model. We can observe that both models deliver similar prediction accuracy for healthy firms, while CART appears to be significantly more precise for bankrupt firms. However, when CART dummies are included in the logit model, its performance becomes even slightly better. Overall, the results clearly indicate the usefulness of CART in variable selection as it appears to successfully capture potential non-linearities present in the data.

Fourth, the most important observation concerns the comparison of logit and the Klein&Spady semi-parametric model. It clearly emerges from Table 5 that logit is better in prediction accuracy of both bankrupt and healthy firms only when prediction is done on a matched sample. Such a situation does not correspond to real-life assessment of firms' creditworthiness. The population of credit applicants is not drawn from a distribution with balanced group shares. The share of bankrupt firms is considerably smaller in the true population of credit applicants. In this respect, the most interesting comparison of models follows from prediction accuracy on the sample with population group shares (label P). Logit is better in predicting bankruptcy cases while the semi-parametric model more successfully captures the characteristics of healthy firms. Since the share of the latter group is considerably larger, this also results in better overall prediction accuracy. The difference is not large, but consistent across different model specifications. The relative merits of the two methods therefore depend on the objectives of the financial institution in credit risk assessment. If the objective is minimization of exposure to risk, then the logit model would deliver better results as it would

deliver fewer bankrupt firms to the portfolio. However, this also implies that the institution would reject a very large number of potentially good risks. With the objective of profit maximization, the semi-parametric model seems to be preferable, because it offers a better overall prediction accuracy. The difference is particularly pronounced when financial institutions estimate their models on relatively small and choice-based samples.

Table 5: Prediction accuracy

		Sample					
		Matched			Larger		Full
Model		M→M	M→P	M→P*	L→P	L→P*	
Logit model 1	Healthy	84.8	87.9	14.7	98.1	86	99.9
	Bankrupt	79.7	79.7	98.6	54.7	80.4	20.9
	Overall	82.8	87.8	16	97.5	85.9	98.7
Logit model2	Healthy	91.7	97.2	13.6	98.8	91.4	99.8
	Bankrupt	82.4	82.4	99.3	55.4	79.8	41.9
	Overall	87	96.7	14.9	98.1	91.2	98.9
K&S model 1 e=6	Healthy	77.2	99.8	99.8	99.7	99.4	
	Bankrupt	72.3	9.5	18.9	6.1	10.8	
	Overall	74.7	98.5	98.6	98.3	98.1	
K&S model 1 e=4.3	Healthy	80	99.7	99.7	99.8	99.3	
	Bankrupt	77.7	10.8	10.8	8.1	10.8	
	Overall	78.8	98.4	98.4	98.4	98	
K&S model 2 e=6	Healthy	85.5	99.8	99.7	99.7	99.6	
	Bankrupt	74.3	10.8	11.5	20.3	8.8	
	Overall	79.9	98.5	98.4	98.5	98.2	
K&S model 2 e=4.3	Healthy	84.8	99.8	99.8	99.7	99.4	
	Bankrupt	81.8	12.8	10.1	8.8	12.2	
	Overall	83.3	98.5	98.5	98.3	98.1	
CART	Healthy	77.2	92.4		98.4		99.8
	Bankrupt	85.8	79.7		67.6		39.9
	Overall	81.2	92.2		97.9		98.9

Notes: * denotes the correction of constant for the logit model (see also notes to Table 1) and Zmijewski-type correction for the Klein and Spady model.

Finally, it must be noted that choice-based sampling induces the same type of trade-off as between parametric or semi-parametric methods. Balancing the sample in favor of bankrupt firms obviously increases the prediction accuracy of potential bankruptcy cases. However, extended credit lines in real life have highly unequal shares. Minimization of risk exposure in this respect comes at the expense of overall prediction accuracy and hence profit opportunities. In this respect, both the choice of sampling method and the choice of estimation method should be made conditional on an explicit objective function of the financial institution in assessing credit risk.

6 CONCLUSION

This paper uses data on a full sample of Slovenian firms to assess the effects of choice-based sampling and different estimation methods on bankruptcy prediction accuracy. The results reveal that choice-based sampling significantly affects prediction accuracy. Balancing group shares in the estimation sample in favor of bankrupt firms increases the prediction accuracy of potentially bad risks. However, this does not correspond to the situation financial institutions face in real life. Credit applicants come from a population with heavily unequal group shares with bankrupt firms representing only a small portion of all observations. Using choice-based sampling thus leads to over-rejection of potentially good risks. This implies that choosing to minimize risk exposure should be traded off with profit maximization. Because the share of healthy firms is considerably larger, this problem should not be neglected.

As regards different estimation methods, I find non-parametric CART to be a very useful complementary method of variable selection. Augmenting classic models with variables selected by CART considerably improves forecasting accuracy. The choice between classic parametric method - logit - and the semi-parametric model of Klein and Spady (1993) interestingly induces the similar trade off as choice-based sampling. While logit appears to be more precise in detecting bad risks, it is also true that the semi-parametric model better captures the characteristics of healthy firms. A considerably larger share of the latter group in the population also implies better overall prediction accuracy. Both the choice of sampling method and the choice of estimation method should be thus made conditional on an explicit objective function of the financial institution in assessing credit risk.

A potential problem with these conjectures is the fact that I use 50 percent probability of default as a cut-off point in predicting bankruptcy. For this reason, in our future work we plan to include investigation of an optimal cut-off point that should correspond to the optimal choice of the trade-off described above.

BIBLIOGRAPHY

1. Atiya A.F., (2001), "Bankruptcy Prediction for Credit Risk Using Neural Networks: A Survey and New Results", IEEE Transactions on Neural Networks, Vol.12, No. 4., July.
2. Altman E.I. and R.A.Eisenbeis (1978), "Financial application of discriminant analysis: a clarification" J.Financ.Quantitat.Anal.,13,185-195.
3. Altman E.I., (1983), "Corporate Financial Distress: A Complete Guide to Predicting, Avoiding, and Dealing with Bankruptcy", New York: John Wiley and Sons.
4. Altman, E.I., G. Marco and F. Varetto, (1994), "Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks", J. Banking and Finance, vol. 18, pp. 505—529.
5. Altman, E.I., (2000), "Predicting Financial Distress of Companies: Revisiting the Z-Score and Zeta® Models", mimeo, Stern School, New York University, July.
6. Amemiya T and Powell J, (1980), "A comparison of the logit model and normal discriminant analysis when independent variables are binary", Technical Report, No. 320, Institute for Mathematical Studies in the Social Sciences, Stanford University.
7. Amemiya T., (1981), "Qualitative response models: A survey, Journal of Economic Literature", 19:4, 1483-1536.
8. Andrews D.W.K., (1991), "Asymptotic normality of series estimators for nonparametric and semiparametric regression models". Econometrica 59: 307--345.
9. Breiman L., Friedmann H.L., Olshen R. A., and Stone C.J., (1984) "Classification and regression trees", Wadsworth International Group.
10. Eisenbeis R, (1977), "Pitfalls in the application of discriminant analysis in business, finance, and economics", The Journal of Finance, 32:3, 875-900.
11. Fitzpatrick P, (1932), "A comparison of ratios of successful industrial enterprises with those of failed firms", Certified Public Accountant, October, November, and December, 598-605, 656-662, and 727-731, respectively
12. Frecka T. and W. Hopwood, (1983), "The effects of outliers on the cross-sectional distributional properties of financial ratios", The Accounting Review, 58:1, 115-128.
13. Fried D., A.C. Sondhi, and G.I. White, (1997), "The Analysis and Use of Financial Statements", Jon Wiley and Sons, Inc.
14. Frydman H., E. Altman, D. Kao, (1985), "Introducing recursive partitioning for financial classification: the case of financial distress", The Journal of Finance, Vol. XL No.1, pp.269-91.
15. Horowitz, J.L. and W. Härdle (1996), "Direct Semiparametric Estimation of Single-Index Models with Discrete Covariates", Journal of the American Statistical Association, 91, 1632-1640.
16. Ichimura, H. and L.F. Lee (1991), "Semiparametric Least Squares Estimation of Multiple Index Models: Single Equation Estimation", In: Barnett, W.A., J. Powell, and G. Tauchen (eds.), Nonparametric and Semiparametric Methods in Econometrics and Statistics. Cambridge: Cambridge University Press, pp. 3-49.
17. Ichimura H., (1993), "Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single Index Models", Journal of Econometrics 58(1/2), 71-120..

18. Joy M.O. and J.O.Tollefson, (1975) "On the financial Applications of discriminant Analysis", *Journal of Financial and Quantitative Analysis*.
19. Karson M.J. and T.F. Martell,(1980), "On the onterpretation of individual variables in multiple discriminant analysis", *J. Financ. Quantitat. Anal.*,15, 211-217.
20. Kennedy P, (1991), "Comparing classification techniques, *International Journal of Forecasting*", 7:3, 403-406.
21. Klein R. and R. Spady, (1993), "An Efficient Semiparametric Estimator of the Bynary Choice Model", *Econometrica* 61, 387-421.
22. Lo, A. (1986), "Logit versus discriminant analysis: A specification test and application to corporate bankruptcies", *Journal of Econometrics*, 31:3,151-178.
23. Maddala G, (1983), "Limited-dependent and Qualitative Variables in Econometrics", Cambridge: Cambridge University Press.
24. Maddala, G.S.(1987),"Limited Dependent Variable Models Using Panel Data", *The Journal of Human Resources*, Vol.3,p.307-337.
25. Malhotra N, (1983), "A comparison of the predictive validity of procedures for analyzing binary data", *Journal of Business and Economic Statistics*, 1:4, 326-336.
26. Marais, M.L., J.M. Patell and M.A. Wolfson, (1984)," The experimental design of classification models: an application of recursive partitioning and bootstrapping to commercial bank loan classifications", *Supplement to the Journal of Accounting Research*, 22, 87–118.
27. Odom, M.D., Sharda, R., (1990), "A neural network for bankruptcy prediction", In: *Proceedings of the International Joint Conference on Neural Networks*, vol. 2, 163-168.
28. Ohlson J., (1980), "Financial ratios and the probabilistic prediction of bankruptcy", *Journal of Accounting Research*, 18:1, 109-131.
29. Pagan, A. and A. Ullah (1999) "Nonparametric Econometrics", Cambridge University Press.
30. Pantalone, C. C., and M. B. Platt, (1987)"Predicting Commercial Bank Failure Since Deregulation", *New England Economic Review*, Jul/Aug, 37-47.
31. Santomero A. and J.D. Vinso (1977), "Estimating the Probability of failure for commercial banks and the banking system", *Journal of Banking and Finance.*, September.
32. Shirata C.Y., (1998), "Financial Ratios as Predictors of Bankruptcy in Japan :An Empirical Research", *Proceedings of The Second Asian-Pacific Interdisciplinary Research in Accounting Conference*, pp. 437-445.
33. Sung, T.K., Chang, N., Lee, G. (1999), "Dynamics of modelling in data mining: interpretive approach to bankruptcy prediction", *Journal of Management Information Systems*, Vol. 16 No.1, pp.63-85.
34. Tam K. and M. Kiang, (1992) "Managerial applications of the neural networks: The case of bank failure predictions", *Management Science*, vol. 38, pp. 416—430.
35. Tam K., (1991) "Neural network models and the prediction of bank bankruptcy", *Omega*, vol. 19, pp. 429—445.
36. Zmijewski M, (1984), "Methodological issues related to the estimation of financial distress prediction models", *Journal of Accounting Research*, 22, Supplement, 59-82.