

## PREDICTING BANKRUPTCY WITH SEMI-PARAMETRIC SINGLE-INDEX MODEL

### *Abstract*

*Semi-parametric methods are virtually neglected in the bankruptcy prediction literature. This paper compares the logit model, as the standard parametric model for bankruptcy prediction, to the semi-parametric model developed by Klein and Spady (1993). Special care is devoted to the effect of choice-based sampling prediction accuracy. The choice of the sampling and estimation method lead to a similar trade offs. Using choice-based sampling and logit model leads to minimization of risk exposure. Samples unbalanced across groups and the semi-parametric method allow for better overall prediction accuracy and thus profit maximization.*

**JEL codes:** G32, G33, C14, C25

**Keywords:** *bankruptcy prediction, semi-parametric methods*

### 1 INTRODUCTION

The problem of predicting corporate failure is at the heart of risk management procedures in banks worldwide. An important stimulus towards development of more complex bankruptcy prediction systems came through the implementation of Basel II Accord since it put these systems in the function of managing bank's capital requirements and internal assessment of risk appetite. The interest of academic community has long been present (see Fitzpatrick, 1932) and intensified in recent years with an array of novel methodological approaches.

Methodological approaches used in bankruptcy prediction can be broadly classified into statistical methods and artificial intelligence methods (Min and Jeong, 2009). The first group includes discriminant analysis (used in pioneering studies of Beaver (1996) and Altman (1968)) and binary-choice models like logit or probit. A common statistical property of these methods is that they are fully parametric. The artificial intelligence group comprises of methods that range from artificial neural networks (ANN) and genetic algorithms (GE) to classification and regression trees (CART) (see Paliwal and Kumar, 2009; and Li et al., 2010). A common statistical feature of this second class of models is a fully non-parametric specification of both the distributional form of variables and functional relations among them.

This paper bridges the two model classes in bankruptcy prediction with the application of the semi-parametric binary-choice model. Semi-parametric estimators of single-index binary choice model as proposed by Ichimura (1993) or Klein and Spady (1993) have been successfully applied in other fields of economic analysis, like labor economics, but there exists no published

<sup>1</sup> University of Primorska, Faculty of Mathematics, Natural Sciences and Information Technology, Glagoljaška 8, Koper, Slovenia, Institute of Macroeconomic Analysis and Development, Gregorčičeva 27, Ljubljana

<sup>2</sup> University of Ljubljana, Faculty of Economics, Kardeljeva pl. 17, SI-1000 Ljubljana, Slovenia



application to prediction of corporate failure. The advantage of these models is that they retain all the benefits of parametric binary-choice models in terms of linear functional form among predictors, i.e. the single index assumption, and straightforward interpretability of estimated coefficients. At the same time, they do not impose strict distributional assumptions of the single index, which is the case with logit or probit model. In the case of predicting corporate failure, which is a rare event, such a feature of the model may be very important. In addition, potential heterogeneity of firms may be better captured by models that do not rely on too restrictive distributional assumptions. After all, the application of fully non-parametric data mining methods in bankruptcy prediction is importantly motivated with such a property.

Our comparisons of prediction accuracy are performed out of sample, even though we provide results also for classification accuracy. Besides evaluating the performance of competing models according to overall prediction accuracy only, we put special emphasis to prediction accuracy of bankrupt and healthy firms separately. Finally, special attention is devoted to the construction of the estimation sample. Namely, choice based sampling of the observations into estimation sample that equates the number of bankrupt and healthy firms in the estimation sample is a quite common approach in the literature and practice. While such an approach may be motivated by computational considerations, it is definitely at odds with composition of real data and, consequently, bankruptcy prediction in practice. Choice-based sampling may for this reason lead to significant biases in model's prediction accuracy (Zmijewski, 1983). This paper evaluates prediction accuracy also conditional on the construction of the estimation sample.

On a population of Slovenian listed and non-listed companies our results show that while logit appears to be more precise in detecting bad risks it is also true that the semi-parametric model of Klein and Spady (1993) captures better the characteristics of healthy firms. Considerably larger share of the latter group in the population implies also 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 remainder of the paper is organized as follows. Section 2 introduces competing bankruptcy prediction models. Section 3 presents our data and discusses the formation of the estimation samples. Section 4 presents the results, while Section 5 concludes.

## 2 PREDICTION MODELS

The focus of our analysis is the use of semi-parametric binary-choice models in bankruptcy prediction, which is, to the best of our knowledge, the first such application in the literature. In particular, we use the estimator proposed by Klein and Spady (1993). The semi-parametric approach to bankruptcy prediction is a methodological bridge between parametric and non-parametric methods. For this reason it is natural to compare its forecasting precision to both parametric and non-parametric methods. In the choice of the first we follow the approach of many applications in the literature and use the logit model (see for example, Li et al., 2010; Chen, 2011 and Min and Jeong, 2009, among others). There are several reasons for such a choice. First,

logit model has been widely used and taught.<sup>3</sup> Second, it is relatively easy to understand and readily available in virtually all software packages. Third, it has resulted to be a fairly robust and reliable tool for forecasting financial distress. Last but not least, both two methods take on 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 specification (logistic distribution). The second uses no parametric assumptions.

Because of their relatively common basic structure we treat the exposition of the logit model and the semi-parametric model of Klein and Spady in a similar way. As a starting point consider a single-index binary choice model

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

where

$$P(y = 1|x) = h(\theta'x). \quad (2)$$

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  $\theta$  and the probability transformation function  $h$  need to be estimated. Parametric methods assume a known form of  $h$ , while semi-parametric methods make no parametric assumptions about  $h$ .

## 2.1 LOGIT MODEL

In the logit model  $h$  is a logistic cumulative distribution function

$$h(\theta'x) = \frac{e^{\theta'x}}{1 + e^{\theta'x}}$$

With this assumption the parameter vector  $\theta$  can be estimated consistently and efficiently by maximizing

$$L = \sum_{i=1}^N [y_i \ln(P_i) + (1 - y_i) \ln(1 - P_i)] \quad (3)$$

Common characteristic of all bankruptcy prediction cases is a very small population share of bankrupt firms. This implies that only a limited share of the overall probability mass is accounted for by bankruptcy cases, which may cause the estimated prediction models to describe well the characteristics of healthy firms, but have only limited prediction power for bankruptcy cases. For this reason many empirical applications balance the estimation sample by including equal shares of healthy and bankrupt firms. In such a case, estimation samples a

<sup>3</sup> Among the first to apply logit 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. Notable applications that followed include Zmijewski (1984), and Wilson (1992). 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 use logit analysis to determine the causes of banks bankruptcy in the US after the deregulation.

choice based, i.e. the probability of an observation entering the sample depends on the value of dependent variable, which violates the random sampling assumption. Choice-based sampling in general causes both parameter and probability estimates to be asymptotically biased (Zmijewski,1984).

## 2.2 KLEIN AND SPADY (1993) 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.<sup>4</sup> These models allow for simultaneous estimation of  $h$  and  $\theta$  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 to avoid 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,  $\theta$  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  $\theta$  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  $\theta$ . Several estimators of  $\theta$  are available in the literature. Ichimura (1993) developed a nonlinear least squares estimator. Theoretically superior is the semi-parametric maximum likelihood estimator of Klein and Spady (1993), which in addition to exhibiting  $\sqrt{N}$ -consistency and asymptotic normality achieves also the semi-parametric efficiency bound, assuming that the regressors and the errors are independent.

The estimate of  $\theta$  is obtained by maximizing the quasi-loglikelihood function:

$$\log L(\theta, \lambda) = n^{-1} \sum_{i=1}^n \tau_i \left[ y_i \log(P_i(\theta)) + (1 - y_i) \log(1 - P_i(\theta)) \right] \quad (4)$$

$\tau_i$  represents the trimming function as specified by Klein and Spady (1993) and is needed to weight 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)

<sup>4</sup> 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 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.

show that with this modifications the proposed estimator of  $\theta$  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.

As discussed above choice-based sampling may lead to significantly biased results. This observation applies also to semi-parametric methods. For this reason we considered a modification of the quasi-likelihood function in the spirit of Zmijewski (1994). In particular, we optimize

$$\log L(\theta + \Lambda) = n^{-1} \sum_{i=1}^n \left[ \tau_i \log(P_{1i}(\theta)) + (1 - \tau_i) \log(P_{2i}(\theta)) \right] + \left[ \frac{P_1}{P_2} \log\left(\frac{P_1}{P_2}\right) + \frac{1 - P_1}{1 - P_2} \log\left(\frac{1 - P_1}{1 - P_2}\right) \right] \quad (5)$$

where  $P_1$  and  $P_2$  are proportions of bankrupt firms in population and estimation sample respectively. 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 VARIABLE SELECTION

Selection of predictor variables is an important step in all bankruptcy prediction studies. To date no unified theory has been generally accepted. We propose a three-stage strategy, which combines expert knowledge and evidence on most successful predictors found in the literature with statistical testing. We start by constructing 64 different financial ratios as potential predictors. In the first step, bivariate logistic regressions were run for each of the 64 ratios on ten randomly created matched subsamples of our dataset. The ratios that on average classify correctly at least 60 percent of bankrupt firms and 60 percent of non-bankrupt firms, were kept for further stages. This left us with a group of 27 financial ratios, 14 measuring profitability, 9 solvency and 4 liquidity of firms. Ratios that classify neither bankrupt nor non-bankrupt firms at 60 per cent accuracy are nine in number. Seven describe firm activity and two are profitability measures. The remaining 28 ratios classified at required precision either bankrupt or non-bankrupt firms, but not both. As such they were not considered in subsequent steps of variable selection.

The ratios that passed the first step were in the second step grouped into seven groups of highly correlated indicators, 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 to using the principal component in prediction models in order to avoid the efficiency loss problem associated with generated regressors, and because principal components can be hardly given any direct economic interpretation.<sup>5</sup>

<sup>5</sup> Chen (2011) reports prediction accuracy loss from using PCA.

In the last step 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.

The three-step selection approach resulted in four financial ratios as the most suitable variables for bankruptcy prediction. Two of the ratios measure liquidity, one solvency and one profitability.

#### 4 DATA AND SAMPLE DESIGN

The data come from two databases of Slovenian companies. The first is the database of 592 bankruptcy cases firms for the same time period collected by I d.o.o. (Slovenian franchise of Dun&Bradstreet), from which we are able to obtain 592 bankruptcy cases in the period under analysis. These are data on all bankruptcy cases filed with Slovenian legal authorities. The second are data of annual financial statements for all Slovenian firms for the period 1995-2001 provided by Agency for public legal records and related services (AJ PES).<sup>6</sup> 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). Industries in the sample mainly cover the manufacturing sector with their size ranging from very small to large.

From the databases we construct two estimation samples. The first uses choice-based sampling, which means balancing the shares of bankrupt and healthy firms in the sample. From the initial sample we created a sub-sample with 592 bankrupt firms and corresponding 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 vintage. 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.

The second estimation sample is larger. It contains 3900 healthy firms and 592 bankruptcy cases. This implies that the sample contains 13.2 percent of bankrupt firms. Larger estimation samples were not considered because the estimation of the Klein and Spady (1993) semi-parametric model resulted computationally impossible.

Our goal is to test prediction accuracy of competing models in a pseudo out-of-sample context and not model comparison based on in-sample classification accuracy. In both approaches 75 percent of observations were allocated to a subsample on which the models were estimated, and 25 percent to a subsample on which out-of-sample prediction accuracy

<sup>6</sup> Post-2001 data was not considered because of a change in accounting standards.

was tested.

From the balance-sheet and income statement data we calculated 64 financial ratios as candidate predictors.<sup>7</sup> 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 financial literature and their potential relevance to the study of financial distress. 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  $t - 1$ .

## 5 BANKRUPTCY PREDICTION

Table 1 presents the estimates of the logit model. Explanatory variables in the model are four financial ratios obtained in the three-stage selection procedure. The model is estimated on two different samples. The first is the matched sample with equal number of bankrupt and healthy firms in the sample. We label this sample "Matched". The second sample is the larger sample with considerably higher share of healthy firms. We label the sample "Large".

As seen from Table (1), all coefficients are significant and correctly signed. We can observe some variation of estimated coefficients across estimation samples. With the exception of the constant all coefficients are within one standard error. In line with findings of Zmijewski (1994) for the constant this is not so, which clearly implies that selection of the estimation sample is important in prediction practice, where loan applicants come from a population with a small share of bankrupt firms.

**Table 1: Estimates of the logit model**

Coefficient	Sample	
	Matched	Large
<i>constant</i>	-6.98 (0.69)	-5.14 (0.47)
<i>tfs</i>	-3.86 (0.51)	-3.12 (0.35)
<i>pppo</i>	0.11 (0.01)	0.10 (0.005)
<i>kol</i>	0.60 (0.19)	0.85 (0.15)
<i>cf2d</i>	2.54 (0.74)	3.01 (0.70)

Standard errors in parentheses.

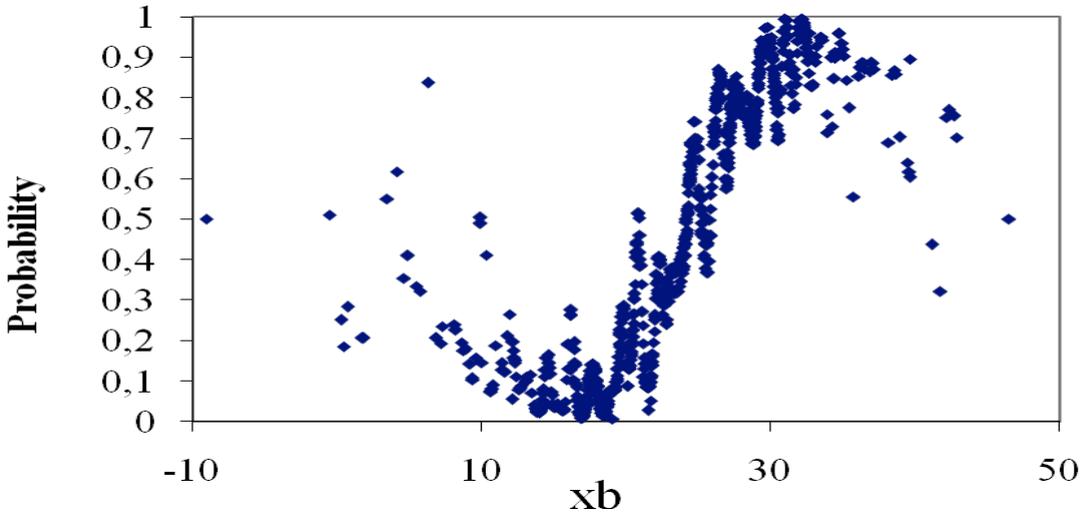
Source: Author calculation

The motivation for considering semi-parametric models in bankruptcy prediction is clearly seen from Figure 1. It plots the distribution function (fitted values) of the estimated Klein and Spady model. It is significantly different from the logistic distribution, which has been confirmed by the Horowitz and Härdle (1994) test (available from the authors upon request). At tails of the distribution it indicates a non-monotonic relation between the estimated single

<sup>7</sup> Financial ratios, by their nature, have the effect of deflating statistics by size, implying that a their potential predictive power is not contaminated by firm size (Altman, 2000).

index and probability of default.

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



Source: Author calculation

Table 2 contains the estimation results for the Klein&Spady semi-parametric model. The model was estimated with two different choices of trimming intensity in optimization of the quasi-likelihood. 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 coefficients result to be statistically significant and with signs similar to the logit model (note that the first coefficient is normalized to unity). What clearly emerge from the table 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 the trimming intensity importantly affects coefficient estimates. As there is no theoretical guidance for the most appropriate choice we consider both degrees of trimming intensity in the evaluation of prediction accuracy of the model.

Comparison of in-sample classification accuracy of the models is given in Table 3. In this respect three 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 finding clearly points to certain robustness and reliability of the logit model. Second, choice of sampling clearly demonstrates the trade-off faced by researchers. Choice-based sampling improves classification accuracy of bankrupt firms, but a smaller precision for healthy firms results in an inferior overall fit of the model. Finally, it can be noted that a higher degree of both likelihood and probably trimming in the estimation of the Klein&Spady model that more intensively down-weights the influence of outlying observations in the sample improves model's classification accuracy. The same holds also for out-of sample prediction accuracy, which we turn to next.

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

	Sample		Sample	
	Matched	trimming intensity (e)	Large	
Coefficient	6	4.3	6	4.3
<i>tfs</i>	1.00	1.00	1.00	1.00
<i>pppo</i>	-5.57	-2.33	-2.38	-3.24
	(0.13)	0.13)	(0.01)	(0.09)
<i>kol</i>	-4.34	-2.94	-6.42	-1.32
	(0.10)	(0.16)	(0.03)	(0.04)
<i>cf2d</i>	-2.26	-2.81	-6.58	-1.34
	(0.10)	(0.24)	(0.02)	(0.05)

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

Source: Author calculation

**Table 3: In-sample classification accuracy**

Model		Sample	
		Matched	Larger
Logit	Healthy	89.0	98.1
	Bankrupt	82.4	54.9
	Overall	85.7	92.4
Klein & Spady $e = 6$	Healthy	81.7	96.6
	Bankrupt	79.7	59.9
	Overall	80.7	91.8
Klein & Spady $e = 4.3$	Healthy	83.6	97.1
	Bankrupt	83.8	55.4
	Overall	83.7	91.6

Source: Author calculation

The results of out-of-sample prediction accuracy are presented In Table 4. Models are tested on two different samples and, as explained in Section 4, both samples were divided so that 75 per cent of observations were used for estimation and 25 per cent for testing out-of-sample prediction accuracy.

Fourth, the most important observation concerns the comparison of logit and Klein&Spady semi-parametric model. It clearly emerges from Table 4 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. Credit applicants are not drawn from a distribution with balanced group shares. 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 turns out to be 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 results also in

better overall prediction accuracy. The difference is not large, but consistent across different model specifications. Relative merits of the two methods therefore depend on the objectives of financial institution in credit risk assessment. If the objective is minimization of exposure to risk then logit model would deliver better results as it would deliver less bankrupt firm to the portfolio. However, this also implies that the rejection of 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 if financial institutions estimate their models on relatively small and choice-based samples.

Table 4: Prediction accuracy

Model		Sample	
		Matched	Large
Logit	Healthy	84.8	98.1
	Bankrupt	79.7	54.7
	Overall	82.8	97.5
Klein& Spady $e = 6$	Healthy	77.2	99.7
	Bankrupt	72.3	6.1
	Overall	74.7	98.3
Klein& Spady $e = 4.3$	Healthy	80.0	99.8
	Bankrupt	77.7	8.1
	Overall	78.8	98.4

Source: Author calculation

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 firm 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

In this paper we give the first evaluation of semi-parametric estimators of binary-choice models in bankruptcy prediction. The method is very appealing, on one hand because its single index restriction facilitates easy interpretation of estimated coefficient, on the other because it does not rely on any parametric assumptions of the distribution of the single index. As such it is a bridging method between standard and fully parametric statistical methods and non-parametric artificial intelligence methods.

On an exhaustive sample of Slovenian firms we show that logit appears to be more precise in detecting bad risks. The semi-parametric model of Klein and Spady (1993) better captures the characteristics of healthy firms. Considerably larger share of the latter group in the population

implies also better overall prediction accuracy.

Present application of the semi-parametric model to bankruptcy prediction is the first such evaluation of the method in the literature and thus represents a starting point for further research. One obvious extension is predictor selection within the semi-parametric context. Namely, in the present application we selected the set of predictors by using the step-wise procedure that builds the testing procedure on the likelihood function of the logit model and then estimated the semi-parametric model using the same variables. In this respect we put the semi-parametric model to an important disadvantage in the comparison of predictive accuracy. Given this constraint we find the presented results even more encouraging. An obvious extension would thus be casting the step-wise variable selection procedure to a semi-parametric likelihood framework.

## References

- Altman, E.** (1968). Financial Ratios. Discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance*, 23:4, 589-609.
- Altman, E., Marco, G., & Varetto, F.** (1994). Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks, *Journal of Banking and Finance*, 18, 505--529.
- Amemiya, T. & 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.
- Amemiya, T.** (1981). Qualitative response models: A survey. *Journal of Economic Literature*, 19:4, 1483-1536.
- Beaver, W.** (1966). Financial ratios as predictors of failure. *Journal of Accounting Research*, 4, Supplement, 71-111.
- Coats, P. & Fant, L.** (1993). Recognizing financial distress patterns using a neural network tool, *Financial Management*, 22, 142-155.
- Chen, M-Y.** (2011). Predicting corporate financial distress based on integration of decision tree classification and Logistic Regression. *Expert Systems with Applications*, 38, 11261-11272.
- Cho, S., Hong, H., & Ha, B-C.** (2010). A Hybrid Approach Based on the Combination of Variables Selection Using Decision Trees and Case-based Reasoning Using the Mahalanobis distance: For bankruptcy prediction. *Expert Systems with Applications*, 37, 3482-3488.
- Eisenbeis, R.** (1977). Pitfalls in the application of discriminant analysis in business, finance, and economics. *Journal of Finance*, 32:3, 875-900.
- 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.
- Horowitz, J. L.** (2001). Nonparametric Estimation of a Generalized Additive Model with and Unknown Link Function. *Econometrica*, 69, 499-513.
- Ichimura Hidehiko:** Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single Index Models. *Journal of Econometrics* 58(1/2), 71-120, 1993.
- Klein, R., & Spady, R.** (1993). An efficient semi-parametric estimator of the binary choice model. *Econometrica*, 61, 387-421.
- Kennedy, P.** (1999). Comparing classification techniques. *International Journal of forecasting*, 7:3, 403-406.
- Leshno, M. & Spector, Y.** (1996). Neural network prediction analysis: The bankruptcy case, *Neurocomputing*, 10, 125--147.
- Li, H., Sun, J., & Wu, J.** (2010). Predicting business failure using classification and regression tree: An empirical comparison with popular classical statistical methods and top classification mining methods. *Expert Systems with Applications*, 37, 5895-5904.
- 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.
- Manski, C.F.** (1988). Identification of Binary Response Models. *Journal of the American Statistical Association*, 83, 729-738.

- McKee, T. E. & Greenstein, M.** (2000). Predicting bankruptcy using recursive partitioning and a realistically proportioned data set. *Journal of Forecasting*, 19, 219--230.
- Min, J. H., & Jeong, C.** (2009). A Binary classification method for bankruptcy prediction. *Expert Systems with Applications*, 36, 5256-5263.
- Ohlson, J.** (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, 18:1, 109-131.
- Paliwal, M. & Kumar, U. A.** (2009). Neural networks and statistical techniques: A review of applications. *Expert Systems with Applications*, 36, 2-17.
- Tam, K.** (1991), Neural network models and the prediction of bank bankruptcy, *Omega*, 19, 429-445.
- Zmijewski, M.** (1984). Methodological issues related to the estimation of financial distress prediction models, *Journal of Accounting Research*, 22, Supplement, 59-82.

## PREDVIĐANJE BANKROTA POMOĆU POLU-PARAMETARSKOG MODELA JEDINSTVENOG INDEKSA

*Sažetak*

*Polu-parametarski modeli su doslovno zanemareni u literaturi o predviđanju bankrota. Ovaj rad uspoređuje logit model, kao standardni parametarski model za predviđanje bankrota, sa polu-parametarskim modelom kojeg su razvili Klein i Spady (1993). Posebna je pažnja posvećena efektu choice-based uzorkovanja na točnost predviđanja. Odabir metode uzorkovanja i procjene dovele su do sličnih balansiranja (trade offs). Korištenje choice-based uzorkovanja i logit modela dovodi do minimaliziranja rizika. Nebalansirani uzorci i polu-parametarska metoda omogućuju generalno bolju kvalitetu predviđanja te tako i maksimizaciju profita.*

**Ključne riječi:** *predviđanje bankrota, polu-parametarske metode, CART*