A Race for Long Horizon Bankruptcy Prediction

Author(s): Altman, Edward I.; Iwanicz-Drozdowska, Małgorzata; Laitinen, Erkki K.; Suvas, Arto

Title: A Race for Long Horizon Bankruptcy Prediction

Year: 2020

Version: Accepted manuscript

Please cite the original version:

Altman, E. I., Iwanicz-Drozdowska, M., Laitinen, E. K. & Suvas, A. (2020). A Race for Long Horizon Bankruptcy Prediction. *Applied Economics* 52(37), 4092-4111. https://doi.org/10.1080/00036846.2020.1730762

A Race for Long Horizon Bankruptcy Prediction

Edward I. Altman, New York University, Stern School of Business
Małgorzata Iwanicz-Drozdowska, Warsaw School of Economics
Erkki K. Laitinen, University of Vaasa
Arto Suvas, University of Vaasa

Abstract

This study compares the accuracy and efficiency of five different estimation methods for predicting financial distress of SME companies. We apply a larger set of financial and non-financial variables than do previous studies, using filter and wrapper selection, among other methods, to predict bankruptcy up to 10 years before the event in an open, European economy. Our findings show that logistic regression and neural networks are superior to other approaches. We document how the cost-return ratio considerably affects the location of optimal cutoff points and attainable profit in credit decisions. Once a loan provider selects a particular prediction model, an effort should be made to find the optimal cutoff score to maximize the efficiency of the technique. Indeed, this often involves determining several cutoff levels where the portfolio of products and services exhibits different cost-return characteristics.

Keywords:
SMEs, estimation technique, variable selection, cutoff, cost-return ratio

1. Introduction

Failure prediction models are used by many stakeholders in predicting or avoiding a failure. High accuracy, a long prediction horizon, and interpretability, as well as low cost are desired properties of these models. The early roots of scientific bankruptcy prediction modeling are usually traced back to the univariate approach (Beaver, 1966), the multivariate approach (Altman, 1968), the use of non-financial variables (Keasey & Watson, 1987) and, at later stages, the application of various estimation techniques to improve prediction accuracy (e.g., Dimitras et al., 1996; Bellovary et al., 2007; du Jardin, 2015).

Beaver (1966) has already observed that certain financial ratios follow a systematic process over time and have some predictive ability up to five years before the failure. A number of studies (Argenti, 1976; Hambrick & D'Aveni, 1988; D'Aveni, 1989; Laitinen, 1991; Ooghe & de Prijcker, 2008) also show that the failure process can take a number of years (even to 5–8 years). Many researchers have focused on longer horizons, analyzing different failure processes (Argenti, 1976; Laitinen, 1991; du Jardin & Severin, 2011, 2012; du Jardin, 2015) and also focusing on non-financial variables to explain different processes in the long term (Ooghe & de Prijcker, 2008; Hambrick & D'Aveni, 1988; D'Aveni, 1989; Moulton et al., 1996; Altman et al., 2010; Altman et al., 2016).

In traditional failure models, the predictive ability of financial variables is high for the one-year horizon, but their predictive ability decreases quickly after that time (du Jardin & Severin, 2011; du Jardin, 2015). For the three-year horizon, most approaches provide inaccurate prediction results. However, it proved to be possible to extend the forecasting horizon up to three years using hazard models (Gepp & Kumar, 2008; Dakovic et al., 2010) or terminal failure processes (du Jardin, 2015). However, development of an effective prediction model for a longer horizon is a very challenging task due to instability of financial ratios and fluctuations alongside the economic cycle. In particular, financial ratios of SMEs are very unstable over time and do not contain reliable annual information, which makes the use of non-financial variables important (Balcaen & Ooghe, 2006). However, the use of non-financial variables is hardly possible on a cross-country basis due to their incomparability and limited availability. Therefore, studies on non-financial predictors are focused on single countries (e.g., the U.S., Finland, Italy, Spain).

Moreover, stakeholders target profit maximization, moving the cutoff scores up or down in order to, on the one hand, avoid extending credit to potential bankruptcies (Type I error) and, on the other hand, to avoid losing attractive customers (Type II error). Therefore, the performance of prediction methods should be assessed through the lens of cost-return analysis (e.g., Altman et al., 1977; Weiss & Capkun, 2005).

The purpose of this study is twofold. We aim to assess the predictive ability of different estimation methods (traditional: logistic regression (LR); data mining: decision tree (DT), gradient boosting (GB), neural network (NN), and support vector machine (SVM)) in a large cross-sectional sample of Finnish firms (59,099 firms) over a ten-year period 2003–2013. Moreover, we assess the profit at each method's optimal cutoff in comparison with a hypothetical "perfect credit decision." The majority of the sample firms are small and medium-sized enterprises (SMEs) domiciled in Finland, for which we have access to both financial and non-financial data.

We contribute to the research on failure prediction in three ways. First, compared to recent studies (e.g., Barboza et al., 2017, Jones et al., 2017), we use a very wide set of input variables (including non-financial variables). Second, we analyze the effect of horizon length (up to 10 years) on the comparative performance of different methods. Third, we assess the value of the methods using the cost-return ratio for classification. Thus, we show the effect of relative costs on the value of different methods. We demonstrate that LR and NN are the most efficient methods in bankruptcy prediction for all the horizons in our data. Similarly, we show that GB and DT lead to the least accurate prediction models.

The paper is organized as follows. In this introductory section, the motivation and objectives of the study are briefly discussed. The second section reviews prior studies on failure prediction with the use of different variable selection and estimation techniques. In this section, we also draw research hypotheses. In the third section, the sample of bankrupt and non-bankrupt firms, variables, variable selection methods, and estimation methods are presented. The fourth section presents and discusses the empirical results on the accuracy and value of different methods, while the last section summarizes our findings and presents limitations of the study as well as challenges for future research.

2. Literature Review

This section presents a review of the literature divided into discussion of the selection of variables and the use of different estimation techniques.

2.1 Selection of Variables

How to select variables properly and how many variables should be used are important questions in failure prediction modeling. Balcaen and Ooghe (2006) reviewed business failure studies from the last 35 years and concluded that there is a lack of any theoretical framework for variable selection and little consensus on which financial variables are the best for discriminating between failed and non-failed firms. In reality, variables selected for a given failure prediction model are often sample and environment specific; thus, the results are difficult to generalize. Bellovary et al. (2007) reviewed more than 150 bankruptcy studies and reported that, in total, 752 different variables were used. Karels and

Prakash (1987) emphasized that careful selection of variables is necessary to improve performance of the models. In the same way, Zavgren and Friedman (1988) indicated three drawbacks of previous studies, one of those being arbitrary selection of variables.

Financial variables are usually highly correlated with one another, which can be disruptive for the accuracy of the model, especially in the validation set (Balcaen & Ooghe, 2006). Therefore, researchers have paid attention to non-financial variables, which typically are not closely correlated with one another or with financial variables. In addition, these variables are not exposed to window-dressing or smoothing in the same way as financial variables. Therefore, non-financial variables are useful in small-business failure prediction studies. These kinds of variables can be derived from different sources, such as credit registers, other official registers related to business activity, market intelligence, press announcements and even annual reports. In practice, their availability depends on the size of the firm, its legal form, and the country-specific legal framework. Keasey and Watson (1987) published one of the earliest tests of the usefulness of non-financial variables, concluding that marginally better predictions may be provided by the use of non-financial variables. Back (2005) also combined financial and non-financial variables, concluding that the number of payment delays was statistically the most important factor. In addition, Altman and Sabato (2007) pointed out that model prediction accuracy may be improved by the use of non-financial variables. This line of research was continued by Altman et al. (2010, 2016), who reported that the default prediction power of risk models built specifically for SMEs significantly exceeded that of generic models.

Because of the multitude of variables available, the selection of predictors for use in bankruptcy prediction models is a relevant issue (du Jardin, 2009). Back et al. (1996a, 1996b) compared stepwise discriminant analysis (DA), stepwise logistic regression (LR), and genetic algorithm (GA) in the selection of predictors. They showed that these methods lead not only to different predictors but also to a different number of predictors. These findings are supported by du Jardin (2010), who compared three different selection methods that are especially suitable for neural network (NN) analysis.

Acosta-González and Fernández-Rodríguez (2014) employed GA for variable selection in order to improve the accuracy of logit modeling for the Spanish building industry. This type of selection improved forecasts for failed firms and presented similar accuracy for non-failed ones. Delen et al. (2013) used exploratory factor analysis (EFA) to identify underlying dimensions of financial ratios that explained the best firms' performance (ROE and ROA). In that case, however, the selection of ratios had not been applied to bankruptcy prediction but rather to explain financial performance.

To improve the accuracy of a prediction model, it may be beneficial to ensure that the criteria used in variable selection are related to the criteria used in modeling (du Jardin, 2010). In this study, we only compare two commonly used selection methods: a filter method (the linear R-squared method) and a wrapper method (the logistic stepwise LR). The filter method neglects potential interactions among variables, while the wrapper method allows us to detect possible interactions among variables. SVM, GB, and DT methods include their own respective mechanisms for variable selection. We can expect that the (linear) filter method will be outperformed by the (non-linear) wrapper method, since we will not use linear estimation methods (for example DA) in this study. The potential superiority of the wrapper method is also supported by the expected interactions between the (financial and non-financial) predictors, which are neglected by the filter method. The weaker the symptoms of bankruptcy are, the more important it is that the variable selection method be consistent with the modeling method and detect potential interactions. These characteristics are of relevance to our research on the prediction accuracy for several horizons, since the longer the horizon, the weaker the symptoms of bankruptcy. We expect that, especially for longer horizons, the filter method will be outperformed by the wrapper method, as the latter is more closely related to the modeling in this study and takes into account variable interactions. Therefore, we present the following two-part research hypothesis (H1):

Hypothesis H1a: *The filter method is outperformed by the wrapper method in selecting variables for failure prediction models.*

Hypothesis H1b: *The difference in performance between the wrapper method and the filter method increases with the prediction horizon.*

2.2 Comparing Estimation Methods

There are a great many studies on failure model estimation methods. One of the first studies providing a wide review of various modeling techniques was carried out by Dimitras et al. (1996). At that time, multiple DA and LR prevailed as tools for bankruptcy prediction. Bellovary et al. (2007), using 165 studies from 1930 onwards, showed trends in bankruptcy prediction tools from the univariate approach in the 1930s through the discriminant analysis of the 1960s and 1970s to the logit analysis and neural networks of the 1980s and 1990s. Kumar and Ravi (2007) underlined that traditional statistical tools have to be replaced by artificial intelligence (AI) techniques. Do Prado et al. (2016) ran a bibliometric analysis of research indexed on Web of Science from 1968 to 2014, showing that the most often cited methods were NN, LR, and DA. On the other hand, after 2008, highly cited research mainly referenced SVM. Moro et al. (2014) applied text mining and Dirichlet allocation to the analysis of business intelligence in banking over the 2002–2013 period. The leading roles were played by credit risk analysis and bankruptcy prediction, with relevant topics dedicated to NN and SVM. Due to the high level of interest of many stakeholders in improving accuracy, there are still new approaches emerging to predict bankruptcies. Some recent studies present hybrid models, combining statistical methods with artificial intelligence (Li et al., 2016; Pal et al., 2016).

There are a number of studies comparing the performance of different estimation methods in bankruptcy prediction. Initially, the analysis was limited to different statistical techniques (e.g., DA, LR); however, it has been moving gradually toward artificial intelligence. Laitinen and Kankaanpää (1999), for example, compared six alternative methods (DA, LR, recursive partitioning, survival analysis, NN, and the Human Information Processing (HIP) approach), reporting that there was a statistically significant difference in predictive accuracy only between LR and survival analysis one year prior to the failure. Two and three years prior to the failure, statistically significant differences were not found. They concluded that none of the statistical methods was superior to any other.

Balcaen and Ooghe (2006) concentrated on classic methods and concluded that despite the extensive research, there seemed to be no superior modeling method in bankruptcy prediction. They further claimed that most studies reach heterogeneous conclusions. As a result of this lack of consensus, the selection of the modeling method is left entirely to the researcher. Du Jardin (2009) analyzed 190 papers on bankruptcy prediction and found that more than fifty methods were used in those studies. The results usually revealed only minor differences in prediction accuracy among the methods.

Chen (2011), using data for Taiwanese companies, employed LDA, LR, DT, NN, SVM, and evolutionary computation techniques, supported by the selection of ratios with the use of principal component analysis (PCA). The conclusions were that traditional statistical methods were better in handling large datasets, while AI techniques performed better with small datasets. Moreover, SVM with evolutionary computation provided a good balance of high accuracy in the short and long term as well as for distressed and non-distressed firms.

Du Jardin (2015) used five popular methods (DA, LR, NN, survival analysis, and self-organizing mapping) to predict bankruptcy for horizons of one, two, and three years for French firms. He did not test the differences in accuracy among the methods statistically, but his research proved intuitively that the differences are minor and dependent on the industry and the prediction horizon. Especially with a short horizon, the differences are negligible. With a very short horizon, the signs of bankruptcy may be so obvious that any statistical method can make an accurate prediction. When the horizon increases and the symptoms of bankruptcy become weaker, the differences in accuracy among the methods can increase.

Gordini (2014) compared performance of GA, LR, and SVM for the prediction of bankruptcy of Italian SMEs. This study showed that GA was more efficient than the two other methods, especially in reducing Type II error. GA outperformed LR and SVM in total predictive performance as well as for bankruptcy and non-bankruptcy cases in different settings (according to size or geographical area).

Li et al. (2016) built up a hybrid model using LR and NN for Finnish SMEs (based only on financial variables) over the 2004–2012 period. This hybrid model provided higher accuracy (84.59%) than models based on each technique taken separately (71.55% and 75.44%, respectively). In the same stream of hybrid models, Pal et al. (2016) combined hybrid regression (powerful stepwise for variable selection and linear for modeling expert rating) with SVM to predict the business health of 198 multinational manufacturing companies. The results of their four-stage experiment generated an accuracy rate exceeding 95%.

Du Jardin (2016) employed DT, DA, LR, and NN for data on French firms from 2002 to 2012 to build single models as well as ensembles of models (using bagging, boosting, and random subspace). He assumed that there is no single bankruptcy path and that different profiles should be taken into account and therefore introduced profile-based models. Previous studies indicated that traditional and hybrid ensemble models outperformed single models in accuracy. However, du Jardin showed that on the one hand, profile-based models outperformed ensemble models in most of the cases and, on the other hand, ensemble models were more accurate than single ones.

Jones et al. (2017) examined the predictive performance of several "simple classifiers" (LR, DA, and probit), "more advanced techniques" (NN, SVM), and "new age" methods (random forests, AdaBoost, and generalized boosting) in a large U.S. public company data set. While LR and DA performed reasonably well, the "new age" methods had superior predictive performance. Similar evidence was provided by Barboza et al. (2017), who tested, on a large set of North American firms, the performance of machine learning methods (SVM, bagging, boosting, and random forest), showing they outperformed traditional LDA, LR, and NN.

So far, the results of studies designed for comparing the performance of different methods have been mixed. However, the overall conclusion has been that artificial intelligence and hybrid models outperform traditional statistical techniques. Not much attention has been paid to the long-term horizon, but keeping in mind the PD lifetime required by IFRS 9, failure prediction modeling should move in this direction. In this study, we try to fill this gap. We expect that with a short prediction horizon, the symptoms of failure are so obvious that the differences in accuracy among alternative methods will not be significant. However, when the horizon increases, the symptoms become weaker and, consequently, the differences in performance among the methods of varying sophistication increase. The longer the horizon, the larger the differences in performance between less and more sophisticated modeling methods. Therefore, we set the following three-part hypothesis (H2):

Hypothesis H2a: *With a short horizon, there are no differences in the prediction performance of different estimation methods.*

Hypothesis H2b: *The difference in the prediction performance of different estimation techniques increases with the prediction horizon.*

Hypothesis H2c: *More sophisticated methods perform better than less sophisticated methods with a longer horizon.*

3. Data and Statistical Methods

3.1 Empirical Data

3.1.1 Firms

Over the analyzed period, the macroeconomic situation in Finland changed considerably due to the consequences of the global financial crisis. Up to 2008, the macroeconomic environment was satisfactory, but in 2009, the GDP fell by 8.5%. In the following years, the Finnish economy was sluggish. Therefore, the number of bankruptcies is significant.

This study is based on financial statements and non-financial data provided by a credit information company (Suomen Asiakastieto Oy (SA)) as of the end of 2003, with the following restrictions and selection criteria. As of the cross-section date (end of 2003), the maximum acceptable age of the most recent financial statement was set to 24 months. Private proprietors, foundations, associations, financial firms, and housing companies were not included. If the company age was not recorded or if turnover or total asset figures were missing, the firm was dropped. Two consecutive most-recent financial statements were needed to calculate the cash-flow and the annual change-based variables, which left out very young startups from the data. We have gathered follow-up data regarding the future statuses and status change years for the 2004–2013 period.

The total number of observations fulfilling our criteria is 59,099. Over the 2004–2013 period, some of those firms (10,183) ceased operations due to a merger, voluntary liquidation, or diffusion. The number of firms that were active at the end of 2013 is 46,949. This comprises 79.4% of the original data set, whereas 1,967 firms (3.3% of the data set) went bankrupt, and the remaining firms dissolved for various reasons.

3.1.2 Variables

This study is based on a large set of both financial and non-financial variables in order to produce results that are as general as possible. The list of variables and their definitions are presented in Appendix 1. We have selected the two sets of variables on the basis of relevant prior studies. When selecting the financial variables, we paid special attention to Edmister (1972), Ohlson (1980), Beaver (1966), Altman (1968), and Bellovary et al. (2007). The available set of variables includes 15 financial variables classified into seven groups (profitability, liquidity, solvency, cash flow, size, growth, and changes in ratios). It should be noted that almost all the firms in the data are private, limited companies. Therefore, market-based ratios are not included in the analyses (as in Hillegeist et al., 2004; Chava & Jarrow, 2004; and Reisz & Perlich, 2007). Since the basic data are cross-sectional, only one year's growth is considered as a measure of growth. The distributions of most financial variables are skewed so that medians provide better insight on average values than do means.

Variables containing obvious outlier observations have been winsorized, generally at 1% and/or 99%.[1]

The second set of variables is composed of 20 non-financial variables with special reference to Keasey and Watson (1987), Flagg et al. (1991), Laitinen (1999), Back (2005), and Altman et al. (2010, 2016). These non-financial variables are also classified into seven groups (firm type, age, industry risk, audit report, disclosure policy, payment behavior, and board members). Since almost all the firms are limited companies, the company form variable is not included in the data set.

3.2 Statistical Methods

In this study, we compare two commonly used variable selection methods: a filter and a wrapper. Filter methods analyze intrinsic properties of the data but do not take the relationships among predictors into consideration. Wrapper methods allow the detection of possible interactions among variables. In variable selection and modeling, we use the SAS Enterprise Miner 14.1 software (hereafter SAS EM).[2] We apply these selection methods only with LR and NN methods, contrasting the results with corresponding all-variables models.

We use the R-squared method as representative of filter methods. This method is performed in two successive steps. First, R-squares between each potential predictor and the dummy target variable (bankruptcy = 1, non-bankruptcy = 0) are calculated. The variables with a correlation above a specified threshold are selected in the first step. The variables which are selected in the first step enter the second step of variable selection in which a sequential forward selection process is used. This process starts by selecting the

[1] The descriptive statistics for financial and non-financial variables are available from the authors upon request.
[2] Exact descriptions of the selected methods can be found in the SAS EM support pages (http://support.sas.com/software/products/miner/index.html). The filter method is included in the variable selection node, whereas the wrapper method is located in the regression node.

predictor variable that has the highest correlation coefficient with the target variable. A regression model is estimated with the selected predictor. Then, at each successive step of the sequence, an additional predictor that provides the largest incremental contribution to the model R-square is added to the regression. If the lower bound for the incremental contribution to the model R-square is reached, the selection process stops.

The wrapper method adopted in this study is the stepwise LR, which is naturally consistent and connected with LR in modeling. This method is performed in several successive steps. First, the method estimates the intercept and computes the chi-square statistic for each variable not in the model and examines the largest of these statistics. If it is significant at the specified level, the corresponding predictor is added to the model. Then, the chi-square statistic for each variable not in the model is again calculated, and a new predictor is potentially entered into the model. However, a variable already in the model does not necessarily remain but can be removed if its significance level is below the specified level after adding a new predictor to the model. Effects are entered into and removed from the model in such a way that each forward selection step can be followed by one or more backward elimination steps. The stepwise selection process terminates if no further effect can be added to the model or if the current model is identical to a previously visited model.

We compare the performance of five different methods: decision tree (DT), gradient boosting (GB), logistic regression (LR), neural network with multi-layer perceptron (NN), and support vector machine (SVM) methods. In fact, SVM is a part of the family of NN algorithms (Andras, 2002). These methods can be classified into three classes with respect to sophistication. DT and GB belong to the lowest level, while NN and SVM belong to the highest level of sophistication. LR can be classified into the middle level of sophistication. The methods are only briefly described here because they have been reviewed in detail in recent studies (Chen, 2011a; Chen, 2011b; Erdogan, 2013; du Jardin, 2015), and the exact descriptions are presented in the SAS EM Internet support pages. If not otherwise explicitly reported, the results were run with default values of the SAS EM options.

DT is a non-linear discrimination method which uses a set of independent variables to split a sample into progressively smaller subgroups. The procedure is iterative at each branch of the tree, and it selects the independent variable that has the strongest association with the dependent variable according to a specified criterion (Chen, 2011). Thus, an empirical tree represents a segmentation of the data that is created by applying a series of simple rules. Each rule assigns an observation to a segment based on the value of one input. Then, one rule is applied after another, resulting in a hierarchy of segments within segments (a tree). The hierarchy is called a *tree*, and each segment is called a *node*. The final nodes are called *leaves*. For each leaf, a decision is made and applied to all observations in the leaf. In predictive modeling such as bankruptcy prediction, the decision is the predicted value. DT of SAS EM is located in the Decision Tree node.

Like decision trees, GB does not make assumptions about the distribution of the data. It is a machine learning technique for classification problems that produces a prediction model in the form of an ensemble of weak prediction models, typically decision trees. It builds the model in a stagewise fashion and generalizes it by allowing optimization of an arbitrary differentiable loss function. It resamples the data several times to generate results that form a weighted average of the resampled data set. Tree boosting creates a series of decision trees that form a single predictive model. Boosting is less exposed to overfitting the data than is a single decision tree. If a decision tree fits the data relatively well, then boosting often improves the fit. GB of SAS EM is found in the Gradient Boosting node.

LR is based on the assumption that the probability of bankruptcy is related to the independent variables through a logistic link function. It can be used to predict a binary dependent variable (bankrupt or non-bankrupt) and to determine the (pseudo) percentage of variance in the dependent variable explained by the independent variables (predictors). This analysis does not require that the distributions of independent variables be multivariate normal or that groups have equal covariance matrices, which are basic assumptions in linear discriminant analysis (Hosmer & Lemeshow, 1989). LR creates a score (logit) for every firm. It is assumed that the independent variables are linearly related to the logit. This (risk) score is used to determine the probability of membership in bankrupt firms through the logistic function. LR of SAS EM can be gotten from the Regression node when the target variable is binary.

NN is a collection of computational elements where neurons are interconnected. The basic computational structure is comprised of three layers of neurons: the input layer (independent variables), the hidden layer, and the output (bankrupt or non-bankrupt) layer. In addition to these neurons, the network is also composed of connections between the layers. The number and patterns of these connections determine the task a network is capable of performing. We apply here a version of NN called multilayer perceptron (MLP), which is a feedforward artificial neural network model. MLP consists of multiple layers of nodes in a directed graph, with each layer fully connected to the next one. Except for the input nodes, each node is a neuron (or processing element) with a nonlinear activation function. MLP utilizes a supervised learning technique called *backpropagation* for training the network. In this study, we run NN using the Neural Network model node in SAS EM.

Support Vector Machine (SVM) is a machine learning algorithm for classifying high-dimensional data. SVM uses a linear model to implement nonlinear class boundaries by mapping input vectors nonlinearly into a high-dimensional feature space (Chen, 2011). SVM has also been shown to be very resistant to the overfitting problem, eventually achieving high-generalization performance in solving various forecasting and classification problems. Training SVM is equivalent to solving a linearly-constrained quadratic programming problem, so that the SVM solution is always unique and globally optimal, unlike other training that carries the risk of getting into local optima. In the SAS EM software, SVM is run by the HP Support Vector Machine node, which uses the high-performance HPSVM procedure for binary classification problems.

When interpreting the results, the multivariate prediction (classification) accuracy for different horizons plays the key role. The models are estimated for four different horizons: short term (1 year), middle term (2–3 years), long term (4–5 years), and very long term (6–10 years). Each of the bankruptcy prediction models is estimated using randomly selected estimation (training), validation, and test samples (40% + 30% + 30% of the total sample). Classification accuracy of the models is measured by the AUC (Area Under Curve) measure extracted from the ROC (Receiver Operating Characteristic curve). These profiles show the trade-offs between Type I and Type II errors and statistically represent the cumulative probability distribution of default events. We report AUCs only for the test data.

We estimate each model so that the bankrupt and non-bankrupt portions are equally weighted in the estimations, whereby the cutoff pseudo-probability of each model is located at 0.50. We find this weighting procedure necessary in order to make the analysis of alternative cutoff points of competing models across different horizons feasible in the profit-maximization results section. In practice, the accuracy and profitability of the adopted prediction model depend on the choice of the specific cutoff point with associated Type I and Type II errors. Weiss and Capkun (2005) stressed that the superiority of one model over another cannot be fully measured unless the costs of errors are taken into account.

It is assumed that the lender receives returns from its non-bankrupt customers but suffers losses from its bankrupt customers. Therefore, we assume that if the user of the model lends to a firm which does not go bankrupt, he or she gets 1 unit of money as return. If credit is given to a firm which goes bankrupt, the user will lose as cost from 1 to 100 units of money. Then, we assume that the user lends money to all the firms scoring below a specific cutoff point, and we calculate the profits of the credit decision (cost-return) for the sample firms for each method and each period, contrasting the profits generated in each case and comparing them with the naive policy of lending to all the firms. If the potential customer's score is above the cutoff, the credit is declined and no transactions are carried out, leading to zero cost or profit. Finally, we search for the cutoff point giving the highest profits (profit maximization). We use all the data to calculate these profits (estimation, validation, and test samples) to increase the dispersion of probability value points in the data. The absolute amounts of cost and return are not relevant to our analysis since only the cost-return ratio is important.

4. Empirical Results

4.1 AUC Results

Table 1 (Panel 1) presents the accuracy of the statistical methods in terms of AUC for each length of the horizon. Comparing the three LR approaches, the LR-ALL versions that are run without a preceding variable selection procedure (thus including all our potential predictor variables) give the highest AUCs for all the prediction horizons. In the same way, the LR model based on the wrapper method (LR-W) outperforms the version employing the filter method (LR-F), which supports H1a. The difference in AUCs between the wrapping and filter methods also increases with the length of the horizon, supporting H1b. For the NN models, the interpretation of the results is similar with respect to the research hypotheses. For each horizon, the wrapper method (NN-W) gives a higher AUC than the filter method (NN-F), thereby supporting H1a. Moreover, the difference in accuracy increases with the horizon, which is consistent with H1b. However, the all-variables version of NN (NN-ALL) outperforms the wrapper version (NN-W) for only the one-year horizon.

(Table 1 here)

The differences in AUCs among the best-performing methods are small. The one-year horizon AUCs of all LR and NN versions are nearly equal and comparable with the AUC of the support vector machine method (SVM). SVM and LR-W give almost equal AUCs for the different horizons. Similarly, the AUCs given by the methods based on the filter methods (LR-F and NN-F) differ only slightly. The AUCs of these methods for the horizon of 4–5 years are exceptionally low (0.6480), reflecting the poor performance of the filtering method for this horizon. On average, DT gives low AUCs of around 0.70 for each horizon. However, classification accuracy does not decline with the prediction horizon but rather is almost stable. Both GB and DT clearly give the lowest AUCs for the sample. These methods give low AUCs already in the one-year horizon, contradicting H2a. The most sophisticated methods (NN, SVM, and also LR) give higher AUCs for the longer periods, supporting H2c, but the difference in AUCs between the more and less sophisticated methods does not increase with the length of the horizon, contradicting H2b.

Panel 2 of Table 1 presents the ranks of the methods based on AUCs for all horizons and the final ranks (according to the sum of ranks) over all the horizons. The final ranks show that the LR-ALL and NN-W methods perform well for each horizon. The rank of NN-W for the one-year horizon is only fifth in the performance order, but the difference compared to the

AUC of the best method is insignificant (0.008). It is somewhat surprising that the all-variables version of LR performs so well. A probable reason contributing to this is that some of those variables not selected by the stepwise procedure still contribute to prediction (significance close to the level required to enter the model) and that due to the winsorization procedures, there are no strongly exceptional, outlying values in the unselected variables. We emphasize, however, that we do not propose abandoning variable selection procedures in model development.

The gradient boosting (GB) and decision tree (DT) methods have low rankings for all the horizons and get ranked to the last positions in overall final accuracy classification. The models based on the filter variable selection method (LR-F and NN-F) are clearly outperformed by the all-variables models (LR-ALL and NN-ALL), the wrapper method versions (LR-W and NN-W), and by the SVM. On the basis of generic classification accuracy (AUC), there are both statistical and data mining method versions among the top ranks.

4.2 Profit-Maximization Results

Detailed results for the selection of the optimal (profit maximizing) cutoff scores are reported separately for the four horizons in Appendices 2–5. The results (cutoff, profit, and percentage of bankrupt and non-bankrupt firms correctly classified) are calculated for six different combinations of cost (lending to bankrupt firms) and return (lending to non-bankrupt firms) so that the cost to return ratio varies from 1 to 100. Table 2 reports a summary of the appendices, presenting the profit at each method's optimal cutoff as a percentage relative to the hypothetical perfect credit decision profit. The perfect decision means here that credit would be granted to each non-bankrupt firm but to none of the bankrupt firms. For each horizon, the number of non-bankrupt firms is constant (46,949), which yields 46,949 units of money as return (maximum profit). For example, Table 4 (Panel

1) shows that for the one-year horizon model, with a cost-return ratio of 100:1, the profit yielded by the LR-ALL model is 79.02% of the maximum profit. This method yields 37,100 units of money with its optimal cutoff score of 0.560 (see Appendix 2), which makes just 79.02% of the maximum profit. It clearly beats the “credit to all” decision (last column), which only gets 24.6% of the maximum profit for this cost-return ratio.

(Table 2 here)

When the cost-return ratio declines from 100:1 toward 1:1, the profit as a percentage of maximum profit will approach almost 100%, irrespective of the method and the horizon. At the same time, the optimal cutoff score rises and approaches the value of 1. Due to the low costs of Type I errors, the “credit to all” policy also yields a profit close to 100% of the maximum when the cost-return ratio is close to 1. The percentage of profit earned for different horizons strongly depends on the number (percentage) of bankruptcies in the sample within the period. If there are plenty of bankrupt firms and the cost-return ratio is high, the percentage of profit earned will be low due to high costs of Type I errors. The numbers of bankrupt firms in this study for the horizons of 1, 2–3, 4–5, and 6–10 years are, respectively, 354, 438, 225, and 950, leading to percentages of 0.75, 0.93, 0.48, and 2.02. These percentages, together with the accuracies of the models, explain why the profit (percentage) figures are so high for the 4–5 year horizon and so low for the 6–10 year horizon. For this reason, the “credit to all” decision leads to negative profits for the period of 6–10 years, when the cost-return ratio is 50:1 or higher. The lower the percentage of bankrupt firms, the smaller the differences in earned profit percentage among different models. For the period of 4–5 years, these differences are very small, except for very high cost-return ratios.

For simplicity, we can limit our analysis and assume that the cost to return ratio is in practice 30:1, which is close to the value (35:1) suggested by Altman et al. (1977). For the one-year horizon, all the statistical models in this case (30:1) clearly beat the naive "credit to all" rule. However, only LR-ALL and NN-W models are able to return more than 90% of the maximum profit. GB is outperformed by DT but, in general, the differences in the profit (as a percentage of the hypothetical perfect model profit) among all the methods are quite small (in the 90.28–84.89 range). Table 3 presents the percentage of profit that a method earns in comparison with the profit of the best method for the cost to return ratio and the horizon in question. Thus, for the ratio of 30:1 and the horizon of one year, GB earns 91.55% of the profit (Panel 1 of Table 3) yielded by the best method (LR-ALL) in that case. This percentage, compared to the best model profit, reflects the loss incurred by not using the best method at hand.

(Table 3 here)

Considering the 2–3 year horizon with the cost to return ratio of 30:1, the methods which give the highest profit are the same as for the one-year horizon (LR-ALL, NN-ALL, LR-W, and NN-W), but here DT performs exceptionally well, earning 97.34% of the profit given by LR-ALL (Panel 2). In this case, DT clearly outperforms GB (89.58%). It also beats the profit yielded by the models based on the filter method (LR-F and NN-F). When the cost-return ratio is 10:1, DT yields the highest profit. For the 4–5 year horizon with the cost to return ratio of 30:1, all methods except DT yield more than 99% of the profit of NN-W (100%) (see Panel 3). DT here earns the same profit as the naive "credit to all" rule. For this horizon, GB earns very high profits for each cost-return ratio. For the ratio of 100:1, GB earns the highest profit of all methods. Then, for the horizon of 6–10 years with the ratio of 30:1, GB performs worst, yielding only 82.60% of the maximum profit earned by the best method (NN-ALL). For this horizon, the models based on the filter method (LR-F and NN-F) also yield exceptionally low profits, whereas DT yields satisfactory results.

Table 4 presents the ranks of different prediction methods in earning profit for different horizons using the (justified) cost-return ratio of 30:1 as an exemplary value. The table shows that these ranks are consistent with the ranks based on AUCs (Table 1). LR-ALL, NN-ALL, and NN-W share the first places, and GB and DT take the last ones. Moreover, the models based on the filter method (LR-F and NN-F) are outperformed by SVM. SVM and DT rank highest for the horizons of 6–10 and 2–3 years, when the percentage of bankrupt firms is the highest. However, GB's best rank (5) is for the 4–5 year horizon, when the percentage is the lowest. LR-F and NN-F reach quite low ranks for each horizon, indicating the obvious inefficiency of the filter method.

(Table 4 here)

In summary, these specific results for profit maximization indicate that for any horizon, the wrapper method performs better than the filter method, conforming to H1a. For the shortest horizon, the differences in profit earning among the methods are small, but they increase with the length of the horizon, which is consistent with H1b. For the one-year horizon, assuming a reasonable cost to return ratio (such as 30:1), the differences in profit earned among the statistical methods are quite small (the worst model earns 94.03% of the highest profit). This result lends support to H2a. The differences in profit earned among the methods increase with the prediction horizon, supporting H2b. The most sophisticated models, NN-ALL and NN-W, perform best with a long horizon, but LR-ALL and LR-W are comparable with SVM, which is more sophisticated than the logistic regression models. GB and DT earn quite low profits in long-term prediction. Thus, the results lend support to H2c.

5. Summary and Conclusions

Failure prediction models are widely used by many stakeholders in predicting a failure emerging in terms of payment default, bankruptcy, or a similar event. The higher the accuracy and the longer the prediction horizon an adopted model has, the more valuable it is to stakeholders. The value is also dependent on the cost of Type I and II errors in classifying bankrupt and non-bankrupt firms (Altman et al., 1977; Weiss & Capkun, 2005). Because the relative cost of Type I error is high, it is important that the model be efficient, especially in classifying bankrupt firms correctly. There are a vast number of different statistical methods developed for failure prediction designed to increase accuracy or to lengthen the horizon or both (Dimitras et al., 1996; Bellovary et al., 2007; du Jardin, 2015). However, the value of these methods is rarely assessed in the context of non-financial variables and small firms (Altman & Sabato, 2007). Therefore, there is a call to compare the performance of different methods based on both financial and non-financial variables, especially in the small business context.

The purpose of our study was to assess the performance of different estimation methods in a large cross-sectional sample of Finnish firms over a 10-year period. We applied five different methods for bankruptcy prediction using both financial and non-financial variables: decision tree (DT), gradient boosting (GB), logistic regression (LR), neural network (NN), and support vector machine (SVM) methods. We also applied two different variable selection methods (a wrapper method and a filter method) in the context of LR and NN. The majority of firms in our sample are small and medium-sized enterprises (SMEs). Thus, our study is a response to the call for small business studies based on financial and non-financial variables. The findings contribute to the previous research on variable selection methods (Acosta-González & Fernández-Rodríguez, 2014; Back et al., 1996a, 1996b; du Jardin, 2010) and estimation methods (Laitinen & Kankaanpää, 1999; Chen, 2011; du Jardin, 2015, 2016; Gordini, 2014, Barboza et al., 2017; Jones et al., 2017).

We divided our empirical analysis into two parts. Firstly, we analyzed the generic accuracy of different methods using the area under the ROC (AUC) as a measure. This AUC measure is independent of the cutoff point, including information over all such points. However, in practice, we know that the user of a failure prediction model selects a certain cutoff point to be applied in credit decisions. Therefore, it is important to the value of the model how the user selects the cutoff point, which is based on relative costs of Type I and II errors. Thus, secondly, we analyzed the effect of these relative costs on the value of different methods. We calculated the profit earned by the methods in six different combinations of cost (a loan is granted to a bankruptcy firm) and return (a loan is granted to a non-bankruptcy firm). We also analyzed in detail the value ranks of methods for the 30:1 cost-return ratio, which is comparable with the ratio suggested by Altman et al. (1977).

We first analyzed the accuracy of the methods in terms of AUCs in the test data. Although we had a large number of variables, the LR model consisting of all the variables outperformed the LR versions where wrapper or filter variable selection methods were used. For both LR and NN, the wrapper method (logistic stepwise selection) gave higher AUCs than the filter method (the R-squared method) for all the horizons. The longer the horizon, the larger was the difference in AUCs among the selection methods. For NN, the naive all-variables model and the version based on the wrapper method gave comparable AUCs, although on average, AUCs of the wrapper version slightly surpassed those of the all-variables model. SVM was very efficient with a very short or a very long horizon. On average, SVM clearly outperformed the filter versions of LR and NN. In our comparison, GB and DT clearly gave the lowest AUCs for most horizons.

The AUC results appear to be in contrast with those in Jones et al. (2017), in which generalized boosting (which is equivalent to our GB), AdaBoost, and random forests (RF) showed superior performance compared to more conventional methods, including NN and stepwise LR. The same conclusions were reached by Barboza et al. (2017). There are several potential sources for these contradicting results. First of all, our data consist of small Finnish SMEs, whereas theirs are composed of public U.S. companies. Secondly, we used, except in the variable selection, default options of the software. Thus, the differences in options (or in software) may have contributed to the disparities. Our poorest results are for the DT. As GB leans on an ensemble of decision tree models, poor DT performance may be a contributing factor to the modest performance of GB in our study. It is worth noting that in three of the four horizons examined, GB was still able to produce remarkably better AUCs than DT. Thirdly, we used rather conventional financial statement variables and many non-financial predictors, whereas they used mainly financial variables, of which the best-performing ones (in the RF model) were highly exceptional, the top three being annual growth in capital expenditures, annual growth in leverage-free cash flow, and earnings per share. Furthermore, only conventional variables were significant in their standard logit model. It would be very interesting to test whether these kinds of exceptional variables can perform well in the bankruptcy prediction of SMEs, for which conventional variables are even less stable than for larger public companies. Unfortunately, such variables are not available in most SME company databases.

We also assessed the value of methods using the cost-return ratio for classification. When this ratio is very small (around 1), all the methods earn almost 100% of the maximum profit for any horizon. In this extreme case, the naive “credit to all” rule is also efficient due to the low relative cost of Type I error. In more normal cost-return cases, we found that the value of a method is strongly dependent on the percentage of bankrupt firms in the sample. This is due to the high relative Type I error cost and to differences in error rates among different methods. For simplicity, we discussed in detail the empirically justified case with the 30:1 cost-return ratio. We found that in this case, the LR and NN all-variables versions and the NN wrapper version shared the best overall rank, while GB, DT, and the filter versions of LR and NN had the lowest ranks. SVM and DT had their best ranks for the horizons of 5–10 and 2–3 years, when the percentage of bankrupt observations was the highest. However, GB reached its best rank for the 4–5 year horizon, when the percentage was lowest.

Thus, our empirical findings imply that for small businesses, an efficient failure prediction model can be based on a very large set of both financial and non-financial variables. Moreover, an LR model can perform as efficiently as the more sophisticated NN and even better than SVM. If a variable selection method is used, the wrapper method seems to be superior to the filter method for all the horizons. For LR and NN, the wrapper method and the all-variables versions give comparable results. Therefore, if the user of the model prefers a simple version with few variables to the all-variables version, a wrapper variable selection method is strongly recommended. These results are not sensitive to the length of the prediction horizon in this study.

In summary, our study has many important implications for failure prediction in the small business context. However, it also has limitations which can be mitigated in further studies. Firstly, we limited our analysis to only a small set of methods. In further studies, the variable selection methods can be expanded to such methods as the chi-square method or a genetic algorithm approach that is naturally connected with NN. In the estimation, more methods should be assessed, including such methods as survival analysis, rough set approach, LARS (least-angle regressions), LASSO (least absolute shrinkage and selection operator), AdaBoost, and other “new age” methods. Secondly, our analysis is based on data in which the percentage of bankrupt firms strongly varies with the horizon, which affects the performance of some methods. In further research, this effect should be eliminated using equal percentages for each horizon. Finally, it would be useful to assess the statistical significance of the differences among different methods. Moreover, studies that screen the sensitivity of performance to alterations in the options that are available for various methods are needed to further improve the ability to predict bankruptcies of large firms, public companies, and SMEs.

References

Acosta-González, E., Fernández-Rodríguez, F. (2014). Forecasting Financial Failure of Firms via Genetic Algorithms. Computational Economics, 43(2), 133–157. https://doi.org/10.1007/s10614-013-9392-9

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

Altman E.I., Haldeman R., Narayanan, P. (1977). Zeta Analysis: A new model to identify bankruptcy risk of corporations. Journal of Banking and Finance 1(1), 29-54.

Altman, E.I., Sabato, G. (2007). Modelling credit risk for SMEs: evidence from the U.S. market. Abacus 43 (3), 332–357.

Altman, E.I., Sabato, G., Wilson, N. (2010). The value of non-financial information in small and medium-sized enterprise risk management. Journal of Credit Risk 6(2), 1-33.

Altman E.I., Iwanicz-Drozdowska M., Laitinen E. K., Suvas A. (2016). Financial and nonfinancial variables as long-horizon predictors of bankruptcy. Journal of Credit Risk, 12(4), 49-78.

András, P. (2002). The equivalence of support vector machine and regularization neural networks. Neural Processing Letters, 15(2), 97–104. https://doi.org/10.1023/A:1015292818897

Argenti J. (1976) Corporate collapse; the causes and symptoms. McGraw Hill.

Back, B., Laitinen, T., Sere, K. (1996a). Neural Networks and Bankruptcy Prediction: Funds Flows, Accrual Ratios and Accounting Data. Advances in Accounting 14(Winter), 23-37.

Back, B., Laitinen, T., Sere, K. (1996b). Neural Networks and Genetic Algorithms for Bankruptcy Predictions. Expert Systems with Applications. An International Journal 11(4), 407-413.

Back, P. (2005). Explaining financial difficulties based on previous payment behavior, management background variables and financial Ratios. European Accounting Review 14(4), 839–868.

Balcaen, S., Ooghe, H. (2006). 35 years of studies on business failure: an overview of the classic statistical methodologies and their related problems. The British Accounting Review 38(1), 63–93.

Barboza, F., Kimura, H., Altman, E. (2017). Machine learning models and bankruptcy prediction. Expert Systems with Applications, 83, 405–417.

Beaver, W.H. (1966). Financial ratios as predictors of failure. Empirical Research in Accounting, Selected Studies. Journal of Accounting Research 4, 71–111.

Bellovary, J.L., Giacomino, D.E., Akers, M.D. (2007). A review of bankruptcy prediction studies: 1930 to present. Journal of Financial Education 1 (1), 3-41.

Chava S., Jarrow R.A. (2004). Bankruptcy prediction with industry effects. Review of Finance 8, 537–569.

Chen, M. Y. (2011a). Bankruptcy prediction in firms with statistical and intelligent techniques and a comparison of evolutionary computation approaches. Computers and Mathematics with Applications, 62(12), 4514–4524.

Chen, M.Y. (2011b) Predicting corporate financial distress based on integration of decision tree classification and logistic regression. Expert Systems with Applications, Volume 38, Issue 9, 2011, Pages 11261-11272

D'Aveni, R.A. (1989). The aftermath of organizational decline: a longitudinal study of the strategic and managerial characteristics of declining firms. Academy of Management Journal 32 (3), 577-605.

Dakovic R., Czado C., Berg D. (2010). Bankruptcy prediction in Norway: a comparison study. Applied Economic Letters 17 (3), 1739-1746.

Delen, D., Kuzey, C., Uyar, A. (2013). Measuring firm performance using financial ratios: A decision tree approach. Expert Systems with Applications, 40(10), 3970–3983. https://doi.org/10.1016/j.eswa.2013.01.012

Dimitras, A.I., Zanakis S.H., Zopounidis C. (1996). A survey of business failures with an emphasis on prediction methods and industrial applications. European Journal of Operational Research 90 (3) (1996), 487–513.

Do Prado, J. W., de Castro Alcantara, V., de Melo Carvalho, F., Vieira, K. C., Machado, L. K. C., Tonelli, D. F. (2016). Multivariate analysis of credit risk and bankruptcy research data: a bibliometric study involving different knowledge fields (1968-2014). Scientometrics, 106(3), 1007–1029. https://doi.org/10.1007/s11192-015-1829-6

Du Jardin P., (2009). Bankruptcy prediction models: How to choose the most relevant variables? Bankers, Markets & Investors, Issue 98, January-February, 39–46.

Du Jardin P., Severin E. (2011). Predicting corporate bankruptcy using a self-organizing map: An empirical study to improve the forecasting horizon of a financial failure model. Decision Support Systems 51 (3), 701–711.

Du Jardin P., Severin E. (2012). Forecasting financial failure using a Kohonen map: a comparative study to improve model stability over time. European Journal of Operational Research 221, 378–396.

Du Jardin, P. (2010). Predicting bankruptcy using neural networks and other classification methods: The influence of variable selection techniques on model accuracy. Neurocomputing, 73(10–12), 2047–2060. https://doi.org/10.1016/j.neucom.2009.11.034

Du Jardin, P. (2015). Bankruptcy prediction using terminal failure processes. European Journal of Operational Research 242 (1), 286-303.

Du Jardin, P. (2016). A two-stage classification technique for bankruptcy prediction. European Journal of Operational Research, 254(1), 236–252. https://doi.org/10.1016/j.ejor.2016.03.008

Edmister, R.O. (1972). An empirical test of financial ratio analysis for small business failure prediction. Journal of Financial and Quantitative Analysis 7 (2), 1477-1493.

Erdogan, B.E. (2013). Prediction of bankruptcy using support vector machines: An application to bank bankruptcy. Journal of Statistical Computation and Simulation 83(8), 1543-1555

Flagg, J.C., Giroux, G.A., Wiggins, C.E. (1991). Predicting corporate bankruptcy using failing firms. Review of Financial Economics, 1(1), 67–78.

Gepp, A., K. Kumar (2008).The role of survival analysis in financial distress prediction. International Research Journal of Finance and Economics 16, 13–34.

Gordini, N. (2014). A genetic algorithm approach for SMEs bankruptcy prediction: Empirical evidence from Italy. Expert Systems with Applications, 41(14), 6433–6445. https://doi.org/10.1016/j.eswa.2014.04.0

Hambrick, D.C., D'Aveni, R. (1988). Large corporate failures as downward spirals. Administrative Science Quarterly 33 (1), 1-23.

Hillegeist, S.A., Keating, E.K., Cram, D.P., Lundstedt, K.G. (2004). Assessing the probability of bankruptcy. Review of Accounting Studies 9, 5–34.

Hosmer, D. W., Lemeshow, S. (1989). Applied logistic regression. Wiley, New York.

Jones, S., Johnstone, D., Wilson, R. (2017). Predicting corporate bankruptcy: an evaluation of alternative statistical frameworks. Journal of Business Finance & Accounting 44(1) & (2), 3-34.

Karels, G., Prakash, A.J. (1987). Multivariate normality and forecasting of business bankruptcy. Journal of Business Finance & Accounting 14(4), 573-593.

Keasey, K., Watson, R. (1987). Non-financial symptoms and the prediction of small company failure: a test of Argenti's hypotheses. Journal of Business Finance & Accounting 14(3), 335-354.

Kumar, P., Ravi V. (2007). Bankruptcy Prediction in Banks and Firms via Statistical and Intelligent Techniques – A Review. European Journal of Operational Research 180, 1–28.

Laitinen, E.K. (1991). Financial ratios and different failure processes. Journal of Business Finance and Accounting. September 18, 649-674.

Laitinen, E.K. (1999). Predicting a corporate credit analyst's risk estimate by logistic and linear models. International Review of Financial Analysis 8 (2), 97-121.

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

Li, K., Niskanen, J., Kolehmainen, M., Niskanen, M. (2016). Financial innovation: Credit default hybrid model for SME lending. Expert Systems with Applications, 61, 343–355. https://doi.org/10.1016/j.eswa.2016.05.029

Moro, S., Cortez, P., Rita, P. (2015). Business intelligence in banking: A literature analysis from 2002 to 2013 using text mining and latent Dirichlet allocation. Expert Systems with Applications, 42(3), 1314–1324. https://doi.org/10.1016/j.eswa.2014.09.024

Moulton, W. N., Thomas, H., Pruett, M. (1996). Business failure path ways: Environmental stress and organizational response. Journal of Management, 22(4), 571–595. https://doi.org/10.1177/014920639602200403

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

Ooghe, H., De Prijcker, S. (2008). Failure processes and causes of company bankruptcy: a typology. Management Decision 46 (2), 223-242.

Pal, R., Kupka, K., Aneja, A. P., Militky, J. (2016). Business health characterization: A hybrid regression and support vector machine analysis. Expert Systems with Applications, 49, 48–59. https://doi.org/10.1016/j.eswa.2015.11.027

Reisz A.S., Perlich, C. (2007). A market-based framework for bankruptcy prediction. Journal of Financial Stability 3, 85–131.

Weiss, L.A., Capkun, V. (2005). The impact of incorporating the cost of errors into bankruptcy prediction models. http://www.hec.unil.ch/urccf/recherche/publications/bankruptcy12004_draft.pdf

Zavgren, C.V, Friedman, G.E. (1988). Are bankruptcy prediction models worthwhile? An application in securities analysis. Management International Review 28 (1), 34-44.

APPENDICES

Appendix 1. List of variables

Notation	Definition
FINANCIAL VARIABLES	
1. Profitability	
ROA	(Profit after Interest and Taxes but before Extraordinary Items) /Total Assets *100
RETA	Retained Earnings/Total Assets *100
2. Liquidity	
SHTDEBTtoASSETS	ShortTerm Debt to Assets * 100
QUICKRATIO	(Current Assets - Inventories)/Current Liabilities
3. Solvency	
EQRATIO	Book value of Equity/Total Assets *100
DEBTtoEBITDA	Debt/EBITDA *100
4. Cash flow	
TradCFtoASSETS	(Profit after Interest and Taxes +Depreciation) /Total Assets *100
CFLOWtoASSETS	Cash Flow /Total Assets *100
5. Size	
SIZEA	LN(Total Assets)
SIZESQ	LN(Total Assets) squared
LNSALES	LN(Sales)
6. Growth	
ASSETG1Y	LN(Assets(t)/Assets(t-1))/(Accounting period length in months)/12); One-year Total Asset growth
SALESG1Y	(Sales(t)/Sales(t-1) - 1) * 100, One-year growth rate of Sales
7. Changes in ratios	
SHTDEBTtoASSETSch	Short Term Debt/Total Assets -Lag(Short Term Debt/Total Assets)
TOTALDEBTchToASSETS	(Change in Debt)/Total Assets
NON-FINANCIAL VARIABLES	
1. Firm type	
DAUGHTER	DUMMY
PARENT	DUMMY
2. Industry	
WRDUMMY	Wholesale-Retail Dummy
CONSTRDUMMY	Construction Firm Dummy
3. Age	
LOGofAGE	LN(Age in years)
AGE3to9	Age 3 to 9 years
4. Industry risk	

IND_BR_RISK	Industry Bankruptcy Risk (% of firms)
IND_PDEF_RISK	Industry Payment Default Risk (% of firms)
5. Audit report and disclosure policy	
AUDITQUALIF	Auditors report is qualified (not severe)
AUDITQUALIF_SEVERE	Auditors report is qualified (severe)
LateFiling	Postponed publication of the financial statement
6. Payment behaviour	
DELAYSDUM	The firm has delays recorded within 9 months, DUMMY
DelaysOver60	Number of payment delays over 60 days late
DELAYStoASSETS	Delays(EUR) /Total Assets
LNPOSPAYOBS	LN(Positive payment observations+1)
PRIOR_PDEFAULTS	Number of payment default events prior to cross-section - all
7. Board members	
BOARD_SIZE	Number of board members (No deputy members)
RESIGNED	Number of resigned persons in charge (the last 6 months)
LNBOARD_OWN_PDEFS	Number of board member's own payment defaults
PDBRCONNECTS	Number of associations with payment default firms (other than bankruptcy) by persons in charge

Appendix 2. Model efficiency comparisons; one year before bankruptcy; all data combined.

Cost=100, Return=1

Cost / Return	Profit	Optimal cutoff	Bankr correct-%	Non-br correct-%	Credit given/All-%
LR-ALL	37100	0,560	83,33	91,59	91,03
LR-W	36546	0,624	79,10	93,60	93,06
LR-F	35691	0,534	77,40	93,06	92,53
NN-ALL	37218	0,657	82,77	92,27	91,71
NN-W	37432	0,471	86,16	90,17	89,59
NN-F	36365	0,532	82,20	90,88	90,33
SVM	36429	0,527	83,33	90,16	89,61
GB	32446	0,599	78,81	85,08	84,61
DT	28947	0,301	52,26	97,65	97,28
Credit to all	11549	1	0	100	100

Cost=70, Return=1

Cost / Return	Profit	Optimal cutoff	Bankr correct-%	Non-br correct-%	Credit given/All-%
LR-ALL	39075	0,664	78,53	94,56	94,01
LR-W	38766	0,624	79,10	93,60	93,06
LR-F	38091	0,534	77,40	93,06	92,53
NN-ALL	39067	0,670	82,20	92,60	92,04
NN-W	39291	0,616	82,49	92,93	92,37
NN-F	38406	0,693	77,12	93,88	93,35
SVM	38199	0,527	83,33	90,16	89,61
GB	34837	0,747	66,67	91,80	91,36
DT	34017	0,301	52,26	97,65	97,28
Credit to all	22169	1	0	100	100

Cost=50, Return=1

Cost / Return	Profit	Optimal cutoff	Bankr correct-%	Non-br correct-%	Credit given/All-%
LR-ALL	40633	0,696	76,84	95,28	94,74
LR-W	40246	0,624	79,10	93,60	93,06
LR-F	39778	0,594	74,58	94,31	93,80
NN-ALL	40347	0,682	81,36	92,97	92,41
NN-W	40616	0,722	79,10	94,39	93,84
NN-F	40026	0,693	77,12	93,88	93,35
SVM	39718	0,575	74,86	94,08	93,56
GB	37200	0,752	66,10	92,01	91,58
DT	37397	0,301	52,26	97,65	97,28
Credit to all	29249	1	0	100	100

Cost=30, Return=1

Cost / Return	Profit	Optimal cutoff	Bankr correct-%	Non-br correct-%	Credit given/All-%
LR-ALL	42387	0,770	71,47	96,73	96,22
LR-W	41868	0,657	76,84	94,42	93,88
LR-F	41578	0,594	74,58	94,31	93,80
NN-ALL	42245	0,828	71,75	96,37	95,86
NN-W	42318	0,851	71,75	96,53	96,02
NN-F	41690	0,753	73,45	94,81	94,29
SVM	41502	0,600	70,06	95,17	94,68
GB	39855	0,829	53,95	95,31	94,94
DT	40777	0,301	52,26	97,65	97,28
Credit to all	36329	1	0	100	100

Cost=10, Return=1

Cost / Return	Profit	Optimal cutoff	Bankr correct-%	Non-br correct-%	Credit given/All-%
LR-ALL	44750	0,910	50,00	98,79	98,42
LR-W	44672	0,915	49,72	98,94	98,58
LR-F	44583	0,940	43,22	99,24	98,92
NN-ALL	44685	0,914	58,19	98,33	97,91
NN-W	44631	0,928	45,20	99,19	98,86
NN-F	44576	0,956	42,66	99,27	98,96
SVM	44622	0,706	51,41	98,71	98,33
GB	43846	0,918	28,25	98,80	98,60
DT	44157	0,301	52,26	97,65	97,28
Credit to all	43409	1	0	100	100

Cost=1, Return=1

Cost / Return	Profit	Optimal cutoff	Bankr correct-%	Non-br correct-%	Credit given/All-%
LR-ALL	46616	0,996	11,02	99,96	99,88
LR-W	46607	0,995	9,89	99,95	99,88
LR-F	46601	0,997	5,08	99,74	99,94
NN-ALL	46608	0,993	12,43	99,93	99,84
NN-W	46615	0,993	14,69	99,93	99,82
NN-F	46610	0,993	8,76	99,97	99,90
SVM	46605	0,855	7,91	99,96	99,90
GB	46596	0,975	0,28	100,00	100,00
DT	46595	0,971	0,00	100,00	100,00
Credit to all	46595	1	0	100	100

Appendix 3. Model efficiency comparisons; years 2-3 before bankruptcy; all data combined.

Cost=100, Return=1

Cost / Return	Profit	Optimal cutoff	Bankr correct-%	Non-br correct-%	Credit given/All-%
LR-ALL	27651	0,508	76,71	80,62	80,09
LR-W	26527	0,471	77,63	77,38	76,87
LR-F	24479	0,473	75,34	75,14	74,68
NN-ALL	28744	0,552	78,31	81,46	80,91
NN-W	27450	0,557	74,20	82,54	82,01
NN-F	24427	0,527	74,66	75,67	75,21
SVM	26220	0,372	68,49	85,24	84,74
GB	23036	0,475	83,33	64,61	64,17
DT	24852	0,332	65,30	85,31	84,84
Credit to all	3149	1	0	100	100

Cost=70, Return=1

Cost / Return	Profit	Optimal cutoff	Bankr correct-%	Non-br correct-%	Credit given/All-%
LR-ALL	30763	0,592	68,26	86,25	85,75
LR-W	30002	0,614	63,24	87,91	87,44
LR-F	28335	0,550	59,59	86,74	86,31
NN-ALL	31594	0,552	78,31	81,46	80,91
NN-W	30840	0,557	74,20	82,54	82,01
NN-F	28517	0,669	59,13	87,43	87,00
SVM	30397	0,372	68,04	85,62	85,12
GB	26040	0,543	59,82	81,71	81,32
DT	29412	0,332	65,30	85,31	84,84
Credit to all	16289	1	0	100	100

Cost=50, Return=1

Cost / Return	Profit	Optimal cutoff	Bankr correct-%	Non-br correct-%	Credit given/All-%
LR-ALL	33771	0,634	64,16	88,65	88,16
LR-W	33314	0,635	61,19	89,06	88,60
LR-F	32048	0,565	56,16	88,71	88,29
NN-ALL	34315	0,716	64,84	89,49	88,99
NN-W	33678	0,659	65,07	88,03	87,54
NN-F	32234	0,689	55,48	89,43	89,01
SVM	33206	0,372	66,44	86,38	85,89
GB	29560	0,543	59,82	81,71	81,32
DT	32452	0,332	65,30	85,31	84,84
Credit to all	25049	1	0	100	100

Cost=30, Return=1

Cost / Return	Profit	Optimal cutoff	Bankr correct-%	Non-br correct-%	Credit given/All-%
LR-ALL	37741	0,804	44,52	95,91	95,54
LR-W	37458	0,763	46,80	94,67	94,29
LR-F	36470	0,774	36,07	95,57	95,28
NN-ALL	37686	0,778	55,71	92,67	92,22
NN-W	37450	0,797	47,72	94,40	94,01
NN-F	36639	0,804	39,04	95,10	94,78
SVM	37182	0,379	46,12	94,28	93,90
GB	33809	0,614	0,00	100,00	100,00
DT	36738	0,716	45,21	93,59	93,23
Credit to all	33809	1	0	100	100

Cost=10, Return=1

Cost / Return	Profit	Optimal cutoff	Bankr correct-%	Non-br correct-%	Credit given/All-%
LR-ALL	42987	0,940	16,44	99,36	99,21
LR-W	42979	0,899	24,89	98,55	98,33
LR-F	42786	0,933	10,73	99,46	99,37
NN-ALL	42976	0,907	21,00	98,91	98,72
NN-W	43024	0,949	16,44	99,44	99,29
NN-F	42853	0,906	14,61	99,24	99,11
SVM	42979	0,388	22,15	98,81	98,61
GB	43569	0,614	0,00	100,00	100,00
DT	43605	0,869	27,85	97,48	97,23
Credit to all	42569	1	0	100	100

Cost=1, Return=1

Cost / Return	Profit	Optimal cutoff	Bankr correct-%	Non-br correct-%	Credit given/All-%
LR-ALL	46511	0,998	0,00	100,00	100,00
LR-W	46511	0,998	0,00	100,00	100,00
LR-F	46511	0,993	0,00	100,00	100,00
NN-ALL	46512	0,983	0,23	100,00	100,00
NN-W	46511	0,976	0,00	100,00	100,00
NN-F	46512	0,975	0,23	100,00	100,00
SVM	46511	0,417	0,00	100,00	100,00
GB	46511	0,614	0,00	100,00	100,00
DT	46511	0,925	0,00	100,00	100,00
Credit to all	46511	1	0	100	100

Appendix 4. Model efficiency comparisons; years 4-5 before bankruptcy; all data combined.

Cost=100, Return=1

Cost / Return	Profit	Optimal cutoff	Bankr correct-%	Non-br correct-%	Credit given/All-%
LR-ALL	31497	0,742	46,22	92,86	92,67
LR-W	31206	0,695	48,44	91,18	90,99
LR-F	29353	0,452	31,11	95,54	95,41
NN-ALL	29992	0,739	45,33	90,08	89,91
NN-W	31158	0,575	66,67	82,34	82,11
NN-F	29535	0,444	31,11	95,54	95,41
SVM	30320	0,002	60,89	83,32	83,11
GB	32031	0,697	56,00	89,31	89,10
DT	31433	0,453	61,33	85,48	85,26
Credit to all	24449	1	0	100	100

Cost=70, Return=1

Cost / Return	Profit	Optimal cutoff	Bankr correct-%	Non-br correct-%	Credit given/All-%
LR-ALL	35150	0,759	44,00	93,65	93,48
LR-W	34686	0,695	48,44	91,18	90,99
LR-F	34043	0,459	30,22	95,92	95,79
NN-ALL	34148	0,804	31,56	95,70	95,57
NN-W	34549	0,738	40,89	93,42	93,25
NN-F	34043	0,451	30,22	95,92	95,79
SVM	34161	0,002	31,56	95,72	95,59
GB	35126	0,715	53,78	90,32	90,11
DT	34183	0,795	40,89	92,64	92,48
Credit to all	31199	1	0	100	100

Cost=50, Return=1

Cost / Return	Profit	Optimal cutoff	Bankr correct-%	Non-br correct-%	Credit given/All-%
LR-ALL	37670	0,759	44,00	93,65	93,48
LR-W	37587	0,822	28,89	97,10	96,98
LR-F	37183	0,459	30,22	95,92	95,79
NN-ALL	37364	0,831	24,89	97,58	97,48
NN-W	37529	0,868	21,78	98,68	98,58
NN-F	37183	0,451	30,22	95,92	95,79
SVM	37241	0,002	31,56	95,72	95,59
GB	37658	0,814	35,44	95,44	95,29
DT	36843	0,795	40,89	92,64	92,48
Credit to all	35699	1	0	100	100

Cost=30, Return=1

Cost / Return	Profit	Optimal cutoff	Bankr correct-%	Non-br correct-%	Credit given/All-%
LR-ALL	40800	0,867	23,11	97,96	97,86
LR-W	40917	0,887	18,67	98,85	98,76
LR-F	40691	0,815	15,56	98,81	98,74
NN-ALL	40833	0,866	19,11	98,60	98,52
NN-W	41049	0,868	21,78	98,68	98,58
NN-F	40683	0,914	12,89	99,18	99,12
SVM	40680	0,002	20,00	98,15	98,06
GB	40709	0,839	31,56	96,55	96,42
DT	40199	0,881	0,00	100,00	100,00
Credit to all	40199	1	0	100	100

Cost=10, Return=1

Cost / Return	Profit	Optimal cutoff	Bankr correct-%	Non-br correct-%	Credit given/All-%
LR-ALL	44720	0,975	3,56	99,87	99,86
LR-W	44763	0,939	10,22	99,59	99,54
LR-F	44713	0,966	4,89	99,80	99,77
NN-ALL	44701	0,957	1,33	99,94	99,93
NN-W	44731	0,961	6,22	99,77	99,74
NN-F	44699	0,953	0,00	100,00	100,00
SVM	44708	0,002	0,44	100,00	100,00
GB	44713	0,944	2,22	99,92	99,91
DT	44699	0,881	0,00	100,00	100,00
Credit to all	44699	1	0	100	100

Cost=1, Return=1

Cost / Return	Profit	Optimal cutoff	Bankr correct-%	Non-br correct-%	Credit given/All-%
LR-ALL	46724	0,996	0,00	100,00	100,00
LR-W	46724	0,996	0,00	100,00	100,00
LR-F	46724	0,967	0,00	100,00	100,00
NN-ALL	46724	0,980	0,00	100,00	100,00
NN-W	46724	0,988	0,00	100,00	100,00
NN-F	46724	0,952	0,00	100,00	100,00
SVM	46724	0,002	0,00	100,00	100,00
GB	46724	0,962	0,00	100,00	100,00
DT	46724	0,881	0,00	100,00	100,00
Credit to all	46724	1	0	100	100

Appendix 5. Model efficiency comparisons; years 6-10 before bankruptcy; all data combined.

Cost=100, Return=1

Cost / Return	Profit	Optimal cutoff	Bankr correct-%	Non-br correct-%	Credit given/All-%
LR-ALL	10695	0,356	88,32	46,42	45,73
LR-W	10180	0,367	88,11	45,75	45,08
LR-F	8287	0,392	89,16	39,59	39,02
NN-ALL	10663	0,356	86,32	50,40	49,67
NN-W	10537	0,326	86,63	49,49	48,78
NN-F	8728	0,342	88,84	41,17	40,57
SVM	7993	0,000	89,05	39,18	38,62
GB	5934	0,478	90,21	32,45	32,00
DT	5817	0,259	78,95	54,99	54,32
Credit to all	-48051	1	0	100	100

Cost=70, Return=1

Cost / Return	Profit	Optimal cutoff	Bankr correct-%	Non-br correct-%	Credit given/All-%
LR-ALL	14538	0,383	85,47	51,54	50,81
LR-W	14198	0,393	85,47	50,82	50,10
LR-F	11530	0,441	78,32	55,27	54,61
NN-ALL	14879	0,461	81,16	58,38	57,60
NN-W	14568	0,342	85,37	51,75	51,02
NN-F	12607	0,417	82,63	61,45	50,78
SVM	13385	0,000	77,58	60,27	59,52
GB	10722	0,489	80,84	49,97	49,36
DT	11854	0,339	77,89	56,56	55,88
Credit to all	-19551	1	0	100	100

Cost=50, Return=1

Cost / Return	Profit	Optimal cutoff	Bankr correct-%	Non-br correct-%	Credit given/All-%
LR-ALL	19219	0,479	73,47	67,77	66,96
LR-W	18444	0,472	73,58	66,02	65,23
LR-F	16066	0,468	71,16	63,40	62,72
NN-ALL	19002	0,514	70,95	69,87	69,06
NN-W	18534	0,428	77,58	62,15	61,37
NN-F	16714	0,497	74,63	61,27	60,55
SVM	18182	0,000	73,05	65,99	65,22
GB	14362	0,489	80,84	49,97	49,36
DT	17308	0,419	67,89	69,35	68,61
Credit to all	-551	1	0	100	100

Cost=30, Return=1

Cost / Return	Profit	Optimal cutoff	Bankr correct-%	Non-br correct-%	Credit given/All-%
LR-ALL	25485	0,607	50,11	84,57	83,88
LR-W	24936	0,542	60,53	77,08	76,33
LR-F	22235	0,540	41,37	82,95	82,47
NN-ALL	25918	0,630	50,42	85,30	84,59
NN-W	25610	0,593	57,05	80,62	79,87
NN-F	22702	0,652	33,58	88,69	86,24
SVM	25465	0,000	52,95	82,80	82,09
GB	21408	0,497	60,63	69,50	68,90
DT	24761	0,468	58,21	78,11	77,39
Credit to all	18449	1	0	100	100

Cost=10, Return=1

Cost / Return	Profit	Optimal cutoff	Bankr correct-%	Non-br correct-%	Credit given/All-%
LR-ALL	37783	0,867	9,68	98,75	98,58
LR-W	37690	0,856	9,47	98,60	98,44
LR-F	37488	0,850	5,79	98,91	98,82
NN-ALL	37870	0,860	14,42	97,98	97,73
NN-W	37763	0,852	13,16	98,01	97,78
NN-F	37595	0,847	9,63	98,56	98,42
SVM	37735	0,000	10,74	98,44	98,25
GB	37449	0,531	0,00	100,00	100,00
DT	37655	0,705	14,53	97,50	97,26
Credit to all	37449	1	0	100	100

Cost=1, Return=1

Cost / Return	Profit	Optimal cutoff	Bankr correct-%	Non-br correct-%	Credit given/All-%
LR-ALL	46000	0,984	0,11	100,00	100,00
LR-W	46000	0,981	0,11	100,00	100,00
LR-F	45999	0,964	0,00	100,00	100,00
NN-ALL	46000	0,904	0,11	100,00	100,00
NN-W	46000	0,890	0,11	100,00	100,00
NN-F	45999	0,864	0,00	100,00	100,00
SVM	46000	0,000	0,11	100,00	100,00
GB	45999	0,530	0,00	100,00	100,00
DT	45999	0,883	0,00	100,00	100,00
Credit to all	45999	1	0	100	100

TABLES

Table 1. The classification accuracy of alternative methods as measured by AUC in the test data.

Panel 1. AUC in the test data.

Horizon	Year 1	Years 2-3	Years 4-5	Years 6-10	Average
LR-ALL	0,9340	0,8380	0,7810	0,7390	0,8230
LR-W	0,9330	0,8250	0,7680	0,7360	0,8155
LR-F	0,9240	0,8220	0,6480	0,7100	0,7760
NN-ALL	0,9400	0,8380	0,7750	0,7350	0,8220
NN-W	0,9320	0,8430	0,7800	0,7360	0,8228
NN-F	0,9220	0,8230	0,6480	0,7150	0,7770
SVM	0,9350	0,8220	0,7620	0,7360	0,8138
GB	0,8930	0,7590	0,7730	0,6810	0,7765
DT	0,7040	0,7110	0,6830	0,6950	0,6983
Average	0,9019	0,8090	0,7353	0,7203	0,7916

Panel 2. Ranks of methods.

Horizon	Year 1	Years 2-3	Years 4-5	Years 6-10	Sum of ranks	Final rank
LR-ALL	3	2	1	1	7	1
LR-W	4	4	5	2	15	4
LR-F	6	6	8	7	27	7
NN-ALL	1	2	3	5	11	3
NN-W	5	1	2	2	10	2
NN-F	7	5	8	6	26	6
SVM	2	6	6	2	16	5
GB	8	8	4	9	29	8
DT	9	9	7	8	33	9

Legend:
LR-ALL = Logistic regression with all variables
LR-W = Logistic regression with wrapper method
LR-F = Logistic regression with filter method (R-Square based variable selection)
NN-ALL = Neural network with all variables
NN-W = Neural network with wrapper method
NN-F = Neural network with filter method (R-Square based variable selection)
SVM = High performance support vector machine
GB = Gradient boosting
DT = Decision tree

Table 2. Profit of alternative methods as percentage of perfect credit decision profit.

Panel 1. Horizon 1 year.

Cost / Return	Cost=100, Return=1	Cost=70, Return=1	Cost=50, Return=1	Cost=30, Return=1	Cost=10, Return=1	Cost=1, Return=1
LR-ALL	79,02	83,23	86,55	90,28	95,32	99,29
LR-W	77,84	82,57	85,72	89,18	95,15	99,27
LR-F	76,02	81,13	84,73	88,56	94,96	99,26
NN-ALL	79,27	83,21	85,94	89,98	95,18	99,27
NN-W	79,73	83,69	86,51	90,14	95,06	99,29
NN-F	77,46	81,80	85,25	88,80	94,95	99,28
SVM	77,59	81,36	84,60	88,40	95,04	99,27
GB	69,11	74,20	79,23	84,89	93,39	99,25
DT	61,66	72,46	79,65	86,85	94,05	99,25
Credit to all	24,60	47,22	62,30	77,38	92,46	99,25

Panel 2. Horizon 2-3 years.

Cost / Return	Cost=100, Return=1	Cost=70, Return=1	Cost=50, Return=1	Cost=30, Return=1	Cost=10, Return=1	Cost=1, Return=1
LR-ALL	58,90	65,52	71,93	80,39	91,56	99,07
LR-W	56,50	63,90	70,96	79,78	91,54	99,07
LR-F	52,14	60,35	68,26	77,68	91,13	99,07
NN-ALL	61,22	67,29	73,09	80,27	91,54	99,07
NN-W	58,47	65,69	71,73	79,77	91,64	99,07
NN-F	52,03	60,74	68,66	78,04	91,28	99,07
SVM	55,85	64,74	70,73	79,20	91,54	99,07
GB	49,07	55,46	62,96	72,01	92,80	99,07
DT	52,93	62,65	69,12	78,25	92,88	99,07
Credit to all	6,71	34,70	53,35	72,01	90,67	99,07

Panel 3. Horizon 4-5 years

Cost / Return	Cost=100, Return=1	Cost=70, Return=1	Cost=50, Return=1	Cost=30, Return=1	Cost=10, Return=1	Cost=1, Return=1
LR-ALL	67,09	74,87	80,24	86,90	95,25	99,52
LR-W	66,47	73,88	80,06	87,15	95,34	99,52
LR-F	62,52	72,51	79,20	86,67	95,24	99,52
NN-ALL	63,88	72,73	79,58	86,97	95,21	99,52
NN-W	66,37	73,59	79,94	87,43	95,28	99,52
NN-F	62,91	72,51	79,20	86,65	95,21	99,52
SVM	64,58	72,76	79,32	86,65	95,23	99,52
GB	68,23	74,82	80,21	86,71	95,24	99,52
DT	66,95	72,81	78,47	85,62	95,21	99,52
Credit to all	52,08	66,45	76,04	85,62	95,21	99,52

Panel 4. Horizon 6-10 years.

Cost / Return	Cost=100, Return=1	Cost=70, Return=1	Cost=50, Return=1	Cost=30, Return=1	Cost=10, Return=1	Cost=1, Return=1
LR-ALL	22,78	30,97	40,94	54,28	80,48	97,98
LR-W	21,68	30,24	39,29	53,11	80,28	97,98
LR-F	17,65	24,56	34,22	47,36	79,85	97,98
NN-ALL	22,71	31,69	40,47	55,20	80,66	97,98
NN-W	22,44	31,03	39,48	54,55	80,43	97,98
NN-F	18,59	26,85	35,60	48,35	80,08	97,98
SVM	17,02	28,51	38,73	54,24	80,37	97,98
GB	12,64	22,84	30,59	45,60	79,77	98,89
DT	12,39	25,25	36,87	52,74	80,20	99,43
Credit to all	-102,35	-41,64	-1,17	39,30	79,77	97,98

Legend:
LR-ALL = Logistic regression with all variables
LR-W = Logistic regression with wrapper method
LR-F = Logistic regression with filter method (R-Square based variable selection)
NN-ALL = Neural network with all variables
NN-W = Neural network with wrapper method
NN-F = Neural network with filter method (R-Square based variable selection)
SVM = High performance support vector machine
GB= Gradient boosting
DT = Decision tree

Table 3. Profit of alternative methods as percentage of best method profit.

Panel 1. Horizon 1 year.

Cost / Return	Cost=100, Return=1	Cost=70, Return=1	Cost=50, Return=1	Cost=30, Return=1	Cost=10, Return=1	Cost=1, Return=1
LR-ALL	99,11	99,45	100,00	100,00	100,00	100,00
LR-W	97,63	98,66	99,05	98,78	99,83	99,98
LR-F	95,35	96,95	97,90	98,09	99,63	99,97
NN-ALL	99,43	99,43	99,30	99,66	99,85	99,98
NN-W	100,00	100,00	99,96	99,84	99,73	100,00
NN-F	97,15	97,75	98,51	98,36	99,61	99,99
SVM	97,32	97,22	97,75	97,91	99,71	99,98
GB	86,68	88,66	91,55	94,03	97,98	99,96
DT	77,33	86,58	92,04	96,20	98,67	99,95
Credit to all	30,85	56,42	71,98	85,71	97,00	99,95

Panel 2. Horizon 2-3 years.

Cost / Return	Cost=100, Return=1	Cost=70, Return=1	Cost=50, Return=1	Cost=30, Return=1	Cost=10, Return=1	Cost=1, Return=1
LR-ALL	96,20	97,37	98,41	100,00	98,58	100,00
LR-W	92,29	94,96	97,08	99,25	98,56	100,00
LR-F	85,16	89,68	93,39	96,63	98,12	100,00
NN-ALL	100,00	100,00	100,00	99,85	98,56	100,00
NN-W	95,50	97,61	98,14	99,23	98,67	100,00
NN-F	84,98	90,26	93,94	97,08	98,28	100,00
SVM	91,22	96,21	96,77	98,52	98,56	100,00
GB	80,14	82,42	86,14	89,58	99,92	100,00
DT	86,46	93,09	94,57	97,34	100,00	100,00
Credit to all	10,96	51,56	73,00	89,58	97,62	100,00

Panel 3. Horizon 4-5 years

Cost / Return	Cost=100, Return=1	Cost=70, Return=1	Cost=50, Return=1	Cost=30, Return=1	Cost=10, Return=1	Cost=1, Return=1

LR-ALL	98,33	100,00	100,00	99,39	99,90	100,00
LR-W	97,42	98,68	99,78	99,68	100,00	100,00
LR-F	91,64	96,85	98,71	99,13	99,89	100,00
NN-ALL	93,63	97,15	99,19	99,47	99,86	100,00
NN-W	97,27	98,29	99,63	100,00	99,93	100,00
NN-F	92,21	96,85	98,71	99,11	99,86	100,00
SVM	94,66	97,19	98,86	99,10	99,88	100,00
GB	100,00	99,93	99,97	99,17	99,89	100,00
DT	98,13	97,25	97,80	97,93	99,86	100,00
Credit to all	76,33	88,76	94,77	97,93	99,86	100,00

Panel 4. Horizon 6-10 years.

Cost / Return	Cost=100, Return=1	Cost=70, Return=1	Cost=50, Return=1	Cost=30, Return=1	Cost=10, Return=1	Cost=1, Return=1
LR-ALL	100,00	97,71	100,00	98,33	99,77	100,00
LR-W	95,18	95,42	95,97	96,21	99,52	100,00
LR-F	77,48	77,49	83,59	85,79	98,99	100,00
NN-ALL	99,70	100,00	98,87	100,00	100,00	100,00
NN-W	98,52	97,91	96,44	98,81	99,72	100,00
NN-F	81,61	84,73	86,97	87,59	99,27	100,00
SVM	74,74	89,96	94,60	98,25	99,64	100,00
GB	55,48	72,06	74,73	82,60	98,89	100,00
DT	54,39	79,67	90,06	95,54	99,43	100,00
Credit to all	-449,28	-131,40	-2,87	71,18	98,89	100,00

Legend:
LR-ALL = Logistic regression with all variables
LR-W = Logistic regression with wrapper method
LR-F = Logistic regression with filter method (R-Square based variable selection)
NN-ALL = Neural network with all variables
NN-W = Neural network with wrapper method
NN-F = Neural network with filter method (R-Square based variable selection)
SVM = High performance support vector machine
Boosting = Gradient boosting
Decision Tree = Decision tree

Table 4. Ranks of methods according to earned profits for cost-return ratio 30:1.

Horizon	Year 1	Years 2-3	Years 4-5	Years 6-10	Sum of ranks	Final rank
LR-ALL	1	1	4	3	9	1

LR-W	4	3	2	5	14	4
LR-F	6	8	6	8	28	7
NN-ALL	3	2	3	1	9	1
NN-W	2	4	1	2	9	1
NN-F	5	7	7	7	26	6
SVM	7	5	8	4	24	5
GB	9	9	5	9	32	9
DT	8	6	9	6	29	8