Int. J. Information and Decision Sciences, Vol. 10, No. 1, 2018 57

A new non-parametric classifier to predict small-business failures in Italy via performance ratios

Francesca di Donato and Luciano Nieddu*

Università degli Studi Internazionali di Roma – UNINT,
Via C. Colombo, 200, 00145, Rome, Italy
Email: francesca.didonato@unint.eu
Email: l.nieddu@unint.eu
*Corresponding author

Abstract: We considered the case of small-medium enterprises (SMEs) in Italy *introducing a new classifier to predict bankruptcy up to eight years prior to failure*. We considered a stratified random sample of 100 non-listed Italian SMEs, 50 of which filed for bankruptcy during the years 2000 to 2011. Results suggest that the proposed method more than holds its own when compared with standard non-parametric classification techniques. The performance of the proposed method based on recognition rate, sensitivity and specificity shows that the proposed technique is effective in predicting the failure of a firm up to eight years prior to the event. The high specificity makes the proposed technique very effective as a warning signal to determine if a firm is in distress with a sufficient enough time to take proper actions. The performance assessment has been achieved via cross-validation to get unbiased estimates of the performances.

Keywords: constrained *k*-means; bankruptcy prediction; discriminant analysis; performance ratios; small-medium enterprises; SMEs; Italy.

Reference to this paper should be made as follows: di Donato, F. and Nieddu, L. (2018) 'A new non-parametric classifier to predict small-business failures in Italy via performance ratios', *Int. J. Information and Decision Sciences*, Vol. 10, No. 1, pp.57–76.

Biographical notes: Francesca di Donato is a Tenured Assistant Professor of Accounting at UNINT University in Rome. She has received her PhD in Information Systems from LUISS University of Rome. Her research interests span from corporate governance to accounting for big and SMEs.

Luciano Nieddu is a Tenured Assistant Professor of Statistics at UNINT University in Rome. He has received his PhD in Operation Research from the University of Rome 'La Sapienza'. His research interests are in pattern recognition and discriminant analysis but he has also publications in statistical modelling and computational statistics.

This paper is a revised and expanded version of a paper entitled 'The effects of performance ratios in predicting corporate bankruptcy: the Italian case' presented at the ICDSST2015, Belgrade, Serbia, 27–29 May 2015.

1 Introduction

Accounting data have the ability to forecast events of interest to their users and, in the form of financial ratios, are commonly used in predicting companies' failures. The role of accounting ratios for predictive purposes (Barnes, 1987), in particular for predicting corporate bankruptcy, has a long history and an important body of accounting research has focused on the topic. Since Beaver's (1966) and Altman's (1968) pioneering works, many studies have been devoted to exploring the use of accounting information in forecasting business failures.

Upon Altman's model, many studies have been conducted in the USA examining the usefulness of models based on accounting information for the prediction of big corporate failure (Deakin, 1972; Blum, 1974; Ohlson, 1980), also in different industries as railroads (Altman, 1973), banks (Sinkey, 1975) and insurance sector (Trieschmann and Pinches, 1973; Pinches and Trieschmann, 1977).

Accounting ratios are indicators constructed from financial reporting information that firms have to file with public and tax authorities. According to several authors (see, e.g., Altman, 1968; Ohlson, 1980; Bhimani et al., 2013; Altman and Saunders, 1998; Allen et al., 2004), the failure of a limited company is related to two strictly connected situations:

- the inability to pay financial obligations when they come due (i.e., lack of liquidity, low solidity and very high leverage)
- the inability to generate operational profits or earnings before interests and taxes (i.e., negative or very low income and profitability).

Most of the initial studies on bankruptcy focused on large firms. In fact, before Storey et al.'s (1987) seminal contribution, only a few studies dealt with the failure of small-medium enterprises (SMEs) (Edmister, 1972; Argenti, 1976). Storey et al. (1987) drew his sample from the small firm sector, also using non-financial variables but without considering a control group of surviving firms. Hall (1994) studied the factors affecting small companies failure distinguishing between small firms that fail from those that survive but only considering the construction sector.

Hence, to our knowledge, there are few studies in the literature focusing on SMEs. For this kind of firms, it would be very helpful to have some tools to predict their failure in advance. In fact, SMEs are often under-capitalised and mostly rely on external financial sources provided by banks and other financial institutions, which then need prediction models to evaluate the risk of a loan loss. Moreover, the company itself could use this kind of analysis to predict its own failure and use this information to make management decisions.

The aim of this paper is to present a new classification algorithm to predict the possible failure of a SME with a sufficient time lag to allow for corrective actions. Following the approach of di Donato and Nieddu (2015), we will be comparing the performance of this new proposed method with that of very well established classification methods w.r.t. correct recognition rate, sensitivity and specificity.

This work is different from analogous works from different points of view: first of all, we introduce a new classification algorithm, which is non-parametric and data driven, therefore, no particular restrictive assumptions on the data are necessary and no design decisions must be made to build the classifier. This allows for replicability of the results and non-subjectiveness in the methodology. The classifier is based on a 'divide etimpera' criterion (Gershenfeld, 1997) and therefore will be compared with standard classifiers and with classification trees (Breiman et al., 1984; Frydman et al., 1985) which are the prototype of such a criterion in classification.

Our contribution has important implications both for regulators and SMEs that do not always keep their financial situation under control. It is also important for banks that using this method could avoid lending to firms that are likely to fail.

It must be stressed that in the literature there is a number of approaches to bankruptcy prediction assessment. Although it was pointed out by Jones (1987) that any methodology to predict bankruptcy should be tested on a hold-out sample of firms not used in the analysis, this suggestion has not been utterly followed in the literature. Any methodology tested on the dataset that it has been trained on, yields biased performance estimates since the methodology is obviously optimal for the dataset it has been trained on (see e.g., Nieddu and Patrizi, 2000 or McLachlan, 2004). Even when the methods have been tested on a hold-out sample the result is dependent on the particular sample that has been selected. In the following experimental setup, the methodologies will be tested using leave-one-out (Efron and Tibshirani, 1995, Nieddu and Patrizi, 2000), i.e., all the firms in turn will be tested.

The layout of the paper is as follows. In Section 2, a brief literature review will be proposed. In Section 3, the proposed classifier will be introduced, together with the other benchmark classification methods that will be used as comparisons. In Section 4, the experimental setup will be outlined and the results will be presented. Finally, in Section 5, some conclusions will be drawn.

2 Literature review

In the past, financial statement analysis was mainly used by credit suppliers to assess the credit worthiness of its borrowers. Today, this analysis is ubiquitous and involves a wide variety of ratios and users. First of all, with an accurate business failure analysis, banks and other financial institutions could avoid lending to firms that are likely to fail, and thus never repay their loans. The financial investment sector could improve the risk return trade-off from investments by not investing in failing businesses. Moreover, companies could establish long-term relationships with other companies (such as suppliers) that will not likely fail in the future, and thus increase the longevity and viability of their business relationships. Finally, regulators could make early identifications of failing business in order to legally handle the failure, preventing illegal activities, such as avoiding taxes or diluting debt (Gepp and Kumar, 2012).

The basic assumption in using accounting ratios is that the failure process is characterised by a systematic deterioration over time of such indicators (Laitinen, 1991). Sharma and Mahajan (1980) presented a general model of failure process according to which ineffective management together with unanticipated events lead to a systematic deterioration in performance indicators. In the absence of an effective corrective action, this situation leads to failure.

One relevant issue is the definition of 'failure' as used in literature. Many studies define failure as actual filing for bankruptcy or liquidation (Altman, 1968; Blum, 1974; Boardman et al., 1981; Castagna and Matolcsy, 1981). Others define it as suffering financial stress or an inability to pay financial obligations (Beaver, 1966; Chen and Lee, 1993). Some studies do not provide the definition of failure used for the research and some other authors use the term 'failed' interchangeably with 'bankrupted' (see Karels and Prakash, 1987). The existence of so many definitions is due to the difficulty to define objectively the start of the financial distress. McKee (2003) stated that a firm goes through various stages of financial distress before bankruptcy, as, for example, inadequate income, inadequate liquid asset position, and difficulties with paying the invoices and other contractual obligations. When financial distress cannot be relieved, it will lead to bankruptcy even if it is difficult to discern the precise moment that bankruptcy occurs. It appears to be a subjective decision in which financial failure persists until the firm or the creditor decides to file legal action.

Concerning the prediction models for bankruptcy, Beaver (1966) used only univariate statistics on US market data to determine the effect of financial ratios on the probability of bankruptcy, finding out a high predictive ability of financial ratios up to five years before failure. As noted by Altman (1968), although the univariate approaches establish certain important generalisations regarding the performance and trends of particular measurements, the adaption of their results for assessing potential bankruptcy of firms, both theoretically and practically, is questionable. The univariate nature focuses on individual signals of impending problems.

For this reason, Altman in 1968 applied linear discriminant analysis (LDA) (Fisher, 1936; McLachlan, 2004) and his approach was popularised as the Z-score model. In the Z-score model no cash flow ratios were found to be significant, which contrasts the focus on cash flow ratios by Beaver. Altman's LDA outperformed Beaver's univariate model for one year prediction intervals. He went back up to five years prior to failure but the results deteriorated already at three years yielding results below 50%. Before these seminal works, the use of discriminant analysis in economic research was illustrated in studies by Tintner (1946), Gales (1965), and Brown (1967).

According to Altman's model, many studies have been conducted in the USA (Deakin, 1972; Blum, 1974; Ohlson, 1980), also in different sectors as railroads (Altman, 1973), banks (Sinkey, 1975) and insurance (Trieschmann and Pinches, 1973; Pinches and Trieschmann, 1977).

The interest in this approach has spread outside the USA in the late '70s. In 1977, Taffler (Taffler and Tisshaw, 1977; Taffler, 1982, 1983) developed a UK-based Z-score model for analysing the financial health of firms listed on the London Stock Exchange. Other studies were drawn for French (Altman et al., 1974) and Australian companies (Castagna and Matolcsy, 1981).

Altman and Saunders (1998) and Allen et al. (2004) reviewed the vast literature on the influence of accounting indicators on corporate distress (bankruptcy and default) in detail. These reviews identify the predominant use of discriminant analysis and logistic models in corporate distress prediction and the influence of several ratios.

For the failure prediction method to be effective and worth using, it is important to consider how far ahead in time the model is able to accurately predict bankruptcy. Many studies have a high accuracy rate one year prior to failure, because ratios from the year immediately preceding it are generally very representative of a firm's deteriorating condition. Unfortunately, it could be argued that one year prior to failure the health of the firm could be so far compromised that any action to recover could be useless.

Some models are able to predict bankruptcy much sooner. For example, Deakin's (1972) model could predict bankruptcy with 96% accuracy two years prior to the failure.

Nevertheless, prediction of failure only two years in advance was judged, by an expert banking panel, not to be sufficiently timely, in general, to allow a lending institution to extricate itself without incurring the risk of a significant loan loss (Casey, 1980).

In general, the farther back in time we go the less the accuracy level of the model is and therefore its usefulness. Blum's (1974) model, for example, is able to predict bankruptcy six years prior to failure but with only 57% accuracy. Laitinen (1991) did the same with around 60% accuracy six years before failure. Those accuracies are very close to the random recognition rate that could be obtained assigning each firm randomly to the failed or sound firms. In 2015, di Donato and Nieddu applied LDA, quadratic discriminant analysis (QDA) and CART to a sample of 100 SMEs from the Italian public industry, being able to obtain a recognition rate around 65% for eight years prior to failure using LDA.

3 The algorithms

In a standard classification problem, a set of measurements are available on a set of objects. Therefore, for each object *i*, a pattern vector (p.v.) $\boldsymbol{x}_i$ (usually $\boldsymbol{x}_i \in \mathbb{R}^p$) of measurements is available. The aim is to map those measurements into one of K unordered and mutually exclusive classes. The mapping is obtained via function $\psi(\cdot)$ which is called 'classifier' and maps from the pattern space to the class space.

The aim of any classification algorithm is then to determine an estimate of $\psi(\cdot)$, based on the data at hand (training set). If the training set, on whom the classifier is estimated, is composed of objects of known classes, i.e., an instance of the function $\psi(\cdot)$ is available, then the problem is known as 'supervised classification problem', otherwise it is an unsupervised classification problem that is usually tackled via clustering or finite mixture models. We will be dealing with supervised classification problems, i.e., a dataset of previously classified objects is available and the classifier must be trained on that dataset.

Once the classifier has been determined, its performance must be assessed. It goes without saying that any classification rule is optimal for the training set it has been estimated on. Therefore to get a reliable non-biased estimate of the performance each classifier must be tested on a test set which is independent of the training set (cross-validation).

In 1996, Wolpert proved that there is no universally optimal classification algorithm and that each problem can have its tailor-made classifier that works best on that particular data. The classifier that we are introducing in this paper is an optimal subset detection classifier, i.e., it partitions the pattern space into homogenous and mutually exclusive regions, determining in such way prototypes of subclasses. The idea is closely related to cluster weighted modelling (Gershenfeld, 1997). Most classification algorithms assume that the data or a transformation of the data could be linearly separated. Our method works by assuming that it is true locally, in each region of the pattern space.

In a problem like that of predicting bankruptcy, it is fairly acceptable to imagine that failing firms must have some common features that tend to group the firms together and that those features are reflected in their performance ratios. It goes without saying that not all performance ratios must behave in the same way: some firms may fail for instance because they do not repay their loans and some others may fail because they have very high running costs not matched by equal revenues. Either of these behaviours can lead to bankruptcy even if they reflect two different situations. Therefore, a method that is based on subclass detection should be appropriate in predicting failure of companies using prototypes of previously failed companies, where these prototypes should represents the various behaviours of sound and failing companies. In the following, the proposed method will be introduced more formally.

3.1 The proposed method

Given a dataset of n p.v.s in $\mathbb{R}^p$, let us assume a partition defined on the dataset, i.e., each p.v. is assigned to one and only one of K known classes. Let us assume a Euclidean norm defined on the dataset and let $\psi(\cdot)$ be a function from $\mathbb{R}^p$ onto the set $C = \{1, 2, \ldots, K\}$ which maps each p.v. $\boldsymbol{x}_j$, $j = 1, \ldots, n$, into the class $c \in C$ that it belongs to, i.e., this function (classifier) assigns each object to one of mutually exclusive classes on the basis of a set of measurements (p.v.) on the object.

The proposed algorithm works as follows:

1 INPUT: p.v.s in the training set $\boldsymbol{x}_j$, $j = 1, \ldots, n$.

a function $\psi(\boldsymbol{x}_j)$ that assigns to each p.v. in the training set its class

initial set of K barycentres $\boldsymbol{B}_0$, one barycentre for each class

OUTPUT: a set of final barycentres $\boldsymbol{B}_t$ ($\#\boldsymbol{B}_t \geq \#\, \boldsymbol{B}_0$).

2 Compute the distance of each p.v. in the training set from each barycentre in $\boldsymbol{B}_0$.

3 If each vector is closer to the barycentre of its class the algorithm stops, otherwise there will be a non-empty set $\mathcal{M}$ of p.v.s which belong to a class and are closer to a barycentre of a different class. These objects, although they belong to a class, are closer to a barycentre of a class different from the one they belong to: in a minimum distance classification criterion they would be misclassified. Therefore, in the set $\mathcal{M}$ select the p.v. $\boldsymbol{x}_w$ that is farthest from the barycentre of its class, i.e., that one that is the most different from the elements of its own class. This p.v. will be used as a seed for a new barycentre for class $\psi(\boldsymbol{x}_w)$.

4 A *k*-means algorithm (Hastie et al., 2001) will then be performed for all the p.v.s in class $\psi(\boldsymbol{x}_w)$ using, as starting points, the set of barycentres for class $\psi(\boldsymbol{x}_w)$ and the vector $\boldsymbol{x}_w$.

5 Once the *k*-means has been performed, the set of barycentres for class $\psi(\boldsymbol{x}_w)$ will be composed of $k + 1$ barycentres determined by the *k*-means algorithm, where k is the number of barycentres in class $\psi(\boldsymbol{x}_w)$ to the *k*-means iteration. Let $\boldsymbol{B}_t$ be the union of the sets of barycentres for classes different from $\psi(\boldsymbol{x}_w)$ and the new set of $k + 1$ barycentres for class $\psi(\boldsymbol{x}_w)$. Note that the barycentres at the new iterations need not be computed for all classes, but only for class $\psi(\boldsymbol{x}_w)$ since the barycentres for the other classes have remained unchanged.

6 Determine the distances of each p.v. in the training set from all the barycentres in $\boldsymbol{B}_t$, and update the set $\mathcal{M}$.

7 If $\mathcal{M}$ is not empty then the p.v. $\boldsymbol{x}_w$ in $\mathcal{M}$ which is farthest from a barycentre of its own class is once again selected to serve as a seed for a new barycentre. Go to Step 4 and iterate the procedure until the set $\mathcal{M}$ is empty.

Upon convergence, the algorithm yields a set of barycentres which, in the worst case, are in a number equal to the number of elements in the training set and which has a lower bound in the number of classes. It is worth noticing that if the partition defined on the dataset is consistent with the features considered, i.e., if the p.v.s are linearly separable, then the algorithm generates a number of barycentres equal to the number of classes. On the other hand, if the classes in the dataset are not linearly separable, then the algorithm continues splitting the classes until the subclasses obtained are linearly separable. It is obvious that it can continue splitting until all the subclasses are composed of only one vector (singleton).

Therefore, the aim of this algorithm is to find subclasses in the dataset which can be used to classify new vectors of unknown class.

Definition: A training set is *coherent* if $\boldsymbol{x}_s = \boldsymbol{x}_j \Rightarrow \psi(\boldsymbol{x}_s) = \psi(\boldsymbol{x}_j)$.

I.e., in a coherent training set, if two objects have the same p.v.s they must belong to the same class.

If the proposed method is applied on a non-coherent training set it will not converge. In this case there will always be an object in the set $\mathcal{M}$ and the algorithm will keep generating barycentres. Such a problem can be caused by to different sets of motives:

- The object *s* and *j* are actually different and belong to different classes, therefore there must be some measurement that can be done on them in order to get $\boldsymbol{x}_s \neq \mathbf{x}_j$, i.e., by increasing the dimension of the pattern space such anomaly should be fixed.
- The objects *s* and *j* indeed indistinguishable according to any measurements that can be carried out. It could then be argued that they are the same object, therefore they belong to the same class and the original classification is wrong.

Either way the problem can be fixed and it can be assumed that the training set is coherent.

Upon convergence, the sets of barycentres can be considered as prototypes for the subclasses they represent and can be used to classify new objects (*query points*) assigning the new element to the class of the barycentre it is closest to. Ties can be broken randomly.

Proposition: if the barycentres are used to classify the elements in the training set then the classification error is zero.

If elements from the training set are used as query points, then the algorithm always classify them correctly in a minimum distance framework, because, once converged, all p.v.s in the training set are closer to a barycentre of their own class. This can be also stated saying that the *apparent error rate* (Nieddu and Patrizi, 2000; McLachlan, 2004) of the proposed classifier is equal to zero.

The algorithm can be generalised allowing for impurity in the data, i.e., the recursive partitioning of the feature space can be performed until the percentage of elements that are closer to a barycentre of another class has decreased under a certain threshold which can be set to a value different from zero. This can be easily achieved allowing for the set $\mathcal{M}$ to be non-empty upon convergence. This can be helpful when the training set has been classified with error (imperfect supervisor): in this case allowing for impurity in the subclasses can prevent the algorithm from overfitting the data.

To evaluate the predictive ability of the proposed method, three among the most used non-parametric classification techniques in the sector have been applied on the same data, namely LDA, QDA and classification trees (CART). Since 1966 until recently, LDA was the dominant method in failure prediction. To make the paper self-contained, a very brief overview of the theory of these classification algorithms follows.

3.2 Benchmark classification algorithms

3.2.1 LDA and QDA

The problem that is posed in LDA is to determine which linear combination **wx** (projection) of the *p* variables of a p.v. **x** yields the best separation of the classes in the projected space (i.e., separates the sample of failed firms from that of healthy ones for the problem we are considering). The best projection is usually evaluated with respect to a criterion involving some ratio of the within-groups to between-groups variances.

The solution outlined by Fisher (1936) consists in determining a projection such that the ratio of the difference between sample means to the within-groups standard deviation is maximised, i.e.:

$$\arg\max_{w} \frac{w^t S_b w}{w^T S_w w} \tag{1}$$

The solutions to this problem can be found in the eigenvectors of the matrix $S_w^{-1} S_b$ (at most $j - 1$ non-zero eigenvectors).

The original approach due to Fisher can be considered a dimensionality reduction method that preserves as much of the class separation as possible. Once the points have been projected onto the new space, a minimum distance criterion can be used to classify new points in one of the known classes. Although Fisher did not make any parametric assumption on the distribution of the p.v.s, his approach can be proven to optimal in the case of multivariate Gaussian data.

Namely, if the data have been drawn from multivariate normal populations with equal covariance matrices, the solution to the maximisation problem (1) also provides the best Bayesian classifier (see Nieddu and Patrizi, 2000; McLachlan, 2004) where units are assigned to the class k that maximises the posterior probability, i.e.:

$$\arg\max_{k} \left\{ -\frac{1}{2} \ln|\Sigma_k| - \frac{1}{2} (\mathbf{x} - \mu_k)^t \, \Sigma_k^{-1} (\mathbf{x} - \mu_k) + \ln \pi_k \right\}$$

where Σ_k is the covariance matrix for class *k* and π_k is the prior probability of vector **x** to belong to class *k*. If the data are drawn from homoscedastic populations (equal covariance matrices Σ_k for all classes) this rule reduces to a *linear discriminant rule* which is equivalent to the LDA classifier. When the homoscedasticity assumption is violated (Smith, 1946) a quadratic classifier can be estimated (QDA).

Since LDA can be considered a special case of QDA (i.e., quadratic classifier with equal covariance matrices for every class), it goes without saying that the latter allows for more flexibility for the covariance matrix and therefore tends to better fit the data. This increased flexibility is paid in term of a larger number of parameters to estimate since a separate covariance matrix for each class must be estimated from the data. This could be a problem when the number of elements in the training set is not large as it will be the case in the application we present in this paper.

3.2.2 CART

Classification and regression trees (CART) (Breiman et al., 1984) are based on the concept of cluster-weighted modelling (Gershenfeld, 1997) where a separate model can be fit for a subset of objects in the data (clusters). The idea is to partition the pattern space into a disjoint set of orthogonal regions and to fit a very simple model in each subregion. CART constitute an alternative approach to LDA and QDA especially in presence of missing data or when the number of features is too large when compared to the number of cases in the dataset.

Classification and regression trees have been popularised by Breiman et al. (1984); in both classification and regression trees the value of an outcome variable must be predicted according to the values assumed by a set of predictors. If the outcome is continuous then we have regression trees, otherwise the outcome is considered categorical and the method is known as classification trees.

The basic idea for CART is to partition the feature space in non-overlapping regions and then fit a constant model in each region for the outcome variable. The partitioning is obtained using one covariate at a time, the one that yields the best separation for the data at hand, and then recursively partitioning the subsets. The recursive partitioning is carried out until the number of objects in the subsets has decreased below a certain threshold or until an optimality criterion, usually based on the impurity on the subsets, is met. The problem of detecting the best partition on the pattern space that optimises a measure of impurity on the subsets is NP-hard. Various heuristics exit to get a local optimal solution to the problem: usually the greedy algorithm is used in such cases.

4 Experimental results

4.1 The data

The performances of the proposed technique will be tested on a random sample of 100 non-listed SMEs, including 50 companies that filed for bankruptcy during the years 2000 to 2011 and 50 companies still operating at the end of 2011. Therefore, for each company, in the sample more than one financial statement is available: for those still active at the end of the observational period a statement for each year was available, while for those that failed during the follow up only the statements up to the time of failure were available. The sample was randomly selected from the firms operating in Italy. We considered only companies with a turnover from 2 million to 50 million euros at the beginning of the analysed period. As in Abdullah et al. (2008), companies that were classified as financial and property industries were not considered in the analysis since their ratios are highly volatile. Besides, the interpretation of the ratios is slightly different since financial companies, for example, have different nature of income and expenses from non-financial companies.

The data used in this work were collected through CERVED (http://www.cerved.com) database, related to economic and financial data of Italian non-listed companies. Using the available financial statements for each of the firm in the dataset, we have computed the most common ratios for every year in the period 2000 to 2011. According to Barnes (1987), we selected the ratios throughout the criterion of popularity, meaning their frequency of appearance in the literature (Bellovary et al., 2007).

Moreover, we grouped the selected ratios w.r.t. the dimension of the company they refer to (economic or financial) and for this reason they have been classified into two groups:

- Profitability ratios alone, related to the economic dimension of the company:
  1 return on equity
  2 capital turnover
  3 net income/total assets
  4 return on investment
  5 earning/sales
  6 return on sales
  7 financial interests/Ebitda
  8 financial interest/sales
- Leverage and liquidity ratios alone, related to the financial dimension of the company:
  1 financial debts/equity
  2 short-term bank loan/working capital
  3 cash flow/total debt
  4 structure ratio 1
  5 structure ratio 2
  6 working capital/total assets
  7 quick ratio
  8 working capital cycle
  9 financial debt/working capital
  10 current ratio
  11 retained earnings/total assets.

Therefore, for each year, each firm is represented as a p.v. in $\mathbb{R}^8$ w.r.t. the profitability ratios and as a vector in $\mathbb{R}^{11}$ w.r.t. the financial ratios.

Although the data is longitudinal in nature, we have conducted a cross-sectional study: all the failed companies have been considered at various years prior to failure (up to eight years). Each distressed company was randomly matched with a healthy company belonging to the same industry sector and had the closest total assets. These criteria were set as control factors to ensure minimum bias in the selection of the control sample used in the estimation of the classifier. In Table 1, the number of firms available at each time lag has been displayed.

Table 1 Number of firms available at each time lag prior to failure

Time lag (in years)	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
Number of firms	100	100	100	98	92	66	46	34

Since the matching of failed firms with healthy ones was carried out randomly, the selection mechanism has been repeated 300 times for each year lag to get an average estimate of the performance of each prediction technique. Failed firms have been considered from one up to eight years prior to failure. It goes without saying that the sample size of failed firms decreased as the number of years prior to failure increased.

Bellovary et al. (2007) make a detailed review on similar studies mostly considering data up to six years prior to failure.

4.2 Performance assessment

Once a classifier has been trained on the available training set, its performance on an independent test set must be determined.

The performances of the proposed model will be compared with those of very well-known classifiers with respect to correct recognition rate, sensitivity and specificity. Correct recognition rate is an overall measure of the ability of the classifier to correctly classify new objects while sensitivity and specificity are related to the ability of the classifier to correctly classify failed firms or to correctly classify non-failed firms. They are related to type II and type I statistical errors:

- *Correct recognition rate*: is the probability that the classifier correctly assigns a class to a query point.
- *Sensitivity*: is the statistical power of the test and is related to the type II error. It is the probability of correctly identifying a firm that has failed.

  Sensitivity alone, without considering correct recognition rate, cannot be used to evaluate the performance of a classifier since sensitivity does not take into account false-positives (i.e., sound firms that have been classified as failed).
- *Specificity*: is related to the type I error of a statistical test. It is the probability of correctly classifying a firm that has not failed. Again, specificity alone cannot be used to evaluate the performance of a classifier without considering correct recognition rate. If a firm tests positive (failed) to a highly specific test than it would have a great probability of being a failed firm therefore a highly specific classifier with a good correct recognition rate could be useful as a warning signal.

Unbiased estimates (McLachlan, 2004) of these quantities can be obtained via cross-validation. A standard approach usually considers a *k*-fold cross-validation scheme (Hastie et al., 2001) where the performance of a classifier is then an average of the performances obtained on the *k* test sets. This technique allows for all the elements in the dataset to be used at least once in testing.

A special version of the *k*-fold cross-validation is the leave one out scheme (LOO), where in turn each element of the sample is singled out to be tested on the classifier trained on the remaining $n - 1$ elements. The routine is repeated *n* times and the performance of the classifier is then a synthesis of the outcomes on each unit. LOO is particularly useful when the dataset is not large and therefore as much as possible elements of the dataset should be used in training.

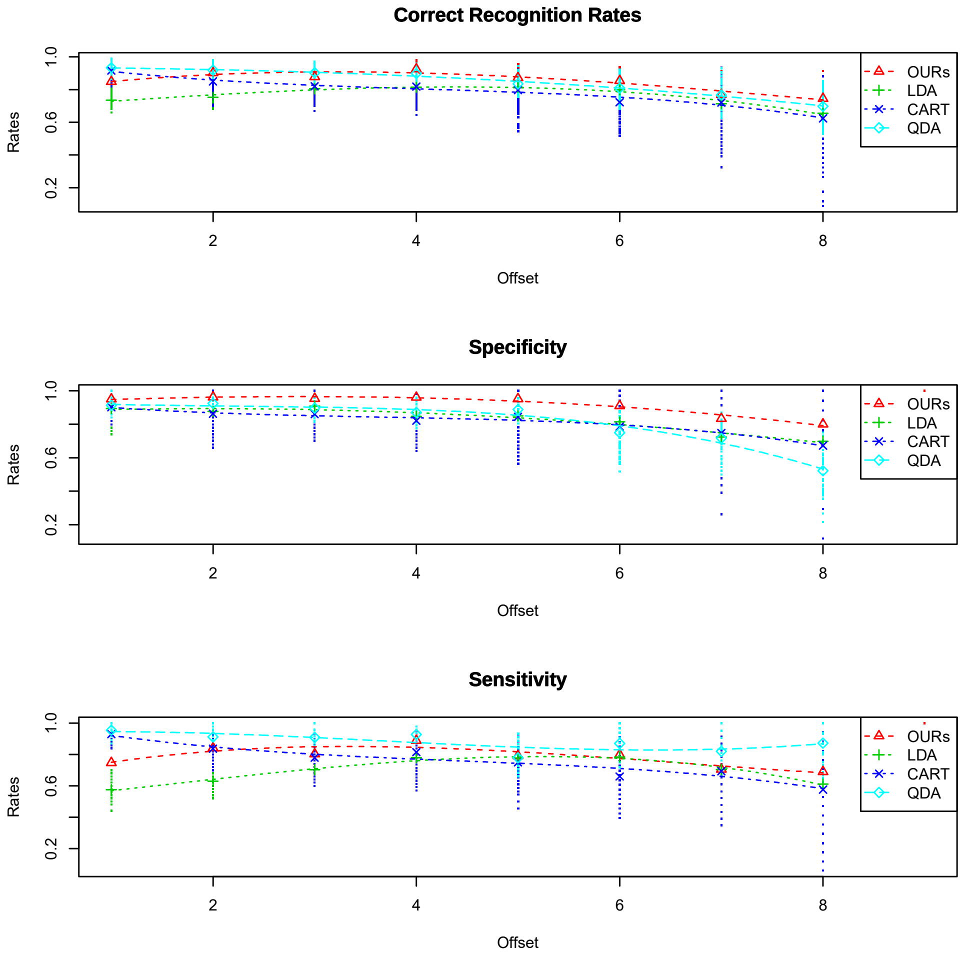
The use of an independent test sample to assess the performance of the classifier is not so frequent in the literature related to bankruptcy prediction as it should be. Although in the specialised literature, it was suggested the need of an independent sample to test the classifier (see Jones, 1987), several papers have been published that continued testing the performance of various classifiers using only resubstitution error, i.e., the error rate that is obtained applied the classifier on the data it has been trained upon. For instance, the apparent error rate of the classifier we propose in this paper is zero, but that does not imply that the classifier we have introduced will perform perfectly on new query points. Bellovary et al. (2007) have considered the problem of assessment of bankruptcy prediction methods and they report that, from 1987 to 2007, out of 90 papers on the topic, almost 45% still did not use hold-out sample or cross-validation to test the performance of the classification algorithm yielding results based on the training set. To obtain an unbiased estimate of how the classifier will likely perform on new query points a hold out sample or cross-validation approach should be used. Cross-validation provides a nearly unbiased estimate (Efron and Tibshirani, 1995) of the future error rate. With LOO, this small bias is further reduced. One appreciable feature of LOO or *k*-fold cross-validation over hold-out-sample (see e.g., Nieddu and Patrizi, 2000) is that, with the formers, all the elements of the dataset will in turn be tested, avoiding any subjectivity in the choice of the hold-out sample.

In our experiments, we will be using LOO, since the sample size of failed firms decreases as the time lag before failure increases (Table 1). Correct recognition rate, sensitivity and specificity have then been calculated. Since the matching between the failed firms and the healthy ones was carried out randomly 300 times, the performances that will be reported are averages over 300 trials.

4.3 Results

In Table 2, the results of the performances using financial ratios have been depicted. In the table, the average correct recognition rates, specificities and sensitivities and their standard deviations (std.) over 300 trials for the proposed classifier ('our method' in Table 2) and for the three considered standard methods (LDA, CART and QDA) have been displayed. Results have been detailed by years prior to failure ('offset').

Figure 1 Trends for the correct recognition rates, sensitivity and specificity for the proposed method and the comparing classifiers obtained using financial ratios (see online version for colours)



In Figure 1, the average recognition rates, sensitivity and specificities have been plotted vs. time lag. For each year prior to failure, the recognition rates and their averages have been displayed together with a spline interpolating the trend of the recognition rates, sensitivities and specificities. As it was to be expected, in general, the performances of all the considered methods decrease as the time lag increases and the corresponding standard deviations increase. Up to three years prior to failure, the best performance in terms of correct recognition rates is obtained using QDA. LDA gives the worst performance over such a short period of time prior to failure. CART and the proposed method tend to perform fairly well although not as well as QDA. With respect to specificities QDA shows the best performance over all possible time lags (except seven years prior to failure for whom the best performance is obtained by the proposed method).

The proposed method shows the best specificities over all time lag prior to failure, therefore it can be efficiently used as a warning signal to determine if a company is in distress. The proposed method is the one showing the best recognition rates and the best specificities for time lags from four to eight years prior to failure. It also shows the best sensitivity rate five years prior to failure. QDA still performs fairly well, followed by LDA and CART. With increasing time lag, the variability of the results increases and this can be partially due to a decreased sample size (there are only 34 firms for whom we have data eight years prior to failure) that increases the standard errors of the estimates. On this account, the increase of the variability of the recognition rate for CART is the greatest observed (from 0.026 to 0.157), while the other methods seem to be a bit more robust as the time lag increases.

In Table 2, we have highlighted in bold the best performances for each combination of time-lag and classification algorithm. From the results, it seems that QDA represents the best prediction algorithm in the immediate proximity of a failure and it performs fairly well especially when it comes to detecting firms that will fail. In the long run, the best prediction method is the one proposed in this paper, which shows the best correct recognition rate, and the lowest type I error. More specifically, the proposed method is very reliable to assess the solidity of the firm, since it has the best specificity, i.e., if the proposed algorithm classifies a company as failed it will probably fail in the future.

In Table 3, the average correct recognition rates, specificities and sensitivities over the 300 trials for the proposed method and the standard classification techniques, using only profitability ratios, have been displayed together with their standard deviations. In Figure 2, the corresponding plots for each time lag prior to failure have been displayed. In the plot, both the averages for correct recognition rates, specificity and sensitivity together with the distribution of the performance over the 300 trials have been displayed. Splines have been added to give an idea of the trend over time. With respect to correct recognition rates, the best performance is obtained using the proposed method over all time lags. QDA, which for financial indicator performed pretty well up to three years prior to failure, has dropped to a recognition rate around 60%. CART more then hold their own, still performing quite well and so does LDA.

The best specificity is obtained using the proposed method for large time lag (from five to eight years) and QDA for time lags lower than 5. It must be stressed that QDA is the method that shows the worst power, yielding very low sensitivity results up to five years prior to failure. Therefore, QDA on profitability ratios performs very badly when trying to correctly classify companies that have failed.

The best recognition rates are still those obtained using financial ratios up to three years with QDA and then the proposed method from four to eight years.

It is worth noticing that the performance obtained using profitability information is almost always worse than the corresponding obtained using financial ratios, implying that the financial information they carry is more informative to determine if a firm is going to survive or fail.

Table 2 Average correct recognition rates, sensitivities and specificities and corresponding standard deviations over 300 replications obtained using financial ratios

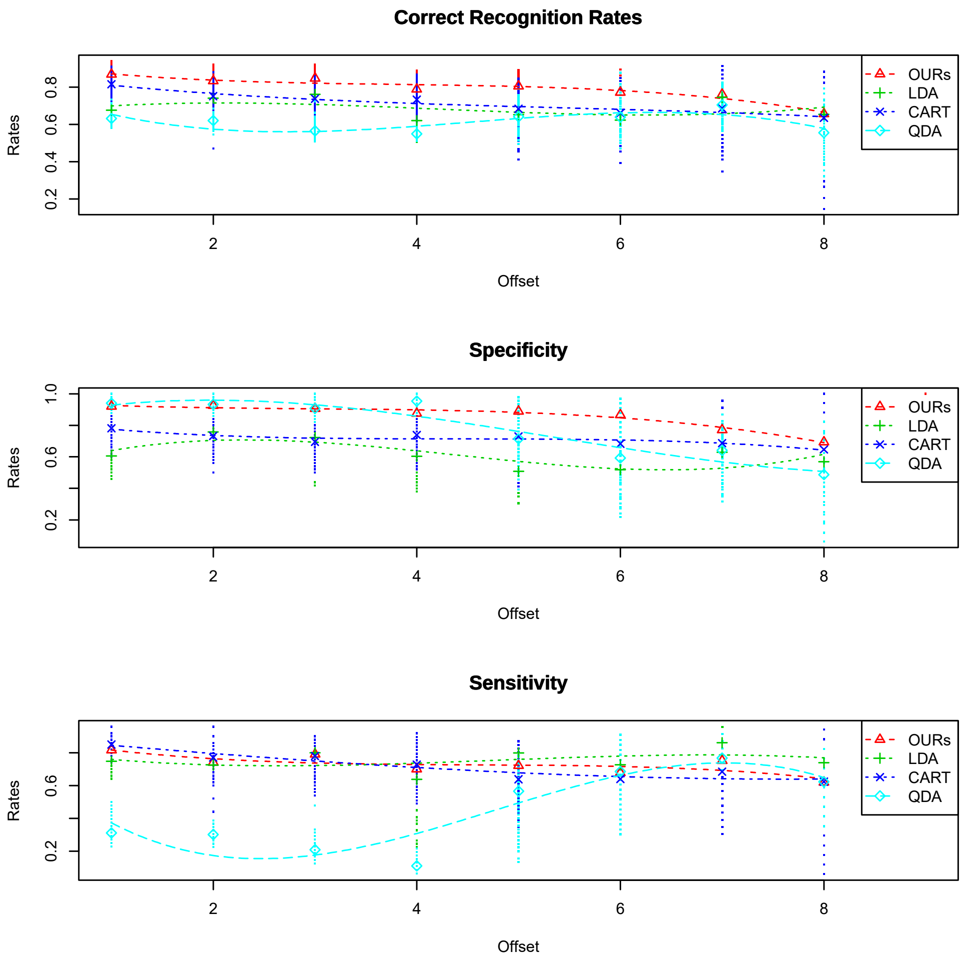
FIN	*Offset*	*Our method*		*LDA*		*CART*		*QDA*	
	(Years)	*Average*	*Std.*	*Average*	*Std.*	*Average*	*Std.*	*Average*	*Std.*
Correct recognition	1	0.8486	0.0335	0.7345	0.0283	0.9159	0.0258	*0.9330*	0.0207
	2	0.9023	0.0236	0.7533	0.0283	0.8500	0.0535	*0.9187*	0.0244
	3	0.8784	0.0296	0.8056	0.0263	0.8200	0.0482	*0.9027*	0.0234
	4	*0.9252*	0.0189	0.8141	0.0307	0.8181	0.0578	0.8983	0.0214
	5	*0.8717*	0.0298	0.8136	0.0311	0.7972	0.0680	0.8345	0.0336
	6	*0.8548*	0.0347	0.7987	0.0402	0.7228	0.0711	0.8096	0.0439
	7	*0.7694*	0.0561	0.7140	0.0590	0.7206	0.1113	0.7666	0.0601
	8	*0.7456*	0.0609	0.6549	0.0755	0.6243	0.1568	0.6974	0.0722
Sensitivity	1	0.7468	0.0558	0.5749	0.0397	0.9295	0.0233	*0.9546*	0.0271
	2	0.8423	0.0358	0.6301	0.0299	0.8367	0.0583	*0.9120*	0.0403
	3	0.8035	0.0476	0.7027	0.0272	0.7793	0.0671	*0.9087*	0.0301
	4	0.8888	0.0261	0.7799	0.0317	0.8148	0.0814	*0.9279*	0.0185
	5	0.7916	0.0479	0.7772	0.0328	0.7502	0.0757	*0.7791*	0.0572
	6	0.7991	0.0465	0.7833	0.0413	0.6597	0.0964	*0.8702*	0.0508
	7	0.7052	0.0788	0.7036	0.0651	0.6932	0.1368	*0.8199*	0.0748
	8	0.6902	0.0880	0.6112	0.0722	0.5761	0.2053	*0.8724*	0.1048
Specificity	1	*0.9503*	0.0272	0.8941	0.0485	0.9022	0.0417	0.9132	0.0267
	2	*0.9623*	0.0236	0.8765	0.0422	0.8633	0.0645	0.9249	0.0260
	3	*0.9533*	0.0283	0.9085	0.0409	0.8607	0.0586	0.8968	0.0297
	4	*0.9609*	0.0243	0.8477	0.0419	0.8213	0.0642	0.8704	0.0341
	5	*0.9517*	0.0293	0.8499	0.0587	0.8443	0.0910	0.8878	0.0325
	6	*0.9105*	0.0423	0.8140	0.0785	0.7859	0.0935	0.7499	0.0745
	7	*0.8337*	0.0642	0.7243	0.1009	0.7480	0.1217	0.7166	0.0747
	8	*0.8010*	0.0783	0.6986	0.1070	0.6725	0.1572	0.5221	0.0799

Table 3 Average correct recognition rates, sensitivities and specificities and corresponding standard deviations over 300 replications obtained using profitability ratios

PROF	*Offset*	*Our method*		*LDA*		*CART*		*QDA*	
	(years)	*Average*	*Std.*	*Average*	*Std.*	*Average*	*Std.*	*Average*	*Std.*
Correct recognition	1	*0.8691*	0.0264	0.6768	0.0356	0.8157	0.0377	0.6320	0.0235
	2	*0.8339*	0.0276	0.7416	0.0408	0.7516	0.0545	0.6206	0.0204
	3	*0.8482*	0.0301	0.7619	0.0397	0.7380	0.0582	0.5654	0.0275
	4	*0.7893*	0.0377	0.6208	0.0486	0.7336	0.0641	0.5494	0.0147
	5	*0.8056*	0.0335	0.6533	0.0408	0.6841	0.0708	0.6414	0.0460
	6	*0.7722*	0.0442	0.6231	0.0389	0.6616	0.0827	0.6394	0.0591
	7	*0.7614*	0.0530	0.7453	0.0503	0.6860	0.1028	0.7045	0.0560
	8	*0.6569*	0.0717	0.6541	0.0832	0.6355	0.1367	0.5556	0.0816
Sensitivity	1	0.8175	0.0350	0.7475	0.0379	*0.8504*	0.0292	0.3109	0.0408
	2	0.7421	0.0461	0.7252	0.0500	*0.7707*	0.0663	0.3016	0.0274
	3	0.7907	0.0368	*0.8007*	0.0347	0.7806	0.0784	0.2088	0.0407
	4	0.7007	0.0499	0.6380	0.0789	*0.7281*	0.0904	0.1100	0.0168
	5	0.7219	0.0473	*0.7988*	0.0361	0.6383	0.0825	0.5661	0.1653
	6	0.6783	0.0628	*0.7275*	0.0251	0.6401	0.1037	0.6857	0.1237
	7	0.7518	0.0675	*0.8614*	0.0409	0.6851	0.1156	0.7677	0.0489
	8	0.6201	0.0774	*0.7394*	0.0940	0.6239	0.1613	0.6227	0.0606
Specificity	1	0.9207	0.0324	0.6061	0.0638	0.7811	0.0646	*0.9403*	0.0238
	2	0.9256	0.0302	0.7579	0.0487	0.7325	0.0677	*0.9333*	0.0293
	3	0.9056	0.0359	0.7231	0.0754	0.6955	0.0754	*0.9078*	0.0346
	4	0.8759	0.0425	0.6039	0.0964	0.7391	0.0863	*0.9537*	0.0242
	5	*0.8893*	0.0382	0.5078	0.0741	0.7298	0.0941	0.7151	0.1390
	6	*0.8662*	0.0526	0.5187	0.0696	0.6830	0.1038	0.5930	0.1705
	7	*0.7710*	0.0740	0.6291	0.0857	0.6870	0.1319	0.6404	0.1092
	8	*0.6936*	0.1070	0.5688	0.1168	0.6471	0.1799	0.4875	0.1397

The proposed method has very good recognition rates on financial ratios from four to eight years prior to failure, performing better than LDA, CART and QDA, showing very high specificity. From one to three years prior to failure, although it performs pretty well, QDA performs better, showing higher recognition rates and higher sensitivity, but lower specificity. Considering the profitability dimension of the firm, the proposed method shows robustness in the results, presenting very good recognition rates and high specificities while QDA drops its performance w.r.t. recognition rates and sensitivity, while showing good specificities in the short run.

Figure 2 Trends for the correct recognition rates, sensitivity and specificity for the proposed method and the comparing classifiers obtained using profitability ratios (see online version for colours)



5 Conclusions

In this paper, we have proposed a new classification algorithm to predict bankruptcy of a SME based on its financial and profitability performances. To test the proposed method, we have performed a cross sectional study based on a sample of 100 Italian non-listed SMEs whose financial and profitability dimensions have been monitored over the period from 2000 to 2011, considering 50 firms that have declared bankruptcy during this time period and 50 still active on the market over the same period. The performance of the proposed classifier has been compared with that of very well-known classifiers. The results on the different sets of indicators are consistent with the literature. In more details, although the best warning signal for bankruptcy is given by QDA on financial indicators (Table 2) for the short time and by the proposed method still on financial ratios for the long run (four to eight years prior to failure). The proposed method shows the best specificity over all time periods that, together with the high recognition rates, make it a very effective method to assess the health of a firm.

When the failure prediction is based on profitability ratios (Table 3), the performances of all the classifiers get worse. The proposed method more than holds its own and so does CART and also the decrease in performance of LDA is not large. QDA, on the other hand, shows a large drop on correct recognition rates and very noticeably on sensitivities that drop below 50%. This very low sensitivity put any prediction method based on QDA on profitability indicators at risk of non-detecting firms that are likely to fail.

This method could be used properly by whatever kind of SMEs in order to monitor their financial situation and also by banks in order to monitor companies' financial situation before lending money.

References

Abdullah, N.A.H., Halim, A. and Rus, R.M.D. (2008) 'Predicting corporate failure of Malaysia's listed companies: comparing multiple discriminant analysis, logistic regression and the hazard model', *International Research Journal of Finance and Economics*, Vol. 1, No. 15, pp.201–217.

Allen, A., Delong, G. and Saunders, A. (2004) 'Issues in the credit risk modeling of retail markets', *Journal of Banking and Finance*, Vol. 28, No. 1, pp.727–752.

Altman, E.I. (1968) 'Financial ratios, discriminant analysis and prediction of corporate bankruptcy', *Journal of Finance*, Vol. 23, No. 4, pp.589–609.

Altman, E.I. (1973) 'Predicting railroad bankruptcies in America', *Bell Journal of Economics and Management Science*, Vol. 4, No. 1, pp.184–21l.

Altman, E.I. and Saunders, A. (1998) 'Credit risk measurement: developments over the last 20 years', *Journal of Banking and Finance*, Vol. 20, No. 1, pp.1721–1742.

Altman, E.I., Margaine, M., Schlosser, M. and Vernimmen, P. (1974) 'Financial and statistical analysis for commercial loan evaluation', *Journal of Financial and Quantitative Analysis*, Vol. 9, No. 2, pp.195–211.

Argenti, J. (1976) *Corporate Collapse, The Causes and Symptom*, McGraw-Hill, New York.

Barnes, P. (1987) 'The analysis and use of financial ratios: a review article', *Journal of Business Finance & Accounting*, Vol. 14, No. 4, pp.449–461.

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

Bellovary, J., Giacomino, D. and Akers, M. (2007) 'A review of bankruptcy prediction studies: 1930–present', *Journal of Business and Economics Research*, Vol. 5, No. 5, pp.9–29.

Bhimani, A., Gulamhussen, M.A. and Da Rocha Lopes, S. (2013) 'The role of financial, macroeconomic, and non-financial information in bank loan default timing prediction', *European Accounting Review*, Vol. 22, No. 4, pp.739–763.

Blum, M. (1974) 'Failing company discriminant analysis', *Journal of Accounting Research*, Spring, Vol. 12, No. 1, pp.1–25.

Boardman, C.M., Bartley, J.W. and Ratliff, R.B. (1981) 'Small business growth characteristics', *Amer. J. Small Bus*., Winter, Vol. 5, No. 3, pp.33–42.

Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J. (1984) *Classification and Regression Trees*, Chapman & Hall/CRC, Wadsworth.

Brown, W.J. (1967) 'Factors influencing installment loan repayment tardiness at commercial banks', *Proceedings of the Business and Economic Statistics Section*, American Statistical Association, pp.469–72.

Casey, C.J. (1980) 'The usefulness of accounting ratios for subjects' predictions of corporate failure: replication and extensions', *Journal of Accounting Research*, Vol. 18, No. 2, pp.603–613.

Castagna, A.D. and Matolcsy, Z.P. (1981) 'The prediction of corporate failure: testing the Australian experience', *Australian Journal of Management*, Vol. 6, No. 1, pp.23–50.

Chen, K.C. and Lee, C.J. (1993) 'Financial ratios and corporate endurance: a case of the oil and gas industry', *Contemporary Accounting Research*, Vol. 9, No. 2, pp.667–694.

Deakin, E.B. (1972) 'A discriminant analysis of predictors of business failure', *Journal of Accounting Research*, Spring, Vol. 3, No. 1, pp.167–179.

di Donato, F. and Nieddu, L. (2015) 'The effects of performance ratios in predicting corporate bankruptcy: the Italian case', *Business Information Processing (LNBIP) ICDSST 2015 Conference Proceedings Decision Support Systems V: Big Data Analytics for Decision Making*, Springer.

Edmister, R. (1972) 'An empirical test of financial ratio analysis for small business failure prediction', *Journal of Financial and Quantitative Analysis*, Vol. 7, No. 2, pp.1477–1493.

Efron, B. and Tibshirani, R. (1995) *Cross-Validation and the Bootstrap: Estimating the Error Rate of a Prediction Rule*, Technical Report 176, Stanford University.

Fisher, R.A. (1936) 'The use of multiple measurements in taxonomic problems', *Annals of Eugenics*, Vol. 7, No. 2, pp.179–188.

Frydman, H., Altman, E.I. and Kao, D.L. (1985) 'Introducing recursive partitioning for financial classification: the case of financial distress', *The Journal of Finance*, Vol. 40, No. 1, pp.269–291.

Gales, K. (1965) 'Discriminant functions of socio-economic class', *Applied Statistics*, Vol. 6, No. 2, pp.123–132.

Gepp, A. and Kumar, K. (2012) 'Business failure prediction using statistical techniques: a review', in Kumar, K. and Chaturvedi, A. (Eds.): *Some Recent Developments in Statistical Theory and Applications*, pp.1–25, Brown Walker Press, Boca Raton, Florida, USA.

Gershenfeld, N. (1997) 'Nonlinear inference and cluster-weighted modeling', *Annals of the New York Academy of Sciences*, Vol. 808, No. 1, pp.18–24.

Hall, G. (1994) 'Factors distinguishing survivors from failures amongst small firms in the UK construction sector', *Journal of Management Studies*, Vol. 31, No. 5, pp.737–760.

Hastie, T., Tibshirani, R. and Friedman, J. (2001) *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*, Springer, New York.

Jones, F.L. (1987) 'Current techniques in bankruptcy prediction', *Journal of Accounting Literature*, Vol. 6, pp.131–164.

Karels, G.V. and Prakash, A.J. (1987) 'Multivariate normality and forecasting of business bankruptcy', *Journal of Business Finance & Accounting*, Vol. 14, No. 4, pp.573–593.

Laitinen, E.K. (1991) 'Financial ratios and different failure processes', *Journal of Business Finance & Accounting*, Vol. 18, No. 5, pp.649–673.

McKee, T.E. (2003) 'Rough sets bankruptcy prediction models versus auditor signaling rates', *Journal of Forecasting*, Vol. 22, No. 8, pp.569–586.

McLachlan, G.J. (2004) *Discriminant Analysis and Statistical Pattern Recognition*, Wiley, Hoboken, New Jersey.

Nieddu, L. and Patrizi, G. (2000) 'Formal methods in pattern recognition: a review', *EJOR*, Vol. 120, No. 3, pp.459–495.

Ohlson, J.A. (1980) 'Financial ratios and probabilistic prediction of bankrucy', *Journal of Accounting Research*, Vol. 18, No. 1, pp.109–131.

Pinches, G.E. and Trieschmann, J.S. (1977) 'Discriminant analysis, classification results and financial distressed P-L insurers', *Journal of Risk and Insurance*, Vol. 44, No. 2, pp.289–298.

Sharma, S. and Mahajan V. (1980) 'Early warning indicators of business failure', *Journal of Marketing*, Vol. 44, No. 4, pp.80–89.

Sinkey Jr., J.F. (1975) 'A multivariate statistical analysis of the characteristics of problem banks', *Journal of Finance*, Vol. 30, No. 1, pp.21–36.

Smith, C.A.B. (1946) 'Some examples of discrimination', *Annals of Eugenics*, Vol. 13, No. 1, pp.272–282.

Storey, D., Keasey, K., Watson, R. and Wynarczk, P. (1987) *The Performance of Small Firms*, Croom Helm, London.

Taffler, R.J. (1982) 'Forecasting company failures in the UK using discriminant analysis and financial ratio data', *Journal of Royal Statistical Society*, Vol. 145, No. 3, pp.342–358.

Taffler, R.J. (1983) 'The assessment of company solvency and performance using a statistical model: a comparative UK based study', *Accounting & Business Research*, Vol. 13, No. 52, pp.295–307.

Taffler, R.J. and Tisshaw, H. (1977) 'Going, going, going – four factors which predict', *Accountancy*, Vol. 88, No. 1003, pp.50–54.

Tintner, G. (1946) 'Some applications of multi-variate analysis to economic data', *Journal American Statistical Association*, Vol. 41, No. 236, pp.472–500.

Trieschmann, J.S. and Pinches, G.E. (1973) 'A multivariate model for predicting financially distressed P-L insurers', *Journal of Risk and Insurance*, Vol. 40, No. 3, pp.327–347.

Wolpert, D. (1996) 'The lack of a priori distinctions between learning algorithms', *Neural Computation*, Vol. 8, No. 7, pp.1341–1390.