A non parametric approach to firms' failures in Italy: a case study from 2000 to 2011

Author Details: [(1)]**Francesca di Donato** - Assistant Professor - UNINT University of Rome [(2)]**Luciano Nieddu** - Assistant Professor - UNINT University of Rome

Abstract—*In this paper the problem of firms' failures will be addressed. The aim is to determine which are the trigger factors that can predict the inability of a firm to cover its obligations. Various methods are available in the literature in order to analyze this problem. The aim of this paper is to use two non parametric robust classification methods to determine the variables that can affect the probability of failure. The study will be carried out on an Italian sample of non listed small-medium firms (both failed and still on the market) randomly selected over a period of 12 years (2000-2011).*

Keywords-component; Outliers; Business Failures, Classification Trees, Cross-section Study, Discriminant Analysis.

I. INTRODUCTION

Accounting information can be used in predicting companies' failure. Mostly, financial ratios have been used by many authors as an important tool to forecast corporate bankruptcy due to their predictive power of this phenomenon.

Bankruptcy is defined as the lack of capital to cover the obligations of a business as they mature (Boardman et al. 1981) [9]. Beaver [6] Horrigan [27] Altman [3] Daniel [18] and Deakin [19] are the first exploring the use of financial ratios in predicting business failure and bankruptcy.

Generally, financial ratios play an important role in forecasting the default because these are precise indicators and are constructed from financial reporting information that firms have to file with public and tax authorities. For example, a firm's inability to generate operational profits or earnings before interest and tax (currently and in the future) to service debt can increase the hazards of default. Analogously, insufficient resources in the long (solidity) and short (liquidity) term can also increase the hazards of default (Bhimani et al., [8]).

Beaver [6] used univariate statistics on US market data to determine the effect of financial ratios on the probability of bankruptcy. Altman [2] stated that, even if the univariate approach is important in generalizing about the performance and trends of particular measurements, the adoption of the results for assessing potential bankruptcy of firms, is questionable. The univariate nature focuses on individual signals of impending problems.

For this reason, Altman introduced multivariate analysis applying the Multiple Discriminant Analysis (MDA) to predict firms bankruptcy, but his analysis does not take into account the evolution of the financial ratio over time.

In 1980 Ohlson, [41] tried to overcome some of the limitations of MDA applying a conditional logistic regression and using information of the performance of each firm at various stages prior to bankruptcy.

It is clear that the first studies did not take into account longitudinal information and have been focused on using financial ratios at a given time prior to the occurrence of the event to determine the probability of bankruptcy.

Only in the mid 1980s, there was a shift on the use of longitudinal models and semi-parametric and non parametric approaches such as recursive partitioning algorithms (see [23],[26]), neural networks techniques (see [40],[15],[46],[48]), survival analysis ([16],[34],[17],[14],[5],[47],[38]) and classification trees [28].

Moreover, the beginning studies on bankruptcy mainly focused on large firms. In fact, before Storey [45] seminal contribution, a few studies dealt with the failure of small-medium firms ([21],[4]).

Storey (1987) identified its sample from the small firm sector, also using non financial variables but without a control group of survivors. Hall [25] studied the factors affecting small companies failure distinguishing between small firms that fail from those that survive but only considering the construction sector.

In 2005 Huarng et al. [28] applied classification trees to business failure in a study that did not produce reliable results due to a very small sample size.

Despite the increasing use of survival analysis and longitudinal statistical techniques to model financial distress, little attention has been given to the use of time varying covariates to estimate these models. Shumway [44] considered longitudinal data and a semi-parametric model to determine the probability of failure of a firm, allowing for time dependent covariates to influence the hazard function, defined as the probability of a firm to experience bankruptcy at time t given the fact that it has survived until that time. The use of time dependent covariate allows the varying financial indicator to vary their effect on the probability of bankruptcy, therefore yielding a dynamic model. More recent studies using the hazard function are Romer [42], Chanchrat et al. [13], Kim and Parkington [30], Nieddu and Vitiello [38].

In this paper we study the phenomenon from a cross-sectional point of view using a non parametric approach (classification trees) to determine the conditional probability of bankruptcy of a firm at various time prior to occurrence of the event.

The approach will be applied on original data collected over 12 years (2000-2011) for a stratified sample of non listed small – medium Italian companies. The reason for the analysis is that we want to determine which are the financial statement items that influence bankruptcy at various points in time using a robust non parametric technique which allows to mine the information on the data without requiring any prior assumption.

This paper differs from analogous papers on the topic for the following reasons. First of all, we use two very robust classification techniques technique to test if there is a real relation between data at hand and firms' survival.

Moreover, we have used an original stratified random sample of small – medium Italian companies using business sectors as stratifying variable selecting only firms with revenues from sales from euro 2 millions to 50 millions. In the previous literature small – medium firms are not studied very often and the Italian sample is totally new.

The results concern a retrospective study since the aim of the paper is not to determine the proportion of failed firms but to determine the factors affecting the failure. Therefore 50 active firms and 50 failed firms have been selected and their financial statements have been studied during a period of 10 years.

The layout of the paper is as follows. In Section 2 the data will be described. Section 3 presents the classification methods. Section 4 presents empirical results concerning the application of the classification trees and the Discriminant Analysis. Finally in Section 5 some conclusion will be drawn.

II. THE DATA

In prior studies, default is related to the capital structure of firms: firms default on their obligations if the market value of their assets falls below a threshold determined by the respective default model. Restricted liability creates incentives for partners to default and to shift ownership to lenders and consequently ensure a minimum limit in the settlement of their equity (Duffie and Singleton, 2003).

Altman and Saunders (1998) and Allen et al. (2004) reviewed the vast literature on the influence of financial 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 financial accounting ratios on corporate distress.

According to many authors (Bhimani et al., 2013; Altman and Saunders, 1998; Allen et al., 2004), the failure of a limited company is connected to two different situations strictly connected: the inability to pay financial obligations when they come due (meaning lack of liquidity, low solidity and very high debt ratios) and the inability to generate operational profits or earnings before interest and tax (meaning negative income and negative or very low profitability ratios ROE, ROI, ROS).

According to Beaver (1966) we state that ratios are not the only predictors of failure, but that also the simple financial statement data have a predictive ability concerning business failure.

For this reason, we used at the same time financial statement data and financial ratios as predictors of business bankruptcy, in particular the ones mainly focusing on financial performance, liquidity and solidity, and on economic strength including various kinds of margins and profitability and returns (Laitinen and Kankaanpaa, 1999; Neves and Vieira, 2006).

We selected financial-statement data of all the non listed companies in the sample (50 failed and 50 still active on the market) over the period 2000 – 2011. The companies were randomly chosen in the sectors of tourism, agriculture, industry, services, construction and retail according to their revenues from sales within the range 2 – 50 million euros in the first year of the analysis (2000). The companies where all active at the year 2000 and then we followed them until 2011 or until they failed. Therefore, for each company more than one financial statement is available. Due to the longitudinal feature of the data, the sample dimension will decrease from 100 companies at 2000 to a plateau of 50 companies at 2011 since we have decided to select an equal representation of failed and non-failed companies. So the aim of the study is not to determine an estimate of the fraction of failed companies in Italy during the early years of 2000 but to describe which factors can be considered as indicators of possible distress for a firm.

All the data were collected through CERVED database, assembling all economic and financial data related to Italian non listed companies.

Then, from financial statement, the most common financial ratios have been computed, for every year in the period 2000 – 2011. The most important criterion used in order to select the ratios was popularity, meaning their frequency of appearance in the literature.

In Table 1 a brief summary of the financial statement items that have been used in the analysis, together with the financial ratios, have been displayed. For each item its quantile has been reported (minimum, 1st quartile, median, 3rd quartile and maximum). From a quick glance it is clear that there is a big variability and heterogeneity in the data. Some of the items seem to exhibit peculiar values: this is probably due to the fact that some of the items in the financial statement start to loose their informative power when the firm is very close to bankruptcy.

TABLE I. PERFORMANCE RATIO AND DATA

Financial Statement Data	min	Q1	median	Q3	max
ROE	-71.315098	0.00144582	0.06174602	0.28264578	42.0361842
ROI	-235.79793	0.00046384	0.07148847	0.19494863	60.2710678
ROS	-3445.4514	0.00335971	0.02954216	0.0615699	19.5367822
CT	-5142.0123	0.35615148	1.93750378	4.84041303	1386.27519
StructureRatio1	-29295.48	0.13318499	0.4882849	1.23870849	2940.65713
StructureRatio2	-21458.44	0.56129976	1.175466	2.71810232	6745.22509
FD/E	-2185.2716	0	0.44316189	3.84172697	1745.56447
QUICK.RATIO	-3.9202031	0.54143462	0.84645883	1.16973262	60.354118
FinancialDebt/Working Capital	0	0	0.10032628	0.38072007	10.379214
ShortTermBankLoan/WorkingCapital	0	0	0.06412584	0.30159147	10.379214
Ebitda/Sales	-393.72873	0.01749757	0.05409041	0.1042129	19.5405452
FinancialInterests/Sales	0	0.00312117	0.0138575	0.03161644	120.311214
FI/Ebitda	-39.319212	0.00495928	0.1511666	0.41984525	111.524002
WorkingCapitalCycle	-263849.94	-30.239141	8.85981108	67.1221417	171114.952
FixedAssets	0	125025	519667	2510838.35	651078661
Receivables	6941.47	763630	1752681	4336010	1110382896
Liquidity	0	20498	94518	351519	55699427
Equity	-62419890	68926	305556.13	1162205.28	1060343146
LongTermLiabilities	0	96620.58	295794	1531810.77	165926786
FinancialDebt	0	0	456045	2336685	175179965
ShortTermDebt	14995	1213229.31	2628571	6582242.38	919874000
BanksLoan	0	0	191716	1491869.26	175179965
Other Financial Debts	0	0	0	0	88101654
SuppliersPayables	0	159413	939029.51	2483518	468800669
sales	0	2526546	4532270	8670265	2296328639
Ebitda	-5413449	57472	224621	802208	349570094
ebit	-60281618	12282.91	134323	406355	229270523
NetIncome	-160863184	-38821	8050	63475	159478841
year	2000	2002	2004	2007	2011

We intend to perform a cross-section study: therefore we will be studying the firms at various years prior to failure. Since the sample size reduces going back in time, we have considered financial statements on up to 8 years prior to failure. To maintain the balance between failed and active firms for each year prior to failure, the data of the financial statement of each failed firm has been randomly matched with the data of an analogous active firm at the same year.

III. CLASSIFICATION METHODS

A. C&RTs

Classification and regression trees (C&RT) are a non-parametric method to partition the covariate space X into a set of rectangles and then fit an appropriate model (usually a constant value) in each rectangle. If the response variable is continuous then a regression tree can be grown on the data, otherwise usually a classification problem is considered. In this paper only classification trees will be considered being the outcome a binary variable (failed, non-failed).

In each supervised learning algorithm a dataset of previously classified units is available. The units are usually classified by an expert or have experience and event. When the classification is carried out by an expert a subjective component could be drown into the study since the classification supplied by the expert non necessarily is 100% correct. In this case the problem is known to be of classification with imperfect supervisor **Error! Reference source not found.**. In our case the firms have been classified as "failed" or "in the marked" depending on weather they have declared bankruptcy or not, i.e. they have experienced the event. So we will be dealing with classification with perfect supervisor.

Given the dataset of previously classified units, the goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data.

Tree-based methods partition the covariate space X into disjoint set of rectangular regions, and then classify the observations according to which partition element they fall in. The partitioning of X is usually carried out according to an impurity measure (the Gini index) or according to the information gain (entropy) that a particular partition could achieve. The growth of a tree is a top-down recursive process, therefore, starting with a single node (root) we look for the binary partition of one of the covariates that yields the most information about the class. The same is done on the derived subsets and the process stops either when the units in a node have all the same value of the variable indicating the class or when the splitting no longer adds value to the predictions. The iterative partitioning process is called "growing a tree" or "learning".

When dealing with more than one covariate, the one leading to the split with the lowest impurity is first selected. This process is continued until some stopping criterion is met. For example, we might stop when every partition element has less than a certain number of elements. The bottom nodes of the tree are called the leaves. Each leaf is assigned a class according to a majority rule based on the classes of the elements that belong to that leaf. This majority rule criterion is also used in classifying new objects.

Various problems are connected to classification trees:

- the problem of learning an optimal decision tree is known to be NP-complete therefore decision-tree learning algorithms are based on heuristics such as the greedy algorithm where locally-optimal decisions are made at each node. Such heuristics cannot guarantee that the results be the globally-optimal decision tree;
- classification trees algorithms can create over-complex trees. The complexity of the tree doesn't necessarily imply a good accuracy of the tree. A too complex tree can be due to the nature of the data which represent a complex phenomenon or to the fact that maybe the data do not represent correctly the phenomenon. In both cases the algorithm will try to fit the data growing a rather large tree trying to over fit the data. A too complex tree will clearly perform well on the training data, but this not necessarily means that it will be able to correctly classify new objects of unknown class. To avoid over complex trees, pruning techniques usually based on cross validation (i.e. on their performance on new data) can be used.

Once the tree has been grown, each new unit will be classified into one of K classes according to the most frequent class in the leaf where the unit belongs to.

B. *Linear Discriminant Analysis*

Given an already classified dataset on which k covariates have been collected, the aim of Discriminant Analysis (DA) is to determine a decision rule based on the data at hand. Give K groups, if the conditional distribution of $\mathbf{X}|y$ (y=1,..,K) is known with at most unknown parameter ϑ, i.e.:

$$\boldsymbol{X}|y \sim f(\boldsymbol{x}, \vartheta) \qquad y = 1, .., K$$

The a maximum a posteriori decision rule (MAP) can be formulated to assign each element to one of K classes according to the highest posterior probability that that element belongs to a certain class. Such a discriminant rule is known as Bayes classifier [36]. When the conditional distribution of X is multivariate normal with equal covariance matrices for each class, i.e.:

$$\boldsymbol{X}|y \sim MVN(\boldsymbol{\mu}_y, \boldsymbol{\Sigma})$$

then the Bayes classifier reduces to a linear discriminant rule.

The same result was suggested by Fisher [23] following a *non parametric* approach to the problem. Fisher's suggestion was to try to determine the linear combination $\boldsymbol{a}'\boldsymbol{x}$ of the covariates which maximizes the ratio of the between-groups sum of squares to the within-groups sum os square

$$\frac{\boldsymbol{a}'\boldsymbol{B}\boldsymbol{a}}{\boldsymbol{a}'\boldsymbol{W}\boldsymbol{a}}$$

In other words Fisher discriminant analysis tries to determine the linear combination that optimizes the difference between the means of the groups normalize by the within class variability. The solution proposed by Fisher [23] was to determine the linear combination (LDA) that optimizes the difference between the means of the groups normalized by the within class variability. The solution proposed by Fisher can be easily generalized to more than two classes. It is a robust technique even if the normality hypothesis does not hold [35]. In the case of varying covariance matrices a quadratic discriminant function can be used.

The dataset on which the classification algorithm has been trained cannot be used to assess the performance of the algorithm since it will perform optimally on that dataset. The error rate obtained on the training set is called resubstitution error rate [35] and is a biased estimator of the true error rate. To obtain an unbiased estimator of the error rate of the algorithm it must be tested on an independent sample. Therefore the dataset is usually partitioned into two subsets: the training set and the verification set. The algorithm is trained on the first set and its performance is evaluated on the verification set. This is usually repeated more than once. In the k-fold cross validation approach the original dataset is partitioned into k disjoint subsets and in turns one of the k subsets is singled out, the algorithm is trained on the remaining data and tested on that subset. This is repeated k times.

IV. Results

A. *Training and Cross-Validation results*

In Table 2 the result of the fitting of classification trees on the datasets at various years prior to failure have been reported.

The resubstitution error rate and the variables actually used in the growing of the trees have been displayed. Those variables are those that display the highest discriminatory power in the dataset.

The resubstitution error rate can be considered a goodness of fit measure for the considered tree. It is usually smaller than the true error rate for which the cross-validation error is a non biased estimator. The tree 2 years prior to failure displays the best performance with 25% error on new unknown units that must be classified. The worst performance is obtained 8 years prior to failure, where the only variable relevant to discriminate between firms is the liquidity. The cross validation error rate 8 years prior to failure is 44% on the available sample, implying that the obtained tree is not good enough to classify new firms. This can be ascribed to the fact that probably 8 years prior to failure is too early to accurately predict the outcome.

This is consistent with some studies in literature. In fact, ratios from the third through fifth year prior to failure had been shown (Beaver, 1966 [6]; Deakin, 1972 [19]) to contain information content adequate for predicting failure with greater than random accuracy, especially for the third year prior to the event.

In Table 3 the results of the cross validation for the LDA have been displayed together with the variables relevant for classification. To determine the variables to be used in the training the function stepclass() available in the package KlaR has been used. This function uses 10-fold cross validation to determine the variables in a dataset that have the highest discriminatory power when used for LDA. The performance of LDA is fairly comparable to the one obtained used classification trees. Both techniques underline the importance of financial variables as predictive variables for bankruptcy.

Although LDA seems to perform better than classification trees 8 years prior to failure, it must be stressed that the sample size (17 failed and 17 still active) is too small to actually determine if the difference in performance is relevant.

It should be pointed out that the non optimal performances of the two techniques could be ascribed to a lack of predictive power of the variables available in the financial statement. The non optimal performance of LDA could be due to the fact that LDA is optimal under normality of the covariates and homoscedasticity of the covariance matrices. On the other hand, classification trees are a robust technique that does not assume any particular distribution on the data.

In fact, according to some authors (Hall, 1994 [25]; Kennedy, 1975 [29], the financial statement data are not enough informative in terms of predicting corporates failure.

TABLE II. PERFORMANCE PARAMETER OF THE FITTED TREES AFTER PRUNING

Years prior to failure	Variables used in classificaion	Resubstitution error	cv error
1	FD/E; Liquidity; ShortTermDebt; StructureRatio1; StructureRatio2	0.100	0.390
2	Liquidity	0.240	0.250
3	Liquidity	0.290	0.390
4	FD/E; Liquidity; Other Financial Debts; Receivables	0.192	0.283
5	Liquidity; QUICK.RATIO; ROI; ROS; Sales	0.098	0.370
6	QUICK.RATIO; ROE	0.197	0.348
7	QUICK.RATIO; Receivables	0.152	0.283
8	Liquidity	0.235	0.441

TABLE III. PERFORMANCE OF LDA

Years prior to failure	Variables used in classification	cv error
1	Equity	0.370
2	Equity	0.320
3	Quick.Ratio	0.350
4	STBC/WC; FI/EBITDA	0.313
5	STBC/WC	0.293
6	ROS; FI/EBITDA	0.258
7	FD/E; FI/S	0.196
8	Liquidity	0.265

Concerning Table 2 and 3 we can highlight that liquidity, that is a financial variable, is always relevant as predictor of bankruptcy. In particular, it is the only discriminant variable in the short term, that is 2 and 3 years prior to failure, consistent with the role of financial variables in measuring short term performance. The relevance of liquidity (item also included in quick ratio) can be explained by the companies need of having financial resources in order to finance their business, because it is not always easy for them to get financial funds from the banks.

On the contrary, the economic variables like ROE and ROI are relevant in the medium - long run, that is 5 and 6 years prior to failure. This is consistent with the role of economic variables related to medium – long term performance.

In the following the results of classification trees 8, 7, 6, 5, 4, 3, 2 and 1 years prior to failure will be considered.

In all the models that have been fitted, the financial variables are the ones constantly influencing the risk of failure, providing the higher information gain as a first split in the classification tree. Liquidity and in two cases quick ratio, that includes the item liquidity, are always the first discriminant variables affecting the failure/survival of companies and representing their capability to meet the financial obligations when they come due.

In particular, liquidity is an important discriminant of firms failure 8 years prior to bankruptcy and in the shorter years standing from failure (5, 4, 3, 2 and 1 years before bankruptcy). This highlights the importance of this kind of resource for Italian small and medium-sized enterprises, in order not to be dependent on the market and in particular on the banks.

Instead, 7 and 6 years prior to bankruptcy, the discriminant variable for the failure is represented by quick ratio, which is another financial ratio representing the relation between current activities and current liabilities, meaning the capability of companies to pay operating debts with its cash plus receivables (cash + receivables/current liabilities).

Financial ratios and liquidity in particular show the capability of a company to be autonomous in paying financial and operating obligations without needing to resort to other costly financial funds.

The existence of a high degree of liquidity is moreover a sign of companies' health because firms with a lot of cash regularly collect trade receivables and therefore do not have problems of uncollectable. But a high degree of liquidity can also depend on a good access to credit, meaning that banks trust the companies and lend them money, meaning that they are healthy and solvent.

The significance of liquidity is consistent with the importance, for Italian small-medium companies, of having financial resources in order to finance their business, because they are not able to get financial funds on the market and some of them can have problems in getting resources from the banks.

Moreover, it is also consistent with the global financial crisis that has produced many problems for small-medium companies that are not very capitalized. They suffered from a situation of illiquidity due to the difficulties to collect receivables from customers and due to the credit crunch by the banks drastically reducing their funding due to the lack of trust.

In Figure 1 the classification tree for firms 8 years prior to failure has been displayed. In this case, liquidity is the only item discriminating between failure and survival. Namely, liquidity tends to entirely explain the phenomenon, in fact, if it is higher than 0.02619 there are 12 active over 3 failed firms. A low value of liquidity, on the other hand, is more common for failed companies (5 active over 14 failed).

In Figure 2 and 3 the classification trees for firms 7 and 6 years prior to the event have been displayed. In these situations quick ratio is the discriminant variable to classify a firm as active or failed. 7 years prior to failure, quick ratio's values higher than 1.045 tend to explain very well the phenomenon because they are associated with 14 active firms over 1 failed firm, while values of quick ratio lower than 1.045 tend to be associated with failed companies (8 active over 22 failed). To further refine the classification 7 years prior to failure, receivables could be used for firms with low quick ratio. Receivables values lower than 0.1568 tend to be associated with active companies (6 active over 4 failed), while higher values of receivables are associated to failed companies (2 active over 18 failed). This is also consistent with the fact that if the amount of receivables is low, it means that the companies collect them regularly reducing cash problems.

Also 6 years prior to failure quick ratio is a discriminant variable well explaining the probability of bankruptcy. In this case values higher than 1.019 tend to be associated with active firms (19 active over 3 failed firms) while low values tend to be associated with failed companies (14 active over 30 failed). To further refine the classification 6 years prior to failure, Return on Equity (ROE) could be used for firms with low quick ratio. ROE values higher than 0.1267 are connected to active companies (11 over 7 failed) while ROE values lower than 0.1267 tend to be connected with failed companies (23 over 3 active). This is consistent with the fact that if the ROE is positive and high the companies are creating value for the shareholders.

In Figure 4 the classification tree for firms 5 years prior to the event has been shown. Liquidity is again the discriminant variable to classify a firm as active or failed. Very high values of liquidity tend to be associated with all active firms (21 active over 0 failed firms) while low values tend to be associated with failed firms (25 active over 46 failed). The classification can be further refined for firms with low liquidity considering quick ratio, sales, Return on Investment (ROI) and Return on Sales (ROS), ratios mainly measuring the economic performance and always showing that if they are positive and higher than certain threshold, they tend to be associated with active firms.

In Figure 5 the classification tree for firms 4 years prior to the event has been shown. Liquidity is still the discriminant variable to classify a firm as active or failed. Values of liquidity higher than 0.03992 tend to be associated with active firms (30 active over 8 failed firms) while low values tend to be associated with failed firms (20 active over 41 failed). The classification can be further refined considering financial debt/Equity (FD/E) for firms with high liquidity. In this situation, companies with low FD/E are connected to active firms (25 over 1 failed) consistent with the importance of controlling this ratio that implies companies solvency and financial independence.

On the contrary, 4 years prior to failure, the classification can be refined considering other financial debts and receivables for firms with low liquidity.

In Figure 6 and 7 the classification trees for firms 3 and 2 years prior to the event have been displayed where liquidity is the only discriminant variable to classify a firm as active or failed. In particular, values of liquidity higher than 0.04176 (7 years) and 0.02025 (6 years) tend to be associated with active firms (30 active over 9 failed firms 7 years prior to failure and 35 active over 9 failed 6 years prior to bankruptcy) while low values tend to be associated with failed firms (20 active over 41 failed and 15 active over 41 failed ones).

In Figure 8 the classification tree for firms 1 year prior to the event has been shown. Liquidity is still the discriminant variable to classify a firm as active or failed. Values of liquidity higher than 0.07744 tend to be associated with active firms (23 active over 3 failed firms) while low values tend to be associated with failed firms (27 active over 47 failed). For companies with low liquidity, the classification can be further refined considering structure ratio1. In this case values of this ratio higher than 0.1502 are associated with active companies (26 over 25 failed) while values lower than 0.1502 are related to failed companies (22 over 1 active). One year prior to failure the classification can be further refined for companies with good structure ratio 1 using short term debt, and then Financial Debt/Equity and structure ratio 2.

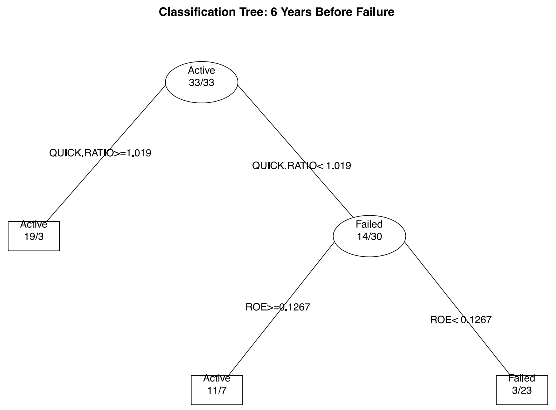

Figure 3: classification tree for 6 years prior to failure

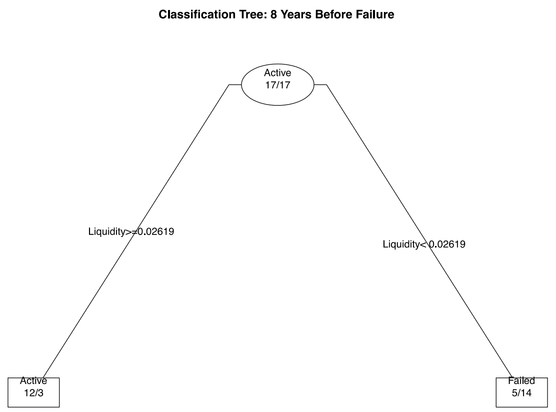

Figure 1: classification tree for 8 years prior to failure

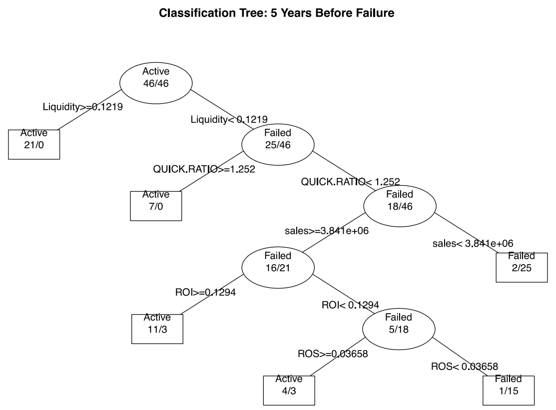

Figure 4: classification tree for 5 years prior to failure

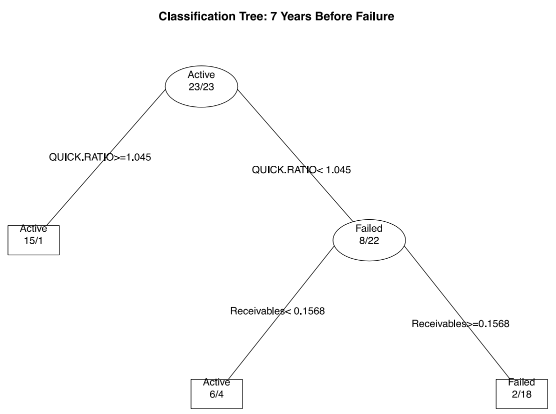

Figure 2: classification tree for 7 years prior to failure

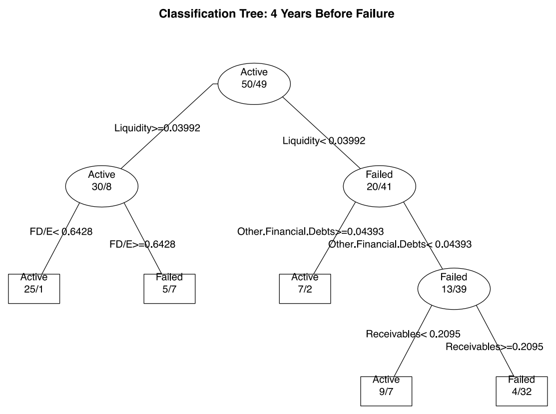

Figure 5: classification tree for 4 years prior to failure

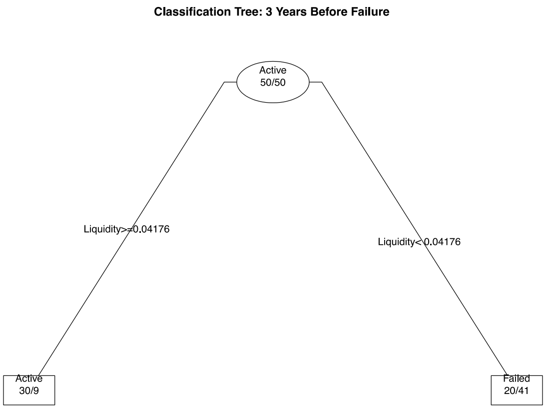

Figure 6: classification tree for 3 years prior to failure

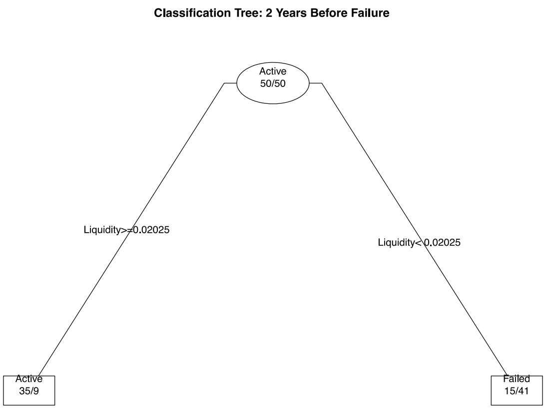

Figure 7: classification tree for 2 years prior to failure

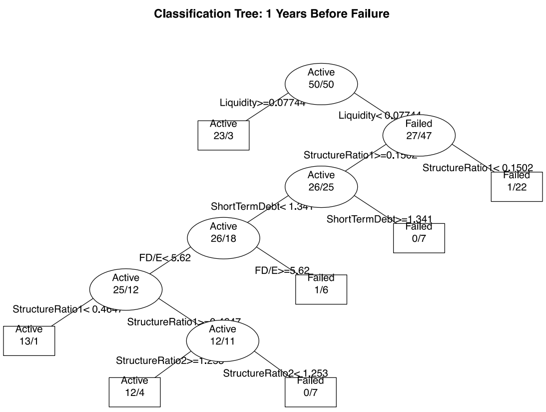

Figure 8: classification tree for 1 year prior to failure

V. Conclusions

We performed a cross sectional analysis based on a sample of 100 Italian non listed companies out of which 50 are bankrupted and 50 are still active on the market over the period 2000 - 2011.

The results of the analysis clearly show that in Italy the ratios measuring the economic performance, relying on financial statement data and based on estimations, such as ROE, ROI and ROS are not sufficient to accurately predict companies failure.

On the contrary, ratios measuring the financial performance are much more significant in predicting companies bankruptcy because they really measure the capability of companies to face financial obligations with autonomous financial resources, in order not to depend on the banks or on other lenders.

The poor relevance of economic variables could be also explained by the fact that Italian companies tend to apply estimations to economic margins in order to minimize the net income for tax purpose. This means that the economic ratios, calculated through financial statement, do not reflect the real company's performance, do not measure his real health and for this reason they are not good predictor of their potential failure.

On the contrary, ratios and items connected to solidity and liquidity are much more important and more accurate in predicting companies failure because they are an important indicators for small medium companies in having financial resources necessary for firms' survival.

From the results obtained using a very robust classification method it is clear that the only information available in the financial statement is not sufficient enough to discriminate between companies that are going to fail in the immediate future. This is probably due to the fact that the financial statement reflects only a part of the firm status.

References

[1] Allen A., Delong G. and Saunders A., "Issues in the credit risk modelling of retail markets", Journal of Banking and Finance, 28(1), 2004, 727-752.

[2] Altman E.I, "Financial Ratios, discriminant analysis and prediction of corporate bankruptcy", Journal of finance, 23(4), 1968, 589-609.

[3] Altman E.I, Saunders, A. "Credit risk measurement: developments over the last 20 years", Journal of Banking and Finance, 20(1), 1998, 1721-1742.

[4] Argenti, J., "Corporate Collapse, the Causes and the Symptoms", McGraw-Hill, NY, 1976.

[5] Bandopadhyaya, A., "An estimation of the hazard rate of firms under Chapter 11 protection", The Review of Economics and Statistics **76**, 1994, 346-350.

[6] Beaver W.H., "Financial Ratios as predicots of Failure", Journal of Accounting Research 4, 1966, 71-111.

[7] Bednarski, T., "On sensitivity of Cox's estimator" Statistics and Decisions, **7,** 1989, 215-228

[8] Bhimani, A., Gulamhussen M. A., Da Rocha Lopes S., "The role of Financial, Macroeconomic, and Non-Financial Information in bank loan default timing prediction" European Accounting Review, 22:4, 2013, 739-763

[9] Boardman, C. M., J. W. Bartley, R. B. Ratliff. "Small business growth characteristics", Amer. J. Small Bus. 5(Winter), 1981, 33–42.

[10] Breiman L., Friedman J. H., Olshen R. A., and Stone, C. J., Classification and Regression Trees. 1984, Wadsworth.

[11] Breslow, N.E., "Covariance analysis of censored survival data" Biometrics, **30**, 1974, 89-99

[12] Cain, K.C., Lange, N.T., "Approximate case influence for the proportional hazards regression model with censored data", Biometrics, **40**, 1984, 493-499

[13] Chancharat, N., Davy, P., McCrae, M., and Tian, G., "Firms in financial distress, a survivaldmo el analysis" School of Accounting and finance, University of Wollongong, 2007

[14] Chen, K. C. and Lee, C. J., "Financial ratios and corporate endurance: A case of the oil and gas industry", Contemporary Accounting Research **9**(2), 1993, 667-694.

[15] Coats, P.K and Fant, L.F., "A neural network approach to forecasting financial distress", The Journal of Business Forecasting, 1992, 9-12.

[16] Cox, D. R., "Regression and life tables", Journal of the Royal Statistical Society **34**, 1972, 187-220.

[17] Crapp, H. and Stevenson, M., "Development of a method to assess the relevant variables and the probability of financial distress", Australian Journal of Management **12**(2), 1987, 221-236.

[18] Daniel, T., E., "Discriminant Analysis for the Prediction of Business Failure", Ann Arbor,Michigan, University Microfilms, Inc. 1969

[19] Deakin, E., B., "A Discriminant Analysis of Predictors of Failure", Journal of Accounting Research, 10, 1972, 167-79

[20] Duffie D., Singleton, K., "Credit risk: Pricing, Measurement and Management", Princeton University Press, 2003

[21] Edmister, R., O., "An Empirical test of financial ratio analysis for small business failure prediction", Journal of Financial and Quantitative Analysis, **7**, 1972, 1471-93

[22] Efron, B., "The efficiency of Cox's likelihood function for censored data", JASA, **72**, 1977, 557-565

[23] Fisher, R. A., "The Use of Multiple Measurements in Taxonomic Problems". Annals of Eugenics 7 (2):, 1936, 179–188. doi:10.1111/j.1469-1809.1936.tb02137.x. hdl:2440/15227.

[24] Frydman H., Altman E.I., Kao D.L., "Introducing Recursive Partitioning for Financial Classification: The Case of Financial Distress", The Journal of Finance, Vol. XL, N. 1. 1985, 269-291.

[25] Hall, G., "Factors distinguishing survivors from failures amongst small firms in the UK construction sector", Journal of Management Studies, 31:5, 1994.

[26] Hastie, T., Tibshirani, R., Friedman, J., The Elements of Statistical Learning. Data Mining, Inference, and Prediction. Springer, New York, 2001.

[27] Horrigan, J., O.,"The determination of long term credit standing with financial ratio", Empirical Research in Accounting; Selected Studies, 1966.

[28] Huarng, K., Yu, H.K., and Chen, C.J., "The application of decision trees to forecast financial distressed companies", In 2005 International Conference on Intelligent Technologies and Applied Statistics Proceeding, Taipei, Taiwan, 2005

[29] Kaplan, E. L.; Meier, P., "Nonparametric estimation from incomplete observations", J. Amer. Statist. Assn. **53**, 1958,457–481

[30] Kennedy, H.A. "A behavioral Study of the Usefulness of Four Financial Ratios", 1975.

[31] Kim, M.H. and Partington, G. The dynamic prediction of company failure. The 21st Australasian Finance and Banking Conference, pp. 41-47. Sydney, 2008.

[32] Laitinen, E.K. and Luoma, M., "Survival analysis as a tool for company failure prediction", International Research Journal of Finance and Economics, **16** 2008, 673-678

[33] Laitinen, T. and Kankaanpaa, M., "Comparative analysis of failure prediction methods: the Case", European Accounting Review, 8:1, 1999, 67-92

[34] Lane, W. R., Looney, S. W., Wansley, J. W., "An application of the Cox proportional hazards model to bank failure", Journal of Banking and Finance **10**(4), 1986, 511-531

[35] McLachlan, G. J. (2004). Discriminant Analysis and Statistical Pattern Recognition. Wiley Interscience. ISBN 0-471-69115-1. MR 1190469

[36] Mardia, .K.V., Kent, J.T., Bibby, J.M., "Multivariate Analysis", Academic Press, 1989

[37] Neves J. C., Vieira, A., "Improving bankruptcy prediction with hidden layer learning vector quantization" European Accounting Review, 15(2), 2006, 253-271

[38] Nieddu, L. and Vitiello, C. "Determinants of the duration of contracts: a longitudinal study in the Italian fishery sector" Italian Review of Economics, Demography and Statistics, LXVII(1), 2013, 135-142

[39] Nieddu L. and Patrizi G., "Formal properties of pattern recognition algorithms: A review". European Journal of Operational Research, 120:459–495, 2000

[40] Odom, M. D. and Sharda, R. "A neural network model for bankruptcy prediction", Proceedings of the IEEE International Joint Conference on Neural Networks *(IJCNN)* **2**, 1990, 163-168.

[41] Ohlson J.A., "Financial Ratios and probabilistic prediction of bankrucy", Journal of Accounting Research, 18(1), 1980, 109-131.

[42] Rao, R. C., "The utilization of multiple measurements in problems of biological classification".Journal of the Royal Statistical Society, Series B 10 (2):, 1948, 159–203.

[43] Romer,A., "A comparative analysis of determinats of financial distress in French, Italian and Spanish firms" Working paper, Danmarks Nationalbank, Copenhagen, 2005.

[44] Shumway T. "Forecasting bankruptcy more accurately: A simple hzard model", Journal of Business **74**, 101-124.

[45] Storey, D., Keasey, K., Watson, R. and Wynarczk, P. "The Performance of Small Firms", London: Croom Helm, 1987.

[46] Tam, K. Y. and Kiang, M. Y., "Managerial applications of neural networks: The case of bank failure predictions", Management Science **38**, 1992, 926-947.

[47] Therneau T. M., Grambsch P. M., Modeling Survival Data, Extending the Cox Model, Springer, 2000.

[48] Wilson, R. L. and Sharda, R., "Bankruptcy prediction using neural networks", Decision Support Systems **11**, 1994, 545-557.

[49] Zavgren, C.V., "Assessing the Vulnerability to Faiulure of American Industrial Firms: a logistic analysis", Journal of Business Finance and Accounting, **12**, 1985, 19-46