TRIGGER FACTORS THAT INFLUENCE BANKRUPTCY: A COMPARATIVE AND EXPLORATORY STUDY

Leonardo Di Marco, Luciano Nieddu

1. Introduction

The phenomenon of bankruptcy has significantly prejudiced the Italian productive fabric in the last few years. The impact of the world economic crisis on the Italian economy has generated devastating effects for the wealth of the Nation, causing strong unbalances in terms of employment, productivity and investments.

Business failure prediction is one of the most essential problems in the field of finance. The research on developing business failure prediction models has been focused on building classification models to distinguish among failed and non-failed firms. The pioneer for corporate failure prediction models was William Beaver (1966). He applied a univariate model in which a classification model was carried out separately for each financial ratio, and an optimal cut-off point was identified so that the percentage of misclassifications was minimized.

Altman (1968) applied Linear Discriminant Analysis introduced by Fisher in 1936 to the problem of predicting bankruptcy. This technique dominated the literature on corporate failure models until the 1980s and is commonly used as benchmark for comparative studies.

Altman’s study involved 66 manufacturing companies with equal number of failures and survivors, and a total of 22 ratios from five categories, namely liquidity, profitability, leverage, solvency, and activity. From this set of ratios, five were finally chosen on the basis of their predictive ability.

Since these seminal studies not much work has been done to apply different methodologies to the problem of failure predictions, until the early 80s when Ohlson (1980) and Zmijewski (1984) applied logistic regression to the problem of predicting bankruptcy.

The goal of this paper is twofold: determine if the information obtainable from the financial statement can be used to predict the failure of a company and analyze financial statement items and ratios to find out which variables are the most determinant for the failure or the survival of a company.
This research focuses on the situation of firms located in the province of Lazio and covers a time frame of eleven years, from 2000 to 2011. The study has been carried out on a stratified sample of 100 firms from various economic sectors.

This paper differs from previous works since we used all the items of a financial statement as covariates and not only some performance ratios. Since the data at hand suffers from some serious issues we have decided to apply a very robust non-parametric classification technique such as classification trees (Breiman et al., 1984) in order to select those items that should be helpful in determining the failure of a company.

The paper unfolds as follows: in section 2 a brief description of the main issues related to the available data will be carried out. In section 3 the methodology used to analyze the data will be presented and the results will be presented in section 4. In section 5 some conclusions will be drawn.

2. Sample’s Issues

The stratified sample is composed of 100 companies randomly selected from those that, at year 2000, had revenue from sales between 2 million euros and 50 million euros. Of these firms, 50 were still active at 2011 while 50 had filed for bankruptcy sometime during the considered time frame.

The sample was stratified with respect to economic sector. No firms operating in the financial sector were included in the analysis, since they are known to react and operate differently from the other firms in case of distress. For each company the financial statement was available.

A financial statement describes the activities and the performances of a business throughout a specific period of time. It is composed by three main parts:

a) **Balance sheet**, that provides detailed information about assets, liabilities and shareholders

b) **Income statement** (or profit and loss account) shows the company’s revenues and expenses during a specific period of time and it is made of revenues and expenses

c) **Notes to financial statement** are additional information that further detail specific items as well as provide a more comprehensive assessment of a company’s financial condition.

In addition to the data of the financial statements, further information can be obtained from the financial statement to evaluate the profitability, solvency, liquidity and stability of a business.

Considering this very short summary of the content of any financial statement, it is clear that, using all the items of a financial statement as variables, yields a
dataset with an enormous number of variables and with a sparse structure in term of non-missing values.

Some other limitations of the data are:

a) Financial statements, prepared by the companies themselves, are the main source of external information, and most of the firm's performance evaluation is based on it. The final financial statement, therefore, may be the result of an adjustment that is performed within the boundaries of existing legislation to make it suitable to the particular and contingent needs of a company. Therefore some balance sheets not only reflect the financial and economic status of the firm but are also the outcome of a particular need that the firm is facing.

b) Although the firms that we analyzed are only small-medium firms, this category includes companies that can present a wide range of values for each item of the balance sheet: some with few employees and low absolute turnover, and some with several employees and high-volume business. This makes the sample data very heterogeneous and diverse.

Considering all these limitations and the small sample size, the application of a robust non-parametric methodology was necessary. Among the many models available in data mining we have opted for classification trees (Breiman et al., 1984), which are known to perform quite well in the presence of missing data and as a tool for feature selection when a large number of variables are available.

3. C&RTs

Classification and Regression Trees are a non-parametric statistical method, conceptually simple and yet powerful. They split the feature space into rectangles and then fit a simple model in each subset. Regression trees deal with continuous outcomes while classification trees consider a polytomous response variable (class) that is supposed to depend on a set of covariates.

They are particularly useful in data mining when there is a plethora of covariates with missing data. They are invariant to monotone transformations of the variables and are very robust with respect to outliers.

The goal of classification trees is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data at hand that has been previously classified (supervised classification).

Tree-based classification methods split the covariate space X into disjoint set of rectangular regions, and then classify the observations according to the mode of the class of the elements that belong to that region. The partitioning is usually performed according to an impurity measure (usually the Gini index) or according
to the information gain (entropy) that can be achieved once the covariate space has been partitioned. Therefore, starting with a single node (root) containing all the elements in the training data, we look for the binary partition that yields the best information gain or impurity reduction. The data are then partitioned according to the values of that covariate, yielding two subsets that stem from the root and which should be more homogenous according to the outcome than the set that generated them. This partitioning is recursively performed on the derived subsets and it stops either when the units in a node have all the same value of the response variable or when splitting no longer adds value to the predictions or the number of elements in the derived subset decreases under a pre-specified threshold.

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. The terminal 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.

Two main issues 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 will clearly perform well on the training data (overfitting), 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.

Classification trees, to our knowledge, have only been applied once to business failure in a study that did not produce reliable results due to a very small sample size (Huarng et al., 2005).

4. Experimental Setup and Results

The collected data refers to balance sheets of companies from 2000 to 2011. For the companies that are still active at 2011 all the balance sheets are available. The study that has been carried out is a cross-sectional study: companies have been considered at various years prior to failure and for each failed company the balance sheet up to eight years prior to failure (if available) has been considered. Each failed company has been matched to a non-failed company that, in the same year,
presented a balance sheet, operated in the same economic sector and was comparable in size. Therefore eight datasets were generated with equal number of failed and non-failed companies.

The performance of each classification tree was assessed via 10-fold cross validation. In Table 1 the resubstitution error rate (R-ER) and cross validation error rate (CV-ER) have been reported.

<table>
<thead>
<tr>
<th>YEARS PRIOR TO FAILURE</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>R-ER</strong></td>
<td>0.051</td>
<td>0.051</td>
<td>0.071</td>
<td>0.083</td>
<td>0.078</td>
<td>0.081</td>
<td>0.065</td>
<td>0.029</td>
</tr>
<tr>
<td><strong>CV-ER</strong></td>
<td>0.092</td>
<td>0.143</td>
<td>0.102</td>
<td>0.135</td>
<td>0.100</td>
<td>0.081</td>
<td>0.109</td>
<td>0.088</td>
</tr>
</tbody>
</table>

In Figures 1 and 2 some of the eight classification trees have been displayed. One variable that is always influential in determining the failure of a firm for all the trees that have been grown is the “financial charges on sales”: companies that are still active at 2011 present a ratio between financial expenses and revenues greater than those that have failed. This could at first seem counterintuitive since financial charges are considered a negative “asset” for the firms. This result must be embedded in the Italian economic framework, where access to the credit system is not as flexible as it should be in a healthy efficient economic system. “Financial charges on sale” can then be considered a proxy of the ability of a firm to access the credit system: high values for this variable indicate the ability of a company to access the credit system and therefore survive even after turnover's reductions.

In Figure 1, the decision tree shows that companies with a percentage of financial charges on sales over 15% are those who remained active, while those with less than 15% failed after eight years. These results are very similar to the seventh, sixth and fifth year preceding the bankruptcy, where the only discriminating variable was the relationship between financial expenses and revenues. The situation becomes more interesting from the fourth year up to one year before the bankruptcy, as other discriminating variables come into play.

Figure 2 shows the tree for one year prior to failure. Only a few years prior to failure the standard performance ratios come into play. Among the companies with the item "financial charges on sales" higher than 7%, 40 are still in business whereas 6 have started bankruptcy procedures: the next subdivision is given by ROS index that measures the return on sales. Firms with a ROS higher than 14.5% remained in activity, but those with a ROS less than 14.5% failed. This is a natural conclusion because the last balance sheet of a company before the bankruptcy
represents a situation where the actual activity of the company is already finished, so the "sales revenue" of these companies has a value of zero.

**Figure 1 – Decision tree 8 years before bankruptcy**

![Decision tree 8 years before bankruptcy](image1)

**Figure 2 – Decision tree 1 year before bankruptcy**

![Decision tree 1 year before bankruptcy](image2)

On the other hand, companies with a "financial charges on sales" less than 7%, are mostly companies that then have started a bankruptcy procedure (43), whereas few companies remained in business (9). The additional discriminating variable is given by the “financial proceeds and charges”: a very positive difference between
financial proceeds and charges has brought the survival of companies, whereas those who have had a minor discrepancy between financial proceeds and charges are going to fail.

The bottom-line of all the classification trees is that some companies continued to have access to credit and this allowed them to stay in business despite the crisis period; on the other hand, other companies were not able to be granted credit by banks and then carry out their production. Hence, they have been compelled, due to a sharp drop in sales resulting from a drastic reduction in production, to declare the state of crisis, and then bankruptcy.

5. Conclusions

In this paper we have used classification trees to predict firm bankruptcy based on all the items of the financial statement and some performance ratios at various years prior to failure. This is an unusual approach to the problem that has not been properly studied in business failure literature. We have found that the discriminant of business failure in Italy has been the inability of the companies to access the credit system in order to continue production. One conclusion is that the timely financial help of the banks is essential for all those companies who are in distress.

Bibliography

SUMMARY

Trigger Factors that influence Bankruptcy: a comparative and exploratory study

The phenomenon of bankruptcy has significantly influenced Italian productive environment in the last few years. The impact of the world financial crisis on the Italian economy has generated devastating effects for the wealth of the Nation, causing strong unbalances in terms of employment, productivity and investment in all sorts of industries.

The following study aims to study the financial and economic factors that have cause the failure of many firms in Italy, focusing the attention on the companies’ balance sheets.

This research focuses on the situation of Lazio’s companies and covers a time frame of eleven years, from 2000 to 2011 on a stratified sample of 100 firms, 50 of which are still in activity and 50 declared bankruptcy during the period 2000-2011. The attention will be focused on a cross-sectional study, considering firms at various years prior to failure.

Leonardo Di MARCO, UNINT, leonardo.dimarco88@gmail.com
Luciano NIEDDU, UNINT, l.nieddu@unint.eu