Default Prediction for Small-Medium Enterprises in France: A comparative approach

Sami BEN JABEUR
Youssef FAHMI

http://www.ipag.fr/fr/accueil/la-recherche/publications-WP.html

IPAG Business School
184, Boulevard Saint-Germain
75006 Paris
France
Default Prediction for Small-Medium Enterprises in France: A comparative approach

Sami BEN JABEUR, Adjunct Professor, IPAG Business School

Youssef FAHMI, Professor, University of South Brittany

sbenjabeur@gmail.com
fahmi.youssef@yahoo.fr

Abstract: The aim of this paper is to compare between three statistical methods in predicting corporate financial distress. We will use the Discriminant Analysis, Logit model and Random Forest. These approaches are based on a sample of 800 companies during the period from 2006 to 2008, as well as on the use of 33 financial ratios. The results show a superiority of the random forest approach.

1. Introduction

In actual economic situation an increasing number of firms are facing economic and financial difficulties which can, in certain cases, drive to failure. In fact, difficulties do not happen suddenly. Before a firm is declared bankrupt, it is confronted with financial difficulties of growing seriousness: default in payment of a debt, temporary insolvency, scarcity of liquidity, etc.

Identifying the causes of the failure is not obvious, since one cannot exhaustively enumerate the factors that cause it. The causes are multiple and overlapping they compromise even more for the company's survival. The importance of this phenomenon and its impact on the overall economy justifies the need to understand, and explain it by analyzing the causes and origins.

This paper joins within the framework of the research that works on the models used to forecast bankruptcies, which could be used to detect the money troubles of the SME (small and medium-sized enterprises).

The research on the forecast of the financial difficulties of companies are of a big importance for all the partners of a company. From a manager point of view, having tools of forecasting the financial failures allows to take in time, strategic measures and management corrective appropriate to prevent these failures (Hooted and al 2007). For other partners, such tools contribute to reduce the informative asymmetry with the company, to detect quickly the vulnerable companies and to optimize the allocation of their capital (financial, human and social). The forecast of the financial difficulties constitute a means of diagnostics of the performance of companies and ends in a classification of them among the failing or not failing companies. The forecast of failing companies can help in particular to prevent the difficulties before they are translated by a financial crisis and to set up the necessary measures (assistants, restructuring) before it is too late.

Before presenting the main methods of forecast used in this research work, it is useful to call back quickly the definitions of the financial distress.

2. Financial Distress

It was only after the crisis of the 1930s and the early work of writers such as Fitzpatrick (1932) that the problem of failure has become a field of investigation of research in its own right. According to Franks and Sussman (2005), a firm is defined as being in distress once the local branch or regional credit manager decides to transfer a status report to the monitoring unit of economic enterprises or responsible financial diagnosis. Such decisions may occur, especially for SMEs, in the case of violations of certain terms (non-payment of interest exceeding the overdraft limit ...), or following a poor assessment of the future of the firm by directors of credit (by reference to indicators such as high debt and low profitability).

Several surrounding areas of forecast of the financial distress of companies have been proposed since the end of the sixties, using financial ratios and accounting data. The purpose is to differentiate between the failing companies and the not failing companies and to build an explanatory model for the failure of companies. The most used techniques are the univariate approach, the multivariate approach, the discriminant analysis, the multiple regression, the logit regression, the models based on the artificial neuronal approach, etc.

In this work, we are going to apply the discriminant analysis, logit regression and random forest approach in the detection of the failing companies. The three approaches will be based on a
sample of 800 companies during the period between 2006 and 2008, as well as on a recourse to 33 financial ratios.

3. Methodological choice

3.1. The construction of the sample

Our approach to data collection involves several steps: the choice of the database, the selection of companies and the indicators of financial failure.

3.1.1. Presentation of data

We use the database DIANE (instant access to data from French companies for economic analysis) for our sample. This database provides access to a fund composed of more than one million businesses, as these are the companies that publish their annual accounts within the registries of registers commercial courts. Information on companies’ accounts is collected by coface services and is enriched with lots of related information, with ten years of historical accounts, Diane is the comprehensive basis of financial and general information about French companies.

3.1.2. The survey sample

A company was considered deficient if it has been a first event declaration to the judicial tribunal of commerce during 2009. The data studied is organized so that the accounting year is available for 2008, 2007 and 2006. The final sample obtained upon completion of this rigorous selection process consists of 800 companies split into two sub-samples: 400 healthy firms and 400 failing firms. This choice was self-imposed, because of constraints arising from the availability of identifying information of companies.

3.2. The selection of indicators and financial ratios

Selecting financial ratios has been a more logical and methodological choice in order to provide a relevant and credible battery capable of meeting the objectives and expectations of external reviews. Financial ratios were selected according to:

- their recurrence in French literature (Bardos, 1995, Bank of France’s work) and international literature (Altman, 1968, 1984, Conan and Holder 1979; Rose and Giroux, 1984, Remade, 2004).
- their relevance to financial analysis, incorporating the basic ratios in most existing models of detecting bankruptcy: liquidity ratios, profitability, management, productivity and financial structure. Thus, a series of 33 ratios (R01 to R33) was retained among those commonly used in literature and which have a significant informational content in the analysis of the financial situation of the company.

4. Results

With reference to the following table, there are some elements that help to choose the best model for predicting the failure of SMEs.

4.1. The classification of companies by the method of discriminate analysis (DA)

Table 1 shows the result obtained with discriminant analysis:

<table>
<thead>
<tr>
<th>Group Trust real</th>
<th>Nature of the firm</th>
<th>Allocation group under the model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>T-1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H</td>
</tr>
<tr>
<td>Number</td>
<td>healthy</td>
<td>376</td>
</tr>
<tr>
<td>Defaulting</td>
<td>9</td>
<td>391</td>
</tr>
<tr>
<td>%</td>
<td>healthy</td>
<td>94</td>
</tr>
<tr>
<td>Defaulting</td>
<td>2,2</td>
<td>97,8</td>
</tr>
</tbody>
</table>

H: healthy / D: defaulting
T-1: According to this table obtained using the DA method, the rate of correct classification of healthy companies is about 94%, an error of type 1 (percentage of failing firms considered healthy) of 2.2% and a type 2 error (percentage of non-failing companies considered risky) of 6%. According to this method, DA was able to identify 376 healthy firms, a rate of 94%, with nine failing firms classified as healthy.

T-2: In the group of healthy companies, the model holds 134 companies as being dysfunctional; in fact, they are really not. Also in the group of failing firms, 148 of them are considered healthy, which is not the case.

The study of the “two years before failure” model, reveals a decrease in classification rate of healthy firms from 94% to 66.5%; and 97.8% of failing companies to 63%. So the more the forecast is, the less the accuracy of the model is.

T-3: Looking at this table, the rate of correct classification of healthy companies is about 61.2%, and 57.2% for failing companies. A second type error of 38.8% and a first type error is about 42.8%.

Regarding the predictive power of the model in the span of time from one to three years before the failure, we find a drop in clearance rates in healthy firms from 94% (T-1) to 66.5% (T-2) and 61.2% (T-3). For the failing companies, the drop in the rates is from 97.8% (T-1) to 63% (T-2), and 57.2% (T-3). Thus, the greater the forecast horizon is far, the more the ability of the model is reduced.

4.2. The classification of companies by the logit method

The results of the classification of companies by the logit method are presented in the following table:

<table>
<thead>
<tr>
<th>1, 2 or 3 year Before the failure</th>
<th>Nature of the firm</th>
<th>Allocation group under the model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>healthy</td>
<td>S</td>
</tr>
<tr>
<td>GROUP REAL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trust</td>
<td>healthy</td>
<td>388</td>
</tr>
<tr>
<td></td>
<td>Defaulting</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>healthy</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td>Defaulting</td>
<td>1</td>
</tr>
</tbody>
</table>

H: healthy / D: defaulting

T-1: According to this table obtained using the logit method, the rate of correct classification of healthy companies is about 97%, an error of type 1 (percentage of failing firms considered healthy) of 4% and a type 2 error (percentage of non-failing companies considered risky) of 3%. According to this method, logit method was able to identify 386 healthy firms, a rate of 97%, with four failing firms classified as healthy.

T-2: In the group of healthy companies, the model holds 140 companies as being dysfunctional; in fact, they are really not. Also in the group of failing firms, 128 of them are considered healthy, which is not the case. The study of the “two years before failure model, reveals a decrease in classification rate of healthy firms from 97% to 65%; and 99% of failing companies to 66.5%. So the more the forecast is, the less the accuracy of the model is.

T-3: Looking at this table, the rate of correct classification of healthy companies is about 58%, and 63% for failing companies. A second type error of 42% and a first type error is about 37%.

4.3. The classification of companies by the random forest model

The results of the classification of companies by the random forest model are presented in the following table:
Table 3: Validation of random forest model

<table>
<thead>
<tr>
<th>1, 2 or 3 year Before the failure</th>
<th>Nature of the firm</th>
<th>Allocation group under the model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>T-1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S</td>
</tr>
<tr>
<td>Group real number</td>
<td>healthy</td>
<td>400</td>
</tr>
<tr>
<td></td>
<td>Defaulting</td>
<td>0</td>
</tr>
<tr>
<td>%</td>
<td>healthy</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Defaulting</td>
<td>0</td>
</tr>
</tbody>
</table>

H: healthy / D: defaulting

T-1: According to this table, the rate of correct classification of healthy companies is 100%, an error of type 1 (percentage of failing firms considered healthy) of 0% and a type 2 error (percentage of non-failing companies considered risky) of 0%. According to this method, random forest method was able to identify 400 healthy firms, a rate of 100%, with no failing firms classified as healthy.

T-2: In the group of healthy companies, the model holds no company as being dysfunctional. Also in the group of failing firms, only one of them is considered healthy, which is not the case. The study of the “two years before failure” model, reveals almost the same classification rate of healthy firms and of failing companies. So for (T-1) and (T-2), the model presents a good accuracy.

T-3: The results of (T-3) are similar to those of (T-2), a single healthy company ranks among the failing companies. On the other hand, failing companies are well categorized.

4.4. Comparison of three methods of classification

The following table summarizes the results of three methods: discriminant analysis, logit model and random forest method.

Table 4: The performance indicators and predictive power of the models chosen

<table>
<thead>
<tr>
<th></th>
<th>T-1</th>
<th>T-2</th>
<th>T-3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Discriminant analysis</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Bankrupt</td>
<td>94 %</td>
<td>66,5 %</td>
<td>61,2 %</td>
</tr>
<tr>
<td>Bankrupt</td>
<td>97,8 %</td>
<td>63 %</td>
<td>57,2 %</td>
</tr>
<tr>
<td>Per cent of correct classifications</td>
<td>95,975 %</td>
<td>64,475 %</td>
<td>59,2 %</td>
</tr>
<tr>
<td><strong>Logit</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Bankrupt</td>
<td>97 %</td>
<td>65 %</td>
<td>58 %</td>
</tr>
<tr>
<td>Bankrupt</td>
<td>99 %</td>
<td>66,5 %</td>
<td>63 %</td>
</tr>
<tr>
<td>Per cent of correct classifications</td>
<td>98 %</td>
<td>66,5 %</td>
<td>60,5 %</td>
</tr>
<tr>
<td><strong>Random Forest</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Bankrupt</td>
<td>100 %</td>
<td>100 %</td>
<td>99,75 %</td>
</tr>
<tr>
<td>Bankrupt</td>
<td>100 %</td>
<td>99,75 %</td>
<td>100 %</td>
</tr>
<tr>
<td>Percentage of correct classifications</td>
<td>100 %</td>
<td>99,875 %</td>
<td>99,875 %</td>
</tr>
</tbody>
</table>
The comparison of the overall performance of the three methods shows a clear superiority to the random forest (RF) method. A year before the failure, the RF method is more effective than discriminant analysis (DA) and logit model. The RF method leads to a correct classification rate of 100%, against 98 for the logit method and 95.975% for the DA approach. Two years before the failure rates were 99.875% for the RF method, and 66.5% for the logit method, and 64.475 for DA approach. Finally, three years before failure, rates are respectively 99.875, 60.5% and 59.2% for the three methods.

On the other hand, if the RF method has the best results compared to the other two methods, one notices that the logit method gives the best results over the DA method in terms of classification of failed or healthy companies.

5. Conclusion

In this study, we tested three methods of forecasting the bankruptcy of enterprises. We conducted our research on a sample of 800 SMEs, composed of 400 healthy companies and 400 defaulting companies. The results showed a clear superiority of the random forest method on the discriminant analysis and logit methods. At the same time, we found better results of the logit method compared with discriminant analysis. The random forest model gives better results in terms of classification of performing or defaulting companies. It allows a better forecast accuracy, since it minimizes the error type 1 and error type 2.

The results of this work have academic implications in terms of research on prediction models, but also empirical implications to assist stockholders to be informed about the financial health of the companies in which they have a stock.

Bibliography
