Predicting Bankruptcy of Belgian SMEs: A Hybrid Approach Based on Factorial Analysis

Loredana Cultrera¹, Mélanie Croquet¹, Jérémy Jospin²

¹Warocqué School of Business and Economics, University of Mons, Belgium and Human Org Research Institute, Belgium
²MAINSYS, Belgium

Correspondence: Loredana Cultrera, Warocqué School of Business and Economics, University of Mons, Belgium and Human Org Research Institute, Belgium. E-mail: loredana.cultrera@umons.ac.be

Received: December 8, 2016 Accepted: January 12, 2017 Online Published: February 8, 2017
doi:10.5539/ibr.v10n3p33 URL: http://dx.doi.org/10.5539/ibr.v10n3p33

Abstract

The aim of this paper is to verify the relevance of technical data analysis which seems to be useful for identifying predictors of bankruptcy of Belgian SMEs. To do so, a sample of 1,860 Belgian companies, including healthy and bankrupt firms, was used. The sample was constituted using Belfirst software (2015). A mixed method data analysis, coupling the Ward aggregation criterion, the method of mobile centres and principal component analysis, was performed on the variables commonly cited in the literature as predictive of corporate bankruptcies. The results of this study show that the use of these methods is not relevant in the context of bankruptcy prediction using this sample, but the results of the logistic regressions did not question the discriminatory power of the introduced active variables.

Keywords: bankruptcy, principal component analysis, bankruptcy prediction models, classification, SME

1. Introduction

The prediction of the bankruptcy phenomenon is based on the work of Fitzpatrick (1932), but the pioneers in this field are Beaver (1966) with one-dimensional methodology, and Altman (1968) with discriminant analysis. Subsequently, discriminant analysis and other techniques, such as logit and probit models (Ohlson, 1980; Zmijewski, 1984), or neural networks (Odom and Sharda, 1990), have exploited the various financial data necessary to predict bankruptcy. In Belgium, different models have been applied, but the discriminant analysis of Ooghe and Verbaere (1982) and of Ooghe et al. (1991) are the two best known. These models aim to predict bankruptcy as soon as possible to prevent the company's failure and minimise the costs associated with a possible bankruptcy (Van Caillie and Dighaye, 2002).

Many Belgian authors have also focused on the preventive aspect, rather than on the predictive aspect, of bankruptcy (Van Caillie, 2004, Crutzen and Van Caillie, 2010). Indeed, the financial approach to bankruptcy is mainly based on bankruptcy prediction models relying on accounting ratios that can be considered and analysed as they reflect the symptoms arising from the real causes of the failure. Three other approaches have been developed (Guilhot, 2000): the economic approach, the strategic/organisational approach and finally, the managerial approach. These appear at earlier stages of the process of bankruptcy, and therefore characterise deeper causes. These bankruptcy processes allow the identification of the causes and their patterns, as well as the determination of ways to prevent bankruptcy (Levratto, 2011).

However, despite the many existing bankruptcy prediction models, some methodologies are not used by researchers. This is particularly the case for multivariate statistical methods, such as factorial analysis or classification algorithms. It was only recently that some authors investigated the case of SMEs for the application of these prediction models (Donckels, 1984; Keasey and Watson, 1991) as the models are mainly used on larger firms (Levratto, 2011). However, while the study of SMEs is attractive because of their characteristics, such as their ability to respond quickly and creatively thanks to their greater flexibility (Bellanca et al., 2015), analysing SMEs requires facing many obstacles, including the lack of data, which is certainly the biggest problem (Van Caillie, 1993).
The objective of this research is twofold:

- Improve the knowledge of bankruptcy prediction for Belgian SMEs
- Analyse the predictive capacity of factorial analysis using in particular the principal components analysis, coupled with classification algorithms. The goal here is to identify if the obvious underuse of these techniques is the result of a lack of recognition from the scientific community or the lack of performance of these methods.

2. Literature Review

Different prediction models have been used for decades when it comes to predicting the risk of bankruptcy. These models include parametric statistical classification models (Beaver, 1966; Altman, 1968; Deakin, 1972, Ohlson, 1980; Mensah, 1984; Platt and Platt, 1991; Zmijewski, 1984; Gentry et al., 1985; etc.) non-parametric ones (Frydman et al., 1985; Calia and Ganucci, 1997; etc.), models based on artificial intelligence (Varetto, 1998; Barney et al., 1999), on the risk index (Tamari, 1966), on the behaviour analysis of titles (Beaver, 1968; Aharony et al., 1980) and SIM models (Ooghe et al., 2005). Although this list is not exhaustive, it represents most of the models used.

Alongside the more "famous" models briefly presented above, other existing methods are less used in the field of failure prediction due to the fact that they are less recognised or have lower performance. This is particularly the case for factorial analysis, including principal components analysis (PCA) and multiple correspondence analysis (MCA). Although they are very similar in their objectives, that is to say they aim to reduce a large number of observed variables into a small number of latent variables, these methods are characterised by the continuous or nominal character of the observed variables. Indeed, PCA is conducted as part of the continuous reduction of observed variables while MCA is used to reduce the nominal observed variables. In the following section, certain studies into bankruptcy predictions that use factorial methods, such as PCA or MCA, will be identified.

PCA is mainly used in combination with neural networks (Mensah, 1984; Zavgren, 1985; Gombola et al. 1987; Skogsvik 1990; Alici, 1996; Sharma and Iselin, 2003; Shin and Lee, 2003; Wang, 2004; Canbas et al., 2005; Min and Lee, 2005; Tang and Chi, 2005; Shin et al., 2006; Sookhanaphibarn et al., 2007; Yao, 2007; Ravi and Pramodh, 2008; Chen and Du 2009; Balas et al., 2010). Indeed, PCA is one of the most popular data selection methods (Zhang, 2000; Li and Sun, 2011). The reason why PCA is very often associated with neural networks is that the linear limitation of this method can be overcome thanks to the neural networks that can capture nonlinear relationships (Zhang, 2000). Pump and Bilderbeek (2005) based their analysis on PCA to predict the bankruptcy of Belgian companies using a neural network.

Other authors, such as Armeanu et al. (2012), have used PCA to determine different stages of insolvency (monthly repayments delays, payment default, default fee or default interest, etc.). Takahashi and Kurokawa (1984) and Laitinen (1991) also used PCA on failing firms that filed accounts three years before bankruptcy in order to achieve the failing companies’ profiles.

The authors Li and Sun (2009) associated PCA with the case-based reasoning method (CBR) in the context of the bankruptcy prediction of listed Chinese companies. They mixed these methods to improve the performance of CBR. In a subsequent study, Li and Sun (2011) together developed a hybrid model using PCA, multivariate discriminant analysis (MDA) and logistic regression. Having compared their hybrid model to models that do not use PCA, the authors concluded that PCA gives better results.

Van Caillie (1993) sought to assess the contribution of MCA to the detection of bankruptcy warning signs. He said that this method applies rather well for SMEs, as it allows the drawbacks of conventional models to be overcome. In his study two axes were used to synthesise of all the variables, while losing as little information as possible. Furthermore, a factorial representation of the financial behaviour of SMEs was developed. The author was able to go from prediction to prevention by analysing the movement over time between different classes of companies, identified by MCA, and by observing the change in various components.

Finally, Crutzen and Van Caillie (2010) conducted a typology of bankrupted Belgian SMEs with MCA. A cluster analysis (conducted using the method of Ward) first allowed them to gather businesses with similar characteristics. Correspondence analysis was then used to determine the variables that could be associated with each group. Two dimensions were used: the first opposed internal and external failure factors, and the second characterised the adaptability of the firms to their environment.

Thus, factorial analysis (through MCA, PCA or other factorial methods), when mobilised by researchers to improve bankruptcy prediction models are often used together with other methodologies. As far as is known, the principal components analysis have never been used before using the Ward criterion and mobile centres as part
of bankruptcy prediction. But these methodologies, if used simultaneously, could help to identify classes of homogeneous businesses, when also combined with other features used in the analysis. The bankrupt or non-bankrupt nature of the companies would therefore be a feature that could potentially explain the obtained classes. The joint use of PCA, the Ward criterion and mobile centres could therefore serve the cause of variable selection in the field of forecasting bankruptcy, and subsequently allow the calculation of bankruptcy probability.

3. Method

3.1 Data

A sample of 1,860 unlisted bankrupt and non-bankrupt Belgian SMEs was developed with the Belfirst software, which is published annually by the Bureau van Dijk. The objective was to focus on SMEs, as such only firms employing fewer than 100 people were selected. Thereafter, only the firms filing their annual accounts in abbreviated format were kept to improve the uniformity of the data (many SMEs opt to present their accounts in an abbreviated format because it is less restrictive than the full format).

The bankrupt companies were selected on the basis of the previous year’s annual accounts. In the case of this study, the last year was 2012. To analyse the five years before the bankruptcy, only companies that filed accounts between 2007 and 2011 were selected. The final sample included 930 bankrupt SMEs.

To ensure the integration of healthy SMEs in the sample the applied logic was the same. However, here, the last year of filed accounts was chosen as 2014, to avoid the possibility of the 2012 filings coming from firms which went on to fail in 2013 or 2014. The year 2014 was the last year of account available in the Belfirst software. The aggregate sample included 2,113 healthy SMEs among which 930 SMEs were randomly selected to obtain a balanced final sample.

3.2 Methodological Choices

To achieve the objective of the study, a mixed classification method was chosen. This method combines a hierarchical ascending method, in this case the Ward aggregation criterion, and a non-hierarchical method, namely the mobile centres, to take the advantages of both methods and to reduce their disadvantages. The application of these methods was preceded by PCA on the active variables. These variables are those most commonly cited in the literature as predictors of bankruptcy:

- Equity / Total Assets,
- Cash Flow / Debts,
- Current ratio,
- Tax, wage and social debts / Net value added,
- EBIT / Total Assets.

These variables characterise the solvency, profitability, corporate liquidity, and also the tax, which is relatively high in Belgium. These ratios were used by Belgian authors (Ooghe and Van Wymeersch, 2000; Pump and Bilderbeek, 2005) and international authors (Altman, 1968; Taffler 1982; Frydman et al., 1985) and directly apply to the case of SMEs because of their availability in abbreviated accounts.

These variables were retained for the years 2007, 2009 and 2011 for all 1,860 firms in the sample. For firms that went bankrupt in 2012, the years 2007, 2009 and 2011 are the years which precede the bankruptcy by five years, three years and one year respectively.

The final database included 15 variables. Finally, the main variable that was retained to characterise the obtained classes was the nominal variable bankrupt / not bankrupt. This is the variable that eventually identifies a failing company’s profile.

SPAD software was used to carry out this classification.

4. Results

Six principal components were selected after the analysis of eigenvalues (Table 1). These values explain 57.15% of the total variance.
Table 1. Eigenvalues of factorial axes (extract)

<table>
<thead>
<tr>
<th>Principal Component</th>
<th>Eigen value</th>
<th>%</th>
<th>% cumulative</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.2362</td>
<td>14.91</td>
<td>14.91</td>
</tr>
<tr>
<td>2</td>
<td>1.8123</td>
<td>12.08</td>
<td>26.99</td>
</tr>
<tr>
<td>3</td>
<td>1.4208</td>
<td>9.47</td>
<td>36.46</td>
</tr>
<tr>
<td>4</td>
<td>1.0608</td>
<td>7.07</td>
<td>43.53</td>
</tr>
<tr>
<td>5</td>
<td>1.0412</td>
<td>6.94</td>
<td>50.48</td>
</tr>
<tr>
<td>6</td>
<td>1.0009</td>
<td>6.67</td>
<td>57.15</td>
</tr>
<tr>
<td>7</td>
<td>0.9979</td>
<td>6.65</td>
<td>63.80</td>
</tr>
</tbody>
</table>

The analysis of the coordinates of the active variables on the principal component allows the characterisation of these axes. The principal components were retained if their eigenvalue was above 1. Then, the Ward aggregation criterion and mobile centres were combined on the factorial scores of the active variables on the first six principal components. These methods identified five classes. Nevertheless, it appears that the class sizes are very different (see Table 2).

Table 2. Class sizes

<table>
<thead>
<tr>
<th>Class</th>
<th>Number of firms</th>
<th>Percentage of the total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1781</td>
<td>95.75%</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0.11%</td>
</tr>
<tr>
<td>3</td>
<td>69</td>
<td>3.71%</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>0.215%</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>0.215%</td>
</tr>
</tbody>
</table>

This imbalance in terms of numbers is a bad sign for either the relevance of these variables in forecasting bankruptcies, which would go against the empirical literature, or the suitability of this methodology in identifying predictors of bankruptcy.

Still, the classification was continued by attempting to identify the active variables of the two largest groups, those that characterise most of these classes. This table only contains the most characteristic variables, those whose test-values are greater than 2. That is, in order to be considered as relevant characteristics, the test-values of the variables is the principal criteria that will be used. The test value makes it possible to evaluate the interest of a variable in the characterization of a class of individuals on the basis of a statistic calculated on the sample and on the class. This statistic is often the mean. Thus, if the difference between the average of a variable calculated on the individuals of the class and the mean of this variable calculated on the sample is due to chance then, this variable does not characterize the class. The test value is expressed as the number of standard deviations of a normal distribution. For example, a test value of 8 for a variable can be interpreted as follows: the observed deviation is equal to the probability of drawing an observation at 8 standard deviations in a normal distribution.

Table 3. Classes 1 and 3: description by the active variables

<table>
<thead>
<tr>
<th>V. Test</th>
<th>P-value</th>
<th>Means</th>
<th>Standard deviations</th>
<th>Characteristic variables</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Class</td>
<td>General</td>
<td>Class</td>
</tr>
<tr>
<td>8.51</td>
<td>0.000</td>
<td>0.15</td>
<td>-0.45</td>
<td>2.60</td>
</tr>
<tr>
<td>7.67</td>
<td>0.000</td>
<td>0.16</td>
<td>0.02</td>
<td>1.02</td>
</tr>
<tr>
<td>6.49</td>
<td>0.000</td>
<td>0.01</td>
<td>-0.04</td>
<td>0.30</td>
</tr>
<tr>
<td>6.25</td>
<td>0.000</td>
<td>0.03</td>
<td>-0.15</td>
<td>0.27</td>
</tr>
<tr>
<td>6.16</td>
<td>0.000</td>
<td>0.07</td>
<td>-0.42</td>
<td>1.56</td>
</tr>
<tr>
<td>4.32</td>
<td>0.000</td>
<td>0.07</td>
<td>-0.03</td>
<td>3.15</td>
</tr>
<tr>
<td>2.45</td>
<td>0.007</td>
<td>0.00</td>
<td>0.37</td>
<td>0.42</td>
</tr>
<tr>
<td>-3.36</td>
<td>0.000</td>
<td>1.23</td>
<td>1.23</td>
<td>2.69</td>
</tr>
<tr>
<td>-6.80</td>
<td>0.000</td>
<td>1.87</td>
<td>1.87</td>
<td>2.24</td>
</tr>
<tr>
<td>-26.98</td>
<td>0.000</td>
<td>2.74</td>
<td>2.74</td>
<td>2.93</td>
</tr>
<tr>
<td>-27.21</td>
<td>0.000</td>
<td>2.87</td>
<td>2.87</td>
<td>2.80</td>
</tr>
<tr>
<td>-28.09</td>
<td>0.000</td>
<td>2.85</td>
<td>0.85</td>
<td>2.69</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>V. Test</th>
<th>P-value</th>
<th>Means</th>
<th>Standard deviations</th>
<th>Characteristic variables</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Class</td>
<td>General</td>
<td>Class</td>
</tr>
<tr>
<td>29.79</td>
<td>0.000</td>
<td>27.61</td>
<td>2.85</td>
<td>21.86</td>
</tr>
<tr>
<td>29.17</td>
<td>0.000</td>
<td>28.97</td>
<td>2.74</td>
<td>23.32</td>
</tr>
<tr>
<td>28.88</td>
<td>0.000</td>
<td>27.98</td>
<td>2.87</td>
<td>23.60</td>
</tr>
<tr>
<td>-4.62</td>
<td>0.000</td>
<td>-3.53</td>
<td>-0.03</td>
<td>24.47</td>
</tr>
</tbody>
</table>
Consequently, the variables for which the test value will be greater than 2 in absolute value will be considered as significantly characteristic of a class because the p-value will be smaller than the significance threshold of 5%. The higher the test value associated with a variable in absolute value, the more this variable characterizes the class of individuals.

Table 3 shows the description of classes 1 and 3 by their active variables.

It appears that the firms in the first class had significantly higher levels of solvency than the general mean calculated using all the firms. Note that these average levels are negative in 2007 and 2011, and slightly positive in 2009. This finding is certainly the translation of the consequences of the recent financial crisis because half the companies in the sample went bankrupt in 2012.

It was also noted that the firms in class 1 had mean liquidity levels lower than average levels calculated using all the firms for the three years. Similarly, it appears that these companies had a lower average level of cash flow / debts than the general average for the years 2007 and 2009, five years and three years before the bankruptcy of the firms in the sample. Conversely, the class 3 companies had higher levels of liquidity than the rest of the sample for the three selected years.

Finally, it seems that the variable “taxes, wages and social debts / Net value added” is also characteristic of classes 1 and 3. Indeed, the firms in class 1 appear to have a higher mean value than the average for this ratio in 2009, while the opposite occurred for the class 3 companies.

To complete the typology, a dichotomous variable “bankruptcy or not” was integrated as an illustrative variable. Table 4 characterises class 3 by the two modalities of this dichotomous variable. Modality 1 of this variable represents failing firms while modality 2 represents the healthy firms. Thus, it seems that the most liquid firms in 2007, 2009 and 2011 showed a higher probability of not defaulting in 2012. Liquidity is therefore an important variable in the field of bankruptcy prediction.

Table 4. Characterisation of Class 3 by the modality of the illustrative variable « Bankrupt »

<table>
<thead>
<tr>
<th>V.Test</th>
<th>P-value</th>
<th>Percentages</th>
<th>Modality</th>
<th>Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CLA/MOD</td>
<td>MOD/CLA</td>
<td>GLOBAL</td>
</tr>
<tr>
<td>3.71</td>
<td>0.000</td>
<td>6.67</td>
<td>89.86</td>
<td>50.00</td>
</tr>
<tr>
<td>7.05</td>
<td>0.000</td>
<td>6.67</td>
<td>89.86</td>
<td>50.00</td>
</tr>
<tr>
<td>-7.05</td>
<td>0.000</td>
<td>0.75</td>
<td>10.14</td>
<td>99.25</td>
</tr>
</tbody>
</table>

After this typology, it can be noted that the results are quite limited in the context of bankruptcy prediction, with respect to the previous empirical results. This finding may be explained in two different ways. First, it could potentially qualify the low predictive power of the bankruptcy variables selected, even if this goes against the empirical literature on the subject (in particular, the findings of Altman, 1968; Frydman et al., 1985; Ooghe and Van Wymeersch, 2000; Pompe and Bilderbeek, 2005). Then, the relevance of using this methodology in the context of bankruptcy prediction could also be questioned. To remove any doubt about the preliminary variable selection, a parallel study using logistic regression was conducted.

STATA software enabled this logistic regression to be carried out on the variables from one, three and five years before bankruptcy. Based on a sample of 930 failing firms and 2,113 healthy companies, that is to say 3,043 raw data, the regression used a final sample of 2,668 exploitable companies after data cleansing.

Table 5. Results of logistic regression – one year before bankruptcy

<table>
<thead>
<tr>
<th>Variables</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equity/Total Assets 2011</td>
<td>-0.372*** (0.039)</td>
</tr>
<tr>
<td>Cash Flow/Debts 2011</td>
<td>-0.281*** (0.108)</td>
</tr>
<tr>
<td>Current ratio 2011</td>
<td>-0.274*** (0.040)</td>
</tr>
<tr>
<td>EBIT/Total Assets 2011</td>
<td>-3.072*** (0.314)</td>
</tr>
<tr>
<td>Tax, wage and social debts/Net value added 2011</td>
<td>-0.018*** (0.004)</td>
</tr>
<tr>
<td>Sig. model (p-value)</td>
<td>0.000</td>
</tr>
<tr>
<td>Number of observations firm-year</td>
<td>2,668</td>
</tr>
</tbody>
</table>

Note. Standard errors are reported in brackets; ***, **, * significant at 1, 5 and 10% respectively.
According to the results of previous studies, the results of logistic regressions highlight the predictive nature of the five variables introduced into the initial model. Therefore, the poor results obtained, following the classification of factors, confirms the relative irrelevance of this methodology in the context of bankruptcy prediction.

Table 5 lists the results of the logistic regression, one year before the bankruptcy. The results are substantially identical for the logistic regressions conducted for three and five years before bankruptcy.

The results of the logistic regression for the year 2011 appear to be very satisfactory. Indeed, the chi square maximum likelihood related to the model is very high, and the resulting p-value indicates that the model is statistically significant at 1%. The five variables included in the model are statistically significant, at a threshold of 1%. The sign associated with each coefficient is negative, which means that the higher the significant value for each variable for a company, the lower the likely risk of bankruptcy will be.

5. Discussion

Owing to its importance for the economy, bankruptcy is one of the most discussed topics in the literature. The costs that bankruptcy may generate have led many authors to develop bankruptcy prediction models (Beaver, 1966; Altman, 1968; Ohlson, 1980).

However, these models have the disadvantage of limiting companies to being categorised as either bankrupt or not, and do not consider the different stages the company can go through before bankruptcy is declared. These stages have also led to different definitions of failure, it may be economic, financial, organisational or legal (Crucifix and Derni, 1992; Gresse, 1994; Quintart, 2001; Hol et al., 2002; Grégoire, 2012).

When it comes to studying the phenomenon of bankruptcy, different approaches have also been developed because of the multitude of factors that can be involved. These are economic, financial, strategic/organisational, and managerial approaches.

By focusing on the readily available financial statements, problems of profitability, solvency and liquidity, which a company may face, can be highlighted. It is from these that most predictive models, such as univariate analysis, linear discriminant analysis, logistic regression, recursive partitioning and neural networks, are built. In Belgium, the most successful models are the linear discriminant analysis of Ooghe and Verbaere (1982) and the logit model of Ooghe et al. (1991). These models were firstly used for the case of large companies with greater visibility. It is only later that SMEs, representing the majority of the economic landscape, have attracted the attention of researchers.

In recent years, various authors (Li and Sun, 2009, 2011) have continued to develop new methods to better predict bankruptcy. PCA is one of the methods that had never been used before as a bankruptcy prediction tool. However, this technique can be applied to qualitative variables and, therefore, adapts quite well to the deeper causes of bankruptcy. Therefore, it was used to determine different types of bankruptcy (Crutzen, 2010).

This article aims to combine this technique with the Ward classification methods and mobile centres, making principal component analysis a real bankruptcy prediction tool. Prior to this study, this method had never been used to discriminate failing firms from non-failing ones. This is, therefore, a real innovation in the field of bankruptcy prediction. This combination of methods was performed using SPAD software and aimed to form various classes. Unlike conventional prediction models, more than two final classes were obtained.

To validate this methodology, a sample of 930 Belgian failing firms and 930 non-failing firms was analysed. Five ratios representing the solvency, profitability, added value, liquidity and the Belgian tax system were identified five years, three years and one year before the bankruptcy. The methodology has identified six factorial axes having eigenvalues greater than 1. Representation on the first two axes (for axes 2 and 3 also) pointed out a high concentration of companies at the origin of the axes, not allowing for healthy firms to be adequately discriminated from failing ones, from the variables initially selected. However, the choice of variables should not be called into question because, on the one hand, they are used in many articles in the literature (Altman, 1968; Taffler, 1982; Ooghe and Van Wymeersch, 2000; Pompe and Bilderbeek, 2005), and on the other hand, the discriminating power of these variables was confirmed using logistic regression (most variables one year, three years and five years before bankruptcy were significant). The principal components analysis method associated with the Ward criterion and mobile centres applied to the prediction of bankruptcy was not found to be effective for determining a sufficient number of classes composed of a majority of healthy or failing firms.

One way to extend this study could be to use the pairing method, to pair firms by sector or by legal form. This should provide a more homogeneous sample, with the hope of improving the results.
References


**Copyrights**

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).