Received 7 April 2018
Revised 27 July 2018
3 November 2018
Accepted 9 January 2019

A combined approach based on robust PCA to improve bankruptcy forecasting

Marianna Succurro
Department of Economics Statistics and Finance "Giovanni Anania", University of Calabria, Rende Cosenza, Italy

Giuseppe Arcuri
University of Paris Nanterre, Nanterre, France, and

Giuseppina Damiana Costanzo
Department of Business and Judicial Sciences, University of Calabria, Rende Cosenza, Italy

Abstract

Purpose – Starting from a series of financial ratios analysis, this paper aims to build up two indices which take into account both the firm's debt level and its sustainability to investigate if and to what extent the proposed indices are able to correctly predict firms' financial bankruptcy probabilities.

Design/methodology/approach – The research implements a statistical approach (tandem analysis) based on both an original use of principal component analysis (PCA) and logit model.

Findings – The econometric results are compared with those of the popular Altman Z-score for different lengths of the reference period and with more recent classifiers. The empirical evidence would suggest a good performance of the proposed indices which, therefore, could be used as early warning signals of bankruptcy.

Practical implications – The potential application of the model is in the spirit of predicting bankruptcy and aiding companies' evaluation with respect to going-concern considerations, among others, as the early detection of financial distress facilitates the use of rehabilitation measures.

Originality/value – The construction of the indebtedness indices is based on an original use of Robust PCA for skewed data.

Keywords Bankruptcy, Robust PCA, Logit, Z-score

Paper type Research paper

1. Introduction

Owing to the international financial crisis, both the number and the average size of bankrupt firms have increased dramatically with consequent greater interest from governments, financial institutions and regulatory agencies.

A correct measure of firms' insolvency risk is very important for both an internal monitoring purpose and the potential investors, stockholders and firm's competitors. The purpose of this study is to construct, analyze and test a new bankruptcy prediction model which can be easily applied as an early warning instrument. The potential application of our

JEL classification – C40, D22, G33, M41

The authors are particularly grateful to Nadine Levratto and two anonymous referees for their useful comments and suggestions.

Review of Accounting and Finance
Vol. 18 No. 2, 2019
pp. 296-320

1475-7702
DOI 10.1108/RAF-04-2018-0077

model is in the spirit of predicting bankruptcy and aiding companies' evaluation with respect to going-concern considerations, among others, since the early detection of financial distress facilitates the use of rehabilitation measures. Insolvency is mostly a consequence of a sharp decline in sales which can be caused by several and different factors like a recession, management deficiencies, important changes in market dynamics, shortage of a raw material, changes in lending conditions, etc. An early warning signal of bankruptcy will allow for the adoption of preventive and corrective measures. Hence, our study aims to contribute to the elaboration of efficient and effective corporate failure prediction instruments to prevent bankruptcy through the adoption of reorganization strategies. Failure, indeed, is not identifiable in a specific episode but in a process of progressive worsening of the financial health of a company.

Our study contributes to the literature in several ways. First, we attempt to improve the research model by implementing a statistical approach (tandem analysis) based on both an original use of robust principal component analysis (RPCA) and logit model. We demonstrate that our combined method of RPCA and logit estimation is promising in evaluating firms' financial conditions. Second, we keep and then analyze separately the debt structure of the firm and its sustainability to avoid a masking effect potentially resulting from an analysis that does not discriminate among variables related to different aspects of the same phenomenon. Third, we attempt to evaluate both the effectiveness and the efficiency of our model, i.e. its economic and organizational usability in an operational context (Cestari *et al.*, 2013). Fourth, logistic regression estimates are compared with those of the popular Altman Z-score for different lengths of the reference period. In addition to several models that have been tested by the relatively short one-year prediction horizon, we test the predictive power of our model several years prior to bankruptcy. Hence, we propose an approach which can be used to catch early warning signals of bankruptcy. Finally, the paper reports an application to Italian manufacturing firms. As a small sample size appears to be a limitation and "... any new model should be as relevant as possible to the population to which it will eventually be applied" (Altman, 1977), we consider the whole population of Italian manufacturing companies including small, medium and large firms.

The paper is organized as follows: Section 2 summarizes the related literature; Section 3 illustrates our methodology; Section 4 shows an application to Italian manufacturing firms and illustrates the empirical findings; Section 5 shows the reliability of the model; Section 6 illustrates the hazard model estimates; and Section 7 concludes.

2. Literature review

Bankruptcy has been the subject of numerous studies over the past few years[1]. Researchers have investigated both the causes and the legislative and financial tools available to start a process of recovery/rehabilitation of the firm. Especially after the international financial crisis, there has been a general need to predict insolvency and financial failure on time to take corrective and remedial measures for protecting business from the problem of bankruptcy.

A broad international field of study has focused on predicting bankruptcy using statistics and financial indicators. Prior to the development of quantitative measures of company performance, agencies were established to supply qualitative information assessing the creditworthiness of firms. During the 1930s many models were developed to help banks decide whether or not to approve credit requests (FitzPatrick, 1932; Smith and Winakor, 1935; Wall, 1936). Bellovary *et al.* (2007) traced a brief historical summary of the early studies (1930 to 1965) concerning ratio analysis for bankruptcy prediction that laid the groundwork for the studies that followed.

At the end of the 1960s, several applications of univariate and multivariate statistical methods were developed. One of the classic works in the area of ratio analysis and bankruptcy classification was performed by Beaver (1968). His univariate analysis of a number of bankruptcy predictors set the stage for the multivariate analysis applied by Altman (1968) and Deakin (1972), followed by several authors (Blum, 1974; Elam, 1975; Libby, 1975; Wilcox, 1976; Altman, 1977; Taffler, 1982; Appetiti, 1984; Flagg *et al.*, 1991; Shumway, 2001; Agarwal and Taffler, 2011). In his seminal study on bankruptcy detection, Altman (1968) improved research methodology by use of multiple discriminant analysis (MDA) where the discrimination was determined by a score – the «*Z-score*»–calculated on the basis of five accounting ratios.

After Altman's seminal study, the linear discriminant analysis has been intensively used in practice mainly because of the simplicity of its application. However, Joy and Tollefson (1975) have criticized the excessive broadness of the so-called gray area and the difficulty of application in predicting bankruptcy *ex ante*. Others have questioned that MDA implies the respect of some strict statistical restrictions such as the normality of the distribution of the explanatory variables and requirement for the same variance-covariance matrices for both groups of bankrupt and non-bankrupt companies.

As a consequence, later studies have tried to upgrade the methodology and improve the predictive power of the models. Several authors have used logit and probit models – instead of MDA – according to whether the residuals follow a logistic or normal distribution. Ohlson (1980) was the first to use the logit model, followed by several authors (Mensah, 1984; Zavgren, 1985; Burgstahler *et al.*, 1989; Flagg *et al.*, 1991; Platt and Platt, 1991; Johnsen and Melicher, 1994; Nam and Jinn, 2000; Foreman, 2003; Vuran, 2009). In other studies, the probit models were implemented (Zmijewski, 1984; Gentry *et al.*, 1985; Lennox, 1999). Similar methodologies – like duration models – have been developed to consider several periods in the analysis (Shumway, 2001; Beaver *et al.*, 2005; Duffie *et al.*, 2007). The recent empirical evidence indicates that prediction of insolvency and credit risk management can be improved by incorporating also nonfinancial information (management, employees, clients, industry, etc.) in failure prediction models. Nevertheless, only few papers (Grunert *et al.*, 2004; Berk *et al.*, 2010; Pervan and Kuvek, 2013) explicitly use non-financial variables to predict failure.

More recently, some authors have resorted to artificial intelligence expert system (AIES) models for bankruptcy prediction. Several types of AIES models have been implemented such as recursively partitioned decision trees, case-based reasoning models (Kolodner, 1993), neural networks (Odom and Sharda, 1990; Yang *et al.*, 1999; Kim and Kang, 2010), genetic algorithms (Varetto, 1998; Shin and Lee, 2002), rough sets model (Dimitras *et al.*, 1999) or "new age" classifiers. Ravi Kumar and Ravi (2007) present a comprehensive review of the work done in the application of intelligent techniques showing the basic idea, advantages and disadvantages for each methodology. In some circumstances these AIES models would slightly outperform classical models like discriminant and logistic analysis (Behr and Weinblat, 2016; Jones *et al.*, 2017) but they are based on complex underlying model structures.

Note that another contemporaneous branch of literature, likewise important but different from a methodological point of view, aims at forecasting default risk for publicly traded companies.

More specifically, the literature on bankruptcy includes two different classes of default models.

A first class of models, based on Merton (1974) and more recently on Vassalou and Xing (2004), compute default likelihood indicators (DLI) for listed companies. They use the market value of a firm's equity and an estimate of the market value of debt – instead of the book value of debt, as the accounting models do – in calculating its default risk. Market prices reflect investors' expectations about a firm's future performance. As a result, they contain forward-looking information.

A second class of methods includes the accounting models, based on the seminal study by Altman's (1968) Z-score model and Ohlson's (1980) conditional logit model, which use information derived from financial statements and can be applied to different forms of firms (both listed and non-listed companies). Such information is backward looking since financial statements aim to report a firm's past performance, rather than its future prospects.

The accounting models, on their turn, can be roughly divided into two main groups both exposed to some potential criticism. The first category includes bankruptcy risk scores based on multivariate statistical methods or econometric techniques which are easy to apply but show a relatively low performance in terms of prediction capacity. Moreover, Eisenbeis (1977) and Ohlson (1980) found that there were some inadequacies in MDA with respect to the assumptions of normality and group dispersion. These assumptions are often violated in MDA, therefore biasing the test of significance and estimated error rates.

The second group includes more sophisticated forecasting methods which are usually characterized by a very high performance but are very difficult to apply in evaluating the financial health of a firm. The main disadvantages of these complex models are the difficulty of building up the underlying data structure, the required time to carry out the iterative process and the effort for model interpretation.

Our approach, based on both an original use of robust principal component analysis (RPCA) and logit model, is positioned between the two families of accounting methods. More specifically, it can be part of the first group of methods with respect to relative simplicity underlying the model structure, but improves the bankruptcy predictive power in line with the second group of studies. PCA is able to identify clearly patterns in data and highlight their similarities and differences. It is a powerful tool for analyzing data especially when patterns are difficult to find because of the high dimension of data or because graphical representation is not available. Another advantage of PCA is that – after finding patterns in the data and reducing the number of dimensions – the loss of information is minimal since the first principal component maximizes variance. In addition, the robust principal component analysis (RPCA) for skewed data proposed by Hubert *et al.* (2009) and used in our analysis allows coping with the common asymmetry that typically characterizes the distribution of financial ratios, overcoming the issue of normal distribution that is required by several multivariate statistical methods.

The review of the literature also suggests additional areas for model improvement, explicitly incorporated in our analysis. First, much past research has employed relatively small samples of firms; recent evidence suggests that large samples are critically necessary to generalize empirical results. Second, financial ratios have been dominant explanatory variables in most research to date; it may be worthwhile to include nonfinancial variables and corporate governance structure in addition to financial variables. Third, several models have been tested by the relatively short one-year prediction horizon; it would be desirable to test the predictive power several years prior to bankruptcy. It is very important to consider how far ahead the model is able to predict bankruptcy accurately. Clearly, a model that is able to predict bankruptcy accurately earlier becomes more valuable for the investors and, at the same time, useful for the adoption of effective policies.

3. Methodology

This section describes our methodology including conceptual and operational definition of the variables used in the study. The basic idea is to maintain and treat separately the debt level of a firm and its sustainability. Indeed, companies might be characterized by a similar level of indebtedness but different degrees of vulnerability.

For this reason, in the first step of the analysis, we independently define and estimate a debt index and a sustainability index of such debt. In the second step, we evaluate the reliability of our indices as early warning signals of financial bankruptcy by applying a logistic regression technique which allows us to specify the probability of default as a function of our indices and other explanatory variables.

3.1 Assessment of the financial health of the firms

The financial and accounting literature suggests that a firm's financial condition is better evaluated by considering several aspects of the indebtedness phenomenon (leverage, indebtedness capacity, form of the financial debt, net financial position, etc.). In formal terms, DEBT = $f\,(x_1, x_2, \ldots, x_n)$ for a set of variables x_i, $i = 1.\,n$ related to the indebtedness condition of a firm, and f an unknown function. Following this approach (Bartoli, 2006; Brealey and Myers, 2001; Fridson, 1995), we build up a debt index which considers the multifaceted aspects of debt and assumes the following linear specification:

$$DEBT_{INDEX} = \alpha_1 \frac{FD}{N} + \alpha_2 \frac{CL}{FD} + \alpha_3 \frac{FD}{CF} + \alpha_4 \frac{CL}{CA} + \alpha_5 \frac{NTCA}{N} + \alpha_6 \frac{TFA}{LTD + N};$$

$$\alpha_i \in \mathrm{R};\ i = 1, 2, \ldots, 6$$

FD/N is the ratio of the Total Financial Debt (FD, given by Current Liabilities (CL) + Non current Liabilities) to the Shareholders Funds (N); it indicates the leverage of the firm. CL/FD is the ratio of the Current Liabilities (Payables due within 12 months + Total accrued expenses and deferred income) to the Total Financial Debt and gives information on the form of financing of the firm. FD/CF is the ratio of the Total Financial Debt to the Cash-Flow (CF, given by Profit for period + Depreciation) and represents the inability of firm's internal finance to cover the total debt. CL/CA is Current Liabilities over Current Assets (CA, given by Total current assets + Total accrued income and prepaid expenses), that is the inverse of the current ratio. NTCA/N is the ratio of the Net Technical Assets (NTCA, that is Tangible fixed assets) to the Shareholders Funds and indicates the inverse of the capitalization rate of technical assets. Finally, TFA/(LTD+N) is Total Fixed Assets (TFA, given by Intangible fixed assets + Tangible fixed Assets + Other fixed Assets) over the sum of Long-Term Debt (LTD, measured as Bonds beyond 12 months + Convertible bonds beyond 12 months + Because of banks beyond 12 months + Because of other lenders beyond 12 months) and Shareholders Funds and represents the equilibrium between fixed assets and long term liabilities. High values indicate that the firm may be forced to find more financial sources through short-term debt, usually subject to higher interest rates.

While a moderate level of debt can spur firm performance, an important element to consider when assessing firms' creditworthiness is the vulnerability of such debt. The maturity structure of assets and liabilities can provide valuable information about their vulnerability to changes in financing conditions. However, on the euro area level, short-term funding accounts for a small proportion of total funding, thus the maturity structure has a limited informative power (European Central Bank, 2013). Hence, an important factor for the assessment of the sustainability of debt is the debt service burden of firms, which indicates the proportion of their income needed for servicing debt. As for the $DEBT_{INDEX}$, we assume a linear specification to define the following index describing firm's weakness to cover the amount of interests on debt:

$$WKN_{INDEX} = \delta_1 \frac{IP}{EBIT} + \delta_2 \frac{IP}{EBITDA} + \delta_3 \frac{IP}{CF}; \; \delta_i \epsilon \mathrm{R}; \, i = 1, 2, 3$$

where IP is the Interest Paid (Total financial charges), EBIT the Earnings Before Interest and Taxes (Operating Profit/Loss), EBITDA the Earnings Before Interest, Taxes, Depreciation and Amortization (Operating Profit/Loss + Depreciation). CF indicates cash-flow.

Note that higher values of the WKN index indicate lower sustainability of debt, hence higher firms' debt vulnerability.

3.1.1 Robust statistical estimation of DEBT and WKN indices. The classical multivariate statistical methods are based on the assumption of normal distribution of variables, but financial data are often characterized by asymmetric distribution. For this reason, to estimate the DEBT and WKN indices, we use a new robust version of PCA, through which we obtain the values of the coefficients α_i and δ_i associated with each financial ratio.

PCA is a dimension reduction technique which transforms the observed variables into a small number of new variables while retaining as much information as possible. PCA is often the first step of a data analysis, followed by other statistical and/or econometrics techniques.

Traditional PCA makes use of eigenvalues and eigenvectors of the classical (from sample or population) covariance matrix; hence, it is sensitive to outliers and asymmetric distribution of variables. Various robust alternatives have been proposed in the literature (Hubert *et al.*, 2005 for a review). Here, to estimate robustly the α and δ coefficients of the DEBT and WKN indices, we apply a robust PCA technique – called modified *RPCA for skewed data* – suggested by Hubert *et al.* (2009). As in the classical case, these new PC_s are linear combinations of original variables, they are uncorrelated and they are extracted according to their importance in terms of explained variance of the original variables. Hence, the first principal component explains a percentage of variance greater than the second one and so on. The number of extractable PCs is equal to the number of original variables, but eigenvectors and eigenvalues solutions of PCA problem is based on a robust estimate of the covariance matrix of the data (Hubert *et al.*, 2005; 2009).

In real applications, when a PCA analysis is performed, if the original variables have a good degree of correlation, so that a high percentage of the original variance can be explained by few PCs, the first principal component (PC_1) is considered a good approximation of the data matrix **X**. Indeed, the explained variance represents a measure of the summary power of the data given by the first component and it is high if there is a good degree of correlation between the original variables. Usually, a percentage around 50-60 per cent of variance explained by the first principal component is considered a good value of summary power (Soares *et al.*, 2003, p. 128; Hair *et al.*, 2010, p. 109).

As accounting data tend to move in the same direction, and more or less proportionately, it is believed that collinearity is always present (Horrigan, 2000). Therefore, we expect the first PC of the two sets of financial ratios to explain a proper percentage of variability, so that $DEBT_{INDEX}$ and WKN_{INDEX} can be properly estimated with the coefficients given by the eigenvector defining the first robust principal component (RPC_1) of the firm's financial ratios data matrix.

3.2 Assessment of the probability of default

To evaluate the reliability of the proposed indices as early warning signals of financial bankruptcy, we apply a logistic regression technique which allows us to specify the probability of default as a function of a set of explanatory variables. Specifically, the dependent variable is a dichotomous variable that takes value 1 for defaulting firms (the firm is under bankruptcy procedure, it has filed for bankruptcy or it is subject to liquidation in 2011), 0 otherwise (the firm is still active in 2011). In formal terms:

$$p_{i,t} = \Pr\left(Y_{i,t} = 1\right) = F\left(x_{i,t-n}\beta\right) \tag{1}$$

where $p_{i,t}$ is the probability that the dependent variable Y = 1 for individual firm at time t = 2011, F(_) is the logistic cumulative distribution function, $x_{i,t\text{-}n}$ is the set of explanatory variables thought to affect $p_{i,t}$ with $n = 1\ldots5$; β are the regression coefficients. The explanatory variables are expressed as follows:

$$\begin{aligned} \Pr\left(Y_{i,t} = 1\right) = F\big(&\beta_0 + \beta_1 DEBT_{i,t-n} + \beta_2 WKN_{i,t-n} + \beta_3 SIZE_{i,t-n} \\ &+ \beta_4 AGE_{i,t-n} + \beta_5 D_own_{i,t-n} + \beta_6 D_mult_{i,t-n} \\ &+ \beta_7 PROD_{i,t-n} + \beta_8 X_region_{i,t-n} + \beta_9 Y_sector_{i,t-n}\big) \end{aligned} \tag{2}$$

i = 1. . . m where i is the i*th* firm, $n = 1\ldots5$.

In accordance with the general literature on bankruptcy, the model considers the financial structure of the firm. The first two explanatory variables, given by the DEBT and WKN scores computed in the first step of the analysis, take into account the financial health of the firm by measuring both the debt level and its vulnerability. As expected, several works find a significant relation between the financial structure of the firms and their probability of default or exit from the market (Molina, 2005; Graham *et al.*, 2011; Hovakimian *et al.*, 2012).

Following recent literature (Chava and Jarrow, 2004; Bhimani *et al.*, 2014, among the others), the model includes other regressors to control for additional non-financial characteristics of the firms, expected to be important in determining their probability of default. Both the theoretical and empirical literature suggest that age and size of the firms impact significantly on their performance (Klepper and Thompson, 2006). More recent studies also analyze the effects of productivity, industrial organization and ownership structure on firm performance (Dunne *et al.*, 1989; Disney *et al.*, 2003; Beck *et al.*, 2006; Foster *et al.*, 2006).

Therefore, equation (2) includes additional nonfinancial variables reported hereafter.

The variable $SIZE_i$ is computed in terms of a firm's annual turnover and measured in hundred thousands of Euros.

The variable AGE_i is the age of a firm since its foundation.

D_own_i is a dummy variable equal to 1 for fully concentrated ownership (unique partner), 0 otherwise (fragmented ownership, several partners). It is a signal of corporate governance since firms in countries with weaker investor protection also have more concentrated ownership (La Porta *et al.*, 1999).

D_mult_i is a dummy variable equal to 1 for multinational firms, 0 otherwise. Multinational firms have been identified through the analysis of ownership data, by selecting companies owning foreign subsidiaries (ownership share equals 51 per cent by default).

The variable $PROD_i$ indicates labor productivity and it is given by value added per employee.

Finally, to take into account the characteristics of the institutional and financial environment in which the firms operate and the specificities of the industrial sectors, we consider both regional dummies and sector dummies as explanatory variables, included in the vectors X and Y respectively. The manufacturing sectors are defined to include firms in the NACE Rev.2 primary codes 10-32. Hence, the model includes 20 regional dummies and 23 sector dummies.

4. An application to Italian firms

This section illustrates the results of our empirical analysis applied to Italian manufacturing firms, based on accounting data taken from the Aida Database, published by Bureau Van Dijk. After dealing with missing data, we build up an appropriate database including 31958 small, medium and large firms (see the Appendix for a description of the sample).

The work is carried out on the balance sheet and income statement over the 2006-2010 period to analyze the characteristics of firms affecting their probability of default after 5 years, in 2011. A firm is considered to have defaulted if it is under bankruptcy procedure, if it has filed for bankruptcy or it is in liquidation; we exclude firms with temporary financial problems or companies which have voluntarily chosen liquidation for economic opportunity, mergers or acquisition.

4.1 Estimation of DEBT and WKN indices

In this paragraph, we present the results obtained by applying the Robust PCA analysis to the Italian case[2].

To estimate DEBT and WKN coefficients, the Robust PCA algorithm has been applied to average values of financial ratios 2006-2010 to increase the stability and the reliability of our financial indices.

After applying the Robust PCA method, we obtain new RPCs variables that are a linear combination of original financial ratios; they are uncorrelated and maximize variance.

As expected, the first robust principal component represents the most important dimension in explaining changes of financial conditions since it explains 72.5 per cent of the total variance. Thus, we retain RPC_1 to estimate the coefficients α_i for $DEBT_{INDEX}$

$$DEBT_{INDEX} = 0.9192\frac{FD}{N} + 0.0045\frac{CL}{FD} + 0.0885\frac{FD}{CF} + 0.0254\frac{CL}{CA} + 0.3706\frac{NTCA}{N} + 0.0657\frac{TFA}{LTD+N}$$

With reference to financial ratios included in the WKN_{INDEX}, the first robust principal component is also the most important dimension in explaining changes in the sustainability of firms' debt. It explains 56.2 per cent of the total variance of the financial ratios. As for $DEBT_{INDEX}$, we estimate the coefficients δ_i for WKN_{INDEX} by retaining only RPC_1:

$$WKN_{INDEX} = 0.1572\frac{IP}{EBIT} + 0.2515\frac{IP}{EBITDA} + 0.9550\frac{IP}{CF}$$

Robust principal components and eigenvalues for $\mathrm{DEBT_{INDEX}}$ and $\mathrm{WKN_{INDEX}}$ are reported in the Appendix (section A.2).

4.2 Econometric results

Table I shows the logistic regression estimates for different lengths of the reference period, in particular for 1, 2, 3, 4 and 5 years before failure [3].

Those variables performing well in the latest year before failure will not necessarily perform well in the other years prior to failure. Some variables, however, can play an

Table I.
Probability of default: Logit estimates

	Year -1 2010		Year -2 2009		Year -3 2008	
	Coeff. β	Odds Ratio e^{β}	Coeff. β	Odds Ratio e^{β}	Coeff. β	Odds Ratio e^{β}
DEBT	0.365*** (0.057)	1.440*** (0.083)	0.338*** (0.047)	1.402*** (0.067)	0.286*** (0.037)	1.331*** (0.049)
WKN	0.518*** (0.045)	1.679*** (0.075)	0.469*** (0.035)	1.599*** (0.057)	0.513*** (0.032)	1.671*** (0.054)
SIZE	−0.134* (0.065)	0.874* (0.057)	−0.063 (0.053)	0.938 (0.050)	0.013 (0.044)	1.014 (0.045)
AGE	−0.262*** (0.066)	0.769*** (0.050)	−0.201*** (0.052)	0.817*** (0.043)	−0.186*** (0.044)	0.829*** (0.037)
D_own	−0.034* (0.150)	0.965* (0.145)	0.144 (0.123)	1.155 (0.142)	0.157 (0.104)	1.170 (0.122)
D_mult	−0.258* (0.141)	0.772* (0.109)	−0.297** (0.120)	0.742** (0.089)	−0.581*** (0.106)	0.559*** (0.059)
PROD	0.074 (0.124)	1.077 (0.134)	0.139 (0.097)	1.149 (0.112)	−0.032 (0.082)	0.967 (0.079)
Regional dummies	Included	Included	Included	Included	Included	Included
Sector dummies	Included	Included	Included	Included	Included	Included
Constant	−3.415** (1.258)		−4.272*** (1.210)		−2.846*** (0.946)	
N of obs.	14,486		14225		15466	
Log-likelihood	−1529.44		−2071.53		−2790.83	
Pseudo R^2	18.77		16.59		15.84	
LR Chi-square(50)	530.06		596.74		749.93	
Prob>Chi-square	0.000		0.000		0.000	

Notes: All variables in logs. Standard errors in parenthesis. Significance levels: *10%; **5%; ***1%

(*continued*)

	Year -4 2007		Year -5 2006	
	Coeff. β	Odds Ratio e^{β}	Coeff. β	Odds Ratio e^{β}
DEBT	0.373*** (0.040)	1.452*** (0.058)	0.275*** (0.042)	1.317*** (0.055)
WKN	0.562*** (0.034)	1.755*** (0.059)	0.529*** (0.034)	1.698*** (0.057)
SIZE	−0.001 (0.042)	0.999 (0.041)	0.075 (0.041)	1.077 (0.044)
AGE	−0.128*** (0.043)	0.879*** (0.038)	−0.060 (0.044)	0.940 (0.042)
D_own	0.232 (0.096)	1.261 (0.121)	0.244 (0.097)	1.276 (0.124)
D_mult	−0.403*** (0.095)	0.668*** (0.063)	−0.547*** (0.094)	0.578*** (0.054)
PROD	−0.087 (0.084)	0.915 (0.077)	−0.127 (0.086)	0.880 (0.076)
Regional dummies	Included	Included	Included	Included
Sector dummies	Included	Included	Included	Included
Constant	−2.260*** (0.813)		−3.280*** (0.791)	
N of obs.	16674		15809	
Log-likelihood	−3159.92		−3095.08	
Pseudo R^2	16.34		15.31	
LR Chi-square(50)	889.94		789.07	
Prob>Chi-square	0.000		0.000	

Table I.

important role in more than one regression given the long-run nature of some factors leading to failure.

Given the non-linearity of the first-order conditions with respect to parameters, a solution of numerical approximation is adopted that reaches convergence after five reiterations. Table I reports the maximized value of the log-likelihood function for all the regressions.

To avoid the risk of multicollinearity among variables, the computed bivariate correlation test was carried out. It does not reveal any linear relation among variables. To further corroborate this result we have computed two additional measures, namely the "tolerance" (an indicator of how much collinearity a regression analysis can tolerate) and the VIF (variance inflation factor, an indicator of how much of the inflation of the standard error could be caused by collinearity). Since both measures are close to 1 for the considered variables, we can exclude any multicollinearity.

Turning to the analysis of the estimates, our empirical findings show that both the DEBT score and the WKN score are statistically significant at 1 per cent level with the expected positive sign. An increase in firm's debt level and/or in its unsustainability significantly increases the probability of default.

Table I also reports the odds ratio of the logistic regression, which coincides with the exponential value of the estimated parameters. Considering one year prior to failure (2010), for a unit increase in the DEBT score, the odds of bankruptcy increases by 44 per cent, holding the other variables constant. Likewise, a unit increase in the WKN score raises the odds by 67.9 per cent. In other words, firms that are exposed to high debt are more than 1.44 times ($e^{0.365}$) more likely to fail than the other firms; firms with an unsustainable debt are more than 1.68 times ($e^{0.518}$) more likely to go to bankrupt than the other firms.

From these results, it is clear that the level of indebtedness and its nature are important factors in explaining firms' default risk. Interestingly, both indices enter with the highest coefficients in all the regressions, that is for different lengths of the reference period. Moreover, the coefficient associated with the vulnerability of debt is always greater than that related to the absolute level of debt[4]. Hence, it is certainly true that total amount of debt and its composition signal the financial health of the company, but the capacity/potential of the firm to sustain such debt is a more important factor to consider in firms' creditworthiness evaluation.

With reference to the other explanatory variables, firm size enters with negative sign at 10 per cent level of significance; therefore, larger companies would face lower probability of default. Note, however, that firm size is not significant when we consider a long period prior to failure. Age enters at 1 per cent level with negative sign, suggesting that younger firms are more likely to go to bankruptcy than older companies. These results confirm previous empirical findings on the impact of age and size on firm performance (European Central Bank, 2013; Hurst and Pugsley, 2011; Fort *et al.*, 2013). In a recent work on Italian manufacturing firms, Ferretti *et al.* (2016) obtain similar results.

Ownership concentration would enter with negative sign in the first year prior to failure suggesting that alignment of interests in fully concentrated ownership firms reduces the probability of financial instability and default. The variable, however, is not significant in explaining the probability of default in the majority of regressions.

On the contrary, being a multinational firm would impact significantly and negatively on the probability of bankruptcy, presumably because of the diversification of risk among different markets worldwide.

Labor productivity, on the contrary, does not seem to influence the probability of default.

As expected, the pseudo R-square increases when the reference period before failure reduces.

Moreover, both the coefficients (thus the odds ratios) and, for some regressors, the significance levels decrease when an increasing number of years is considered before failure. However, the estimates suggest that while some variables (like the annual turnover) are strongly significant in the latest year before failure but less significant – or not significant – in the other years prior to failure, the DEBT and WKN scores always enter at 1 per cent level of significance with the expected positive sign. They play an important role in determining the probability of default for several years before bankruptcy, mainly because of their long-run nature within the process leading to failure.

For a comparison, following Altman *et al.* (2013) and Altman, *et al.* (1994), we have re-estimated the Altman (1983) Z-score for our Italian data. We have then estimated the logit model including the Altman Z-score instead of the DEBT and WKN scores. Empirical findings, reported in Table II, show that the Altman Z-score enters significantly with the expected negative sign. The rest of the results are quite similar in both sign and level of significance. Paragraph 5 compares the reliability of both models.

4.3 Interaction effect between DEBT and WKN

In this paragraph, we estimate the interaction effect between DEBT and WKN to infer how the effect of DEBT (WKN) on the dependent variable depends on the magnitude of WKN (DEBT). We compute the interaction term in our logit model following Ai and Norton (2003). The correct marginal effect of a change in the two interacted variables and the correct standard errors has been computed in accordance with Norton *et al.* (2004). Estimates are based on the same variable list reported in equation (2) plus the interaction term between DEBT and WKN.

The interaction effects and the *z*-statistics are illustrated in Figure 1 and Figure 2, respectively. Both DEBT and WKN are statistically significant at conventional levels, as well as their interaction. Hence, the effect of DEBT (WKN) on the probability of default depends on the level of WKN (DEBT), as well as on other covariates.

The main effects imply that firms with higher debt and vulnerability are more likely to go bankrupt and the mean interaction effect is positive (0.0028482) (Table III). Note, however, that the interaction effect varies widely. For some observations, it is positive and for others it is negative. For firms whose predicted probability of bankruptcy is low (toward the left end of Figure 1), the interaction effect between DEBT and WKN is positive; thus, the association between one of the two predictors and the dependent variable increases if the other predictor increases. Hence, the more positive DEBT is, the more positive effect of WKN on probability of default becomes.

Where firms have a relatively higher predicted probability of bankruptcy, their interaction effects are all negative. That means there is "negative synergy" between the two interacted variables, so their presence at the same time dampens the effect. As debt increases, the effect of WKN on the probability of bankruptcy gets lower and lower. To put it differently, debt and WKN behave like substitutes: it is sufficient that one of them increases – for a given level of the other – to increase bankruptcy probability.

Note that Figure 1 refers to year 2010, but we obtain similar graphs for previous years (available upon request).

5. Reliability of the model

To evaluate the model we compute the percentage of overall correct classifications, which gives us the per cent of correct predictions of our model (Table IV). In total, 97.24 per cent of predicted probability is correctly classified in 2010. More specifically, in 2010, 400 firms are misclassified, consisting of 389 non-failed firms, and 11 failed firms. Hence, the estimated

Table II.
Probability of default: Logit estimates, Z-score

	Year -1 2010		Year -2 2009		Year -3 2008	
	Coeff. β	Odds Ratio e^{β}	Coeff. β	Odds Ratio e^{β}	Coeff. β	Odds Ratio e^{β}
Z-score	−0.611*** (0.050)	0.542*** (0.027)	−0.470*** (0.040)	0.624*** (0.025)	−0.576*** (0.038)	0.561*** (0.021)
SIZE	−0.124** (0.063)	0.883** (0.056)	−0.049 (0.052)	0.951 (0.050)	0.006 (0.042)	1.006 (0.042)
AGE	−0.365*** (0.063)	0.693*** (0.043)	−0.353*** (0.048)	0.702*** (0.034)	−0.244*** (0.041)	0.783*** (0.032)
D_own	−0.018* (0.149)	0.981* (0.146)	0.054 (0.121)	1.056 (0.128)	0.115 (0.098)	1.122 (0.110)
D_mult	−0.580*** (0.140)	0.559*** (0.078)	−0.483*** (0.116)	0.616*** (0.072)	−0.739*** (0.100)	0.477*** (0.048)
PROD	0.151 (0.119)	1.163 (0.139)	0.146 (0.099)	1.157 (0.114)	0.063 (0.080)	1.065 (0.086)
Regional dummies	Included	Included	Included	Included	Included	Included
Sector dummies	Included	Included	Included	Included	Included	Included
Constant	−2.477* (1.259)		−3.228*** (1.172)		−2.422*** (0.904)	
N of obs.	14491		14129		16165	
Log-likelihood	−1587.12		−2193.33		−3111.58	
Pseudo R^2	14.30		11.30		10.82	
LR Chi-square(49)	364.60		345.35		455.79	
Prob>Chi-square	0.000		0.000		0.000	

Notes: All variables in logs. Standard errors in parenthesis. Significance levels: *10%; **5%; ***1%

(*continued*)

Table II.

	Year -4 2007		Year -5 2006	
	Coeff. β	Odds ratio e^{β}	Coeff. β	Odds ratio e^{β}
Z-score	−0.687*** (0.039)	0.503*** (0.019)	−0.614*** (0.040)	1.698*** (0.057)
SIZE	0.033 (0.039)	1.033 (0.040)	0.073 (0.039)	1.077 (0.044)
AGE	−0.156*** (0.039)	0.855*** (0.033)	−0.130*** (0.041)	0.940 (0.042)
D_own	0.215 (0.089)	1.240 (0.111)	0.136 (0.095)	1.276 (0.124)
D_mult	−0.592*** (0.088)	0.553*** (0.048)	−0.677*** (0.090)	0.578*** (0.054)
PROD	−0.145* (0.080)	0.864* (0.069)	−0.120 (0.084)	0.880 (0.076)
Regional dummies	Included	Included	Included	Included
Sector dummies	Included	Included	Included	Included
Constant	−1.532*** (0.761)		−2.782*** (0.850)	
N of obs.	17307		16342	
Log-likelihood	−3622.11		−3381.13	
Pseudo R^2	11.37		10.57	
LR Chi-square(49)	576.66		475.82	
Prob>Chi-square	0.000		0.000	

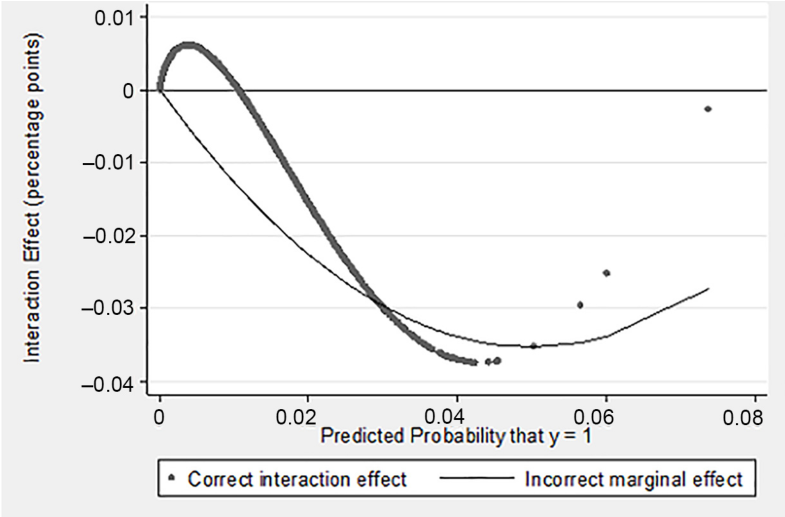

Figure 1.
Interaction effects after logit

Source: Own elaborations

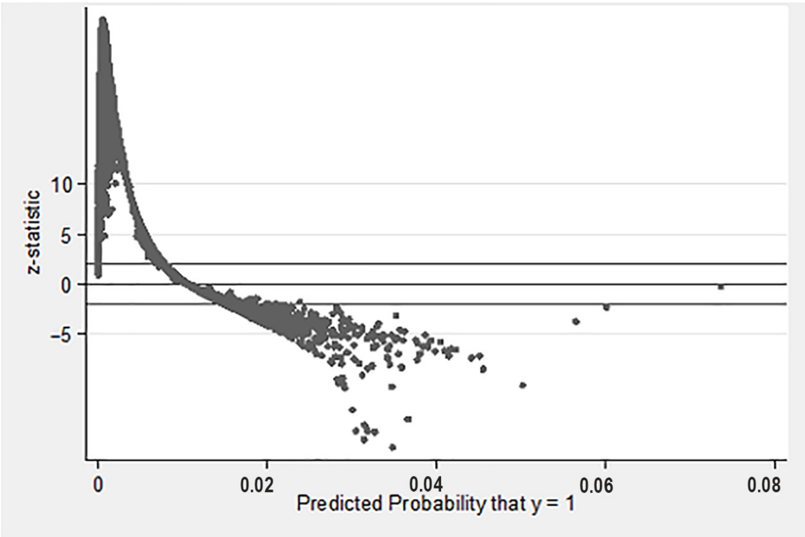

Figure 2.
Z-statistics of interaction effects after logit

Source: Own elaborations

chance of *misclassification* is 2.76 per cent. Misclassification increases when the length of the reference period increases. For the second, third, fourth and fifth years prior to failure, the estimated chance of misclassification is 4.23, 5.30, 5.74 and 5.84 per cent, respectively.

Note that, in terms of classification accuracy, our model and the Altman Z-score perform similarly in the first two years before failure. However, a greater discrepancy occurs in the third, fourth and fifth years prior to failure with expected overall accuracy rates of 94.71 per cent, 94.28 per cent and 94.17 per cent for DEBT-WKN scores versus 94.65 per cent, 93.92 per cent and 94.11 per cent for the z-score.

Upon deeper analysis, the empirical findings indicate that our model and the Altman Z-score show different percentages of first and second type errors. Type I errors refer to firms that are actually defaulting, but are classified as non-default firms. Type II errors refer to non-defaulting firms that are incorrectly classified by the model as default firms. As argued by Bottazzi *et al.* (2011) and Modina and Pietrovito (2014), it is standard to prefer prediction models that reduce the Type I error, that is, models that maximize the percentage of correctly classified defaults. For a bank, and also from a social point of view, it is more costly to fail to predict a default than to classify a non-default firm as a default firm.

Interestingly, our empirical findings show that the first type crucial error rates for misclassifying failed firms, as non-failed firms for the first five years prior to failure are always lower in our model in comparison with the Altman Z-score.

We have further assessed the model's ability to classify accurately observations using a receiver operating characteristic (ROC) curve. The area under the ROC curve (AUC) is a measure of discrimination; a model with a high area under the ROC curve suggests that the model can accurately predict the value of an observation's response. Our model provides outstanding discrimination since the AUC for the first five years prior to failure is 0.83, 0.80, 0.79, 0.78 and 0.77 (Table IV). Note that the area under the ROC curve computed with the DEBT-WKN scores is always greater than the area computed with the Altman Z-score.

Finally, to test the model fit, Hosmer and Lemeshow's test was evaluated. A good fit will yield a large p-value. With a p-value of 0.42, our model fits the data well.

The overall evidence suggests that, in terms of classification accuracy and reliability, our model would outperform Altman Z-score for prediction of corporate failure. This is especially true in the third, fourth and fifth years prior to failure indicating our DEBT–WKN indices to be good early warning signals of probable bankruptcy.

Variable	Obs	Mean	SD	Min	Max
_logit_ie	27702	0.0028482	0.0042808	−0.037341	0.0060973
_logit_se	27702	0.0007213	0.0010507	1.44e-09	0.0117573
_logit_z	27702	11.9979	8.075172	−16.42664	26.45633

Source: Own elaborations

Table III. Interaction effect, standard error and z-statistic – summary statistics

	Year -1 2010		Year -2 2009		Year -3 2008		Year -4 2007		Year -5 2006	
	DEBT-WKN	Z-score	DEBT-WKN	Z-score	DEBT-WKN	Z-score	DEBT-WKN	Z-score	DEBT-WKN	Z-score
Correctly classified (%)	97.24	97.22	95.98	96.00	94.71	94.65	94.28	93.92	94.17	94.11
Type I error (%)	2.69	2.74	3.94	3.99	5.18	5.27	5.58	5.95	5.72	5.80
Type II error (%)	0.07	0.04	0.09	0.02	0.12	0.09	0.16	0.15	0.12	0.10
AUC	0.83	0.78	0.8	0.73	0.79	0.72	0.78	0.73	0.77	0.72

Notes: A firm is classified as default whenever its estimated probability of default (p_i) is higher than 0.5; it is classified as non-default otherwise. We refer to first type errors when the model classifies as healthy a critical firm. We refer to second type errors when the model classifies as critical a healthy firm

Table IV. Model reliability

In Figure 3, we compare our methodology with one of the best performing "new age" classifiers, which is the random forests model, that belongs to the same branch of literature. Figure 3 reports the ROC curves one year before bankruptcy, but the curves of each preceding year are available upon request. Empirical evidence shows that the AUC for our model (on the left) is 0.8320; it is quite similar to the estimated AUC for random forests (on the right), which is 0.8365. Random forests, however, are characterized by complex underlying model structures and are more difficult to interpret.

In brief, compared to alternative accounting models belonging to the same branch of literature, our method would be more reliable than the traditional ones, the implementation feasibility being similar; it would be easier to construct than the newer techniques, the reliability being similar.

Hence, our method would contribute to solve the common tradeoff between implementation feasibility and accuracy characterizing bankruptcy prediction models. As suggested in Jones *et al.* (2017), a simpler more interpretable model should be preferred to a complex model, particularly if there is little difference in predictive performance.

6. Hazard model estimation

As robustness check, Table V shows the hazard model estimates for different lengths of the reference period.

The hazard models allow the evaluation of whether the analyzed explanatory variables determine the probability of default conditioned on the fact that no bankruptcy procedure has been implemented before (by the same firm). The hazard function h(t) – also known as the conditional failure rate – is the instantaneous rate of failure[5] and can be indicated as follows:

$$h(t) = h_0(t)\exp(\beta_0 + x_j\beta_x)$$

where $h_0(t)$ is the baseline hazard and β_x are the regression coefficients of the explanatory variables x_j.

In Table V, we report the common semiparametric Cox proportional hazard model estimates, but the Weibull model and the Cox model with shared frailty give similar results. The Cox (1972) model, which assumes that the covariates multiplicatively shift the baseline hazard function, is the most popular because of its computational feasibility. The baseline hazard, $h_0(t)$, is given no particular parameterization, but whatever the shape of the hazard over time, it is the same for everyone.

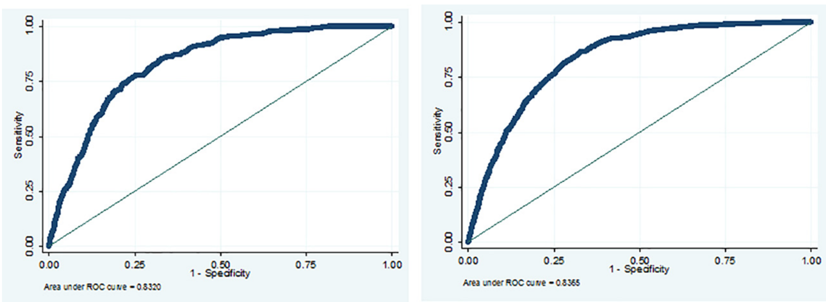

Figure 3.
A comparison between our model and random forests – ROC curves

Table V. Cox regression - Breslow method for ties

	Year -1 2010		Year -2 2009		Year -3 2008	
	Coeff. β	Hazard ratio e^{β}	Coeff. β	Hazard ratio e^{β}	Coeff. β	Hazard ratio e^{β}
DEBT	0.589*** (0.045)	1.802*** (0.033)	0.517*** (0.035)	1.676*** (0.027)	0.500*** (0.027)	1.648*** (0.019)
WKN	0.369*** (0.032)	1.446*** (0.025)	0.258*** (0.024)	1.294*** (0.017)	0.310*** (0.022)	1.363*** (0.014)
SIZE	−0.336*** (0.059)	0.714*** (0.057)	−0.292*** (0.048)	0.746*** (0.040)	−0.167*** (0.044)	0.846*** (0.035)
D_own	−0.058** (0.146)	0.944** (0.135)	−0.036** (0.108)	0.965** (0.102)	−0.078* (0.160)	0.925* (0.112)
D_mult	−0.165* (0.135)	0.848* (0.109)	−0.227** (0.107)	0.797** (0.089)	−0.296** (0.086)	0.744** (0.058)
PROD	0.110 (0.110)	1.116 (0.094)	0.111 (0.090)	1.117 (0.082)	−0.065 (0.078)	0.937 (0.059)
Regional dummies	Included	Included	Included	Included	Included	Included
Sector dummies	Included	Included	Included	Included	Included	Included
N of obs.	19127		19507		21930	
Log-likelihood	−3547.78		−5433.14		−8082.84	
LR Chi-square(50)	768.13		948.96		1385.55	
Prob > Chi-square	0.000		0.000		0.000	

Notes: All variables in logs. Standard errors in parenthesis. Significance levels: *10%; **5%; ***1%

(*continued*)

	Year -4 2007		Year -5 2006	
	Coeff. β	Hazard ratio e^{β}	Coeff. β	Hazard ratio e^{β}
DEBT	0.582*** (0.029)	1.789*** (0.018)	0.452*** (0.031)	1.571*** (0.025)
WKN	0.350*** (0.022)	1.419*** (0.019)	0.361*** (0.023)	1.434*** (0.017)
SIZE	−0.192*** (0.036)	0.825*** (0.031)	−0.174*** (0.037)	0.840*** (0.024)
D_own	−0.094* (0.083)	0.910* (0.061)	−0.098* (0.109)	0.906* (0.094)
D_mult	−0.211** (0.085)	0.810** (0.063)	−0.298* (0.124)	0.742* (0.054)
PROD	0.070 (0.073)	1.072 (0.067)	−0.033 (0.077)	0.968 (0.056)
Regional dummies	Included	Included	Included	Included
Sector dummies	Included	Included	Included	Included
N of obs.	22901		21473	
Log-likelihood	−9111.68		−8773.69	
LR Chi-square(50)	1575.10		1330.04	
Prob > Chi-square	0.000		0.000	

Table V.

We have tested the proportional-hazards assumption[6] by interacting analysis time with the covariates to verify that the effects of these interacted variables have no significant effect. We actually find – throughout the Stata stcox, tvc () option – that neither term significantly interacts with time, hence the model is correctly specified.

The empirical findings showed in Table V confirm previous results on the impact of all explanatory variables on the probability of default.

7. Concluding remarks

The aim of this study is to develop a new bankruptcy prediction approach which can be used in practice to signal the risk of failure of a firm. In this context, we first derive the firms' debt level and its vulnerability based on a new RPCA for skewed financial ratios. Second, we estimate a more complex logit model, based on both the first step computed indebtedness indices and additional non-financial firms' characteristics, which allows specific bankruptcy scores (predicted probabilities of default) to be computed for each firm included in the analysis.

The main findings of our application to Italian manufacturing firms show that the level of indebtedness and its sustainability are significant factors in explaining firms' default risk. The coefficient associated with the vulnerability of debt, however, is always greater than that related to the absolute level of debt indicating that the capacity of the firm to sustain a certain amount of debt is an important factor to consider in firms' creditworthiness evaluation. Moreover, the interaction effect between debt and its sustainability varies widely. For firms whose predicted probability of bankruptcy is low, the interaction effect is positive, while where firms have a relatively higher predicted probability of bankruptcy, their interaction effects are all negative. The majority of the other non-financial explanatory variables enters significantly with the expected sign. In addition to several models that have been tested by the relatively short one-year prediction horizon, we test the predictive power of the model several years prior to bankruptcy and compare it with the popular Altman z-score. The empirical evidence suggests a good performance in terms of both classification accuracy and reliability. Hence, the proposed approach is an efficient alternative to the Altman z-score and can be used as an *early warning signal* of financial bankruptcy. An early warning signal of over-indebtedness assumes a pivotal role in the adoption of effective reorganization procedures. In brief, our method would be more accurate in predicting default than traditional models (like z-score), it would be easier to construct than recent models (like random forests).

The practical use of the empirical results is valuable for entrepreneurs, managers and financiers. However, the research can be developed following several directions. First, it would be interesting to compare the proposed approach with other rating systems to evaluate companies' financial stability and their creditworthiness. Second, our analysis could be extended by applying real time recursive estimation methods. Finally, it may be worthwhile developing a more general model of company default prediction including also managerial practices and other qualitative information.

Notes

1. For comprehensive reviews on predicting corporate bankruptcy methodologies, see Aziz and Dar (2006); Bellovary *et al.* (2007) and Ravi Kumar and Ravi (2007).
2. Since we consider both large companies and SMEs, to mitigate the effect of firm size on selected variables, we first consider large, medium and small enterprises separately; we then divide each financial variable by the average turnover of the corresponding group and, finally, we build up the financial ratios.

3. Note that we have run a standard Logit, a Rare Events Logit and a Linear Probability Model and they produce similar results. We report the standard logistic regression estimates in the paper, while the other results are available upon request.
4. Note that the relatively higher coefficient associated with the variable WKN cannot be ascribed to scale differences because financial ratios have been standardized.
5. It is the (limiting) probability that the failure event occurs in a given interval, conditional upon the subject having survived to the beginning of that interval, divided by the width of the interval.
6. The proportional-hazards assumption states that the effects do not change with time except in ways that have been already parameterized.

References

Agarwal, V. and Taffler, R.J. (2011), "Twenty-five years of the taffler z-score model: does it really have predictive ability?", *Accounting and Business Research*, Vol. 37 No. 4, pp. 285-300.

Ai, C. and Norton, E.C. (2003), "Interaction terms in logit and probit models", *Economics Letters*, Vol. 80 No. 1, pp. 123-129.

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

Altman, E.I. (1977), "Predicting performance in the savings and loan association industry", *Journal of Monetary Economics*, Vol. 3 No. 4, pp. 443-466.

Altman, E.I. (1983), *Corporate Financial Distress*, Wiley Interscience, New York, NY.

Altman, E.I., Danovi, A. and Falini, A. (2013), "Z-score models' application to italian companies subject to extraordinary administration", *Journal of Applied Finance*, Vol. 23 No. 1, pp. 1-10.

Altman, E.I., Haldeman, R. and Narayanan, P. (1977), "ZETA analysis: a new model to identify bankruptcy risk of corporations", *Journal of Banking and Finance*, Vol. 1 No. 1, pp. 29-54.

Altman, E.I., Marco, G. and Varetto, F. (1994), "Corporate distress diagnosis: comparisons using linear discriminant analysis and neural networks (the Italian experience)", *Journal of Banking and Finance*, Vol. 18 No. 3, pp. 505-529.

Appetiti, S. (1984), "Identifying unsound firms in Italy: an attempt to use trend variables", *Journal of Banking and Finance*, Vol. 8 No. 2, pp. 269-279.

Aziz, M.A. and Dar, H.A. (2006), "Predicting corporate bankruptcy: where we stand?", *Corporate Governance: The International Journal of Business in Society*, Vol. 6 No. 1, pp. 18-33.

Bartoli, F. (2006), *Tecniche e Strumenti per L'Analisi Economico-Finanziaria*, Franco Angeli. Milano.

Beaver, W.H. (1968), "Alternative accounting measures as predictors of failure", *The Accounting Review*, Vol. 43 No. 1, pp. 113-122.

Beaver, W.H., McNichols, M. and Rhie, J.W. (2005), "Have financial statements become less informative? Evidence from the ability of financial ratios to predict bankruptcy", *Review of Accounting Studies*, Vol. 10 No. 1, pp. 93-122.

Beck, T., Demirgüç-Kunt, A., Laeven, L. and Maksimovic, V. (2006), "The determinants of financing obstacles", *Journal of International Money and Finance*, Vol. 25 No. 6, pp. 932-952.

Behr, A. and Weinblat, J. (2016), "Default patterns in seven EU countries: a random forest approach", *International Journal of the Economics of Business*, doi: 10.1080/13571516.2016.1252532.

Bellovary, J., Giacomino, D. and Akers, M. (2007), "A review of bankruptcy prediction studies: 1930 to present", *Journal of Financial Education*, Vol. 33 No. 5, pp. 1-42.

Berk, J., Stanton, R. and Zechner, J. (2010), "Human capital, bankruptcy, and capital structure", *The Journal of Finance*, Vol. 65 No. 3, pp. 891-926.

Bhimani, A., Gulamhussen, M. and Lopes, S. (2014), "Owner liability and financial reporting information as predictors of firm default in bank loans", *Review of Accounting Studies*, Vol. 19 No. 2, pp. 769-804.

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

Bottazzi, G., Grazzi, M. and Secchi, A. (2011), "Financial and economic determinants of firm default", *Journal of Evolutionary Economics*, Vol. 21 No. 3, pp. 373-406.

Brealey, R.A. and Myers, S.C. (2001), *Fundamentals of Corporate Finance*, 3rd ed., McGraw Hill, New York, NY.

Burgstahler, D., Jiambalvo, J. and Noreen, E. (1989), "Changes in the probability of bankruptcy, and equity value", *Journal of Accounting and Economics*, Vol. 11 Nos 2/3, pp. 207-224.

Cestari, G., Risaliti, G. and Pierotti, M. (2013), "Bankruptcy prediction models: preliminary thoughts on the determination of parameters for the evaluation of effectiveness and efficiency", *European Scientific Journal*, Vol. 9 No. 16, pp. 265-290.

Chava, S. and Jarrow, R. (2004), "Bankruptcy prediction with industry effects", *Review of Finance*, Vol. 8 No. 4, pp. 537-569.

Cox, D.R. (1972), "Regression models and life-tables (with discussion)", *Journal of the Royal Statistical Society, Series B*, Vol. 34, pp. 187-220.

Deakin, E.B. (1972), "A discriminant analysis of predictors of business failure", *Journal of Accounting Research*, Vol. 10 No. 1, pp. 167-179.

Dimitras, A.I., Slowinski, R., Susmaga, R. and Zopounidis, C. (1999), "Business failure prediction using rough sets", *European Journal of Operational Research*, Vol. 114 No. 2, pp. 263-280.

Disney, R., Haskel, J. and Heden, Y. (2003), "Entry, exit and establishment survival in U.K. manufacturing", *Journal of Industrial Economics*, Vol. 51 No. 1, pp. 91-112.

Duffie, D., Saita, L. and Wang, K. (2007), "Multi-period corporate default prediction with stochastic covariates", *Journal of Financial Economics*, Vol. 83 No. 3, pp. 635-665.

Dunne, T., Roberts, M.J. and Samuelson, L. (1989), "The growth and failure of US manufacturing plants", *The Quarterly Journal of Economics*, Vol. 104 No. 4, pp. 671-698.

Eisenbeis, R. (1977), "Pitfalls in the application of discriminant analysis in business, finance and economics", *Journal of Finance*, Vol. 32 No. 3, pp. 875-900.

Elam, R. (1975), "The effect of lease data on the predictive ability of financial ratios", *The Accounting Review*, Vol. 50 No. 1, pp. 25-43.

European Central Bank (2013), "Structural issues report 2013: corporate finance and economic activity in the euro area", Occasional Papers No.151.

Ferretti, S., Filippone, A. and Micucci, G. (2016), "Le imprese uscite dal mercato nel corso della crisi", Occasional Papers, n.317 Banca d'Italia.

FitzPatrick, P.J. (1932), *Symptoms of Industrial Failures as Revealed by an Analysis of the Financial Statements of Failed Companies, 1920-1929*, The Catholic University of America, Washington, DC.

Flagg, J.C., Giroux, G.A. and Wiggins, C.E. (1991), "Predicting corporate bankruptcy using failing firms", *Review of Financial Studies*, Vol. 1 No1, pp. 67-78.

Foreman, R.D. (2003), "A logistic analysis of bankruptcy within the US local telecommunications industry", *Journal of Economics and Business*, Vol. 55 No. 2, pp. 135-166.

Fort, T.C., Haltiwanger, J., Jarmin, R.S. and Miranda, J. (2013), "How firms respond to business cycles: the role of firm age and firm size", NBER Working Papers, 19134.

Foster, L., Haltiwanger, J. and Krizan, C.J. (2006), "Market selection, reallocation and restructuring in the US retail trade sector in the 1990s", *Review of Economics and Statistics*, Vol. 88 No. 4, pp. 748-758.

Fridson, M.S. (1995), *Financial Statement Analysis: A Practitioner's Guide*, 2nd ed., John Wiley, New York, NY.

Gentry, J.A., Newblod, P. and Whiteford, D.T. (1985), "Classifying bankrupt firms with funds flow components", *Journal of Accounting Research*, Vol. 23 No. 1, pp. 146-160.

Graham, J., Hazarika, S. and Narashimhan, K. (2011), "Financial distress in the great depression", NBER Working Papers, 17388.

Grunert, J., Norden, L. and Weber, M. (2004), "The role of nonfinancial factors in internal credit ratings", *Journal of Banking and Finance*, Vol. 29 No. 2, pp. 509-531.

Hair, J.F., Black, W.C., Babin, B.J. and Anderson, R.E. (2010), *Multivariate Data Analysis*, Pearson Education, Upper Saddle River, NJ.

Horrigan, J.O. (2000), "Sonic empirical bases of financial ratios", *Accounting Review*, Vol. 40, pp. 16-21.

Hovakimian, A., Kayan, A. and Titman, S. (2012), "Are corporate default probabilities consistent with the static trade-off theory?", *The Review of Financial Studies*, Vol. 25 No. 2, pp. 315-340.

Hubert, M., Rousseeuw, P.J. and Vanden Branden, K. (2005), "ROBPCA: a new approach to robust principal components analysis", *Technometrics*, Vol. 47 No. 1, pp. 64-79.

Hubert, M., Rousseeuw, P. and Verdonck, T. (2009), "Robust PCA for skewed data and its outlier map", *Computational Statistics and Data Analysis*, Vol. 53 No. 6, pp. 2264-2274.

Hurst, E. and Pugsley, B. (2011), "What do small business do?", *Brooking Papers on Economic Activity*, Vol. 4 No. 2, pp. 73-142.

Johnsen, T. and Melicher, R.W. (1994), "Predicting corporate bankruptcy and financial distress: information value added by multinomial logit models", *Journal of Economics and Business*, Vol. 46 No. 4, pp. 269-286.

Jones, S., Johnstone, D. and Wilson, R. (2017), "Predicting corporate bankruptcy: an evaluation of alternative statistical frameworks", *Journal of Business Finance and Accounting*, Vol. 44 Nos 1/2, pp. 3-33.

Joy, M.O. and Tollefson, J.O. (1975), "On the financial applications of discriminant analysis", *Journal of Financial and Quantitative Analysis*, Vol. 10 No. 5, pp. 723-739.

Kim, M.J. and Kang, D.K. (2010), "Ensemble with neural networks for bankruptcy prediction", *Expert Systems with Applications*, Vol. 37 No. 4, pp. 3373-3379.

Klepper, S. and Thompson, P. (2006), "Submarkets and the evolution of market structure", *The RAND Journal of Economics*, Vol. 37 No. 4, pp. 861-886.

Kolodner, J. (1993), *Case-Based Reasoning*, Morgan Kaufmann Publishers, San Mateo, CA.

La Porta, R., Lopez-de-Silanes, F. and Shleifer, A. (1999), "Corporate ownership around the world", *Journal of Finance*, Vol. 54 No. 2, pp. 471-517.

Lennox, C. (1999), "Identifying failing companies: a revaluation of the logit, probit and DA approaches", *Journal of Economics and Business*, Vol. 51 No. 4, pp. 347-364.

Libby, R. (1975), "Accounting ratios and the prediction of failure: some behavioral evidence", *Journal of Accounting Research*, Vol. 13 No. 1, pp. 150-161.

Mensah, Y.M. (1984), "An examination of the stationarity of multivariate bankruptcy prediction models: A methodological study", *Journal of Accounting Research*, Vol. 22 No. 1, pp. 380-395.

Merton, R.C. (1974), "On the pricing of corporate debt: the risk structure of interest rates", *Journal of Finance*, Vol. 29 No. 2, pp. 449-470.

Modina, M. and Pietrovito, F. (2014), "A default prediction model for italian SMEs: the relevance of the Capital structure", *Applied Financial Economics*, Vol. 24 No. 23, pp. 1537-1554.

Molina, C.A. (2005), "Are firms underleveraged? An examination of the effect of leverage on default probabilities", *The Journal of Finance*, Vol. 60 No. 3, pp. 1427-1459.

Nam, J.H. and Jinn, T. (2000), "Bankruptcy prediction: evidence from korean listed companies during the IMF crisis", *Journal of International Financial Management and Accounting*, Vol. 11 No. 3, pp. 178-197.

Norton, E.C., Wang, H. and Ai, C. (2004), "Computing interaction effects and standard errors in logit and probit models", *The Stata Journal: Promoting Communications on Statistics and Stata*, Vol. 4 No. 2, pp. 154-167.

Odom, M. and Sharda, R. (1990), "A neural network model for bankruptcy prediction", in: *IEEE International Conference on Neural Network*, Vol. 2, pp. 163-168.

Ohlson, J. (1980), "Financial ratios and the probabilistic prediction of bankruptcy", *Journal of Accounting Research*, Vol. 18 No. 1, pp. 109-131.

Pervan, I. and Kuvek, T. (2013), "The relative importance of financial ratios and nonfinancial variables in predicting of insolvency", *Croatian Operational Research Review*, Vol. 4 No. 1, pp. 187-198.

Platt, H.D. and Platt, M.B. (1991), "A note on the use of industry-relative ratios in bankruptcy prediction", *Journal of Banking and Finance*, Vol. 15 No. 6, pp. 1183-1194.

Ravi Kumar, P. and Ravi, V. (2007), "Bankruptcy prediction in banks and firms via statistical and intelligent techniques – a review", *European Journal of Operational Research*, Vol. 180 No. 1, pp. 1-28.

Shin, K.S. and Lee, Y.J. (2002), "A genetic algorithm application in bankruptcy prediction modeling", *System Experts with Applications*, Vol. 23 No. 3, pp. 637-646.

Shumway, T. (2001), "Forecasting bankruptcy more accurately: a simple hazard model", *The Journal of Business*, Vol. 74 No. 1, pp. 101-124.

Smith, F.R. and Winakor, A.H. (1935), "Changes in financial structure of unsuccessful corporations", *Bureau of Business Research*, University of IL, Vol. 31.

Soares, J.O., Lourenço Marquês, M.M. and Monteiro, C.M.F. (2003), "A multivariate methodology to uncover regional disparities: a contribution to improve European union and governmental decisions", *European Journal of Operational Research*, Vol. 145 No. 1, pp. 121-135.

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

Varetto, F. (1998), "Genetic algorithms applications in the analysis of insolvency risk", *Journal of Banking and Finance*, Vol. 22 Nos 10/11, pp. 1421-1439.

Vassalou, M. and Xing, Y. (2004), "Default risk in equity returns", *The Journal of Finance*, Vol. 59 No. 2, pp. 831-868.

Vuran, B. (2009), "Prediction of business failure: a comparison of discriminant and logistic regression analyses", *Istanbul University Journal of the School of Business Administration*, Vol. 38 No. 1, pp. 47-65.

Wall, A. (1936), *How to Evaluate Financial Statements*, Harper, New York, NY.

Wilcox, J.W. (1976), "The gambler's ruin approach to business risk", *Sloan Management Review*, Vol. 18 No. 1, pp. 33-46.

Yang, Z.R., Platt, M.B. and Platt, H.D. (1999), "Probabilistic neural networks in bankruptcy prediction", *Journal of Business Research*, Vol. 44 No. 2, pp. 67-74.

Zavgren, C. (1985), "Assessing the vulnerability of failure of American industrial firms: a logistic analysis", *Journal of Business Finance and Accounting*, Vol. 12 No. 1, pp. 19-45.

Zmijewski, M. (1984), "Methodological issues related to the estimation of financial distress prediction models", *Journal of Accounting Research*, Vol. 22, pp. 59-82.

Further reading

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

Brédart, X. (2014), "Bankruptcy prediction model: the case of the United States", *International Journal of Economics and Finance*, Vol. 6 No. 3, pp. 1-7.

Corresponding author

Marianna Succurro can be contacted at: m.succurro@unical.it