



# Bankruptcy prediction using Partial Least Squares Logistic Regression



Sami Ben Jabeur

IPAG Business School, 184, bd Saint-Germain, 75006 Paris, France

## ARTICLE INFO

### Keywords:

Bankruptcy prediction  
Forecasting  
Discriminant analysis  
Logistic regression  
Partial Least Squares

## ABSTRACT

In the current conditions of economy there is an increasing number of companies that are facing economic and financial difficulties which may, in some cases, lead to bankruptcy. This research is motivated by the inadequacies of traditional forecasting models. The Partial Least Squares Logistic Regression (PLS-LR) allows integrating a large number of ratios in the model; in addition, it solves the problem of correlation, and taking into account the missing data in the matrix. Indeed, the results obtained are very satisfactory and confirm the superiority of this method compared to conventional methods. The proposed model gives the opportunity to consider all the indicators in predicting financial distress, the reduction of the environment's uncertainty, the control's improvement and the coordination between the different company stakeholders.

## 1. Introduction

The evaluation of the risk of business failure has long been a central concern of researchers and professionals in the field. Failure situations affect the existence of a company and represent a very high cost to banking institutions in the event of partial or total loss of loans, also they represent a major risk for other creditors which, therefore, can in turn become defective. In its common and legal sense, failure covers a range of situations that contribute to the collapse of a particular firm because of serious financial problems that lead to insolvency. [Liang and Wu \(2005\)](#) argue that this failure is a situation in which cash flows generated by companies fail to meet obligations toward their financial and non-financial partners. [Pompe and Bilderbeek \(2005\)](#) suggest that the concept of failure is normally linked to credit risk, since it is the banks which declare that a specific company is insolvent when they refuse to provide credit. [Modigliani and Miller \(1959\)](#) showed that only efficient, effective and profitable firms are able to support a relatively large debt, since they can meet their financial commitments without problems. [Jensen \(1986\)](#) argues that debt acts as a disciplinary mechanism on managers and has a positive effect on company performance. According to [Finet \(2001\)](#), the difficulties experienced by a company become an obstacle of operating performance, its capital and its reputation. [Opler and Titman \(1994\)](#) estimate that debt is a factor of financial distress, and likely to endanger a company's survival. [John \(1993\)](#) asserts that financial distress results from a poor relationship between the availability of cash assets and contractual obligations for a given period. [Tinoco and Wilson \(2013\)](#) propose a combination of accounting data, stock market information and proxies for changes in the macro-economic environment to explain financial distress.

[Chiaromonte and Casu \(2015\)](#) explain that the likelihood of bankruptcy and distress decreases with increased liquidity holdings. [Richardson et al. \(2015\)](#) indicate that financial distress is positively associated with tax aggressiveness. [Amendola et al. \(2015\)](#) show that some characteristics, such as age, legal form and size influence the probability of being inactive and liquidated. In other words, it is impossible for a company to face its current liability with its available assets. The risk of insolvency, as well as a breach of any contractual debt, is a signal of financial distress. In order to treat the failure phenomenon, the work done in this area has developed models that predict more accurately the company's financial health ([Altman, 1968, 1994](#); [Bardos, 1998](#); [Li and Sun, 2009](#)). Banks must provide a systematic assessment of the risks they face, which implies in particular an accurate estimation of the probability of failure of their corporate clients, and accordingly a possible revision of their evaluation methods. If the analysis of the causes of bankruptcy is older, the analytical methods of bankruptcy were not fully developed until the late sixties. Since the work of [Beaver \(1966\)](#), many authors have been successful assessed the risk of corporate failure based on financial analysis. Various tools are available to researchers; the most frequently used is the linear discriminated analysis and logit model. This paper attempts to apply a new approach which is the Partial Least Squares Logistic Regression (PLS-LR). This method finds its success in several areas including finance. To better understand this phenomenon, the problem is formulated as follows: What are the advantages of PLS regression compared to the traditional techniques of bankruptcy prediction?

The objective of this work is to improve the LR in the presence of highly correlated data, by using a PLS-LR that offers an significant alternative by allowing, among other advantages, in considering the

E-mail address: [sbenjabeur@gmail.com](mailto:sbenjabeur@gmail.com).

action of the existing correlation. In LR there is an exclusion of potentially important variables; in the PLS-LR all the key variables are retained and unimportant variables are excluded, or they are assigned a lower weight in the model. To address this issue, the following tracking was adopted: In Section 2, a literature review is presented. In Section 3, the method and the sample used in our study is described. In Section 4, the results are discussed and presented and, in Section 5, the conclusion is provided.

## 2. Literature review

The first work on predicting bankruptcies of companies from data is the work of Beaver (1966) and Altman (1968). It appears to have been the real starting point and the reference number of empirical studies published from thence forward. Beaver (1966) developed a one-dimensional dichotomous classification, that is to say one based upon a single ratio. Subsequently, Deakin (1972) and Edmister (1972), have shown that the predictive power of financial ratios is additive and that individual ratios has less predictive power than a small number of independent ratios used simultaneously. Multivariate analyses allow a richer description of the situation of the company, and are now used systematically.

The choice of Altman's model (1968) is justified by the fact that it has particular significance since he followed the same methodology and resumed the same ratios (Atiya, 2001; Grover and Lavin, 2001). Shirata (1998) and Taffler (1983) tried to determine the financial ratios that predict bankruptcy of the company, using only discriminant analysis (DA). However, the new model for predicting the failure inspired by DA such as multi-criteria discriminant analysis were applied by Zopounidis and Doumpos (2002). These latter conclude that this new alternative dominates the DA, and that it is a comparable analysis to the logit model.

Due to the constraint of multinormality which is rarely empirically observed, some authors have preferred other methods. One possibility is the use of other parametric techniques, which involve a different distribution of accounting variables: the logit model and the probit model. Among the pioneering studies of logistic regression, Ohlson's (1980) was the first in this area to look at the prediction of failure. The probit model was rarely used (Zmijewski, 1984; Grover and Lavin, 2001). If Grover and Lavin (2001), comparing the DA to probit models using the same data, suggest the superiority of the first, *a priori* thanks to its greater reliability' however, more recent studies have led to the opposite conclusion.

Indeed, Lennox (1999) obtains high prediction accuracy by probit and logit models. Lisa et al. (1990) applied two logit models, the first of which distinguishes the failing companies of a group of randomly selected companies among the healthy businesses, while the second establishes a discrepancy between the failing firms and those in difficulty. Lisa et al. (1990), suggest that the first model is more efficient than the second by offering the highest prediction accuracy in the order of 90.80% one year before bankruptcy. Through the logistic regression, Tinoco and Wilson (2013), predicted the failure of the British companies between 1980 and 2011 one year before failure, using macroeconomic indicators as general measures of economic conditions, with capital structure and the financial indicators showing solvency, liquidity, profitability and the management of the company. The authors concluded that their model was efficient and offered a high level of classification. After estimating, they found that the absence of profitability and liquidity were the root causes of business failure, and the explanatory power of the multivariate conditional probability approach exceeded the power of the univariate. Du Jardin (2015), uses the LR with a sample of French companies between 2005 and 2010, wherein he obtained a classification rate better than DA. Through logistic regression, Lopez Iturriaga-and-Pastor Sanz (2015), predicted the failure of the American banks between 2002 and 2012, after estimating that logistic regression provided a high prediction accuracy

of 82.69% one year before the failure while the DA offers a rate of 78.85%. The study realized by Fedorova et al., (2013), recommended primarily retaining the neuronal approach compared to the LR. However, among the disadvantages of dichotomous models, which are extremely sensitive to multicollinearity, the inclusion of highly correlated variables should be avoided (Zopounidis and Doumpos, 2002; Serrano-Cinca and Gutiérrez-Nieto, 2013) because as the logistic regression is based on the analysis of financial ratios, these variables are generally correlated because they often share the same numerator or denominator. Similarly, variables used in accounting must indeed follow a multinomial distribution and their matrixes of variance-covariance must be the same for both the sample of failing companies and that of non-failing companies. In the face of the constraint of homoscedasticity, some authors have resorted to quadratic discriminant analysis, which requires only the hypothesis of multinormality ratios (Rose and Giroux, 1984; Yang et al., 2011). However despite the validity for the hypothesis of multinormality, quadratic discriminant analysis is effective only if it is applied to a very large sample. In light of this situation some authors have preferred to use other techniques: neural networks (Varetto, 1998; Chen and Du, 2009; Yu et al., 2014), support vector machines (Li and Sun, 2009; Li et al., 2010; Horta and Camanho, 2013; Chen and Li, 2014; López Iturriaga and Pastor-Sanz, 2015), the random forest method (Yeh Lin and Hsu, 2012; Booth et al., 2014; Calderoni et al., 2015) and models of life time (Bauer and Agarwal, 2014; Tian et al., 2015).

The application of the LR poses a problem. Indeed, most studies deals with the probability of the default of companies that abandon the ratios and the most correlated accounting magnitudes, although they seem to make a strong contribution to analysis. Several authors (Nguyen and Rocke, 2004; Bastien et al., 2005) have proposed an effective alternatives, namely the PLS-LR which finds its success in several other areas and that will be tested for effectiveness in the case of the prediction of failure. The empirical literature result demonstrates competitiveness of PLS-LR compared with other classification methods. In logistic regression the likelihood estimation of the parameter function is very inaccurate due to the high correlation between explanatory variables. The PLS regression resolve the problems arising from quasi or complete data separation in LR. A PLS algorithm regression can keep a constant good performance on observations with or without missing data.

## 3. Methodology

### 3.1. Partial Least Squares Logistic Regression

PLS univariate regression is a nonlinear model relating a single dependent variable  $Y$  to a set of quantitative or qualitative independent variables  $X_1, \dots, X_k$ . This can be obtained by a series of simple and multiple regressions. By exploiting the statistical tests associated with the framework of classical linear regression, it is possible to select significant predictors to maintain PLS regression and choose the number of PLS components to retain. There are several versions of PLS univariate regression algorithms. They differ in terms of standardization and intermediate calculations. Nguyen and Rocke (2004, 2002) have shown that PLS components are linear combinations of the original variables, and that the coefficients of the linear combination are expressed by the coefficient of the linear regression of the response's variable for each predictor. They therefore potentially undermine the definition and replace the linear regression with a logistic regression (Nguyen and Rocke, 2004; Bastien et al., 2005).

The algorithm of the logistic regression PLS is as follows:

1. *Step 1*: obtain  $h$  PLS components, PLS,  $T=(t_1, \dots, t_h)$ , using the PLS algorithm or Simples in the case of a single explanatory variable.
2. *Step 2*: After calculating the components of step 1, proceed to a logistic regression of  $Y$  on the components selected by a cross-validation:

Download English Version:

<https://daneshyari.com/en/article/5111309>

Download Persian Version:

<https://daneshyari.com/article/5111309>

[Daneshyari.com](https://daneshyari.com)