Available online at www.sciencedirect.com

ScienceDirect

Procedia Economics and Finance 39 (2016) 634 – 641

Procedia Economics and Finance

www.elsevier.com/locate/procedia

3rd GLOBAL CONFERENCE on BUSINESS, ECONOMICS, MANAGEMENT and TOURISM, 26-28 November 2015, Rome, Italy

# The logistic lasso and ridge regression in predicting corporate failure

Jose Manuel Pereira[a]*, Mario Basto[a], Amelia Ferreira da Silva[b]

[a]*IPCA - Polytechnic Institute of Cavado and Ave, Campus do IPCA, 4750-810 Barcelos, Portugal*
[b]*IPP - Polytechnic Institute of Oporto, Rua Dr. Roberto Frias, 4200-465 Oporto, Portugal*

**Abstract**

The prediction of corporate bankruptcy is a phenomenon of interest to investors, creditors, borrowing firms, and governments alike. Many quantitative methods and distinct variable selection techniques have been employed to develop empirical models for predicting corporate bankruptcy. For the present study the lasso and ridge approaches were undertaken, since they deal well with multicolinearity and display the ideal properties to minimize the numerical instability that may occur due to overfitting. The models were employed to a dataset of 2032 non-bankrupt firms and 401 bankrupt firms belonging to the hospitality industry, over the period 2010-2012. The results showed that the lasso and ridge models tend to favor the category of the dependent variable that appears with heavier weight in the training set, when compared to the stepwise methods implemented in SPSS.

Peer-review under responsibility of the Organizing Committee of BEMTUR- 2015

*Keywords:* Corporate Bankruptcy; Prediction Models; Lasso; Ridge Regression

## 1. Introduction

There are several undesirable consequences of business failures. Its economic and social cost can be significant. So, it is quite natural that this issue has occupied a significant part of researcher’s agenda. In spite of recent growing interest on non-financial attributes in explaining business failures, traditionally investigation on this issue has been focused on financial attributes. In most of the works statistical or artificial intelligence techniques were applied to the accountancy data of the companies, aiming at obtaining prediction models that would indicate whether the

* Jose Manuel Pereira. Tel.: +351-963-621-955; fax: +351-253-821-111.
*E-mail address:* jpereira@ipca.pt


Peer-review under responsibility of the Organizing Committee of BEMTUR- 2015
doi:10.1016/S2212-5671(16)30310-0

company would or would not reach a bankruptcy situation in the future (Beaver, 1966; Altman, 1968; Martin, 1977; Tam and Kiang, 1992).

In a study on corporate bankruptcy prediction, one of the aspects we immediately need to clarify is the concept of bankruptcy we shall use. In specialized literature the term has been used in different ways by different authors: legal bankruptcy, insolvency, inability to do payments or continued losses. As we lack a general theory on corporate bankruptcy, there is also no unique definition for this concept. This is an important limitation, since the sample's selection, both in terms of firms that have and have not “bankrupt”, depends on the definition of corporate bankruptcy used.

Throughout the last four decades several techniques were used to design models regarding this issue. The Altman model is the classical model on prediction of corporate bankruptcy most referred to in literature. It was developed in the end of the 70's using the discriminant analysis. The purpose of this technique was to obtain an indicator or «Z» score (variable dependent on a function) that was the result of the linear combination of several independent variables (ratios or financial indicators).

After the publication of Altman's Z-Score in 1968, most studies published in the decade that followed also used the discriminant analysis (Blum, 1974; Elam, 1975; Altman et al., 1977; Moyer, 1977; Norton and Smith, 1979). The emergence of critiques emphasizing the limitations of this theory may have influenced researchers to try new techniques, namely logit and probit. With the application of logit models to corporate bankruptcy prediction, it is possible to estimate the probability that a certain event will happen, as well as the probability of failure or corporate bankruptcy, considering the values of certain indicators of the company (Ohlson, 1980; Keasey and Watson, 1987).

The probit model is associated to the cumulative function of normal probability. Although this model is not as popular as the logit one in this area of research, there are several studies that have used this methodology (Zmijewski, 1984; Lennox, 1999) with similar outcomes as with other techniques. The evolution of ICT has created adequate conditions for the development and application of other techniques that, despite their limitations, do not demand certain conditions, unlike statistical techniques. Among artificial intelligence techniques, the most used ones have been neural networks and the induction of rules and decision trees. During the last few years, studies that use the theory of Rough Sets have emerged. Several neural networks models have been used throughout the last decade in studies related to corporate solvency prediction. Among them we can highlight Bell et al. (1990) and Koh and Tan (1999), using the multilayer perceptron model; Coats and Fant (1993) and Lacher et al. (1995), using the Cascor method (cascade correlation) as a learning algorithm; and Serrano and Martin (1993), using a multilayer perceptron net and Kohonen's self-organizing maps. Min and Lee (2005) applied support vector machines to the bankruptcy prediction problem. Some of the recent research on predicting corporate failure focuses on ensemble models (Kim and Kang, 2010) and hybrid models (Ahn and Kim, 2009).

The least absolute shrinkage and selection operator (Lasso) is a variable selection technique that has been recently applied on corporate bankruptcy forecasts (Tian et al., 2015). For the present study the lasso and ridge approaches were undertaken, since they deal well with multicolinearity and display the ideal properties to minimize the numerical instability that may occur due to overfitting. The models were employed to a dataset of 2032 non-bankrupt firms and 401 bankrupt firms belonging to the hospitality industry, over the period 2010-2012.

## 2. The Ridge and Lasso logistic regression

The task of determining which predictors are associated with a given response is not a simple task. When selecting the variables for a linear model, one generally looks at individual p-values. This procedure can be misleading. For instance, if the variables are high correlated, the p-values will also be high, driving the researcher to mistakenly deduce that those variables are not important predictors. On the other hand, irrelevant variables may be included in the model that are not associated with the response, adding an unnecessary complexity and interpretability to the model. Also, if the number of observations is not much larger than the number of variables then there can be a lot of variability, resulting in overfitting (increased likelihood by adding more parameters but poorer predictions on future observations not used in the model training). There are some approaches for automatically performing variable selection.

Linear regression by least squares is not applied for a binary response coded as zero and one, but although some of the estimates of the binary response might be outside the interval (0,1), the output obtained can be seen as a crude