Knowledge-Based Systems 30 (2012) 67–77

Contents lists available at SciVerse ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Bankruptcy prediction models based on multinorm analysis: An alternative to accounting ratios

Javier de Andrés *, Manuel Landajo, Pedro Lorca

University of Oviedo, Spain

ARTICLE INFO

Article history:
Received 9 February 2011
Received in revised form 2 October 2011
Accepted 3 November 2011
Available online 30 December 2011

Keywords:
Bankruptcy prediction
Classification techniques
Nonparametric methods
Quantile regression
Accounting ratios

ABSTRACT

In this paper we address the bankruptcy prediction problem and outline a procedure to improve the performance of standard classifiers. Our proposal replaces traditional indicators (accounting ratios) with the output of a so-called multinorm analysis. The deviations of each firm from a battery of industry norms (computed by nonparametric quantile regression) are used as input variables for the classifiers. The approach is applied to predict bankruptcy of firms, and tested on a representative data set of Spanish firms. Results indicate that the approach may provide significant improvements in predictive accuracy, both in linear and nonlinear classifiers.

1. Introduction

Under the current economic conditions, bankruptcy early warning systems have become tools of key importance in order to guarantee the stability of the economy, as a consequence of their potential to avoid losses to stockholders, creditors, managers and other interested parties. The passing of the Basel II accord makes the need for accurate systems to predict financial distress even stronger, as this agreement establishes that the reserve capital of banks will directly depend on the use of updated models to estimate the probability of default. Entities whose credit scoring systems meet certain requirements (advanced internal ratings-based approach to credit risk) would face lower capital requirements and therefore have a competitive advantage.

Since the late 1960s, many approaches to bankruptcy prediction systems have been proposed. Most of them share the common feature of relying on multivariate statistical/intelligent techniques whose input variables are mainly financial descriptors of the credit applicant. Although several bankruptcy prediction models have been developed which rely on market information (e.g., [20]) and experts' decisions analysis (e.g., [26,48]), the available evidence (e.g., [1]) suggests that these alternative approaches have not significantly outperformed multivariate-based techniques.

Regarding the use of multivariate techniques for bankruptcy prediction, the majority of research efforts during the last three decades have been devoted to test the accuracy of several kinds of classifiers, proposed by researchers in the fields of statistics or artificial intelligence. Apart from well-known models such as linear discriminant analysis (LDA; [2]), logistic regression [44], or probit analysis [64], several systems have been tested: these include classification trees (e.g., [17]), neural networks (e.g., [3]), rough sets (e.g., [46]), Case-Based Reasoning (CBR) (e.g., [23]), genetic algorithms (e.g., [56]), fuzzy sets (e.g., [13]), Multivariate Adaptive Regression Splines (MARS) (e.g., [33]) and Support Vector Machines (SVM) (e.g., [21]). More recently, some researchers have successfully used hybrid systems, which integrate several single classifiers (e.g., [61]). For a review of research on financial distress prediction models, see, e.g., [35].

Bankruptcy prediction models aimed at the prediction of firm failure typically use as independent variables a number of financial ratios which are computed on the basis of the financial statements of firms. This expedient may be problematic, as a well established stream of research (e.g., [34,59,6]) has evidenced that the use of ratio relationships to assess the relative position of a firm is unsuitable unless certain—considerably restrictive—conditions are met: in particular, a strictly proportional and linear relationship between the components of the ratio is required. Empirical evidence provided by these authors indicates that these conditions typically fail to hold.

A number of alternatives to ratios have been proposed in the literature, all of them relying on several kinds of regression analyses. Most of these proposals may be regarded as "uninorm" in nature, in the sense that the output of a single regression line (typically, a *least squares regression* line) is used as the "industry norm" or

* Corresponding author. Address: Faculty of Economy and Business, University of Oviedo, Avda. Del Cristo s/n, 33006 Oviedo, Spain. Tel.: +34 985104855.
E-mail address: jdandres@uniovi.es (J. de Andrés).

0950-7051/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.knosys.2011.11.005

benchmark, and the relative position of each firm is assessed by reference to this regression line. The proposals include OLS linear regressions [59], loglinear models [47] and neural networks [53]. More recently, a flexible *multinorm* approach [30] has been proposed. It relies on computing a representative battery of (nonparametrically estimated) quantile regression lines, each one providing a different "norm" (e.g., excellence norm, median performance norm, poor performance norm) which allows the relative position of the firm to be assessed naturally by comparison of the financial situation of each firm with this set of norms (details are provided in next section).

Despite the technical superiority over ratios proclaimed by all these alternatives, none of them has been used in order to devise improved classification tools for bankruptcy prediction. In this paper we purport to fill this gap. Our approach is a straightforward one: we shall rely on standard (parametric/nonparametric) classification techniques, and try to assess how the performance of these classifiers may be improved when, instead of accounting ratios, they are fed with the deviations (or "distances") of each firm from a system of multinorms computed by nonparametric quantile regression (NQR).

We rely on two standard linear classifiers, namely, LDA and Logit models, and an artificial neural network-based classifier (ANN). Previous research indicates that nonlinear classifiers may outperform linear devices in our setting, although the small sample size available for our experiment strongly advises against using too complex neural networks. Therefore, we focused on a relatively simple structure, which was in essence an augmented linear classifier, with a single logistic neuron in its hidden layer.

The classifiers were built on the basis of a sample of real data which is made up of both distressed and non-distressed Spanish companies. The predictive performance of these classifiers, when built on the basis of the deviations of each firm with respect to the battery of quantile regression lines, is compared with the benchmark provided by the performance of the same classifiers when accounting ratios are used as covariates instead. The comparisons are carried out by a Monte Carlo experiment on a real world data base.

The remainder of the paper is structured as follows. Section 2 summarizes the drawbacks of ratios and standard regressions as tools for assessing the financial condition of firms, and outlines the multinorm framework we rely on. The design of the study is detailed in Section 3, which includes a summary of the data bases, the ratio relationships analyzed, and the main features of the classifier structures we apply. Section 4 displays the results of the NQR-based multinorm analyses, and then assesses the predictive performance of the quantile-based classifiers, which is compared to that of their ratio-based analogues. A summary of results as well as some further research avenues are included in Section 5.

2. Measuring the firms' economic and financial condition. Some methodological issues

2.1. Accounting ratios versus uninorm and multinorm regressions

It is now well known that the use of accounting ratios may lead to biased results when used in financial analysis. Some early researchers on empirical methods to financial analysis (e.g., [34,59,6]) provided evidence on this fact. These authors demonstrated that the ratio form is only a valid approach when the relationship between numerator (Z) and denominator (X) of the ratio is linear and strictly proportional. In statistical jargon, this amounts to having a null intercept term in the linear regression of Z on X. The argument can be explained by reference to Fig. 1, which represents the Z/X ratio for a sample of firms. We can see that firms i and j have the same value (k) for the ratio (line r). However, j is clearly above the industry norm (line r') whereas i is clearly below.

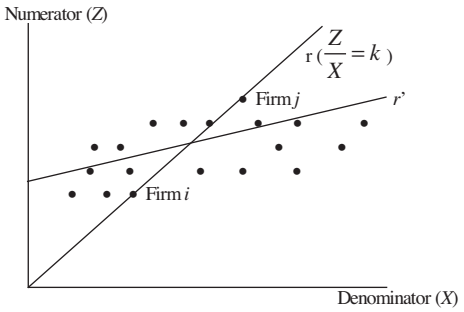

Fig. 1. The inadequacy of the ratio approach (adapted from [34]).

Similar distortions emerge when nonlinearities are present in the relationship between Z and X. Empirical research ([40,32,9,10,47,24], among others) has shown that neither proportionality nor linearity hold for most accounting ratios. Some authors suggested further drawbacks of the ratio method. So, Tippet [51] and Tippet and Whittington [52] indicated that ratios drift upwards or downwards over time, and McLeay and Trigueiros [41] proved that ratio validity requires not only linearity and proportionality, but also that proportionate changes in ratio components should be independent of firm size. Furthermore, there is considerable evidence pointing out the non-Gaussianity of the cross-sectional distributions of financial ratios (e.g., [11,43], among others). It is also well known that, under misspecification (e.g., when fitting a linear model to a nonlinear ratio relationship), many relevant diagnostics (including the standard t and F tests, and the tests for heteroskedasticity) may yield results biased towards the alternative.

As commented above, most authors who have proposed alternative tools for the measurement of the relationship between the components of accounting ratios have relied on regression techniques: these include ordinary least squares (OLS) regression [34,59], weighted least squares (WLS) regression [6,9], loglinear regression [47], and several forms (OLS, WLS, robust) of neural network regressions [53–55,29].

However, even these more sophisticated models may be fairly inadequate to capture the features which are relevant to financial analysts. In particular, these tools are unsuitable for the purpose of estimating a single norm for the industry, which can then be meaningfully used in order to assess for each firm—by computing the firm's deviation from that norm—its financial condition, profitability, or whatever other aspect is under study. The most serious problems are caused by the potential existence of an *heteroskedastic* relationship between numerator and denominator of the ratio. Heteroskedasticity prevents the analyst from achieving a proper interpretation for the deviations from the norm. This is due to the fact that, depending on the size of the firm, the same deviation can be either relevant or not. This is illustrated in Fig. 2.

In this example, both firm i and firm j have the same deviation (α) from the norm of the industry, which is represented by the regression line r, but which could have been obtained similarly by using the ratio approach, or more complex models such as neural networks. However, it is clear that firm i is above the norm while firm j falls in the 'average region' (dotted lines). This implies that, under heteroskedasticity, any attempt of improving on accounting ratios by using the naive expedient of comparing the deviation of each firm from a single-unique industry norm (e.g., a LS regression line) becomes problematic. The same is true no matter what regression technique we may rely on, or its statistical accuracy. A sensible way to tackle this issue is provided by the multinorm approach proposed by Landajo et al. [30]: the researcher estimates a system of