International Review of Financial Analysis 42 (2015) 64–75

Contents lists available at ScienceDirect

International Review of Financial Analysis

Performance evaluation of bankruptcy prediction models: An orientation-free super-efficiency DEA-based framework

Mohammad M. Mousavi [a], Jamal Ouenniche [a,*], Bing Xu [b]

[a] Business School, University of Edinburgh, 29 Buccleuch Place, Edinburgh EH8 9JS, UK
[b] School of Management and Languages, Heriot-Watt University, Edinburgh EH14 4AS, UK

ARTICLE INFO

Available online 21 January 2015

JEL classification:
G33
C19

Keywords:
Bankruptcy prediction
Performance criteria
Performance measures
Data envelopment analysis
Slacks-based measure

ABSTRACT

Prediction of corporate failure is one of the major activities in auditing firms risks and uncertainties. The design of reliable models to predict bankruptcy is crucial for many decision making processes. Although a large number of models have been designed to predict bankruptcy, the relative performance evaluation of competing prediction models remains an exercise that is unidimensional in nature, which often leads to reporting conflicting results. In this research, we overcome this methodological issue by proposing an orientation-free super-efficiency data envelopment analysis model as a multi-criteria assessment framework. Furthermore, we perform an exhaustive comparative analysis of the most popular bankruptcy modeling frameworks for UK data including our own models. In addition, we address two important research questions; namely, do some modeling frameworks perform better than others by design? and to what extent the choice and/or the design of explanatory variables and their nature affect the performance of modeling frameworks?, and report on our findings.

1. Introduction

Corporate failure often occurs when a firm experiences serious losses and/or becomes insolvent with liabilities that are disproportionate to its assets. Corporate failure may result from one or a combination of internal and external factors; e.g., managerial errors due to insufficient or inappropriate industry experience, risk seeking managers, lack of commitment and motivation to lead the company efficiently, refusal or failure to adjust managerial and operational structures of the firm to new realities, inefficient or inappropriate corporate policies, economic climate, changes in legislation, and industry decline — see for example Van Gestel et al. (2006).

Bankruptcy induces substantial costs to the business community such as court costs, lawyer costs, lost sales, lost profits, higher costs of credit, inability to issue new securities, and lost investment opportunities (e.g., Bris, Welch, & Zhu, 2006; Davydenko, Strebulaev, & Zhao, 2012; Elkamhi, Ericsson, & Parsons, 2012) — for a detailed review on the costs of bankruptcy, we refer the reader to Branch (2002). Therefore, the design of reliable models to predict bankruptcy is crucial to audit business risks and assist managers to prevent the occurrence of failure, and assist stakeholders to assess and select firms to collaborate with or invest in (e.g., Ahn, Cho, & Kim, 2000; Balcaen & Ooghe, 2006).

Given the importance of bankruptcy prediction, there is a considerable amount of literature focusing on both financial and non-financial information, and proposing new bankruptcy prediction models to classify firms as healthy or non-healthy (e.g., Aziz & Dar, 2006; Balcaen & Ooghe, 2006; Ravi Kumar & Ravi, 2007). With the increasing number of quantitative models available, one of the challenging issues faced by both academics and professionals is how to evaluate these competing models and select the best one(s).

Our survey of the literature on bankruptcy prediction revealed that although some studies tend to use several performance criteria and, for each criterion, one or several measures to evaluate the performance of competing prediction models, the assessment exercise is generally restricted to the ranking of models by a single measure of a single criterion at a time. For example, Theodossiou (1991) compared the performance of linear probability models, logit models, and probit models using an equally weighted average of Type I and Type II errors as a measure of correctness of categorical prediction, Brier score (BS) as a measure of the quality of the estimates of probabilities of default, and pseudo-R^2 as a measure of information content and found out that logit models outperform both linear probability models and probit models on all measures; however, with respect to pseudo-R^2 and an equally weighted average of Type I and Type II errors, probit models outperform linear probability models, but linear probability models outperform probit models on BS. Bandyopadhyay (2006) compared the performance of several MDA models using Type I errors and Type II errors, and compared the performance of several logit models using overall correct classification (OCC), receiver operating characteristic

* Corresponding author at: Business School, University of Edinburgh, 29 Buccleuch Place, Edinburgh EH8 9JS, UK. Tel.: +44 131 650 3792.
E-mail address: Jamal.Ouenniche@ed.ac.uk (J. Ouenniche).

http://dx.doi.org/10.1016/j.irfa.2015.01.006
1057-5219/

(ROC) measure, pseudo-R^2 statistic, and Log-Likelihood statistic (LL) and found out that the rankings of models differ with respect to different measures. Tinoco and Wilson (2013) compared the performance of several logit models with different categories of explanatory variables using ROC, Gini Index, and Kolmogorov–Smirnov statistic (KS) as measures of discriminatory power and Hosmer–Lemeshow statistic as a measure of calibration accuracy and found out that the rankings of models differ with respect to different criteria and their measures. In sum, the performance evaluation exercise under multiple criteria remains unidimensional in nature, on one hand, and the "big picture" is not taken into account in that a single or a very restricted number of criteria only are used, on the other hand. The drawback of the commonly used approach for the relative performance evaluation of competing bankruptcy prediction models is that the rankings corresponding to different criteria or measures are often different, which result in a situation where one cannot make an informed decision as to which model performs best when taking all criteria and their measures into consideration. In this paper, we address this methodological issue and fill this gap by proposing a data envelopment analysis (DEA)-based framework for the relative performance of bankruptcy prediction models.

DEA is a well-known non-parametric mathematical programming-based framework designed for the performance evaluation of competing entities, commonly referred to as decision making units (DMUs), which could in practice be production units of a manufacturing plant (e.g., Ahn & Neumann, 2014; Debnath & Sebastian, 2014), financial institutions such as a banks (e.g., Chortareas, Garza-Garcia, & Girardone, 2012; Wang, Lu, & Liu, 2014; Zhang, Jiang, Qu, & Wang, 2013), insurance companies (e.g., Kader, Adams, Hardwich, & Kwon, 2014) or mutual funds (e.g., Lozano & Gutierrez, 2008), financial instruments such as stocks (e.g., Lim, Oh, & Zhu, 2014), etc. The relative performance of such DMUs is typically assessed under multiple criteria, where the measures of these criteria are divided into two categories commonly referred to as inputs and outputs, and the most efficient DMUs constitute the so-called efficient frontier and represents an empirical standard of excellence. Note that, unlike other multi-criteria performance evaluation methodologies, DEA benchmarks against the best rather than the average behavior. Note also that the DEA terminology is motivated by an analogy between DMUs and production systems according to the economic theory of production.

Since its early days, DEA witnessed many methodological developments as well as a large number of applications. In the bankruptcy prediction area, DEA has so far been used either to classify firms into healthy and non-healthy categories (e.g., Pradi, Asmild, & Simak, 2004; Premachandra, Bhabra, & Sueyoshi, 2009; Premachandra, Chen, & Watson, 2011; Shetty, Pakkala, & Mallikarjunappa, 2012) or to compute aggregate efficiency scores to be used within statistical or stochastic modeling and prediction frameworks (e.g., Li, Crook, & Andreeva, 2013; Psillaki, Tsolas, & Margarits, 2010; Xu & Wang, 2009; Yeh, Chi, & Hsu, 2010). Unlike these uses of DEA in bankruptcy research, in this paper we propose to use DEA as a performance evaluation framework of competing bankruptcy prediction models.

In sum, the key contribution of this paper is to propose a multi-criteria performance evaluation framework – as a methodological contribution – to assist both academics and practitioners with the ranking of a set of competing bankruptcy prediction models under multiple criteria. In order to assist with the operationalization of the proposed framework, we use the most popular performance criteria for bankruptcy prediction models along with an exhaustive list of typical performance measures. In addition, under the proposed framework, we perform an exhaustive comparative analysis of the most popular bankruptcy modeling frameworks for UK data; namely, statistical and stochastic models including the ones that we designed as part of this research, using the most popular criteria along with a relatively large number of measures of these criteria to find out about the robustness of the results to the choice of the performance measures. Last, but not least, we address two important research questions; namely, do some modeling frameworks perform better than others by design? and to what extent the choice and/or the design of explanatory variables and their nature affect the performance of modeling frameworks?, and report on our findings. Our main findings could be summarized as follows. First, the proposed multidimensional framework provides a valuable tool to apprehend the true nature of the relative performance of bankruptcy prediction models. Second, the multidimensional rankings of the best and the worst models do not seem to be too sensitive to changes in most combinations of performance metrics. Third, numerical results seem to suggest that dynamic models tend be superior to static ones; thus, some modeling frameworks perform better than others by design. Fourth, numerical results seem to suggest that the choice and/or the design of explanatory variables and their nature affect to varying extents the performance of different modeling frameworks.

The remainder of this paper is organized as follows. In Section 2, we survey and classify the literature on bankruptcy prediction models. In Section 3, we present the proposed multi-criteria methodology; namely, an orientation-free super-efficiency DEA framework to evaluate the relative performance of competing forecasting models of bankruptcy. In Section 4, we present and discuss our empirical findings. Finally, Section 5 concludes the paper.

2. Bankruptcy prediction models

Bankruptcy prediction models could be divided into two main categories; namely, accounting-based models and market-based models. Accounting-based models could be further divided into three subcategories; namely, discriminant analysis models, regression models for categorical variables and survival analysis models. Note that the commonly used market-based models are mainly stochastic models. In this paper, we focus on the relative performance of accounting-based models, market-based models, and hybrids. Hereafter, we provide a generic framework for implementing these models followed by a brief description of such models along with a discussion of their main similarities and differences.

2.1. Generic framework of bankruptcy prediction

Most accounting-based and market-based bankruptcy prediction frameworks consist of two main phases. The first phase consists of using a quantitative modeling framework to estimate the probability of default. Then, the second phase classifies firms into two or more risk groups (e.g., risky vs. non-risky or bankrupt vs. non-bankrupt) using one or several cut-off points or thresholds depending on whether one classifies firms into two groups or more than two groups.

2.2. Discriminant analysis models

Discriminant analysis (DA) — first proposed by Fisher (1938), is a collection of classification methods which aim at partitioning observations into two or more subsets or groups so as to maximize within-group similarity and minimize between-group similarity, where "similarity" is measured by some sort of distance between observations (e.g., Mahalanobis distance). Univariate DA was first applied to bankruptcy prediction by Beaver (1966) and multivariate DA (MDA) was first applied to bankruptcy prediction by Altman (1968). A generic MDA model could be summarized as follows:

$$z = f\left(\sum_{j=1}^{P} \beta_j x_j\right),$$

where z is commonly referred to as a score or a z-score, x_js are explanatory variables, β_js represent the coefficients of the explanatory variables in the model, and f denotes the mapping of $\beta^t x$ on the set of real numbers $\mathfrak{R}$ — often referred to as a classifier, and could be either linear or