Performance Evaluation of Bankruptcy Prediction Models:
An Orientation-Free Super-Efficiency DEA-based Framework

Mohammad M. Mousavi[1], Jamal Ouenniche[1,3*], Bing Xu[2]

[1]Business School, University of Edinburgh, 29 Buccleuch Place, Edinburgh EH8 9JS, UK

[2]School of Management and Languages, Heriot-Watt University, Edinburgh, EH14 4AS

[3]Business School, ESC Rennes, Rennes, France

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 modelling frameworks for UK data including our own models. In addition, we address two important research questions; namely, do some modelling 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 modelling frameworks?, and report on our findings.

---

[*] Corresponding author: J. Ouenniche. Email: Jamal.Ouenniche@ed.ac.uk, Tel: +44-131-650-3792.

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, 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 et al., 2006; Davydenko et al, 2012; Elkamhi et al., 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 et al., 2000, Balcaen and 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., Balcaen and Ooghe, 2006, Aziz and Dar, 2006, Ravi Kumar and 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 (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 taken 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., Debnath and Sebastian, 2014, Ahn and Neumann, 2014), financial institutions such as a banks (e.g., Wang et al., 2014, Zhang et al., 2013, Chortareas et al., 2012), insurance companies (e.g., Kader et al., 2014) or mutual funds (e.g., Lozano and Gutiérrez, 2008), financial instruments such as stocks (e.g., Lim et al., 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 behaviour. 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., Shetty et al., 2012; Premachandra et al., 2009, 2011; Paradi et al., 2004) or to compute aggregate efficiency scores to be used within statistical or stochastic modelling and prediction frameworks (e.g., Psillaki et al., 2010; Yeh et al., 2010; Xu and Wang, 2009; Li et al., 2013). 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 modelling 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 modelling 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 modelling frameworks?, and report on our findings. Our main findings could be summarised 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 modelling 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 modelling 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 sub-categories; 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.

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 modelling 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.

2.1 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_j s are explanatory variables, β_j s 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 $\Re$ - often referred to as a classifier, and could be either linear or non-linear. Note that in comparing MDA models to other sub-categories of statistical models, one would typically need to estimate the probability of default (PD), which is used as an input to many performance measures. In this paper, we follow Hillegeist et al. (2004) in using a logit transformation:

$$PD = \frac{e^z}{1+e^z}.$$

Note that, under the normality assumption, MDA and logit approaches are closely related McFadden (1976). For a two-group classification problem, the classifier f is often a simple function that maps all observations or cases with discriminant or z-score values above a certain threshold or cut-off point to the first group and all other cases to the second group, where the cut-off point – often referred to as the *cutting score* or the *critical z-score*, is the average of groups' centroids, if group sizes are equal, or their weighted average, if group sizes are unequal, where a *group centroid* refers to the vector of group means of the explanatory variables. In the literature on bankruptcy prediction, MDA models mainly differ with respect to the choice of the explanatory variables and the form of the classifier – see Appendix A, and are part of most comparative analysis exercises and our comparative analysis is no exception.

2.2 Regression Models for Categorical Variables

As compared to discriminant analysis, regression models for categorical variables – also known as probability models (e.g., logit, probit) allow one to overcome some of the limitations of discriminant analysis. For example, within a regression framework for discrete response variables, the normality and the homoscedasticity assumptions are relaxed, on one hand, and the knowledge of prior probabilities of belonging to each group as well as misclassification costs is not required, on the other hand. The generic model for binary variables could be stated as follows:

$$\begin{cases} PD = Prob(y = 1) \\ PD = F(\beta, x) \end{cases}$$

where y denotes the categorical response variable, x denotes the vector of explanatory variables, β denotes the vector of coefficients of x in the model, and F is a function – commonly referred to as the link function, that maps any real number; e.g., score $\beta^t x$, onto a probability. The choice of F determines the type of probability model. For example, the normal probability model – known as probit, assumes that the link function is the cumulative standard normal distribution, say Φ; that is, $F(\beta, x) = \Phi^{-1}(\beta^t x)$. The logistic probability model – known as logit, assumes that the link function is the cumulative logistic distribution function, say Λ; that is, $F(\beta, x) = \Lambda^{-1}(\beta^t x)$ or equivalently:

$$PD = \Lambda(\beta^t x) = \frac{e^{\beta^t x}}{1 + e^{\beta^t x}}.$$

Finally, the linear probability model assumes that the link function is linear; that is, $F(\beta, x) = \beta^t x$, or equivalently:

$$PD = \beta^t x.$$

In the literature on bankruptcy prediction, logit is the most popular probability model and logit models only differ with respect to the choice of the explanatory variables – see Appendix A, and are part of most comparative analysis exercises.

2.3 Survival Analysis Models

Discriminant analysis models as well as probability models (e.g., linear probability model, logit, probit) are cross-sectional models and as such fail to take account of differences in firms' performance or risk profile over time; in sum, the probability of default (PD)

provided by these static models is time-independent. In order to overcome this issue, one could use a dynamic methodology such as survival analysis. Survival Analysis is concerned with the analysis of time to events. In this paper, we limit ourselves to a single event of interest; namely, bankruptcy or failure. Two functions are of special interest in survival analysis; namely, the survival function and the hazard function. The survival function, say $S(t)$, is a function of time and represents the probability that the time of failure is later than some specified time t; that is: $S(t) = P(T \phi t)$, where T is a random variable describing the time of failure for an observation or firm; in sum, the survival function provides survival probabilities or the probabilities of survival past specified times. On the other hand, the hazard function, say $H(t)$, is also a function of time and represents the failure or hazard rate at time t conditional on survival until t or later; that is:

$$H(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T \leq t + \Delta t | T \geq t)}{\Delta t} = -\frac{S'(t)}{S(t)},$$

where $S'(t)$ denotes the derivative of the survival function S with respect to time and Δt denotes a change in t. As far as the bankruptcy prediction application of survival analysis is concerned, the aim is to model the relationship between survival time and a set of explanatory variables. The most commonly used hazard model for bankruptcy modelling and prediction is the discrete-time hazard model proposed by Shumway (2001), where the survival and hazard functions are defined as follows:

$$S(t, x; \theta) = 1 - \sum_{j \pi t} f(j, x, \theta) \text{ and } H(t, x; \theta) = \frac{f(t, x, \theta)}{S(t, x; \theta)},$$

and $f(t, x; \theta)$ denotes the probability mass function of the discrete random variable "failure time" t defined as the time when a firm leaves the sample, x is a vector of explanatory variables used to predict bankruptcy, and θ is the vector of parameters of the mass function f. Shumway estimated this discrete-time hazard model using an estimation procedure similar to the one used for estimating the parameters of a multi-period logit model – this choice is motivated by a proposition whereby he proves that a multi-period logit model is equivalent to a discrete-time hazard model with a hazard function chosen as the cumulative distribution function of $f(t, x; \theta)$. He compared the performance of the discrete-time hazard model to MDA models, logit models, and probit models based on OCC and proved its

8

superiority for his dataset. Following the lead of Hillegeist et al. (2004), the probability of default at time period t is estimated as follows:

$$PD_t = \frac{e^{H_0(t)+\beta' x_t}}{1+e^{H_0(t)+\beta' x_t}},$$

where $H_0(t)$ denotes the unconditional hazard function – commonly referred to as the baseline hazard.

2.4. Black-Scholes-Merton-based Models

Most bankruptcy prediction models make use of accounting ratios as explanatory variables, which leads to a number of issues or criticisms; e.g., accounting statements only present a firm's historical performance and may not be informative in predicting the future; the "true" asset values may be very different from the book values; and accounting numbers can be manipulated by Management (e.g., Balcaen and Ooghe, 2006, Agarwal and Taffler, 2008). In order to overcome these drawbacks, one could make use of market-based explanatory variables. The rationale behind the use of market-based explanatory variables is that, in an efficient market, stock prices will reflect both the information contained in the accounting statements and the information contained in the future expected cash-flows. Furthermore, market variables are unlikely to be influenced by firm's accounting policies. In this sub-section, a category of such models is presented; namely, Black-Scholes-Merton (BSM)-based bankruptcy prediction models. Before presenting such bankruptcy prediction models, few comments are worthy of consideration. First, in practice, stochastic processes are often used to model stock prices behaviour and a specific type of stochastic processes; namely, Itô process, has proven to be a valid modelling framework for derivatives, where an Itô process refers to a Generalized Wiener process with both drift and variance rate being dependent on the underlying stock price and time. Second, the basic BSM model is concerned with modelling the price of an option as a function of the underlying stock price and time using an Itô process modelling framework. Third, under the Itô process modelling framework, the natural logarithms of stock prices are normally distributed. Last, but not least, the BSM model could be linked to the probability of a firm filing for bankruptcy; to be more specific, based on the observation by Merton (1974) that holding the equity of a firm can be viewed as taking a long position in a call option, the probability of default (*PD*) can be viewed as

the probability that the call option will expire worthless; that is, the value of the firm's assets (V_A) is less than the face value of its liabilities at the end of the holding period. Based on the above mentioned observations, McDonald (2002) derived the following expression for the probability of default or bankruptcy, $P(V_A < D)$:

$$PD = \Phi\left(-\frac{ln(V_A/D) + (\mu - \delta - 0.5\sigma^2) \times T}{\sigma\sqrt{T}}\right),$$

where $\Phi(.)$ denotes the cumulative distribution function of the standard Normal distribution, V_A is the value of the firm's assets, μ is the firm's expected return, σ^2 is the firm's assets volatility, δ is the divided rate and is typically proxied by the ratio of dividends to the sum of total liabilities and market value of equity, D is the firm's debt and is proxied by its liabilities, and T denotes both time to expiry of option and debt maturity and is assumed to be one year. In order to operationalize this BSM-based model of bankruptcy prediction, one would need to estimate V_A, μ, and σ as these parameters are not directly observable. Hillegeist et al. (2004) first estimate V_A and σ by solving the following system of equations:

$$\begin{cases} V_E = V_A e^{-\delta T} \Phi(d_1) - De^{-rT}\Phi(d_2) + (1 - e^{-\delta T})\Phi(d_1)V_A \\ \sigma_E = \dfrac{V_A e^{-\delta T}\Phi(d_1)\sigma}{V_E} \end{cases}$$

where the first equation is referred to as the call option equation, the second equation is referred to as the optimal hedge equation, V_E denotes the market value of common equity at the time of estimation, σ_E denotes the annualized standard deviation of daily stock returns over 12 months prior to estimation, r denotes the risk-free interest rate, and d_1 and d_2 are computed as follows:

$$d_1 = \frac{ln(V_A/D) + (r - \delta - 0.5\sigma^2) \times T}{\sigma\sqrt{T}}; d_2 = d_1 - \sigma\sqrt{T}.$$

Then, μ is estimated as follows and is restricted to lie between r and 100%:

$$\mu = \frac{V_{A,t} + Dividends - V_{A,t-1}}{V_{A,t-1}}.$$

where $V_{A,t}$ denotes the current value of the firm's assets and $V_{A,t-1}$ denotes the previous year value of the firm's assets. Alternatively, Bharath and Shumway (2008) estimate V_A and σ as follows:

$$V_A = V_E + D; \sigma = \frac{V_E}{V_A}\sigma_E + \frac{D}{V_A}\sigma_D$$

where $\sigma_D = 0.05 + 0.25\sigma_E$. As to the firm's expected return μ, it is proxied by either the risk-free rate r or the previous year stock return restricted to lie between r and 100%.

In the next section, we shall describe the DEA framework proposed for assessing the relative performance of bankruptcy prediction models based on these modelling frameworks.

3. A Slacks-based DEA Framework for Assessing Bankruptcy Prediction Models

In this paper, we propose a DEA-based framework for assessing the relative performance of competing bankruptcy prediction models. Hereafter, we first present the basic concepts and models of DEA (see §3.1). Then, we discuss how one might adapt a DEA framework to assess the relative performance of competing bankruptcy prediction models (see §3.2).

3.1. Basic concepts and models

DEA is a mathematical programming-based approach for assessing the relative performance of a set of decision making units (DMUs), where each DMU is viewed as a system and is defined by its inputs, its processes, and its outputs. The basic optimization problem addressed by DEA may be stated as follows:

Basic DEA Optimization Problem: Maximize the performance of a given DMU – as measured by the ratio of a weighted linear combination of outputs to a weighted linear combination of inputs, under the constraints that such ratio is less than or equal to one for each DMU and the weights are non-negative.

The mathematical programming formulation of this basic optimization problem is a fractional program which is typically transformed into a linear program using the Charnes-Cooper transformation (Charnes and Cooper, 1962) and therefore is easy to solve. The mathematical formulations of the basic DEA input- and output-oriented analyses proposed by Charnes,

11

Cooper and Rhodes (1978) and often referred to as CCR models are presented in Table 1, where the parameter $x_{i,j}$ denotes the amount of input i used by DMU$_j$, the parameter $y_{r,j}$ denotes the amount of output r produced by DMU$_j$, and the decision variable v_i (respectively, u_r) denotes the weight of input i (respectively, output r). Note that DEA models where the decision variables are the weights of input and output quantities are said to be stated in a multiplier form. Note also that the optimal value of e_k^{input} (respectively, e_k^{output}) indicates the efficiency status of DMU$_k$; to be more specific, $e_k^{input} = 1$ (respectively, $e_k^{output} = 1$) means that DMU$_k$ is efficient in that its weighted sum of outputs is equal to its weighted sum of inputs, and $e_k^{input} \pi 1$ (respectively, $e_k^{output} \phi 1$) means that DMU$_k$ is inefficient in that it produces less output than the input it requires. The set of efficient DMUs is referred to as the efficient frontier and represents the empirical standard of excellence.

[INSERT Table 1 HERE]

In general, the duals of these multiplier problems – commonly referred to as envelopment problems, are typically used in a relative performance evaluation exercise. The mathematical formulations of the basic DEA input- and output-oriented envelopment problems are presented in Table 2, where the variable θ_k is the dual variable associated with the fixed output amount constraint in the primal and may be interpreted as the technical efficiency ratio of DMU$_k$, the variable λ_j is the dual variable associated with the technical efficiency ratio of DMU$_j$ constraint in the primal and may be interpreted as the weight assigned to DMU$_j$'s inputs and outputs in constructing the ideal benchmark of DMU$_k$, the first set of constraints of, for example, the input-oriented envelopment model state that, for each input i, the amount used by DMU$_k$'s "ideal" benchmark; that is, the projection of DMU$_k$ on the efficient frontier, should at most be equal to the "revised" amount used by DMU$_k$; i.e., amount adjusted for the degree of technical efficiency θ_k of DMU$_k$, and the second set of constraints of, for example, the input-oriented envelopment model state that, for each output r, the amount produced by DMU$_k$'s "ideal" benchmark; that is, the projection of DMU$_k$ on the efficient frontier, should be at least as large as the amount produced by DMU$_k$. It is obvious that envelopment models allow for more appealing

interpretations. In addition, it is not always easy to compute an excess in an input or a shortage in an output from the optimal solution of a model expressed in a multiplier form, whether input- or output-oriented; however, solving the dual would enable one to determine excesses and shortfalls explicitly by the non-zero values of the slack and surplus variables.

[INSERT Table 2 HERE]

Last, but not least, DEA models – whether expressed in multiplier form or in envelopment form – allow one to identify the reference set or peer group used to benchmark each DMU in seeking improvements; for example, the reference set of a specific DMU, say DMU_k, is the set of DMUs with positive dual variables λ_j. Note that if the optimal value of θ_k, say θ_k^*, (respectively ϕ_k, say ϕ_k^*) is equal to 1, then the DMU_k under evaluation is efficient; else, $\theta_k^* \pi 1$ (respectively, $\phi_k^* \phi 1$) indicates that DMU_k is inefficient and the current level of inputs (respectively, outputs) should be decreased (respectively, increased). As our objective is to provide a multidimensional ranking to get rid of the inconsistencies of unidimensional rankings, we only make use of DEA efficiency scores. For a detailed discussion of different DEA models and application areas, the reader is referred to Seiford (1997), Cooper et al. (2005) and Liu et al. (2013).

3.2. Adaptation of DEA framework

DEA is a generic framework and as such its implementation for this specific relative performance evaluation exercise requires a number of key decisions to be made. First, *what are the units to be assessed or DMUs?* In this paper, DMUs are thirty competing bankruptcy prediction models – see Appendix A for a general description of these models. Second, w*hat are the inputs and the outputs?* The inputs and outputs are the performance measures of the relevant criteria for assessing bankruptcy prediction models. In this paper, we focus on the discriminatory power, the calibration accuracy or quality of estimates of the probabilities of default, the information content, and the correctness of categorical predictions criteria and their measures. In addition, inputs (respectively, outputs) are chosen according to the principle of the less (respectively, the more) the better; therefore, inputs (respectively, outputs) refer to the performance metrics to be minimized (respectively, maximized) – see Appendix C for a description of performance metrics. Note that, unless an application of

DEA involves undesirable outputs, the principle of the less (respectively, the more) the better is commonly used across the literature on DEA applications to select inputs (respectively, output) according to the economic theory of production. Third, *what is the appropriate DEA formulation to solve?* Although basic DEA models could be used to classify competing bankruptcy prediction models into efficient and non-efficient ones and rank them according to their scores, one cannot differentiate between efficient ones as they all receive a score of 1. In many application areas, decision makers are interested in obtaining a complete ranking in order to refine DMUs evaluation and our application is no exception. In this paper, we propose an orientation-free super-efficiency DEA framework; namely, a slacks-based super-efficiency DEA framework for assessing the relative performance of competing bankruptcy prediction models. We have deliberately chosen an orientation-free analysis over input-oriented analysis or output-oriented analysis because, in our application of evaluating the performance of bankruptcy prediction models, input-oriented and output-oriented analyses are not relevant. In addition, any type of oriented analysis would be inappropriate for the following reasons. First, under the variable returns-to-scale (VRS) assumption, which is the case with our data on bankruptcy prediction models, input-oriented efficiency scores can be different from output-oriented efficiency scores, which may lead to different rankings. Second, radial super-efficiency DEA models may be infeasible for some efficient decision making units; therefore, ties would persist in the rankings. Third, radial super-efficiency DEA models ignore potential slacks in inputs and outputs and thus may over-estimate the efficiency score by ignoring mix efficiency. The proposed framework is a three-stage process and could be summarized as follows:

Stage 1 – Returns-to-scale (RTS) Analysis: Perform RTS analysis to find out whether to solve a DEA model under constant returns-to-scale (CRS) conditions, variable returns-to-scale (VRS) conditions, increased returns-to-scale (IRS) conditions, or decreased returns-to-scale (DRS) conditions – see Banker et al. (2004) for details.

Stage 2 – Classification of DMUs: For each DMU k (k = 1, …, n), solve the following slacks-based measure (SBM) model (Tone, 2001):

$$\text{Min} \quad \rho_k = \left(1 - \frac{1}{m}\sum_{i=1}^{m}\frac{s_{i,k}^{-}}{x_{i,k}}\right) \Bigg/ \left(1 + \frac{1}{s}\sum_{r=1}^{s}\frac{s_{r,k}^{+}}{y_{r,k}}\right)$$

$$\text{s.t.:} \quad \sum_{j=1}^{n}\lambda_j x_{i,j} + s_{i,k}^{-} = x_{i,k}; \forall i = 1,...,m \tag{1}$$

$$\sum_{j=1}^{n}\lambda_j y_{r,j} - s_{r,k}^{+} = y_{r,k}; \forall r = 1,...,s$$

$$\lambda_j \geq 0, \forall j = 1,...,n; s_{i,k}^{-} \geq 0, \forall i = 1,...,m; s_{r,k}^{+} \geq 0, \forall r = 1,...,s$$

where n denotes the number of DMUs, m is the number of inputs, s is the number of outputs, $x_{i,j}$ is the amount of input i used by DMU_j, $y_{r,j}$ is the amount of output r produced by DMU_j, λ_j is the weight assigned to DMU_j in constructing its ideal benchmark, and $s_{i,k}^{-}$ and $s_{r,k}^{+}$ are slack variables associated with the first and the second sets of constraints, respectively. If the optimal objective function value $\rho_k^{*} = 1$, then DMU_k is classified as efficient. If $\rho_k^{*} \pi 1$, then DMU_k is classified as inefficient. Note that model 1 above is solved as it is if stage 1 reveals that the CRS conditions hold; otherwise, one would have to augment such model with one of the following additional constraints depending on whether VRS, IRS, or DRS conditions prevail, respectively:

$$\sum_{j=1}^{n}\lambda_j = 1; \quad \sum_{j=1}^{n}\lambda_j \geq 1 \; ; \quad \sum_{j=1}^{n}\lambda_j \leq 1. \tag{2}$$

Note that, when model 1 is augmented with one of these constraints, one obtains the BCC model proposed by Banker, Charnes and Cooper (1984).

Stage 3 – Break Efficiency Ties: For each efficient DMU k, solve the following slacks-based super-efficiency DEA model – first proposed by Tone (2002):

$$\text{Min} \quad \delta_k = \left(1 - \frac{1}{m}\sum_{i=1}^{m}\frac{t_{i,k}^{-}}{x_{i,k}}\right) \Bigg/ \left(1 - \frac{1}{s}\sum_{r=1}^{s}\frac{t_{r,k}^{+}}{y_{r,k}}\right)$$

$$\text{s.t.:} \quad \sum_{j=1;\, j\neq k}^{n}\lambda_j x_{i,j} \leq x_{i,k} + t_{i,k}^{-}; \forall i \tag{3}$$

$$\sum_{j=1;\, j\neq k}^{n}\lambda_j y_{r,j} \geq y_{r,k} - t_{r,k}^{+}; \forall r$$

$$\lambda_j \geq 0, \forall j \neq k; t_{i,k}^{-} \geq 0, \forall i; t_{r,k}^{+} \geq 0, \forall r$$

where $t_{i,k}^-$ (respectively, $t_{r,k}^+$) denotes the amount by which input i (respectively, output r) of the efficient DMU$_k$ should be increased (respectively, decreased) to reach the frontier constructed by the remaining DMUs. Note that the model 2 above is solved as it is if stage 1 reveals that the CRS conditions hold; otherwise, one would have to impose an additional constraint from amongst (2) as outlined in stage 2. Use the super-efficiency scores δ_k^*s to rank order the efficient DMUs.

At this stage, it is worth mentioning that unlike radial super-efficiency DEA models (e.g., Andersen and Petersen, 1993), slacks-based super-efficiency models are always feasible (Du et al., 2010, Tone, 2002). Note that Tone (2002) and Du et al. (2010) slacks-based super-efficiency models are identical with respect to their constraints in that one could be obtained from the other using a simple variable transformation. Note, however, that in applications where positive input and output data is a requirement, Du et al. (2010) provide a variant of the model solved in stage 3 to accommodate this situation. In the next section, we shall use the above described methodology to rank order competing bankruptcy prediction models and discuss the empirical results obtained using UK data for the period of 1989-2006.

4. Empirical Investigation

In this section, we first describe the process of data gathering and sample selection (see §4.1). Then, we present the list of models that we used in our comparative analysis (see §4.2). Finally, we discuss our empirical findings on the relative performance evaluation of bankruptcy prediction models under both a single criterion and multiple criteria (see §4.3).

4.1. Data and Sample Selection

In this paper, we first considered all UK firms listed on the London Stock Exchange (LSE) during an 18 years period from 1989 through 2006 and defined the bankrupt firms using the London Share Price Database (LSPD) codes 16 (i.e., firm has receiver appointed or is in liquidation), 20 (i.e., firm is in administration or administrative receivership), and 21 (i.e., firm is cancelled and assumed valueless); the remaining firms are classified as non-bankrupt. Then, we further reduced such dataset by excluding both financial and utilities firms, on one hand, and those firms with less than 5 months lag between the reporting date

and the fiscal year, on the other hand. As a result of using these data reduction rules, the final sample consists of 1414 UK listed firms – 211 of which are bankrupt firms and the remaining 1203 are non-bankrupt firms. In sum, our sample consists of a total of 12452 firm-year observations including 2062 observations related to bankrupt firms and 10390 observations related to non-bankrupt firms.

Within our sample, the average bankruptcy rate is 1.63% per year – which is higher than the 0.67% (respectively, 1.19%) bankruptcy rate of the sample used by Agarwal and Taffler (2008) (respectively, Christidis and Gregory, 2010). This higher rate of bankruptcy in comparison with other studies is due to the period of study being extended. Note however that the actual numbers of observations used to estimate the various models differ depending on the availability of data on each explanatory variable.

4.2. Bankruptcy models to be assessed

In this paper, we have chosen to assess the relative performance of the most popular accounting-based bankruptcy prediction models, market-based bankruptcy prediction models, and hybrid models.

The accounting-based bankruptcy prediction models considered in our comparative analysis include the MDA models proposed by Altman (1968), Altman (1983), and Lis (1972); the logit model proposed by Ohlson (1980); the probit model proposed by Zmijewski (1984); the linear probability model proposed by Theodossiou (1991); along with the MDA models proposed by Altman (1968), Altman (1983), and Lis (1972) reproduced or implemented in a logit framework. The market-based bankruptcy prediction models considered in our comparative analysis include the Black-Scholes-Merton(BSM)-based models proposed by Bharath and Shumway (2008) and Hillegeist et al. (2004). The hybrid models include the survival analysis model proposed by Shumway (2001) and estimated as a multi-period logit model. We refer to these models as the "original" models and represent them in Tables 3 and 4 with white shapes, where the shapes differ from one modelling framework to another – see legends of Tables 3 and 4.

We also include in our comparative analysis three additional categories of models that we refer to as original models refitted, reworking models in a logit framework, and new models.

As the name suggests, original models refitted include the above mentioned models refitted with our sample data (i.e., Altman, 1968, Lis, 1972, Altman, 1983, Ohlson, 1980, Taffler, 1984, Zmijewski, 1984, Shumway, 2001) and represented in Tables 3 and 4 with dotted shapes, where the shapes differ from one modelling framework to another – see legends of Tables 3 and 4. Reworking models in a logit framework refer to original non-logit models implemented or replicated in a logit framework with the same original explanatory variables (i.e., Altman, 1968, Altman, 1983, Lis, 1972, Taffler, 1984, Zmijewski, 1984, Total Liabilities/Total Assets (TLTA) Model of Bemmann, 2005) and represented in Tables 3 and 4 with grey shapes, where the shapes represent the original modelling frameworks – see legends of Tables 3 and 4. Last, but not least, the new models category consists of MDA, Logit, Probit, Linear Probability and Survival Analysis models where the explanatory variables are chosen from a list of variables using stepwise procedures. The list of variables consists of those accounting-based ratios and market-based variables chosen by a repeated use of Factor Analysis to an initial list of 74 accounting ratios and 3 market-based variables, where factors are selected so that both the absolute values of their loadings are greater than 0.5 and their communalities are greater than 0.8, and the stopping criterion is either no improvement in the total explained variance or no more variables are excluded. Note that Factor Analysis was run using Principal Component Analysis with VARIMAX as a factor extraction method. Note also that the list of variables consists of the variables that make up the factors. The new models (see Appendix B) are represented in Tables 3 and 4 with black shapes, where the shapes represent the original modelling frameworks – see legends of Tables 3 and 4.

In sum, a total of 30 models are assessed in our comparative analysis – see Appendices A and B for details on the original models and the new ones. Note that all chosen models are tested out-of-sample and the training period ranges from 1989 to 2001 including 1571 failure and 5615 non-failure firm-year observations. In the next sub-section, we shall assess the relative performance of these models under both a single criterion and multiple criteria and their measures using the proposed DEA framework (see §3.2).

4.3 Performance Criteria and Measures

With respect to performance criteria for evaluating bankruptcy prediction models, the focus of this paper shall be on the most commonly used criteria and their measures; namely, the discriminatory power criterion, the calibration accuracy criterion, the information content criterion, and the correctness of categorical predictions criterion. The discriminatory power criterion refers to the ability of a model to discriminate between the good cases and the bad ones, where a case refers to a firm. The calibration accuracy criterion refers to the quality of estimation of the probability of default. The information content criterion refers to the extent to which the output of a model (e.g., score, PD) carries enough information for bankruptcy prediction. The correctness of categorical prediction criterion refers to the ability of a model to produce forecasts that are consistent with actuals in that forecasts reveal firms as healthy (respectively, non-healthy) when actuals are healthy (respectively, non-healthy).

In our comparative analysis of models, we use Kolmogorov Smirnov (KS) Statistic, Area under Receivable Operating Characteristic (AUC) – also known as c-statistic, Gini Index, and Information Value (IV) to measure the discriminatory power criterion; we use Brier Score (BS) to measure the calibration accuracy criterion; we use log-likelihood statistic (LL) and pseudo-coefficient of determination (pseudo-R^2) to measure the information content criterion; and we use Type I errors (T1), Type II errors (T2), misclassification rate (MR), sensitivity (Sen), specificity (Spe), and overall correct classification (OCC) to measure the correctness of categorical prediction criterion – see Appendix C for descriptions of these measures.

4.4 Performance evaluation of bankruptcy prediction models

In our empirical investigation, we first generated the unidimensional rankings of the 30 models under evaluation (see Table 3 – this is a typical output from existing studies) to highlight the problems with using a unidimensional methodology to rank order competing bankruptcy prediction models; that is, models are ranked in the ascending (respectively, descending) order of the relevant measure of each of the criteria under consideration if the measure is to be minimized (respectively, maximized). Indeed, unidimensional or single criterion rankings tend to have many ties (e.g., the unidimensional rankings corresponding to Type I errors (T1), sensitivity (Sen), and Information value (IV)). In addition, one could clearly see that the unidimensional rankings could be different from one performance

criterion to another – see for example Theodossiou (1991), Bandyopadhyay (2006), and Tinoco and Wilson (2013).

For our dataset, most unidimensional rankings are different; in fact, the unidimensional rankings based on T1 and Sen differ from those based on T2, misclassification rate (MR), overall correct classification (OCC) and specificity (Spe), which differ from those based on area under ROC curve (AUC) and Gini index, which also differ from those based on Kolmogorov-Smirnov statistic (KS), information value (IV), Brier score (BS), log-likelihood statistic (LL) or pseudo-coefficient of determination (pseudo-R^2). Notice that the unidimensional ranking based on IV does not discriminate between the eight worst ranked models, because the probabilities of default produced by these models are all very close to zero and thus belong to the same band in the discrete approximation of the density functions of the good cases and the bad ones.

For our data set, unidimensional rankings suggest that, for all performance measures except IV and BS, the new models outperform both the original models, the original models refitted, and the reworked models with the exception of the logit model of Shumway (2001). Therefore, the selection of explanatory variables using Factor Analysis along with stepwise procedures seems to enhance the performance of models regardless of their underlying modelling framework. In addition, the use of a mixture of accounting-based and market-based information improves bankruptcy prediction. Furthermore, it seems that these new models are doing a better job at classifying firms than at producing their probabilities of default.

Also, for most performance measures, notice that in general refitting models seems to improve their ranks, which suggests that the nature of information within the training sample under consideration along with the period of study do, as expected, tend to affect the performance of bankruptcy models – recall that most original models were fitted to US data; therefore, when refitted to UK data they tend to do better at predicting bankruptcy for UK firms.

On the other hand, for most performance measures, reworking the original MDA, probit and linear probability models with the same explanatory variables in a logit framework seems to improve the ranks – with the exception of the MDA models of Lis (1972) and Taffler (1984)

which were originally fitted to UK data, which suggest that this improvement in the rankings could be due to the change in the training sample, or the modelling framework, or both. Also, for most performance measures, when comparing all logit framework-based models, the multi-period logit model of Shumway (2001) seems to outperform, which suggest as expected that its dynamic nature improves bankruptcy prediction.

Finally, using only market-based data does not seem to provide good enough information to classify a firm as risky or not; in fact, BSM-based models do not make the top 5; however, Hillegeist et al. (2004) model seems to always outperform Bharath and Shumway (2008) model.

At this stage, we would like to remind the reader that unidimensional rankings are not to be discarded as they convey valuable information; however, from both practical and methodological perspectives, one cannot make an informed decision as to which model performs best under multiple criteria. In order to address this issue, one would need a single ranking that takes account of multiple criteria, which we provide using the proposed DEA framework.

The multi-criteria rankings of the above mentioned 30 models are provided in Table 4 for different combinations of measures of the four criteria under consideration, where models are ranked in descending order of the corresponding SBM super-efficiency DEA scores (see §3.2). The empirical results reveal that the multidimensional rankings differ from the unidimensional ones. In addition, the multidimensional rankings have no ties, which suggest that the choice of the SBM super-efficiency DEA framework is an effective one in that it helps to get rid of ties between bankruptcy prediction models. Furthermore, we have considered several measures of the performance criteria under consideration to find out about the robustness of the multidimensional rankings with respect to the choice of measures.

For our data set and regardless of the combination of performance metrics used, multidimensional rankings suggest that some of the new models are always amongst the top ranked ones. In addition, the selection of explanatory variables using Factor Analysis along with stepwise procedures seems to always improve MDA and survival analysis-based bankruptcy prediction. Also, with the exception of combinations of metrics including T1 and

BS or BS and Sen simultaneously, the selection of explanatory variables using Factor Analysis along with stepwise procedures seems to always improve the performance of linear probability models at predicting bankruptcy. However, the new way of selecting explanatory variables does not seem to advantage the logit modelling framework or the probit modelling framework – although for the logit framework, the new models do better than the original ones. In addition, in general, the use of a mixture of accounting-based and market-based information improves bankruptcy prediction in most modelling frameworks.

Also, for most combinations of performance measures, notice that, with the exception of the MDA models of Altman (1968) and Altman (1983) and the logit model of Ohlson (1980), refitting models does not seem to improve their ranks – these conclusions are different from the ones derived from the analysis of the unidimensional rankings. Therefore, under a multicriteria setting, refitting models is not necessarily a mean for improvement.

On the other hand, regardless of the combination of performance metrics, reworking the original MDA models with the same explanatory variables in a logit framework seems to improve their ranks – with the exception of the MDA model of Taffler (1984).

As to reworking the original linear probability models with the same explanatory variables in a logit framework, it seems that for most combination of performance metrics the ranks have improved. Notice however that reworking the original probit model did not lead to any improvement in the multicriteria rankings. Therefore, under a multicriteria setting, reworking models could be a mean for improvement of some modelling frameworks such as MDA models. Also, regardless of the combination of performance metrics, when comparing all logit framework-based models, the multi-period logit model of Shumway (2001) does not seem to perform as well as in the unidimensional case. The refitted logit model of Ohlson (1980) however seems to be superior to the remaining logit models followed by the reworked probit model of Zmijewski (1984).

Finally, using only market-based data does not seem to provide good enough information to classify a firm as risky or not; in fact, BSM-based models do not make the top 5; however, Hillegeist et al. (2004) model seems to always outperform Bharath and Shumway (2008) model. This is amongst the very few findings of the unidimensional analysis that still hold in the multidimensional case.

To sum up, note that the conclusions derived from the analysis of the unidimensional rankings are not always consistent with their multidimensional counterparts. Therefore, multicriteria rankings help to better apprehend the relative performance of bankruptcy prediction models. Notice that 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. However, overall the multicriteria rankings of the models under consideration tend to be sensitive to some extent to the choice of performance measures, which suggest that in practice one would have to carefully select these measures to reflect the application context and the purpose of use of bankruptcy prediction models; in other words, the choice of performance metrics should be "fit for purpose".

Last, but not least, our findings suggest the following answers to our research questions. First, the survival analysis model tends be superior followed by linear probability and multivariate discriminant analysis models; therefore, some modelling frameworks perform better than others by design, as survival analysis models are dynamic and have the modelling ability to take on board both accosting-based and market-based information. Second, 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 modelling frameworks. To be more specific, most modelling frameworks improved in performance by taking account of a mixture of account-based and market-based information, where survival analysis, linear probability, and multivariate discriminant analysis models benefited the most from the new way of selecting explanatory variables.

5. Conclusion

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 results in conflicting rankings of models from one performance criterion to another. In this research, we proposed an orientation-free super-efficiency data envelopment analysis model to overcome this methodological issue; in sum, the proposed framework delivers a single ranking based on

multiple performance criteria. In addition, we performed an exhaustive comparative analysis of the most popular six bankruptcy modelling frameworks resulting in 30 prediction models for UK firms including our own models organized into four categories; namely, original models, original models refitted, reworking models in a logit framework with the same original explanatory variables, and new models. We used four criteria which are commonly used in the literature; namely, the discriminatory power, the calibration accuracy, the information content, and the correctness of categorical prediction. We have considered several measures for each criterion to find out about the robustness of multidimensional rankings with respect to different combinations of measures. Furthermore, we addressed two important research questions; namely, do some modelling 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 modelling frameworks?

Our main findings may 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 the survival analysis model tends be superior followed by linear probability and multivariate discriminant analysis models; therefore, some modelling frameworks perform better than others by design, as survival analysis models are dynamic and have the modelling ability to take on board both accosting-based and market-based information. 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 modelling frameworks. To be more specific, most modelling frameworks improved in performance by taking account of a mixture of account-based and market-based information, where survival analysis, linear probability, and multivariate discriminant analysis models benefited the most from the new way of selecting explanatory variables.

References

AGARWAL, V. & TAFFLER, R. (2008). Comparing the performance of market-based and accounting-based bankruptcy prediction models. *Journal of Banking & Finance,* 32(8), 1541-1551.

AHN, B. S., CHO, S. S. & KIM, C. Y. (2000). The integrated methodology of rough set theory and artificial neural network for business failure prediction. *Expert Systems with Applications,* 18(2), 65-74.

AHN, H. & NEUMANN, L., (2014). Measuring Effectiveness: A DEA Approach Under Predetermined Targets. *International Journal of Business Analytics*, 1(1), 16–28.

ALTMAN, E. (1983). Corporate Financial Distress: A Complete Guide to predicting, Avoiding and Dealing with Bankruptcy. *Edit. John Wiley and Sons*.

ALTMAN, E. I. 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The journal of finance,* 23(4), 589-609.

ANDERSEN, P. & PETERSEN, N. C. (1993). A procedure for ranking efficient units in data envelopment analysis. *Management science,* 39(10), 1261-1264.

AZIZ, M. A. & DAR, H. A. (2006). Predicting corporate bankruptcy: where we stand? *Corporate governance,* 6(1), 18-33.

BALCAEN, S. & OOGHE, H. (2006). 35 years of studies on business failure: an overview of the classic statistical methodologies and their related problems. *The British Accounting Review,* 38(1), 63-93.

BANDYOPADHYAY, A. (2006). Predicting probability of default of Indian corporate bonds: logistic and Z-score model approaches. *Journal of Risk Finance,* 7(3), 255-272.

BANKER, R. D., COOPER, W. W. & CHARNES, A. (1984). Some models for estimating technical and scale efficiencies in Data Envelopment Analysis. *Management Science,* 30(9), 1078-1092.

BANKER, R. D., COOPER, W. W., SEIFORD, L. M., THRALL, R. M. & ZHU, J. (2004). Returns to scale in different DEA models. *European Journal of Operational Research,* 154(2), 345-362.

BEAVER, W. H. (1966). Financial ratios as predictors of failure. *Journal of accounting research*, 4,71-111.

BEMMANN, M. (2005). Improving the comparability of insolvency predictions. Dresden Economics Discussion Paper Series No. 08/2005.

BHARATH, S. T. & SHUMWAY, T. (2008). Forecasting default with the Merton distance to default model. *Review of Financial Studies,* 21(3), 1339-1369.

BRANCH, B. (2002). The costs of bankruptcy: A review. *International Review of Financial Analysis*, 11(1), 39-57

BRIS, A., WELCH, I. & ZHU, N. (2006). Chapter 7 Liquidation versus Chapter 11 Reorganization. *Journal of Finance*, 61(3), 1253-1303

CHARNES, A. & COOPER, W. W. (1962). Programming with linear fractional functionals. *Naval Research logistics quarterly,* 9 (3-4), 181-186.

CHARNES, A., COOPER, W. W. & RHODES, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research,* 2(6), 429-444.

CHORTAREAS , G.E., GARZA-GARCIA, J.G. & GIRARDONE, C. (2012). Competition, efficiency and interest rate margins in Latin American banking. *International Review of Financial Analysis*, 24, 93-103.

CHRISTIDIS, A. & GREGORY, A. (2010). Some new models for financial distress prediction in the UK. *Xfi-Centre for Finance and Investment Discussion Paper* no: 10/04.

COOPER, W. W., SEIFORD, L. M. & TONE, K. (2005). *Introduction to data envelopment analysis and its uses: with DEA-solver software and references*. Springer

DAVYDENKO, S.A., STREBULAEV, I.A. & ZHAO, X. (2012). A market-based study of the cost of default. *Review of Financial Studies*, 25(10), 2955-2999

DEBNATH, R. M. & SEBASTIAN, V.J. (2014). Efficiency in the Indian iron and steel industry − an application of data envelopment analysis. *Journal of Advances in Management Research*, 11(1), 4 -19.

DU, J., LIANG, L. & ZHU, J. (2010). A slacks-based measure of super-efficiency in data envelopment analysis: a comment. *European Journal of Operational Research,* 204(3), 694-697.

ELKAMHI, R., ERICSSON, J. & PARSONS, C.A. (2012). The cost and timing of financial distress. *Journal of Financial Economics*, 105(1), 62-81

FISHER, R. A. (1938). The statistical utilization of multiple measurements. *Annals of Human Genetics,* 8(4), 376-386.

HILLEGEIST, S. A., KEATING, E. K., CRAM, D. P. & LUNDSTEDT, K. G. (2004). Assessing the probability of bankruptcy. *Review of Accounting Studies,* 9(1), 5-34.

KADER , H.A., ADAMS, M., HARDWICH, P. & KWON, W.J. (2014). Cost efficiency and board composition under different takaful insurance business models. *International Review of Financial Analysis*, 32, 60-70.

LIM, S., OH, K.W. & ZHU, J. (2014), Use of DEA cross-efficiency evaluation in portfolio selection: An application to Korean stock market, *European Journal of Operational Research*, 236(1), 361-368.

LOZANO, S. & GUTIERREZ, E. (2008), Data envelopment analysis of mutual funds based on second-order stochastic dominance, *European Journal of Operational Research,* 189(1), 230-244.

Li, Z., Crook, J. & Andreeva, G. (2013). Chinese companies distress prediction: an application of data envelopment analysis. *Journal of The Operational Research Society*, 65(3), 466–479.

LIU, J. S., LU, L. Y. Y., LU, W. M. & LIN, B. J. Y. (2013). A survey of DEA applications. *Omega*, 41(5),893-902.

MCDONALD, R. (2002). *Derivative Markets*, Boston, MA: Addison Wesley.

MCFADDEN, D. (1976). A comment on discriminant analysis 'versus' logit analysis. *Annals of Economic and Social Measurement,* 5(4)*,* 511-523.

MERTON, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *The Journal of Finance,* 29(2), 449-470.

OHLSON, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of accounting research*, 18(1),109-131.

PRADI, J.C., ASMILD, M. & SIMAK, P.C. (2004). Using DEA and worst practice DEA in credit risk evaluation. *Journal of Productivity Analysis*, 21(2), 153–165.

PREMACHANDRA, I. M., BHABRA, G. S. & SUEYOSHI, T. (2009). DEA as a tool for bankruptcy assessment: A comparative study with logistic regression technique. *European Journal of Operational Research,* 193(2), 412-424.

PREMACHANDRA, I.M., CHEN, Y. & WATSON, J. (2011). DEA as a tool for predicting corporate failure and success: A case of bankruptcy assessment. *Omega*, 39(6), 620–626.

PSILLAKI, M., TSOLAS, I.E. & MARGARITS, D. (2010). Evaluation of credit risk based on firm performance. *European Journal of Operational Research*, 201(3), 873-881.

RAVI KUMAR, P. & RAVI, V. (2007). Bankruptcy prediction in banks and firms via statistical and intelligent techniques–A review. *European Journal of Operational Research,* 180(1), 1-28.

SEIFORD, L. M. (1997). A bibliography for data envelopment analysis (1978-1996). *Annals of Operations Research,* 73, 393-438.

SHETTY, U., PAKKALA, T.P.M. & MALLIKARJUNAPPA, T. (2012). A modified directional distance formulation of DEA to assess bankruptcy: An application to IT/ITES companies in India. *Expert Systems with Applications* , 39(2), 1988-1997

SHUMWAY, T. (2001). Forecasting Bankruptcy More Accurately: A Simple Hazard Model. *The Journal of Business,* 74(1), 101-124.

TAFFLER, R. J. (1984). Empirical models for the monitoring of UK corporations. *Journal of Banking & Finance,* 8(2), 199-227.

THEODOSSIOU, P. (1991). Alternative models for assessing the financial condition of business in Greece. *Journal of Business Finance & Accounting,* 18(5), 697-720.

TINOCO, M. H. & WILSON, N. (2013). Financial Distress and Bankruptcy Prediction amongst Listed Companies using Accounting, Market and Macroeconomic Variables. *International Review of Financial Analysis*, 30, 394-419.

TONE, K. (2001). A slacks-based measure of efficiency in data envelopment analysis. *European journal of operational research,* 130(3), 498-509.

TONE, K. (2002). A slacks-based measure of super-efficiency in data envelopment analysis. *European Journal of Operational Research,* 143(1), 32-41.

VAN GESTEL, T., BAESENS, B., VAN DIJCKE, P., GARCIA, J., SUYKENS, J. A. K. & VANTHIENEN, J. (2006). A process model to develop an internal rating system: sovereign credit ratings. *Decision Support Systems,* 42(2), 1131-1151.

WANG ,W. K., LU, W., M. & LIU, P. Y. (2014). A fuzzy multi-objective two-stage DEA model for evaluating the performance of US bank holding companies. *Expert Systems with Applications*, 41(9), 4290-4297.

XU, X. & WANG, Y. (2009). Financial failure prediction using efficiency as a predictor, *Expert Systems with Applications*, 36(1), 366 - 373.

YEH, C.,C., CHI, D., J. & HSU, M., F. (2010), A hybrid approach of DEA, rough set and support vector machines for business failure prediction. *Expert Systems with Applications* , 37(2), 1535–1541.

ZHANG , J., JIANG, C., QU, B. & WANG, P. (2013), Market concentration, risk-taking, and bank performance: Evidence from emerging economies. *International Review of Financial Analysis,* 30, 139-157.

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