Chinese Companies Distress Prediction: An Application of Data Envelopment Analysis

Digital Object Identifier (DOI):
10.1057/jors.2013.67

Published In:
Journal of the Operational Research Society

Chinese Companies Distress Prediction: An Application of Data Envelopment Analysis

Zhiyong Li[*], Jonathan Crook, Galina Andreeva

Credit Research Centre, Business School, University of Edinburgh, UK

Abstract

Bankruptcy prediction is a key part in corporate credit risk management. Traditional bankruptcy prediction models employ financial ratios or market prices to predict bankruptcy or financial distress prior to its occurrence. We investigate the predictive accuracy of corporate efficiency measures along with standard financial ratios in predicting corporate distress in Chinese companies. Data Envelopment Analysis (DEA) is used to measure corporate efficiency. In contrast to previous applications of DEA in credit risk modelling where it was used to generate a single efficiency - Technical Efficiency, we assume Variable Returns to Scale, and decompose Technical Efficiency into Pure Technical Efficiency and Scale Efficiency. These measures are introduced into Logistic Regression to predict the probability of distress, along with the levels of Returns to Scale. Effects of efficiency variables are allowed to vary across industries through the use of interaction terms, whilst the financial ratios are assumed to have the same effects across all sectors. The results show that the predictive power is improved by this corporate efficiency information.

Keywords: Data Envelopment Analysis; efficiency; corporate credit risk modelling; financial distress

[*] *Correspondence: Z Li, Room 3.02, Business School, 29 Buccleuch Place, Edinburgh EH8 9JS, UK, zhiyong.li@ed.ac.uk*

Introduction

The recent financial crisis indicates the importance of credit risk management and the necessity of recognising early warnings of corporate financial distress in order to prevent potential losses. Credit scoring models are such tools to generate early signals of corporate bankruptcy which have received academic attention since at least 1950s and are still widely used.

One of the main problems in failure prediction models is variable selection. Financial ratios which are the quotient of two items in financial statements are the most popular variables that have been considered in the literature. Beaver (1966) was the first author to introduce financial ratios into bankruptcy prediction. In recent decades there have been a great number of bankruptcy prediction studies based on financial ratios using different statistical and machine-learning techniques, these are reviewed in Altman (1993), Balcaen and Ooghe (2006), Kumar and Ravi (2007), Bahrammirzaee (2010), Verikas *et al* (2010). Recent papers (e.g. Wang and Ma, 2011) also demonstrate that financial ratios are still dominating the variable selection. However, it is widely recognized that the main cause of the company's financial failure is its poor management (Gestel *et al*, 2006). The quality of management can be measured by the company's efficiency which compares outputs to inputs.

One way to assess the efficiency of an organisation relative to the most efficient one is to use Data Envelopment Analysis (DEA). A number of papers have used DEA efficiencies in corporate bankruptcy modelling (see next section). In this paper we use DEA to compute various measures of company efficiency that we then input as a variable in a standard classifier to see how well this enables one to predict financial distress. The paper makes a number of contributions. First, unlike previous papers on corporate failure modelling that simply use a single efficiency measure, we decompose this measure – Technical Efficiency (TE) into Pure Technical Efficiency (PTE) which indicates the ability to improve efficiency by wisely allocating resources and applying new technology and Scale Efficiency (SE) which measures the ability to achieve better efficiency by adjusting to its optimal scale, and examine how each of these separately contributes to predicting financial distress. Second, in contrast to most applications of DEA in financial distress prediction we assume variable rather than constant Returns to Scale. Third, DEA can only meaningfully be carried out for a sample of firms that use the same or similar technology (Dyson, 2001) and our study is the first to meet this requirement in the context of mixed-industry bankruptcy prediction. Whilst this reduces our sample size, by modifying the second stage logistic regression we are able to determine the effects of variables that are common across industries. Fourth, we add corroboratory evidence to the very few studies that, regardless of country, have explored the corporate efficiency as a predictive variable in a financial distress model.

The paper is organized as follows. The next section provides a comprehensive review of the application of DEA in corporate distress prediction models. In the third section the methodology adopted in this research is presented. This is followed by the description of the data used in the empirical analysis and the subsequent section reports the results. The paper finishes with some discussion and conclusions.

Literature Review

Data Envelopment Analysis is an optimizing technique which measures the relative efficiencies of a group of companies or Decision Making Units (DMUs) that use multiple inputs and produce multiple outputs. An efficient company uses less inputs to produce more outputs. Such efficiency is evaluated by the distance of a particular DMU to the efficient frontier (ideal position) which is based on its peers (other DMUs in the sample). The main idea and notation will be introduced in the next section, for more comprehensive explanation of DEA see Cooper *et al* (2000).

DEA has been incorporated into the prediction of corporate distress (or bankruptcy) in two different ways. Firstly, DEA has been used to derive a classification algorithm to separate distressed firms from non-distressed firms (Paradi *et al,* 2004; Cielen *et al* 2004; Emel *et al,* 2003). Secondly, the relative efficiency of firms has been computed using DEA and this relative efficiency has been used as a feature of each firm in a subsequently developed classification rule (Xu and Wang, 2009; Yeh *et al,* 2010; Psillaki *et al*, 2010). We consider the former first.

As a classifier DEA has a number of advantages compared with statistical methods. For example it is non-parametric and so does not require any distributional assumptions about error terms or about covariance matrices. But DEA also has some inherent disadvantages such as sensitivity to the selection of inputs and outputs, and issues when dealing with negative values. When the number of variables are close to or larger than the number of companies, efficiency scores tend to be 1 so discriminative power is lost.

It is logical to assume that efficiency is associated with the probability of failure. Barr *et al* (1993) found there are significant differences of scores in a sample of banks between the surviving and failing and the difference increases as the date of failure approaches. Paradi *et al* (2004) used an additive DEA model to compute a worst performance boundary. Output variables are those that reflect poor financial performance such as bad debt, warranty claims etc and input variables represent the opposite, for example profits, sales etc. For each DMU, an inefficiency score is computed. Paradi *et al* (2004) then use the layer technique (or tiered DEA, Barr *et al*. 2000) of removing inefficient companies to find a new boundary, each lower boundary indicating a lower chance of bankruptcy. A similar method is followed by Cielen *et al*

(2004) who apply a cut-off to the estimated efficiency of each DMU (rather than the layer technique). They find, in a comparison of classification accuracy, that the DEA method outperformed decision trees and a linear programming method (Freed and Glover, 1981). However they used the ratio form of the DEA model which is problematic when negative financial ratios are incorporated. Min and Lee (2008) estimated a CCR model (defined in the next section) with constant Returns to Scale and applied a cut-off to the efficiency score for each firm. The DEA score method performed less well than a linear discriminant function. Premachandra *et al* (2009) estimated an additive DEA, which is invariant to data translation (and so can deal with negative data) with varying Returns to Scale. On the training sample DEA had an inferior predictive performance whereas out of sample it was superior. Unfortunately they could not compare the performance of both techniques using the same test dataset. More recently Premachandra *et al* (2011) estimate an additive DEA model to derive efficiency and a bankruptcy frontier and derive a prediction index for each firm from these two. They find the use of a two frontier method improves predictive performance compared to a single bankruptcy frontier. Sueyoshi (1999) proposed a two stage method labelled 'DEA-DA'. In the first stage a linear program is used to predict class membership of each case and to identify cases where the predicted class is ambiguous (since two discriminating functions are computed). In the second stage a model that classifies cases that could fit into either group is estimated. Subsequent work has compared the performance of the two stage classifier with that of other standard methods (Sueyoshi 2001 and 2006; Sueyoshi and Goto, 2009; Tsai *et al,* 2009) with the conclusion that DEA-DA performs at least as well as other techniques for corporate bankruptcy prediction and better in the case of consumer loans.

As the second way of incorporating DEA into distress prediction, many researchers have carried out experiments to incorporate a DEA efficiency score (or Technical Efficiency - TE) as a predictor into other classification models. Xu and Wang (2009) put efficiency score obtained by DEA into Support Vector Machines (SVMs), logistic regression and linear discriminant analysis (MDA). Yeh *et al* (2010) also use efficiency scores into SVMs and neural networks. Both studies found that the inclusion of efficiency scores increased predictive performance of failed companies.

A limitation of many studies that have used DEA efficiency in bankruptcy prediction is that they have estimated TE across a range of industries that use heterogeneous technologies (Cielen *et al,* 2004; Premachandra *et al,* 2009; Premachandra *et al,* 2011). If the technology used by the DMUs in the sample is different then the weights on the inputs and outputs will be different and the concept of efficiency will be somewhat meaningless. Otherwise, the analysis has to use a single industry which obviously limits the sample size (e.g. Shetty *et al,* 2012).

The use of a DEA classifier or an efficiency score computes the relative efficiency of firms in a sample and can be used for in-sample prediction. However, if we wish to predict the failure probability for a case out of the sample, difficulties arise because the addition of a new case may alter the relative efficiencies of all of the firms currently included in the model possibly changing the optimal weights on the inputs and the outputs and so altering the efficient frontier. In principle the addition of a new case would necessitate the re-estimation of the DEA model. Both Emel *et al* (2003) and Min and Lee (2008) estimated a statistical model to predict DEA efficiency using the input and output financial ratios that could be used to classify out of sample cases.

Whilst a large number of papers have estimated models to predict financial distress for Chinese listed companies using financial ratios (for example see Sun *et al,* 2011, Xiao *et al,* 2011 and Ding *et al,* 2008), as far as we are aware only one (Xu and Wang, 2009) has considered DEA efficiency as an explanatory variable.

Stiglitz (1972) emphasized that Returns to Scale (RTS) impacts on the probability of bankruptcy. In practice RTS are typically increasing, or decreasing so it is surprising to see most of applications of DEA in corporate failure prediction have an assumption of constant Returns to Scale (CRS). Examples of papers that assume CRS are Xu and Wang (2009) and Yeh *et al* (2010). The paper of Psillaki *et al* (2010) is one of the few cases which assume VRS to evaluate credit risk. They use the BCC model named by Banker, Charnes and Cooper (1984) but with only one output and two inputs.

The contributions of this research are first, to assume a variable Returns to Scale (VRS) technology rather than CRS which is not common in reality, and second, under the assumption of VRS, to include four additional variables in a model to predict financial distress.

These variables are the Technical Efficiency (CRS efficiency), Pure Technical Efficiency (VRS efficiency), Scale Efficiency and a Returns to Scale parameter (defined in the next section). By incorporating these four variables, our prediction models include variables that are economically directly related to the probability of distress. Unlike most European companies which are relatively small in size, Chinese companies are often much larger and their largest number of employees exceeds 100 thousand and total revenue exceeds £20 billion. Therefore, cases of decreasing returns to scale are often observed and it is expected to have some causality for financial difficulty.

Methodology

DEA

Consider a set of DMUs, each denoted as DMU_j ($j = 1,....,n$), each producing several outputs

$y_r(r=1,\ldots,s)$ by using several inputs $x_i(i=1,\ldots,m)$. For any DMU, DMU_0, we wish to find the weight on each output and on each input that maximises efficiency defined as the ratio of weighted outputs to weighted inputs, subject to the ratio being not greater than 1 for any DMU. This fractional programming problem can be converted into a linear program (Cooper *et al,* 2000) and for convenience the dual program is usually considered:

$$\max \quad v = \mathbf{u}^T \mathbf{y}_0 - u_0 \qquad (1)$$

$$s.t. \quad \mathbf{v}^T \mathbf{x}_0 = 1 \qquad (2)$$

$$-\mathbf{v}^T X + \mathbf{u}^T Y - u_0 \mathbf{e}^T \leq 0 \qquad (3)$$

$$\mathbf{v} \geq 0, \mathbf{u} \geq 0$$

where **u** and **v** are column vectors of weights to be estimated. If (x_0, y_0) is on the efficient frontier then at this point $u_0^* > 0, u_0^* = 0,$ and $u_0^* < 0$ implies and is implied by increasing, constant and decreasing returns to scale respectively (Banker and Thrall 1992). In a one input one output context the u_0 term would be the intercept for the line referred to above. Furthermore, if θ_C^* and θ_B^* denote CCR and BCC efficiency scores of a particular DMU then Scale Efficiency is defined as (Charnes *et al,* 1978)

$$SE = \frac{\theta_C^*}{\theta_B^*} \qquad (4)$$

Intuitively, the BCC model finds the optimal efficiency for a DMU when returns to scale are not necessarily constant. Dividing the efficiency of a DMU when estimated with constant returns to scale by the efficiency when VRS are assumed isolates the Scale Efficiency of the DMU. Thus we can write:

Technical efficiency (TE) = Pure Technical Efficiency (PTE) × Scale Efficiency (SE)

Selecting Inputs and Outputs

Choosing the most appropriate inputs and outputs is of crucial importance when conducting all DEA studies, but so far, there is no generally agreed method for the selection. Different DEA studies have used different inputs and outputs, which is a shortcoming of DEA (Premachandra *et al*, 2009). First of all, inputs and outputs have to be meaningful within the framework of the competitive environment (Oral and Yolalan, 1990). One disadvantage of DEA is that it computes relative efficiency with more discrimination between DMUs when the number of variables is significantly smaller than the number of DMUs (Parkan, 1987). This is normally the case in recent research. It is desirable that the number of input variables is larger than or equal to the number of output variables (Yeh, 1996).

In the few studies that use DEA to model default risk, input variables are selected from Capital, Liability, Human Resources, Technology, Real Estate etc. and the output variables are profit and sales. For example, Psillaki *et al* (2010) used one output (Value Added) and two inputs, Capital Shares and Number of Fulltime Employees. One may argue about the scope of 'Value Added' and how it should be calculated. Similarly, Yeh *et al* (2010) selected R&D Expenses, R&D Designers and the Number of Patents and Trademarks as input variables and the output variables included Gross Profit and Market Share.

When empirically modelling bankruptcy, to eliminate scale or size and unit effects in the values, it is common to use financial ratios rather than physical or monetary items. Min and Lee (2008) include three input ratios which are Financial Expenses to Sales, Current Liabilities Ratio, Bond Payable to Total Assets, an ordinal variable (Total Borrowings) and three output ratios: Capital Adequacy Ratio, Current Ratio and Interest Coverage Ratio. Cielen *et al* (2004) argue that financial ratios with a positive correlation can be used as inputs while those with a negative correlation are output. Premachandra *et al* (2009) propose that the smaller (inferior) values in the financial ratios, which could possibly cause financial distress, are considered to be inputs whereas the larger (superior) values in those ratios, which could cause financial distress, are considered as outputs. Xu and Wang (2009) in a Chinese case study go back to the original definition of efficiency for variable selection. They use Total Assets, Total Liabilities and Costs of Sales as the inputs, Income from Sales as the output.

Our choice of variables has been influenced by the following considerations. Since financial ratios are going to be used in a second stage logistic regression we do not employ them in the first stage so as to reduce possible collinearity. We follow the original idea of DEA that inputs and outputs are measured as absolute amounts rather than as ratios. Thus we have chosen five inputs (Number of Employees, Share Capital, Total Cost, Total Assets and Total Liabilities) and three outputs (Total Sales, Total Profit and Cash Accrued) which are main items in all financial reports.

A key issue regarding DEA is how to deal with negative values in inputs and outputs such as for growth or profits. There are three popular methods which have been used: the Range Directional Measure (RDM) proposed by Portela *et al* (2004), the Modified Slack-Based Measure (MSBM) proposed by Sharp *et al* (2006), and the Semi-Oriented Radial Measure (SORM) proposed by Emrouznejad *et al* (2010). Recently a fourth method, Variant of Radial Measure (VRM), has been introduced by Cheng *et al* (2011).

Our data output matrix, Y, has negative values and we wish to assume VRS, which is both unit invariant and translation invariant and can handle positive and negative mixed data. A suitable model is the slacks based efficiency model which in input orientation can be expressed as (from Cooper *et al,* 2000):

$$\min \quad \rho = 1 - \frac{1}{m}\sum_{i=1}^{m}\frac{s_i^-}{x_{i0}} \qquad (5)$$

$$s.t. \quad \mathbf{x}_0 - X\boldsymbol{\lambda} - \mathbf{s}^- = \mathbf{0} \qquad (6)$$

$$\mathbf{y}_0 - Y\boldsymbol{\lambda} + \mathbf{s}^+ = \mathbf{0} \qquad (7)$$

$$\boldsymbol{\lambda} \geq \mathbf{0}, \mathbf{s}^- \geq \mathbf{0}, \mathbf{s}^+ \geq \mathbf{0}$$

MaxDEA is used to solve the programs for each industry separately.

We deduce a score for each DMU for each type of efficiency and relate these to the probability of distress using logistic regression. However, DEA scores assume a common technology across the DMUs. When we include the four types of efficiency variables we ensure that only DMUs within the same industry sector are accorded the same parameters whilst the financial ratios are assumed to have the same parameters across all sectors. Therefore the specification of the logistic regression is amended to be

$$\text{logit}(p_q) = \alpha + \sum_{l=1}^{L}\sum_{p=1}^{P}\delta_{pl} D_p e_{plq} + \beta_1 w_{1q} + \beta_2 w_{2q} + ..\beta_K x_{Kq} \qquad (8)$$

where p_q denotes the probability of suffering distress for company *q;*

e_{pql} denotes efficiency score type *l* for sector *p* for company *q;*

w_{1q} denotes financial variable 1 for company *q* and so on;

D_p =1 if company *q* is a member of industry *p*, 0 otherwise;

δ_{pl} denotes a parameter for industry *p* for efficiency score *l* to be estimated;

β_1 denotes a parameter for covariate 1 to be estimated.

We compared alternative specifications of equation 8: with only efficiency variables, with only financial variables and with combinations of both.

Data

The data used in this research is from two Chinese security markets, the Shanghai Stock Exchange and the Shenzhen Stock exchange and sourced from the Wind database. The database provides information for those companies listed in both markets (note that no cross listing is allowed) and covers the historic records from 1991. The sample contains the annual data of 2014 listed companies in China between 1998 and 2010. Since one of the important input variables in the DEA models is Number of Employees and it was not until 2001 that the companies started to report this information in their statements, the reports prior to 2001 are excluded from the sample. A few companies with extreme outlying values of input or output variables (mainly caused by unusual or abnormal value changes and rare events) were also excluded because the efficient frontier is very sensitive to outlying values and so their inclusion may have resulted in inaccurate estimates of relative efficiencies. 'Special Treatment' (ST) is the status imposed by the government to give notice of a bad performance to investors and so it is an indicator of financial distress used in this research. A company is ascribed ST status if any of the following conditions holds (Shanghai Stock Exchange, 2008):

- negative profit in the most recent two consecutive years or if the correction of errors yields this result;
- failure to disclose its annual interim report;
- likelihood of being dissolved;
- reorganisation, settlement or bankruptcy liquidation;
- other characteristics determined by the Stock Exchange.

The majority of companies receive ST because of losses in two successive fiscal years.

Since DEA models are estimated from homogeneous production processes (Dyson 2001), we solve DEA programs to compute efficiency scores for separate industry sectors and within the same year to ensure that the companies in the sample share the same productivity process and a similar business environment. To keep as many distressed companies as possible in the sample for modelling, all industries were examined and the second level industrial sectors Raw Materials (code 1510 in Wind), Industrial Equipment (2010) and Real Estate (4040) were found to have the highest frequency of ST cases. In 2002, 2003, 2006 and 2007, there are more ST cases than in other years. Therefore the STs in 2002 or 2003 are grouped together as the training sample and the STs in 2006 or 2007 are grouped into the hold-out sample to test the predictive performance of the logistic regression. Thus efficiency scores and financial covariate data for 2001 with ST/non-ST status taken from 2002 and 2003 were used to train the model, which then was then applied to the data in 2005 to predict the probability of becoming ST in 2006 and 2007. The numbers of ST and non-ST companies are displayed in the Table 1. Some companies were delisted and some new companies entered the sample during the study period. There are 429 cases common in both samples. The predictive accuracy is tested by an out-of-time rather than an out of sample validation, which is in line with the literature (e.g. Shumway, 2001).

Descriptive statistics for the financial variables used in the DEA analysis are shown in Table 2. The occasional negative values for profits and cash flows are apparent.

The financial ratios collected from the database contain 6 groups of measures relating to profitability, operation capacity, growth rates, capital composition, liquidity, cash flow. Those variables with too many missing values were deleted. Variables that were highly correlated

(VIF > 20) were also excluded. For those variables where only a few values were missing, the missing values were replaced by the means in that year. The final list of ratios selected for inclusion in the logistic regression and represented by w_q in equation 8 is in Table 3.

Results

DEA

There are four types of efficiency scores of importance to this paper: Technical Efficiency, Pure Technical Efficiency, Scale Efficiency and Returns to Scale levels. The first three are all continuous scores whereas Returns to Scale is a categorical ordinal variable with three levels: decreasing, constant and increasing.

First, we consider aggregate results. One of the objectives of this paper is to test whether the probability of distress is associated with low efficiency. We consider various efficiency measures where following previous literature (Xu and Wang, 2009) we do not treat each sector separately and then secondly when we do treat each sector separately. Descriptive statistics of efficiency scores are shown in Table 4. As a preliminary analysis we computed two-way ANOVA and found that for each of the three types of efficiency score, there is a significant difference between the mean score for the ST group and the mean score for the non-ST group. But there is a significant difference between the mean efficiency scores between the industry sectors in 2001 only in terms of Technical Efficiency (CCR) and Scale Efficiency and for 2005 only for Scale Efficiency.

From Table 5 we can see that both in 2001 and 2005 there are relatively low numbers of companies with decreasing or constant RTS. We therefore classified the RTS values into two values: decreasing or constant on the one hand and increasing (IRS) on the other and included a dummy variable to represent the existence of IRS in the logistic regressions.

Logistic Regression

We have two objectives. First, to investigate the statistical significance of efficiency measures in explaining the probability of suffering financial distress and second, to evaluate the predictive performance of including efficiency variables in such posterior probability models.

Pre-analysis showed that if efficiency variables and financial ratios are entered together into a stepwise logistic regression, nearly all of the efficiency variables are excluded. However, we are interested in the role specifically of efficiency variables and so we adopted the following procedure. Since values of the efficiency variables were derived from a DEA model where the objective function consisted of financial variables, collinearity is possible between some financial ratios and some efficiency scores. Conscious of this potential collinearity we considered three model specifications. First we have models with only efficiency variables (Models 1-6). Models 1-4 contain only industry specific efficiency variables to try to reduce the heterogeneity in technologies that would otherwise be present. Models 5 and 6 are included simply to show the parameter estimates if, as in previous literature, in the DEA analysis all industrial sectors were assumed to be homogeneous.

Second, we estimated models that included combinations of the industry specific efficiency variables and subsequently uncorrelated financial ratios were entered using a stepwise routine (Models 7-9). Third, we estimated models that included significant financial variables selected from all those available using a forward stepwise routine together with combinations of efficiency scores. Thus the efficiency score was 'force' entered in each model, except Model 10. All of the models were parameterised across all industries with industry specific dummies interacted with each efficiency variable to yield industry specific parameters and the efficiency scores. We therefore assume the marginal effects of the efficiency variables are specific to each industry sector but the marginal effects are the same for each financial variable for all industries. The models are specified in Table 6.

DEA allows one to compute the efficiency of an organisation relative to the most efficient organisations in the dataset. To compute the relative efficiency scores for a new case requires us to solve the program for a different set of DMUs and so could alter the efficiency boundary and thus the efficiencies of the original cases relative to the new efficiency boundary. To assess the discriminatory power of including efficiency variables we computed the relative efficiency for each member of the holdout sample in 2005. We assumed that the marginal effects of relative TE, PTE and SE, and so the logistic regression parameters that were estimated for 2001-3, remained constant over time. We argue that in competitive markets it is relative efficiency rather than absolute efficiency that determines the chance of financial success or, as in our case, financial distress. This is consistent with the approach used in the literature (see Xu and Wang 2009). We then predicted the probability of a new case becoming distressed in 2006-7 using the 2005 efficiencies and 2001-3 parameters.

Parameters and Significance Levels

Table 7 shows that when included alone each of the efficiency variables had the expected sign: an increase in efficiency is associated with a decrease in the probability of distress. This is true when we consider TE alone or PTE and SE together. The effect of a marginal change in relative TE score for Real Estate has a smaller effect on the probability of distress than in the Industrial Equipment industry. Generally an increase in relative PTE has a smaller marginal effect on distress likelihood that does an increase in relative SE. RTS (either constant-decreasing or increasing) have no detectable effect of the probability of distress. A failure to compute relative efficiency for each industry sector separately and so to assume homogeneity of technology across all three sectors not only yields incorrect efficiency scores but if such scores are used masks considerable differences in the effects of each type of efficiency between the industry sectors.

Table 8 shows that when we force the efficiency scores into each logistic regression, and then select financial variables in a stepwise fashion the scale efficiency variables remain significant with the expected signs whilst the RTS variables are never significant. In all sectors improving relative PTE has a smaller effect of the chance of distress than an improvement in relative scale.

The parameters of most of the financial ratios have the expected signs. For example, higher net cash flow per share or higher return on equity or return on assets is associated with a lower chance of distress. In Table 9 we see that if we include the efficiency variables and the financial ratios that would be included if the efficiency variables were not, then only the scale efficiency scores remain significant. Again their parameters have the expected signs.

Predictive Performance

The statistical significance of a covariate does not necessarily imply that predictive performance is increased if the variable is included in a model. We now examine the predictive performance of all of our models. First we compare the predictive performance of using overall efficiency (TE) versus decomposed efficiency (PTE and SE), second we compare models with RTS levels versus models without RTS levels and third we compare models with and without financial ratios. The Area Under ROC curve (AUROC), the Gini coefficient and Error Rates are reported (Table 10 and Figure 1). For Error Rate calculation the proportion of STs that are predicted to be STs is the proportion of the observed STs in the training sample.

In the first comparison (Model 1 v 2 and Model 5 v 6) both pairs show that decomposition of efficiency scores reduces the classification accuracy in the test samples by a noticeable amount. The Gini decreases from 0.841 to 0.797 and from 0.833 to 0.781 if TE is decomposed into PTE and SE. In the second comparison (Models 1 v 3 and Model 2 v 4) we see that inclusion of Returns to Scale decreases predictive performance slightly. For example, without RTS Model 1 has a Gini of 0.841 whilst with RTS this is 0.829 in the test set and the corresponding figures for Models 2 and 4 are 0.797 and 0.791 respectively.

One might notice that for Models 1-6 (with only efficiency variables) the Gini for the test set exceeds that for the training set. We explain this unusual observation with reference to a particular model. Consider Table 4 and industry 4040 (Real Estate). Model 1 consists only of the TE variable. Notice that the difference between the mean TE between the ST and not-ST groups in the training set (0.632 - 0.368 = 0.264) is less than that in the test set (0.578 - 0.207=0.371). In a Kolmogorov-Smirnov diagram (Figure 2) the increase in the difference in the mean TE between the two groups will move the $P_{non\text{-}ST}(s)$ line further from the $P_{ST}(s)$ line in the test set than in the training set, where $P_{non\text{-}ST}(s)$ and $P_{ST}(s)$ denote the cumulative proportions at and below each score, s, of non-STs and STs respectively. Therefore plotting $P_{non\text{-}ST}(s)$ against $P_{ST}(s)$ in a ROC curve graph will result in a more accentuated curve and so the greater difference in means will result in a larger Gini (see Thomas *et al,* 2002).

Turning to the inclusion of financial ratios, we see that they outperform the first six models that contain only efficiency variables. For each performance measure we highlight the model with the greatest predictive power. Generally, in the training sample the models of efficiency variables assisted by ratios are better in predictive accuracy than the models of ratios assisted by efficiency variables. But in the test sample, it is the other way around. In the test sample the highest classification accuracy and the highest discriminatory power is gained by Model 11 that includes industry specific TE together with the most significant of all financial ratios. However the difference between the performance of this model and models 12 and 13 that have the same financial ratios but decompose TE and include RTS (Model 13), is inconsequential.

Discussion and Conclusion

Data Envelopment Analysis is a useful method to measure relative corporate efficiency and corporate efficiency is found to be helpful in credit scoring in previous literature and this paper as well. Rather than assuming Constant Returns to scale, this paper adopts a more realistic assumption, Variable Returns to Scale. It allows the model to decompose overall technical efficiency into Pure Technical Efficiency and Scale Efficiency which actually provides more information for analysis. Practically, these measures indicate that an inefficient company should improve its efficiency of use of inputs or adjust its operating scale to the optimum level to achieve better performance. Our results show not only those less technically efficient firms are at greater risk of becoming financially distressed than more technically efficient firms but that improvements in both pure technical and scale efficiency would reduce the risk. Of these two what really matters is how relatively scale efficient, rather than how pure technically efficient, firms are. This indicates that a firm which wants to perform better, in practice, should pay more attention to optimising its scale of business rather than optimising resources or applying new technology. Increasing scale of operation is likely to have a great effect on reducing risk of distress than moving on an efficiency frontier.

These results are consistent with those of Psillaki et al (2010) who found that technical efficiency was significantly negatively related to the probability of business failure for a sample of French firms in each of three industries. But because no study that models financial distress has decomposed technical efficiency no further comparison can be made.

However, in the prediction of financial distress, decomposition of efficiency variables reduces prediction accuracy. A simpler model is more effective, using Technical Efficiency only to assist financial ratios in logistic regression and that give best results in both model training and out-of-time validation. We also found that the variable level of Returns to Scale had no detectable effect on the probability of being in distress.

In terms of using efficiency as the only predictor, our results show that a group of financial ratios does outperform efficiency scores as they can cover many aspects of business while a DEA score is only based on a limited number on inputs and outputs. That is also the reason why financial ratios have dominated the corporate credit prediction for decades. However, to gain greatest predictive accuracy, financial ratio and efficiency variables should be included. This is consistent with the findings of Yeh et al (2010) and Xu and Wang (2009). Furthermore, in order to test the robustness of the results, in the beginning of modelling, we have also tried another list of financial ratios with less collinearity (VIF < 5). The results do not vary too much, except the selected ratios are different. There is only a little difference on the third decimal places in GINI. The same conclusion remains.

Nevertheless, although predictive accuracy is the main concern in credit risk management, there is also the necessity to understand risk drivers that may give early indications of potential problems. In this respect decomposed efficiency measures, in particular scale efficiency, can provide useful information to a credit analyst interested in relative performance of companies is a credit portfolio.

This paper has also introduced a modified logistic regression model, particularly for DEA variables. This is the first application of DEA in credit scoring to use the dummy variables for different industries to overcome the dilemma that a large sample size and homogeneity of DMUs cannot be achieved at the same time. Industry specification slightly improves prediction accuracy and remarkably increases discriminative power. More importantly, the proposed logistic regression properly handles the assumption of DEA methodology which should be kept all the time when apply it. Such methodology allows employing a large dataset with a mixture of industries, but it needs to be noted that the more industries are involved, the more dummy variables are needed, and the number of companies in each category should still be large enough.

Finally, it has to be mentioned that the data analysed in this paper covers two time periods. It would be beneficial if more years of data are found to be supportive with the above conclusion in cross sectional analysis. Moreover, the recent development of DEA actually can give estimation of time serial efficiency scores which allow panel analysis across a period of time. The panel models and Malmquist DEA scores are the next step in future work.

Table 1 Sample sizes

	Training sample (2001 to 2003)				*Testing sample (2005 to 2007)*			
Sector Code	1510	2010	4040	total	1510	2010	4040	total
non-ST	181	144	95	420	218	185	92	495
ST	17	14	19	50	18	20	22	60
Total	198	158	114	470	236	205	114	555
ST/Non-ST	9.40%	9.70%	20.00%	11.90%	8.30%	10.80%	23.90%	12.10%
ST rate	8.59%	8.86%	16.67%	10.64%	7.63%	9.76%	19.30%	10.81%

Table 2 Statistics of DEA variables

			2001					2005				
		Sector	*N*	*Mean*	*SD*	*Min*	*Max*	*N*	*Mean*	*SD*	*Min*	*Max*
Inputs		1510	198	3925	5388.8	104	45943	236	4240.6	5622.1	140	44421
	employees	2010	158	2498.2	2353	102	15000	205	2281.4	2312.9	129	19676
		4040	114	1252.9	1594.5	56	13319	114	1038.9	1742.2	34	12568
		1510	198	524.5	1182.4	51	12512	236	604.1	1434.3	60.4	17512
	capitals(mCNY)	2010	158	291.7	224.7	80.2	1884.4	205	323.7	296.7	57.6	2689.6
		4040	114	298.3	234.3	66	1867.7	114	369.7	410.9	53.5	3722.7
		1510	198	1410.6	2935	37.6	25497.9	236	3620.3	9098.1	49.3	108422.4
	costs(mCNY)	2010	158	995.1	1820.5	50	19358.6	205	1656.7	2563.9	110.4	19459.7
		4040	114	506.5	607.9	48	4157.4	114	774.4	964.3	12.2	8528.6
		1510	198	2478.9	4881.2	154.6	58042.1	236	4269.4	10471.3	164.4	142024.2
	assets(mCNY)	2010	158	1671.7	1436.8	198.1	9907.9	205	2273.6	2330.4	172.9	18033.6
		4040	114	1659.2	1507	287	9690.3	114	2428.5	2747.5	27.3	21992.4
		1510	198	1120.2	2561.6	43	31752	236	2248.9	4926	22.6	63097.3
	debts(mCNY)	2010	158	805.8	810.5	45.9	4810	205	1318	1491.5	55.6	9517.7
		4040	114	834.4	978.2	6.5	7380.5	114	1459.5	1675.6	7	13411.2
Outputs		1510	198	88.5	355.4	-1797.4	3709.6	236	281.4	1334.9	-997.2	18310.8
	profits(mCNY)	2010	158	54.9	160.9	-1009.8	1011.8	205	60.4	238.2	-696	2057.1
		4040	114	44.1	126.1	-537.6	501.9	114	40.9	293.6	-1142.2	1976.2
		1510	198	16.7	336.9	-3686.2	872.5	236	-6.6	334.4	-2664.6	1784.3
	cash(mCNY)	2010	158	50.1	208.1	-686.4	882.7	205	-15.6	160.9	-953.8	661.3
		4040	114	45.9	150.8	-329.9	585.2	114	-22.9	216.3	-1100.2	597.9
		1510	198	1499.3	3142.1	20.1	29170.8	236	3895.2	10297.5	17.9	126608.4
	sales(mCNY)	2010	158	1037.8	1858.1	51.6	19565.1	205	1706.5	2657.2	0.9	19474.2
		4040	114	536.1	663.6	12.2	4455.1	114	825.4	1154.4	3.5	10558.9

Table 3 List of Eligible Financial Ratios

Group	*Ratio*	*Group*	*Ratio*
Profitability (12)	earnings per share (EPS)	Capital composition (5)	book value per share (BPS)
	operating revenue per share		total assets / total liability
	retained earnings per share		equity multiplier
	return on equity (ROE)		current assets / total assets
	return on assets (ROA)		tangible assets / total assets
	return on invested capital (ROIC)		
	gross margin to total sales	Cash flow (6)	net cash flow from operating per share
	operating profit / total sales		net cash flow per share
	operating expenses / total sales		net cash flow from operating / operating revenue
	financial expenses / total sales		net cash flow from operating / total liabilities
	undistributed profits per share		net cash flow from operating / interest bearing liabilities
	EBIT per share(EBITPS)		net cash flow from operating / current liabilities
Liquidity (8)	current liabilities / total liabilities	Operation capacity (4)	inventory turnover
	current ratio		receivables turnover
	quick ratio		current assets turnover
	cash ratio		total assets turnover
	total liabilities / equity		
	EBITDA / total liabilities	Growth rates (4)	operating revenue growth
	surplus capital per share		total profit growth
	surplus reserve per share		net profit growth
			total assets growth

Table 4 Means and standard deviations of efficiency scores

		Training Sample							*Holdout Sample*						
			TE		*PTE*		*SE*			*TE*		*PTE*		*SE*	
Sector code	*ST*	*N*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*N*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
1510	0	181	.557	.209	.628	.214	.886	.117	218	.493	.219	.597	.214	.824	.163
	1	17	.323	.180	.533	.277	.647	.247	18	.237	.086	.439	.199	.614	.233
	All	198	.537	.216	.620	.221	.866	.148	236	.474	.222	.585	.216	.808	.178
2010	0	144	.556	.239	.694	.214	.792	.171	185	.493	.242	.615	.227	.796	.198
	1	14	.231	.082	.473	.243	.545	.188	20	.201	.084	.439	.182	.497	.194
	All	158	.527	.248	.675	.225	.770	.186	205	.465	.247	.598	.229	.767	.216
4040	0	95	.632	.251	.728	.232	.864	.154	92	.578	.285	.706	.269	.824	.221
	1	19	.368	.271	.574	.264	.665	.300	22	.207	.092	.394	.218	.610	.278
	All	114	.588	.272	.702	.244	.831	.199	114	.506	.298	.646	.287	.782	.247
Total	0	420	.574	.231	.673	.222	.849	.151	495	.509	.243	.624	.233	.813	.188
	1	50	.314	.206	.532	.261	.625	.255	60	.214	.087	.422	.199	.574	.241
	All	470	.546	.242	.658	.230	.825	.179	555	.477	.249	.602	.238	.788	.208

Table 5 Levels of Returns to Scale

	RTS 2001				*RTS 2005*			
	Decreasing	*Constant*	*Increasing*	*Total*	*Decreasing*	*Constant*	*Increasing*	*Total*
ST	69	69	282	420	58	66	371	495
Non-ST	5	0	**45**	50	0	0	**60**	60
Total	74	69	327	470	58	66	431	55

Table 6 Models to be compared

A Efficiency Variables Only	
Model 1	Industry specific TE only
Model 2	Industry specific PTE and SE
Model 3	Industry specific TE and RTS
Model 4	Industry specific PTE, SE and RTS
Model 5	Pooled TE
Model 6	Pooled PTE and SE
B Efficiency Variables force entry, financial ratio variables selected by stepwise routine	
Model 7	Industry specific TE forced entry, financial ratios selected by forward stepwise routine.
Model 8	Industry specific PTE and SE forced entry, financial ratios selected by forward stepwise routine
Model 9	Industry specific PTE, SE and RTS forced entry, financial ratios selected by forward stepwise routine
C Financial variables selected by stepwise and then forced entry with efficiency variables	
Model 10	Financial variable selected by forward stepwise routine.
Model 11	Industry specific TE, financial ratios from Model 10
Model 12	Industry specific PTE and SE, financial ratios from Model 10
Model 13	Industry specific PTE, SE and RTS financial ratios from Model 10

Table 7 Coefficient estimates from efficiency only logistic regressions A

Variable	*Model 1*	*Model 2*	*Model 3*	*Model 4*	*Model 5*	*Model 6*
Technical Efficiency Score					-10.52**	
Raw materials	-11.89**		-12.34**			
Industrial Equipment	-14.23**		-24.41**			
Real Estate	-9.40**		-8.56**			
Pure Technical Efficiency Score						-4.95**
Raw materials		-3.82**		-4.39**		
Industrial Equipment		-6.93**		-9.53**		
Real Estate		-5.89**		-5.77**		
Scale Efficiency Score						-7.33**
Raw materials		-9.79**		-10.16**		
Industrial Equipment		-9.02**		-12.27**		
Real Estate		-7.26**		-6.90**		
Returns to Scale						
Raw materials			-0.97	-5.77		
Industrial Equipment			2.08	3.32		
Real Estate			1.39	-1.54		
Constant	2.58**	7.56**	3.47**	8.59**	2.13**	6.20**

Table 8 Coefficient estimates from logistic regressions B

Variable	*Model 7*	*Model 8*	*Model 9*
Technical Efficiency Score			
Raw materials	-8.35**		
Industrial Equipment	-10.67**		
Real Estate	-7.12**		
Pure Technical Efficiency Score			
Raw materials		-2.81	-1.82
Industrial Equipment		-6.19**	-9.75
Real Estate		-6.24**	-5.25*
Scale Efficiency Score			
Raw materials		-10.82**	-11.08**
Industrial Equipment		-9.66**	-16.17**
Real Estate		-8.21**	-7.93**
Returns to Scale			
Raw materials			-1.68
Industrial Equipment			5.24
Real Estate			-2.09
Net cash flow per share	-5.43**	-6.12**	-4.82**
Return on equity	-0.09*	-0.20**	-0.25**
Return on assets	-0.18**		
Gross margin / total sales	-0.07**		
Operating profit / total sales	0.03*		
Financial expenses / total sales	0.13*	0.12*	0.15*
Tangible assets / total assets	-0.04*	-0.05**	-0.06**
Current ratio		3.22**	
Quick ratio			3.31**
Cash ratio		-6.37**	-7.77**
Net cash flow / interest bearing liabilities			
Net cash flow / current liabilities			-5.53*
Inventory turnover	0.36**	0.66**	0.72**
Total assets growth	-0.08**	-0.09**	-0.09**
Constant	3.47**	10.16**	11.74**

Models 7, 8 and 9: Efficiency variables forced entry, financial variables selected by forward stepwise routine.

Table 9 Coefficient estimates from logistic regressions C

Variable	*Model 10*	*Model 11*	*Model 12*	*Model 13*
Technical Efficiency Score				
Raw materials		-3.88		
Industrial Equipment		-5.2		
Real Estate		-3.1		
Pure Technical Efficiency Score				
Raw materials			-0.63	-1.48
Industrial Equipment			-1.15	-2.79
Real Estate			-1.28	-2.16
Scale Efficiency Score				
Raw materials			-5.37**	-5.97*
Industrial Equipment			-5.95**	-8.19*
Real Estate			-4.20*	-4.95*
Returns to Scale				
Raw materials				-1.73
Industrial Equipment				0.07
Real Estate				-1.3
Net cash flow per share	-3.26**	-3.33*	-3.35*	-3.47*
Return on equity	-0.07*	0.07	-0.07*	-0.07*
Return on assets	-0.18**	-0.18*	-0.19**	-0.18*
Gross margin / sales	-0.05*	-0.06*	-0.06*	-0.06*
Operating profit / total sales				
Financial expenses / total sales	0.11**	0.09*	0.11*	0.11*
Tangible assets / total assets				
Current ratio	-1.25**	-0.96*	-1.17*	-1.25*
Quick ratio				
Cash ratio				
Net cash flow / interest bearing liabilities	-3.23	-2.5	-2.4	-3.16
Net cash flow / current liabilities				
Inventory turnover				
Total assets growth	-0.08**	-0.06*	-0.05**	-0.05**
Constant	1.18	2.57*	5.55**	7.79**

Model 10: All variables selected by forward stepwise routine.

Model 11, 12 and 13: Efficiency variables forced entry, financial variables from Model 10.

Table 10 Model results

	Training sample					*Testing sample*				
	Type I error	*Type II error*	*Overall accuracy*	*AUROC*	*GINI*	*Type I error*	*Type II error*	*Overall accuracy*	*AUROC*	*GINI*
Model 1	40.0%	4.8%	91.5%	0.869	0.738	41.7%	4.8%	91.2%	0.921	0.841
Model 2	52.0%	6.2%	88.9%	0.881	0.761	46.7%	5.5%	90.1%	0.898	0.797
Model 3	42.0%	5.0%	91.1%	0.882	0.765	46.7%	5.5%	90.1%	0.915	0.829
Model 4	52.0%	6.2%	88.9%	0.887	0.775	45.0%	5.3%	90.5%	0.895	0.791
Model 5	42.0%	5.0%	91.1%	0.844	0.687	43.3%	5.1%	90.8%	0.917	0.833
Model 6	50.0%	6.0%	89.4%	0.843	0.686	46.7%	5.5%	90.1%	0.891	0.781
Model 7	20.0%	2.4%	**95.7%**	0.970	0.940	36.7%	4.2%	92.3%	0.952	0.904
Model 8	30.0%	3.6%	93.6%	0.979	0.957	40.0%	4.6%	91.5%	0.935	0.869
Model 9	24.0%	2.9%	94.9%	**0.983**	**0.965**	40.0%	4.6%	91.5%	0.935	0.870
Model 10	26.0%	3.1%	94.5%	0.961	0.923	36.7%	4.2%	92.3%	0.954	0.907
Model 11	22.0%	2.6%	95.3%	0.966	0.933	33.3%	3.8%	**93.0%**	**0.957**	**0.914**
Model 12	22.0%	2.6%	95.3%	0.968	0.937	35.0%	4.0%	92.6%	0.956	0.911
Model 13	20.0%	2.4%	**95.7%**	0.972	0.943	36.7%	4.2%	92.3%	0.954	0.907

Type I error occurs when a distressed company is wrongly classified as a non-distressed company.
Type II error occurs when a non-distressed is wrongly classified as a distressed company.

Figure 1 ROC curves

	Training Set	*Test Set*
Model 1 to Model 6		
Model 7 to Model 9		
Model 10 to Model 13		

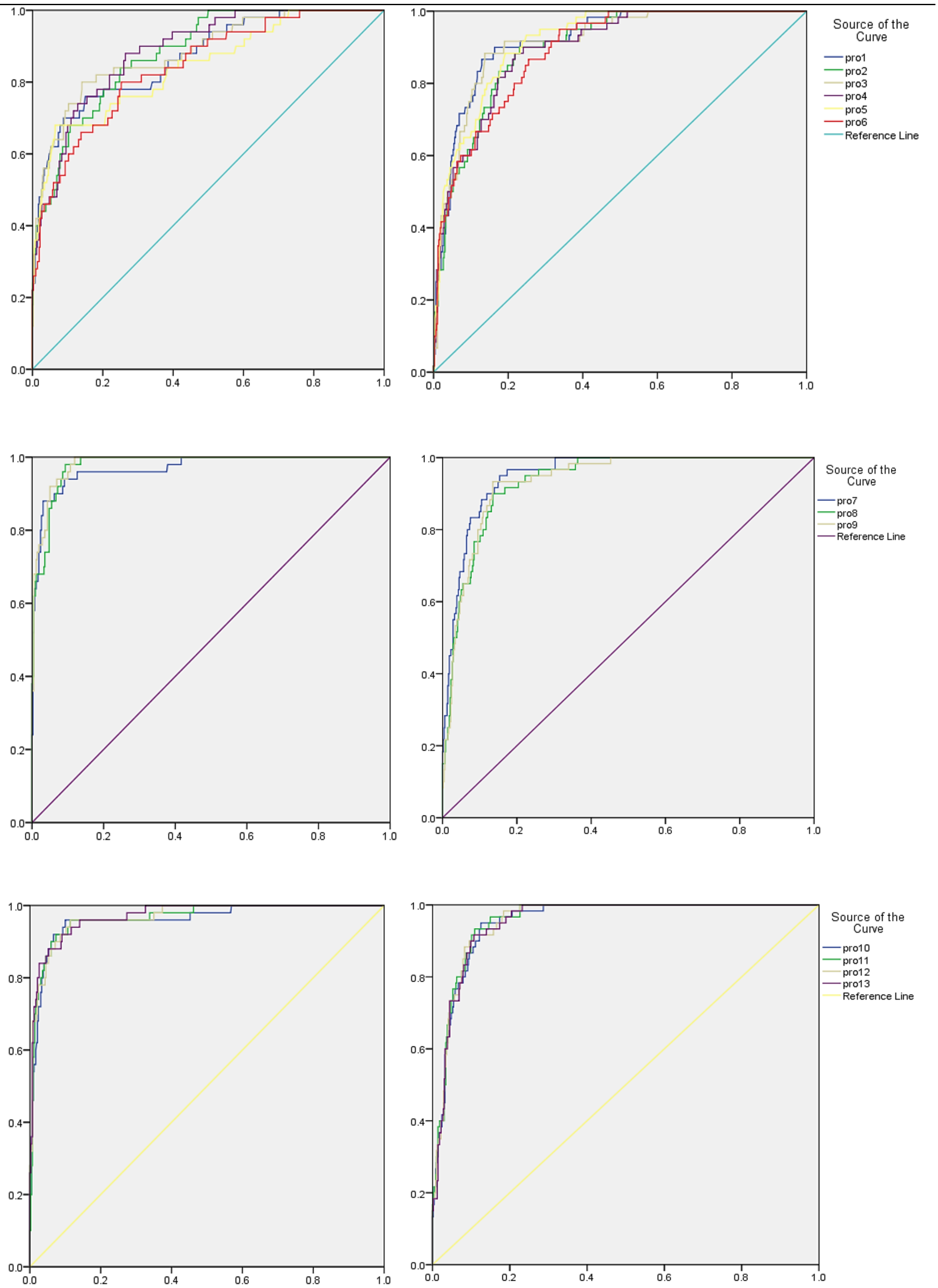
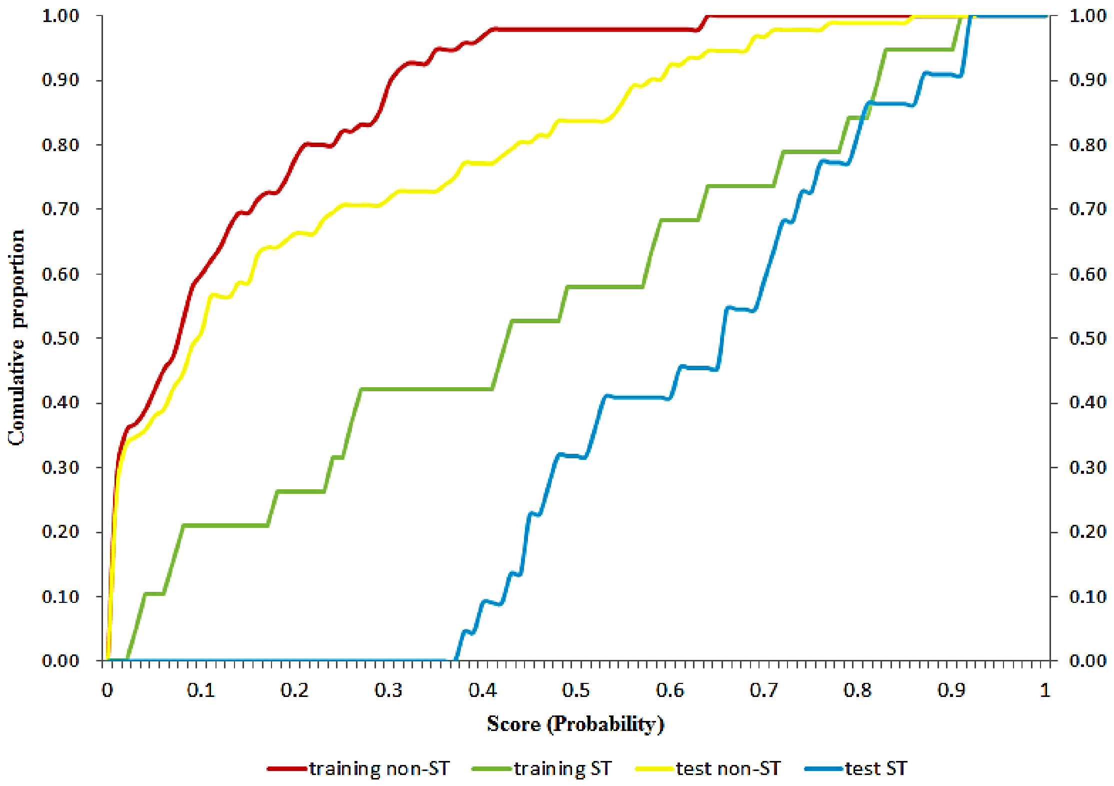


Figure 2 Kolmogorov-Smirnov Plots



Reference

Altman, E.I. (1993). *Corporate Financial Distress and Bankruptcy: A Complete Guide to Predicting and Avoiding Distress and Profiting from Bankruptcy.* New York: Wiley.

Bahrammirzaee, A. (2010). A comparative survey of artificial intelligence applications in finance: artificial neural networks, expert systems and hybrid intelligent systems. *Neural Computing and Applications*, **19** (8): 1165-1195.

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

Banker, R.D., Charnes, A. and Cooper, W.W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science,* **30**: 1078–1092.

Barr, R., Durchholz, M. and Seiford, L., (2000), Peeling the DEA Onion: Layering and Rank-Ordering DMUs Using Tiered DEA, *Southern Methodist University technical report*

Barr, R. S., Seiford L. M. and Siems T. F. (1993). An Envelopment-Analysis Approach to Measuring the Managerial Efficiency of Banks. *Annals of Operations Research.* **45**: 1-19.

Beaver, B.H. (1966). Financial Ratios as Predictors of Failure. *Journal of Accounting Research*. **4**: 71-111

Charnes, A., Cooper, W.W., Rhodes, E.L. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, **2**: 429–444.

Cheng, G., Zervopoulos, P., and Qian, Z., (2011). A variant of radial measure capable of dealing with negative inputs and outputs in data envelopment analysis, MPRA Paper No. 30951

Cielen A., Peeters L. and Vanhoof K. (2004). Bankruptcy prediction using a data envelopment analysis. *European Journal of Operational Research*.**154**:526–32.

Cooper, W.W., Seiford, L.M. and Tone, K. (2000). *Data Envelopment Analysis: A comprehensive text with models, applications, references and DEA-Solver software*, Kluwer Academic Publishers

Ding, Y., Song, X. and Zen, Y. (2008). Forecasting financial condition of Chinese listed companies based on support vector machine. *Expert Systems with Applications*. 34(4): 3081-3089

Dyson, R.G., Allen, R., Camanho, A.S., Podinovski, V.V., Sarrico,C.S. and Shale, E.A. (2001). Pitfalls and protocols in DEA. *European Journal of Operational Research.* **132**: 245–259.

Emel, A.B., Oral, M., Reisman, A. and Yolalan, R. (2003). A credit scoring approach for the commercial banking sector. *Socio-Economic Planning Sciences*. **37**: 103-123.

Emrouznejad A., Anouze A.L. and Thanassoulis E. (2010). A semi-oriented radial measure for measuring the efficiency of decision making units with negative data, using DEA. *European Journal of Operational Research*. **200**: 297-304.

Freed, N. and Glover, F., (1981). A linear programming approach to the discriminant problem. *Decision Sciences*. **12**: 68–73.

Gestel, T. V., Baesens, B., Suykens, J., Poel, D. V., Baestaens, D. E., and Willekens, M. (2006). Bayesian kernel based classification for financial distress detection. *European Journal of Operational Research*, **172**: 979–1003.

Kumar, P. and 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.

Min, J.H. and Lee, Y.C. (2008). A practical approach to credit scoring. *Expert Systems with Applications*. **35**: 1762-1770

Oral, M. and Yolalan, R. (1990). An empirical study on measuring operating efficiency and profitability of bank branches. *European Journal of Operational Research*. **46**: 282-294

Paradi, J.C., Asmild, M. and Simak, P. C. (2001). DEA Based Analysis of Corporate Failure, Working Paper, CMTE, Department of Chemical Engineering, University of Toronto.

Paradi, J.C., Asmild, M. and Simak, P.C. (2004). Using DEA and worst practice DEA in credit risk evaluation. *Journal of Productivity Analysis*. **21**: 153-165

Parkan, C. (1987). Measuring the efficiency of service operations: an application to bank branches. *Engineering Costs and Production Economics.* **12**: 237-242

Portela, M.C.A.S., Thanassoulis, E. and Simpson, G. (2004). Negative data in DEA: a directional distance approach applied to bank branches. *Journal of the Operational Research Society*. **55**: 1111-1121.

Premachandra, I.M, Chen, Y. and Watson, J., (2011), DEA as a tool for predicting corporate failure and success: A case of bankruptcy assessment, *Omega*, **39**: 620-626

Premachandra, I.M., Bhabra, G.S. and Sueyoshi, T. (2009). DEA as a tool for bankruptcy assessment: A comparative study with logistic regression technique. *European Journal of Operational Research*, **193**: 412-424

Psillaki, M., Tsolas, I.E. and Margaritis, D. (2010). Evaluation of credit risk based on firm performance. *European Journal of Operational Research.* **201**: 873-881.

Shanghai Stock Exchange, (2008). *Rules Governing the Listing of Stocks on Shanghai Stock Exchange*, pp. 90

Sharp, J.A., Meng W. and Liu W. (2007). A modified slacks-based measure model for data envelopment analysis with 'natural' negative outputs and inputs. *Journal of the Operational Research Society*. **58**: 1672-1677.

Shetty, U., Pakkala T.P.M. and 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**: 1988-1997

Stiglitz, J.E. (1972). Some Aspects of the Pure Theory of Corporate Finance: Bankruptcies and Take-Overs. *The Bell Journal of Economics and Management Science*. **3**(2): 458-82.

Sueyoshi, T. (1999). DEA-discriminant analysis in the view of goal programming. *European Journal of Operational Research*, **115**: 564-582.

Sueyoshi, T. (2001). Extended DEA-Discriminant Analysis. *European Journal of Operational Research*, **131**, 324-351.

Sueyoshi, T. (2006). DEA-Discriminant Analysis: Methodological Comparison among eight discriminant analysis approaches. *European Journal of Operational Research*, **169**,247-272

Sueyoshi, T. and Goto, M. (2009). Methodological comparison between DEA (data envelopment analysis) and DEA-DA (discriminant analysis) from the perspective of bankruptcy assessment. *European Journal of Operational Research*, **199**, 561-575.

Sun, J. Jia, MY. and Li, H. (2011). AdaBoost ensemble for financial distress prediction: An empirical comparison with data from Chinese listed companies. *Expert Systems with Applications*, **38**(8): 9305-9312.

Thomas, L.C., Edelman, D. E. and Crook, J.N. (2002). *Credit Scoring and its Applications*, Philadelphia: SIAM.

Tsai, MC, Lin, SP, Cheng, CC. and Lin, YP. (2009). The consumer loan default predicting model - An application of DEA-DA, *Expert Systems with Applications*. **36**: 11682–11690

Verikas, A., Kalsyte, Z., Bacauskiene, M. and Gelzinis, A. (2010). Hybrid and ensemble-based soft computing techniques in bankruptcy prediction: a survey. *Soft Computing*, **14**(9): 995-1010.

Wang, G. and Ma, J. (2011). Study of corporate credit risk prediction based on integrating boosting and random subspace. *Expert Systems with Applications*. **38**(11): 13871–13878

Xiao, Z., Yang, X., Pang, Y. and Dang, X. (2012). The prediction for listed companies' financail distress by using multiple prediction models with rough set and Dempster-Shafer evidence theory. *Knowledge Based Systems*, 26: 196-206.

Xu, X. and 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. and Hsu, M.F. (2010). A hybrid approach of DEA, rough set and support vector machines for business failure prediction. *Expert Systems with Applications*, **37**: 1535-1541

Yeh, Q.J. (1996). The application of Data Envelopment Analysis in conjunction with financial ratios for bank performance evaluation. *Journal of the Operational Research Society*, **47**(8): 980-988