View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Insitutional Repository at the National Graduate Institute for Policy Studies

COMBINING SUPPORT VECTOR MACHINE AND DATA ENVELOPMENT ANALYSIS TO PREDICT CORPORATE FAILURE FOR NONMANUFACTURING FIRMS

Xiaopeng Yang
Centre for Management of Technology and Entrepreneurship,University of Toronto,
200 College Street, Toronto, ON M5S 3E5, Canada
Xiaopeng.yang@utoronto.ca

Stanko Dimitrov
Department of Management Sciences, University of Waterloo,
200 University Avenue West, Waterloo, ON, N2L 3G1, Canada

Abstract: Research on corporate failure prediction has drawn numerous scholars' attention because of its usefulness in corporate risk management, as well as in regulating corporate operational status. Most previous research related to this topic focused on manufacturing companies and relied heavily on corporate assets. The asset size of a manufacturing company plays a vital role in traditional research methods; Altman's Z score model is one such traditional method. However, very limited number of research studied corporate failure prediction for nonmanufacturing companies as the operational status of such companies is not solely correlated to their assets. In this manuscript we use support vector machines (SVMs) and data envelopment analysis (DEA) to provide a new method for predicting corporate failure of nonmanufacturing firms. We first generate efficiency scores using a slack-based measure (SBM) DEA model, using the recent three years historical data of nonmanufacturing firms; then we used SVMs to classify bankrupt firms and healthy ones. We show that using DEA scores as the only inputs into SVMs predict corporate failure more accurately than using the entire raw data available.

Keyword: support vector machine (SVM); data envelopment analysis (DEA); corporate failure; nonmanufacturing firms; predictions.

1. INTRODUCTION

Corporate failure prediction is an attractive research topic in the sense that it can provide useful information about the operational status of a company, and it may affect a management team's decision making process. Information of corporate stress or failure may also in turn affect the stock market, customers' choice, business partners, and even competitors' policy. All of these factors lead to intense research efforts within both industry and academia.

A number of methods have been used in corporate failure prediction, most of which use several financial ratios from the financial statements of a company to evaluate the corporate stress or possibility of failure. Among all these methods, Altman's method is predominant and referred in all other studies. Altman used multiple discriminant analysis to create a model that utilizes several ratios in a linear formula to generate a score. This score can classify a company into three categories, namely at the risk of corporate stress or failure, healthy, and the middle status, a "grey area." However, most methods, either Altman's method or other ratio analysis methods, use financial ratios including asset size, and assume it as a crucial factor relative to other factors. For manufacturing companies, this is a valid assumption as many factors need to match the scale of the company asset, such as debt, sales, working capital, earnings, etc., and these factors are important in judging whether a firm may run into stress. In particular, for manufacturing firms where the initial investment occupies a large part of total asset and is a precondition to ensure other factors operating properly, discussing the problem of corporate failure prediction without considering assets is meaningless. However, the total assets of a nonmanufacturing firm usually is not decisive since such firms, in order to enhance their competiveness, pay more money in working capital such as salary, short-term consumables, etc. to provide better service and make more profit. Therefore, using Altman's traditional method to predict corporate failure for nonmanufacturing firms may result in inaccurate conclusions.

Based on his original model, used mainly for predicting bankruptcy for manufacturing firms, which was named the Altman's Z score, Altman then proposed another method that he named the Altman Z'' model (Altman, 2002) to cover the nonmanufacturing industry. Then he assigned appropriate coefficients to variables after determining Z'' score on nonmanufacturing firms in order to allow his previous method to be applicable for both manufacturing and nonmanufacturing companies. Nevertheless, he did not change the status quo, and his method still strongly relies on assets. Unfortunately, most nonmanufacturing companies are mainly focused on services and their most important asset is their people and they do not have a large real asset base (Growth of the Service Sector, 2011). It follows that a new outlet needs to be explored to predict corporate failure for the nonmanufacturing sector, and this is the main contribution of this study.

Since first proposed in 1978 by Charnes et al. (Charnes, Cooper, & Rhodes, 1978), DEA has developed into a prevalent non-parametric approach that is used to evaluate the relative efficiencies of a group of peer units which have the same productive process and inputs/outputs, i.e., decision making units (DMUs). As the first DEA model, CCR model extended Farrell's (Farrell, 1957) prototype model about technical and allocative efficiency. Following this, DEA became a powerful tool which is active in various research fields such as management, finance, agriculture, military, non-profit organizations and many others (Emrouznejad, Parker, & Tavares, 2008; Paradi & Zhu, 2013; Liu, Lu, Lu, & Lin, 2013; Yang & Morita, 2013; Sutton & Dimitrov, 2013).

Comparing to other methods in corporate failure prediction for nonmanufacturing firms, the main benefits to using DEA in our research can be found in the following aspects: (1) It allows us to select inputs/outputs flexibly depending on actual needs, which can eliminate or at least mitigate the influence of the asset factor. (2) DEA is easier to use since it is a nonparametric method and users do not need to handle complicated parameters. Meanwhile, DEA offers more objective analysis results. (3) DEA divides attributes into inputs and outputs and relates them to each other. The efficiency score generated based on such an assumption is more informative compared to barely using raw data. It follows that we propose a method combining DEA and SVM together, which uses the efficiency scores calculated by DEA model to classify healthy and bankrupt firms.

The remainder of this article is structured as follows: Section 2 is the literature review of previous studies in predicting corporate failure. Section 3 introduces the DEA model we are using in this research, and how to combine DEA and SVM. Section 4 provides an application about nonmanufacturing firms covering a number of industries. Section 5 summarizes the research and provides additional discussion.

2. LITERATURE REVIEW

A number of methods and related applications have been broadly studied in the field of bankruptcy prediction. In order to compare our method with others and make a distinct contribution in this field, we summarize and review the main methodologies in the previous published papers in this section.

2.1. Ratio Analysis Methods

William Beaver proposed a method in 1967 (Beaver, 1967) to predict bankruptcy which defined failure as “the inability of a firm to pay its financial obligations as they mature” and a financial ratio as “a quotient of two numbers, where both numbers consist of financial statement items.”The application in Beaver’s study used the data from Moody’s industrial manual between 1954 and 1964. For each bankrupt firm from Moody’s, a healthy firm with the same asset size in the same industry was matched. Beaver argued that firms of different asset-sizes could not be accurately compared (Alexander, 1949). Based on this assumption, he compiled 30 ratios and picked 14 of them to be the most effective in determining the likelihood of bankruptcy, which were *cash flow/total debt, current assets/current liabilities, net income/total assets, quick assets/current liabilities*, etc. Then he claimed that *“cash flow/total debt”* and *“total debt/total assets”* were the best two indicators for bankruptcy prediction. As such univariate method neglect many other ratios which might affect the results in estimating the corporate failure, Edward Altman applied the first multivariate approach, multiple discriminant analysis (MDA) (Altman, 1968), to bankruptcy prediction in 1968. At that time, MDA was usually used in classifying an observation into several previously defined groups. Its main merit was allowing for the entire profile of variables to be analyzed simultaneously rather than individually (Altman, 2002).

Using a similar method to Beaver’s, Altman paired the healthy firms with bankrupt ones, and there were 66 corporations with 33 firms in the bankrupt group and 33 in the non-bankrupt group in Altman’s study. Eventually, the five most influential ratios, as determined by Atlman, in determining the likelihood of bankruptcy, were selected as the main indicators used to predict corporate failure including *working capital / total assets, retained earnings / total assets, earnings before income & taxes / total assets, market value of equity / total liabilities, sales / total assets*. These ratios were selected based on: (1) statistical significance of various potential functions while determining the relative contribution of each individual variable, (2) the inter-correlation between the variables, (3) the predictive accuracy of various profiles and (4) judgement of the analysis (Altman, 1968).

In the same study, Altman next assigned appropriate coefficients to these five ratios and defined the sum of the weighted ratios as the Z score, which relied heavily on the asset size and was considered to be only suitable for the manufacturing industry. Based on Altman’s Z score method, a large number of related studies were developed by employing different ratios (Deakin, 1972; Ohlson, 1980; Zmijewski, 1984; Hsieh, 1993; Grice & Dugan, 2001; Shumway, 2001; Grice & Ingram, 2001), of which the majority still focused on manufacturing companies. Then Altman proposed his lesser known Z'' score method, in which he revised the coefficient and ratio items to make them fit nonmanufacturing industry. Unfortunately, the Z'' score method is still affected by asset size, which motivates us to investigate the corporate failure prediction problem using DEA and SVM in this research.

2.2. Data Envelopment Analysis Methods

Since first introduced via the CCR model, DEA is now a prevalent method in predicting corporate stress and has been used in many studies (Premachandra, Chen, & Watson, 2011; Li, Crook, & Andreeva, 2014; Shetty, Pakkala, & Mallikarjunappa, 2012; Xu & Wang, 2009). Cielen et al. concluded that DEA and linear programming models can outperform decision tree methods based on the result of comparing the three methods, though the authors did not indicate if DEA is more accurate than linear programming models (Cielen, Peeters, & Vanhoof, 2004).On the other hand, Sueyoshi et al. proposed DEA-DA (discriminant analysis) based on DEA models and applied it to bankruptcy prediction. Their research showed that DEA-DA is more appropriate for longitudinal data (Sueyoshi & Goto, 2009). Another study integrated rough set theory (RST) into SVM which is used to increase the accuracy of predicting corporate failure (Yeh, Chi, & Hsu, 2010). Most of the research compares DEA and other methods, and shows that DEA is a better method to use for corporate failure prediction; whereas no study covered nonmanufacturing firms with very small asset sizes. Work on predicting corporate failure, regardless of the method, is of paramount interest to not only banks, but also venture capitalist prior to making any investments. Unlike banking, a firm may be more averse to providing its financial data to an unknown venture capitalist. As such, works on using DEA may allow a firm to only release its DEA score to help a venture capitalist make an investment decision, and not have the firm release all of its closely held information.

3. COMBINE DEA & SVM

SVM is a powerful tool for extracting information from data sets; however, sometimes it may not be an effective method when there are noisy observations or the data is distributed uniformly on the feature space, independent of class. On the other hand, the data points may have multi-attributes, and it is very common that these attributes are correlated or influence each other; therefore, information mining via SVM alone may neglect the inner connection between such attributes. This inspires us to use DEA at first to analyze each data point as a decision making unit (DMU), which consists of input and output attributes and considers the internal transformation from inputs to outputs. Then we use the efficiency scores obtained from DEA to continue extracting further information about the changing trend of these scores. In other words, DEA is a projection-like method that reduces dimensionality for SVMs. Eventually, we use SVM methods to predict corporate failure based only on DEA scores. In a method combing DEA and SVMs, we can utilize the merits of both methods. Also, such an idea provides us more accurate results for corporate failure prediction.

3.1. Data Preparation

The main purpose of this research is to see how accurately bankruptcy can be predicted regardless of the asset size. All of the indicators utilized are inspired by Altman’s research; but due to data availability, some of indicators are not available, such as Earning before Interest and Tax (EBIT). Therefore, we need to reorganize the indicators. In this research, EBIT is substituted for Operating Income which is also considered to be a very valuable indicator of corporate health. Moreover, the attribute “Total Liabilities” was removed, though present in Altman’s method. As we do not have the data for “Working Capital,” this indicator was split into “Current Assets” and “Current Liabilities.”

Unlike manufacturing companies, for which to test the relevance of human capital, which is important to smaller nonmanufacturing firms in our model, the number of employees and the number of shareholders were added to the model. The number of employees was added to introduce the measure of human capital (the most important “asset” in a nonmanufacturing firm) as a contributor to the efficiency of a company. The number of shareholders was added because for many smaller nonmanufacturing firms the shareholders have decision-making power and invest both time and money that contribute to the success of a firm. In this sense, the number of shareholders can also be seen as a reflection of the financial well-being of a company as viewed by the public.

Negative value is a common problem in DEA literature. In our research we have negative values in Retained Earnings (RE), Operating Income (OI) and Book Value of Equity (BVE), to which the SBM model was not applicable. Thus each output was split into positive and negative parts. For example, RE was split into RE^+ and RE^-, where was RE^+ defined as output in its usual meaning, but RE^- was defined as input. This method is essentially saying that RE^+ is an output and therefore should be made as large as possible to improve the company's operating efficiency. However RE^- is viewed as an input which should be minimized. Therefore the inputs/outputs of the model after revision are shown in the following table.

Table 1: Inputs/Outputs classification

Outputs	**Inputs**
Current Assets (CA)	Current Liabilities (CL)
Positive Retained Earnings (RE^+)	Negative Retained Earnings (RE^-)
Positive Operating Income (OI^+)	Negative Operating Income (OI^-)
Positive Book Value of Equity (BVE^+)	Negative Book Value of Equity (BVE^-)
The Number of Shareholders (SH)	The Number of Employees (EM)

3.2. Combining DEA & SVM

As discussed in Section 2.3 we test a variety of kernel functions, $\kappa(x_i, x_j)$, in this study. We list the kernel functions use in Table , their implementations come from the R (Team, 2015) kernel library (Karatzoglou, Smola, Hornik, & Zeileis, 2004).

Table 2: List of Kernel functions used

Kernel Name	**Kernel Generating Functions**	**Parameters**
Gaussian RBF	$\kappa(x_i, x_j) = e^{\left(-\sigma\lvert x_i - x_j\rvert^2\right)}$	σ
Polynomial	$\kappa(x_i, x_j) = \left(s \cdot x_i^T \cdot x_j + c\right)^d$	s, c, d
Hyperbolic tangent	$\kappa(x_i, x_j) = \tanh(s \cdot x_i^T \cdot x_j + c)$	s, c
Laplacian	$\kappa(x_i, x_j) = e^{(-\sigma\lvert x\_i - x\_j\ \rvert)}$	σ
Bessel	$\kappa(x_i, x_j) = -Bessel_{v+1}^{n}\sigma\lVert x_i - x_j\rVert$	σ, n, v
Spline	$\kappa(x_i, x_j) = 1 + x_i \cdot x_j + x_i \cdot x_j \min(x_i, x_j) - \frac{x_i + x_j}{2}\min(x_i, x_j)^2 + \frac{\min(x_i, x_j)^3}{3}$	N/A

We discuss the parameters we consider in our study in the RESULTS SECTION.

As the limited data we have (Table 5 is not listed here.), we use 10-fold cross validation to separate or data into training and testing data, and to test the data. Further, in order to statistically compare the accuracy of using the *raw data*, the firm attributes the first three years of operations, and the *DEA data*, the SBM values computing from the first three years of operations, we bootstrap the 10-fold cross validation by creating 500 instances of the 10-fold cross validation. This means that for every instance we use 5,000 test data instance with approximately 6.9 instances in each instance. When comparing the accuracy resulting from each dataset, we use the Wilcoxon rank-sum test (Wilcoxon, 1945) to see if on average, the SVM using the DEA data is statistically more accurate than the SVM using the raw data. We consider the number of test data instances that are accurately predicted along with the p-value from the rank-sum test.

4. APPLICATION TO BANKRUPTCY PREDICTION FOR NONMANUFACTURING FIRMS

We start this section from data collection in nonmanufacturing industry of North America. From a large number of candidate data points, we select the data which has full records of the recent 3 years. Then we use these records to calculate the DEA efficiency scores for the recent 3 years, and based on this, we classify bankrupt and non-bankrupt firms by different SVMs. By comparing the results of different SVMs, we conclude that combing DEA and SVM is an excellent outlet in predicting corporate failure relative to using raw data for classification alone.

4.1. Data Acquisition

In this research, we collected the data through Mergent Online database (Mergent, 2011) and a professional company which mainly focused on filing bankrupt companies in North America dating back to the 1980sselected by SIC (Standard Industrial Classification) codes. The list of companies was narrowed down to those classified as nonmanufacturing or service-based firms. These companies must also have filed for bankruptcy between the years of 2000 and 2006, as more recent filings would be more easily obtained, and more easily compared to current companies. Due to the economic recession taking place, bankruptcy filings from 2007 to present were not selected, as the data could not reflect the real situation in that period. The companies considered to be bankrupt during that period could be more so for external reasons, which was not the main purpose of the current research.

We used the most recent 3 years data before bankruptcy as we consider such data can reflect the recent trend of the operational status changing of a company, and older data may not be significant in prediction of bankruptcy. Whenever it was possible to identify them, the companies that had filed for bankruptcy but did not fail were excluded from the study. Many of these companies filed for bankruptcy for reasons other than complete insolvency, some liquidations were due to legal issues, and others because they were suffering financial distress, filed in an attempt to reorganize and restructure their corporate strategy and alleviate the debt. Data from the full Balance Sheets, Income Statements, Cash Flow Statements and Retained Earnings were collected. From the Balance Sheet, current assets, total assets, current liabilities, total liabilities, retained earnings and shareholders' equity values were extracted. From the Income Statement, the operating profit was calculated using the formula Net Sales – Cost of goods – Expenses. The number of employees and number of shareholders were also collected.

Once the data was collected for the bankrupt companies, healthy companies were then found. A healthy company was chosen for every bankrupt company based on SIC number and on the years of health. Healthy companies had to be in existence at least 5 years after the bankruptcy of their bankrupt counterpart. Healthy companies also must not have filed for bankruptcy during the time that they are being compared to the bankrupt counterpart. The same financial data was collected for the healthy company as the bankrupt counterpart within the same years. For example, if a bankrupt company filed bankruptcy in 2002, financial data was collected for 1997-2001. The healthy company would have to have been in existence and not to have filed for bankruptcy between the years of 1996 to 2006. In some cases a suitable healthy match could not be found and thus the number of bankrupt companies exceeds the number of non-bankrupt ones.

4.2. Results Analysis

The kernels described in Table 2 each have a set of parameters associated with them, as listed in the third column of Table . In our study we conducted a grid search over the set of parameters to find the best parameters of those considered. For each parameter we considered values between 0 and 10 with varying step size, ranging from 1 to 0.01, we were not able to reach values of 10 for all kernels, for example the polynomial kernel, we only considered degree of 8 or less as the computation time for degree 8 was approaching 5 hours per parameter configuration using a parallelized implementation on a 24 core machine with 128 GB of RAM. For each set of parameters we considered, we used 10-fold cross validations, and kept track of the number of times each the trained kernel correctly predicted the class, bankrupt or not, of the firms that were held out. As there is an exponential number of ways 10-folds may be created, we generated 500 10-folds for each parameters setting, and then conducted a Wilcoxon rank-sum test on the number of correct prediction, comparing the number of correct predictions using the raw data and using only the DEA data. In Table below we show the number parameter configurations we attempted for each kernel, the fraction of tests in which the SVM using DEA data performed statistically better at the 95% confidence level and the SVM using the raw data performed better, at the same confidence level. We also consider the SVM parameters in which each dataset performed best and show that for each kernel the best performing SVM trained on the DEA statistically performs better than the best performing SVM trained on the raw data at the 99% level,

Table . Our results suggest that DEA values, derived from raw data may be more informative, at least in this application than the raw data available in the same applications.

Table 3: SVM performance depending on training data

Kernel	Number of Experiments	Fraction DEA better at 95% CI	Fraction raw better at 95% CI
Gaussian RBF	134	0.92	0.06
Polynomial	851	0.73	0.17
Hyperbolic tangent	538	0.81	0.16
Laplacian	97	0.98	0.02
Bessel	738	0.84	0.07
Spline	1	1	0

Table 4: Comparing best performing SVMs, p-values test if DEA SVM are more accurate than raw data SVM. Meaning we are checking if the number of correctly classified companies using DEA only is greater than the number of correctly classified companies using the raw data.

Kernel	DEA Parameters	Raw Parameters	p-value
Gaussian RBF	$\sigma = 4$	$\sigma = 10$	0.00
Polynomial	$s = 8, c = 8, d = 8$	$s = 3, c = 0, d = 4$	0.00
Hyperbolic Tangent	$s = 2, c = 9$	$s = 6, c = 10$	0.00
Laplacian	$\sigma = 1.85$	$\sigma = 3.8$	0.00
Bessel	$\sigma = 4, \nu = 0, n = 1$	$\sigma = 4, \nu = 0, n = 1$	0.00
Spline	N/A	N/A	0.00

5. CONCLUSIONS

Our research at first surveyed the related studies in bankruptcy prediction, stretching from ratio models to Altman's Z'' model, then proposed the approach of combining DEA and SVM to predict corporate failure. We split the negative factors into positive and negative component, which could be a viable option when needed in DEA analyses. Then the DEA scores were generated via SBM model, and then as the inputs for classification via SVM. From the result comparison, we can conclude that combining DEA and SVM is apparently a more feasible and effective method in predicting corporate failure comparing to using raw data as material of SVM method.

Although our research provide some meaningful findings, there is still a number of suggestions for subsequent future work which includes: (1) employing alternate DEA models or constraint conditions, particularly using the Assurance Region model which will put more restrictions on the variable weights and may obtain more meaningful results; (2) prediction accuracy may be affected by different approaches to selecting inputs/outputs, therefore different or other, related financial factors may bring higher prediction accuracy; (3) due to the lack of available data, the number of DMUs used in this study was insufficient for a more comprehensive assessment of the model. With a larger scale of database, the result will become more realistic and accurate for bankruptcy prediction; (4)the selection of kernels in SVM affects analysis result, therefore we may need to do modification to the existing kernels and improve the accuracy of SVM.

REFERENCES

[1] Alexander, S. S. (1949). The Effect of Size of Manufacturing Corporation on the Distribution of the Rate of Return. Review of Economics and Statistics , 229-235.

[2] Altman, E. (2002). Bankruptcy, Credit Risk and High Yield Junk Bonds. Malden, Massachusetts: Blackwell Publishers Inc.

[3] Altman, E. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. Journal of Finance , 23 (4), 589-609.

[4] Banker, R. D., Charnes, A., & Cooper, W. (1984). Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. Management Science , 30, 1078–1092.

[5] Beaver, W. H. (1967). Financial Ratios as Predictor of Failure. Journal of Accounting Research , 5, 71-111.

[6] Charles, L. M. (1942). Financing Small Corporations in Five Manufacturing Industries. National Bureau of Economic Research , 1926-36.

[7] Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. European Journal of Operational Research , 2, 429-444.

[8] Chava, S., & Jarrow, R. (2004). Bankruptcy Prediction with Industry Effects. Review of Finance , 8, 537–569.

[9] Cielen, A., Peeters, L., & Vanhoof, K. (2004). Bankruptcy prediction using a data envelopment analysis. European Journal of Operational Research , 154, 526–532.

[10] Cooper, W. W., Seiford, L. M., & Tone, K. (2007). Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software (Second Edition). Springer.

[11] Deakin, E. (1972). A Discriminant Analysis of predictors of Business Failure. Journal of Accounting Research , 167-179.

[12] Deng, N., Tian, Y., & Zhang, C. (2012). Support Vector Machines: Optimization Based Theory, Algorithms, and Extensions. Chapman and

Hall/CRC.

[13] Emrouznejad, A., & Amin, G. R. (2009). DEA models for ratio data: Convexity consideration. Applied Mathematical Modelling , 33 (1), 486-498.

[14] Emrouznejad, A., Parker, B. R., & Tavares, G. (2008). Evaluation of research in efficiency and productivity: A survey and analysis of the first 30 years of scholarly literature in DEA. Socio-Economic Planning Sciences , 42 (3), 151–157.

[15] Farrell, M. J. (1957). The measurement of productive efficiency. Journal of the Royal Statistical Society Series A , 120, 253-90.

[16] Grice, J. S., & Dugan, M. (2001). The Limitations of Bankruptcy Prediction Models: Some Cautions for the Researcher. Review of Quantitative Finance and Accounting , 17, 151–166.

[17] Grice, J. S., & Ingram, R. W. (2001). Tests of the generalizability of Altman’s bankruptcy prediction model. Journal of Business Research , 54, 53– 61.

[18] Growth of the Service Sector. (2011, 12). Retrieved from http://www.worldbank.org/depweb/beyond/beyondco/beg_09.pdf

[19] Hsieh, S.-J. (1993). A Note on the Optimal Cutoff Point in Bankruptcy Prediction Models. Journal of Business Finance and Accounting , 30 (3), 457–464.

[20] Karatzoglou, A., Smola, A., Hornik, K., & Zeileis, A. (2004). kernlab -- An S4 Package for Kernel Methods in R. Journal of Statistical Software , 11 (9), 1-20.

[21] Li, Z., Crook, J., & Andreeva, G. (2014). Chinese companies distress prediction: an application of data envelopment analysis. Journal of the Operational Research Society , 65, 466–479.

[22] Liu, J. S., Lu, L. Y., Lu, W.-M., & Lin, B. J. (2013). Data envelopment analysis 1978–2010: A citation-based literature survey. Omega , 41 (1), 3-15.

[23] Mergent, I. (2011). Mergent Online. Retrieved from http://www.mergentonline.com/

[24] Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. Journal of Accounting Research , 18, 109–131.

[25] Paradi, J. C., & Zhu, H. (2013). A survey on bank branch efficiency and performance research with data envelopment analysis. Omega , 41 (1), 61-79.

[26] Premachandra, I., Chen, Y., & Watson, J. (2011). DEA as a tool for predicting corporate failure and success: A case of bankruptcy assessment. Omega , 39, 620–626.

[27] Shetty, U., Pakkala, T., & 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.

[28] Shumway, T. (2001). Forecasting bankruptcy more accurately: a simple hazard model. Journal of Business , 74, 101–124.

[29] Sueyoshi, T., & 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.

[30] Sutton, W., & Dimitrov, S. (2013). The U.S. Navy explores detailing cost reduction via Data Envelopment Analysis. European Journal of Operational Research , 227 (1), 166–173.

[31] Team, R. C. (2015). R: A Language and Environment for Statistical Computing. Vienna.

[32] The Effect of Sample Size on the Mean Efficiency in DEA: Comment2001Journal of Productivity

Analysis 15129–137

[33] Tone, K. (2001). A slacks-based measure of efficiency in data envelopment analysis. European Journal of Operational Research , 130, 498–509.

[34] Wilcoxon, F. (1945). Individual Comparisons by Ranking Methods. Biometrics Bull. , 1 (6), 80–83.

[35] Xu, X., & Wang, Y. (2009). Financial failure prediction using efficiency as a predictor. Expert Systems with Applications , 36, 366–373.

[36] Yang, X., & Morita, H. (2013). Efficiency improvement from multiple perspectives: An application to Japanese banking industry. Omega , 41 (3), 501-509.

[37] 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.

[38] Zmijewski, M. E. (1984). Methodological issues related to the estimation of financial distress prediction models. Journal of Accounting Research , 22, 59–82.