Design and Performance Evaluation Of Failure Prediction Models

Mohammad Mahdi Mousavi

A Thesis Submitted in Fulfilment of the Requirements for the Degree of Doctor of Philosophy in Management

THE UNIVERSITY
of EDINBURGH

2017

Statement of Originality

This thesis has been composed by myself and contains no material that has been accepted for the award of any other degree at any university.

A part of this thesis has been published in the *International Review of Financial Analysis*:

Mousavi, Mohammad M., Jamal Ouenniche, and Bing Xu. "Performance evaluation of bankruptcy prediction models: An orientation-free super-efficiency DEA-based framework." *International Review of Financial Analysis*, 42 (2015): 64-75.

Permission to include text from that paper has been gained from the publisher and the authors.

To the best of my knowledge and belief, this thesis contains no other material previously published by any other person except where due acknowledgment has been made.

Mohammad Mahdi Mousavi

April 2017

Acknowledgment

Pursuing a Ph.D. program was both painful and pleasant experience for me. It was accompanied with lots of frustration, failure, bitterness, stress, insomnia as well as encouragement, trust and success. No need to say that, passing this big project was not possible without the contribution of others.

First of all, I would like to express my appreciation to my supervisor, Prof. Jamal Ouenniche, who accepted me as his Ph.D. student and then generously offered me his idea and advice, and patiently guided me in the right direction. Special thanks are also given to the members of the examination committee, Prof. Ali Emrouznejad, Dr Fernando Moreira and Prof. Chris Carr for their valuable comments and feedbacks on my research.

Special thanks to the college of Social Science and Humanities of the University of Edinburgh for offering the PhD scholarship, which saved and supported my PhD study abroad. I also appreciate University of Edinburgh Business School (UEBS) for offering a proper research equipment and environment where I felt convenient and concentrate on my research.

Last but not the least, I'd like to give my sincere thanks to my beloved parents and family. Their understanding and their love encouraged me to work hard and to finish my PhD abroad. My parents' kind-hearted personality and their prayers have supported me to be persistent and never bend to difficulty. Words cannot express my gratitude.

Table of Contents

List of Tables

List of Figures

List of Appendixes

List of Abbreviations

AIES	Artificially Intelligent Expert Systems
AUROC & AUC	Area Under Receiver Operating Characteristics
BCBS	Basel Committee on Banking Supervision
BCC	Banker, Charnes, and Cooper
BPM	Bankruptcy Prediction Model
BS	Brier Score
BSM	Black-Scholes-Merton
CCA	Contingent Claims Analysis
CCR	Charnes, Cooper, and Rhodes
CDEA	Context-Dependent Data Envelopment Analysis
CIER	Conditional Information Entropy Ratio
CRS	Constant Return-to-Scale
DA	Discriminant Analysis
DD	Duration Dependent
DDWFSB	Duration Dependent with Firm-Specific Baseline
DDWTDB	Duration Dependent with Time-Dependent Baseline
DEA	Data Envelopment Analysis
DIWOB	Duration Independent Without Baseline
DIWTIB	Duration Independent with Time-Independent Baseline
DMU	Decision-Making Unit
DOC	Down-and-Out Call option
DPM	Distress Prediction Model
DRS	Decreasing Return-to-Scale
DTH	Discrete-Time Hazard
EBITDA	Earnings Before Interest, Taxes, Depreciation and Amortization
EC	Efficiency Change
EDF	Expected Default Frequency
EFS	Efficient Frontier-Shift
FA	Financial Accounting
GI	Gini Index
H	Hand Measure
HDR	High Distress Rate
HL	Hosmer-Lemeshow
IO	Input Oriented
IRS	Increasing Return-to-Scale
IV	Information Value
KMV	Kealhofer, McQuown andVasicek
K-NN	k-Nearest Neighbour

KS	Kolmogorov-Smirnov
LA	Logit Analysis
LL	Log-likelihood
LP	Linear Programming
LPA	Linear Probability Analysis
LPD	Last period Probability of Default
LPM	Linear Probability Model
LR	Logistic Regression
LSE	London Stock Exchange
MCDA	Multi-Criteria Decision-Making Analysis
MDA	Multivariate Discriminant Analysis
ME	Mix Efficiency
MI	Macroeconomic Indicator
MIP	Mixed Integer Programming
MLP	Multi-Layer Perceptron
MPI	Malmquist Productivity Index
MR	Misclassification Rate
MV	Market Variables
NN	Neural Network
OCC	Overall Correct Classification
OO	Output Oriented
PA	Probit Analysis
PD	Probability of Default
PTE	Pure Technical Efficiency
RAM	Range Adjusted Measure
ROA	Return on Assets
ROC	Receiver Operating Characteristic
RORWA	Return on Risk-Weighted Assets
SA	Survival Analysis
SBM	Slack-Based-Measure
SE	Scale Efficiency
Sen	Sensitivity
Spe	Specificity
SVM	Support Vector Machines
T1	Type I error
T2	Type II error
UDA	Univariate Discriminant Analysis
VEX	Volatility of Exchange Rate
VRS	Variable Return-to-Scale

Abstract

Prediction of corporate bankruptcy (or distress) is one of the major activities in auditing firms' risks and uncertainties. The design of reliable models to predict distress is crucial for many decision-making processes. Although a variety of models have been designed to predict distress, the relative performance evaluation of competing prediction models remains an exercise that is unidimensional in nature. To be more specific, although some studies use several performance criteria and their measures to assess the relative performance of distress prediction models, the assessment exercise of competing prediction models is restricted to their ranking by a single measure of a single criterion at a time, which leads to reporting conflicting results. The first essay of this research overcomes this methodological issue by proposing an orientation-free super-efficiency Data Envelopment Analysis (DEA) model as a multi-criteria assessment framework. Furthermore, the study performs an exhaustive comparative analysis of the most popular bankruptcy modelling frameworks for UK data. Also, it addresses 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? Further, using different static and dynamic statistical frameworks, this chapter proposes new Failure Prediction Models (FPMs).

However, within a super-efficiency DEA framework, the reference benchmark changes from one prediction model evaluation to another one, which in some contexts might be viewed as "unfair" benchmarking. The second essay overcomes this issue by proposing a Slacks-Based Measure Context-Dependent DEA (SBM-CDEA) framework to evaluate the competing Distress Prediction Models (DPMs). Moreover, it performs an exhaustive comparative analysis of the most popular corporate distress prediction frameworks under both a single criterion and multiple criteria using data of UK firms listed on London Stock Exchange (LSE). Further, this chapter proposes new DPMs using different static and dynamic statistical frameworks.

Another shortcoming of the existing studies on performance evaluation lies in the use of static frameworks to compare the performance of DPMs. The third essay overcomes this methodological issue by suggesting a dynamic multi-criteria performance assessment framework, namely, Malmquist SBM-DEA, which by design, can monitor the performance of competing prediction models over time. Further, this study proposes new static and dynamic distress prediction models. Also, the study addresses several research questions as follows; what is the effect of information on the performance of DPMs? How the out-of-sample performance of dynamic DPMs compares to the out-of-sample performance of static ones? What is the effect of the length of training sample on the performance of static and dynamic models? Which models perform better in forecasting distress during the years with Higher Distress Rate (HDR)?

On feature selection, studies have used different types of information including accounting, market, macroeconomic variables and the management efficiency scores as predictors. The recently applied techniques to take into account the management efficiency of firms are two-stage models. The two-stage DPMs incorporate multiple inputs and outputs to estimate the efficiency measure of a corporation relative to the most efficient ones, in the first stage, and use the efficiency score as a predictor in the second stage. The survey of the literature reveals that most of the existing studies failed to have a comprehensive comparison between two-stage DPMs. Moreover, the choice of inputs and outputs for DEA models that estimate the efficiency measures of a company has been restricted to accounting variables and features of the company. The fourth essay adds to the current literature of two-stage DPMs in several respects. First, the study proposes to consider the decomposition of Slack-Based Measure (SBM) of efficiency into Pure Technical Efficiency (PTE), Scale Efficiency (SE), and Mix Efficiency (ME), to analyse how each of these measures individually contributes to developing distress prediction models. Second, in addition to the conventional approach of using accounting variables as inputs and outputs of DEA models to estimate the measure of management efficiency, this study uses market information variables to calculate the measure of the market efficiency of companies. Third, this research provides a comprehensive analysis of two-stage DPMs through applying different DEA models at the first stage – e.g., input-oriented vs. output oriented, radial vs. non-radial, static vs. dynamic, to compute the measures of management efficiency and market efficiency of companies; and also using dynamic and static classifier frameworks at the second stage to design new distress prediction models.

Keywords: Bankruptcy Prediction; Corporate Distress Prediction; Performance Criteria; Performance Measures; Data Envelopment Analysis; Slacks-Based Measure; Context-Dependent Data Envelopment Analysis; Malmquist Index; Corporate Two-stage Distress Prediction; Efficiency; Malmquist Index

Chapter One

Introduction

1.1 Preamble

Corporate credit and default risk is an extensive terminology in banking and finance. According to the Basel Committee on Banking Supervision (BCBS), default in credit risk refers to a failure of a borrower or counterparty to meet its obligations in accordance with agreed terms (Basel Committee on Banking Supervision, 2000, p. 1). Corporate credit studies have considered different business failure events such as credit default (see, for example, Beaver, 1996), bankruptcy (see for example, Hillegeist et al., 2004; Ohlson, 1980; Shumway, 2001; Wilson and Sharda, 1994) and financial distress (see for example, Bandyopadhyay, 2006; Campbell et al., 2008; Tinoco and Wilson, 2013; Li et al., 2014, 2017).

Corporate distress prediction has received considerable attention and became a major subject of extensive studies after the financial crises in 2007 and the European recession in 2009. Financial distress is defined as a situation that a company cannot generate enough cash flows to fulfill its contractual obligations (Piesse et al., 2006, p. 478). Remaining in this situation for a long time not only could impact adversely on the value of the company and the wealth of stockholders but also causes more financial and operational inefficiencies, and finally, could lead to bankruptcy. Corporate bankruptcy causes significant losses to both business community and the society as a whole - for details about the costs of bankruptcy, we refer the reader to Davydenko et al. (2012), Elkamhi et al. (2012) and Branch (2002). Therefore, early detection of a company deteriorating condition or distress has such economic advantages that motivates both academics and practitioners in developing a range of corporate Distress Prediction Models (DPMs).

Because of considerable increases and an extensive variety of prediction models, a strand of the literature has focused on answering the question which of these models perform better? The performance of bankruptcy and distress prediction models, as data-fitting based empirical studies, is reliant on different factors such as sampling, features selection, modelling, and performance evaluation (Zhou, 2013).

With respect to the performance evaluation, the comparative studies have been criticized because of failure in providing a comprehensive comparison between all types of bankruptcy and distress prediction models (Bauer and Agarwal, 2014; Mousavi et al., 2015). Also, they have used a restricted number of criteria to evaluate the performance of competing models. Further, the nature of the performance evaluation of competing models remains mono-criterion, as they use a single measure of a single criterion at a time. Therefore, under mono-criterion evaluation, the rankings corresponding to different criteria are mostly different, which lead to a situation that practitioners cannot make a well-informed decision as to which model performs best when taken all criteria into account (see, for example, Theodossiou, 1991; Bandyopadhyay, 2006; Tinoco and Wilson, 2013). Another shortcoming of the comparative studies is that the multi-period performance evaluation of DPMs over time has never been considered.

With respect to feature selection as one of the main issues in developing bankruptcy and distress prediction models, financial ratios (Beaver, 1966, 1968; Altman, 1968; Ohlson, 1980; Zmijewski, 1984; Taffler, 1984), market-based information (Shumway, 2001; Hillegeist et al., 2004; Bharath and Shumway, 2008), macroeconomic indicators (Bellotti and Crook, 2007; Nam et al., 2008; Tinoco and Wilson, 2013) and corporate governance indicators (Liang et al., 2016; Darrat et al., 2016; Sueyoshi et al., 2010) are the most popular features that have been used in the literature. However, there are criticisms about the adequate statistical power of these features in developing prediction models.

Recent studies have incorporated corporate managerial efficiency as a feature in developing two-stage DPMs (Xu and Wang, 2009; Psillaki et al., 2010; Yeh et al., 2010; Li et al., 2014, 2017). Data Envelopment Analysis (DEA) is the commonly applied technique to estimate the managerial efficiency of companies in the first stage. In the second stage, the estimated managerial efficiency measure is used as a feature in other classifier techniques to estimate the probability of failure. The main criticism is that only the financial items (e.g., total sales, total assets) and the characteristics of companies (e.g., the number of employees) are used to measure the efficiency of companies. Further, there is no comparative study that considers alternative estimated measures of company efficiency (e.g., managerial and market) using different DEA models (e.g., input oriented versus output oriented, Constant Return to Scale (CRS) versus Variable Return to Scale (VRS), dynamic versus static DEA models).

In sum, the aim of this study is to provide a comprehensive comparative analysis of alternative statistical bankruptcy and distress prediction models. Further, this study proposes new assessment frameworks to compare the relative performance of prediction models. Also, beyond proposing a static multi-criteria framework, this research suggests a dynamic multi-criteria evaluation framework to analyse the multi-period performance of DPMs. Further, this study compares the discriminatory power of different efficiency measures of companies and propose new ones.

1.2 Motivations

The survey of the literature on bankruptcy and distress prediction revealed several gaps. First, although some comparative studies have used several performance criteria and, for each criterion, one or several measures to assess the performance of prediction models, the assessment is generally restricted to the ranking of models by a single measure of a single criterion at a time. Thus, the evaluation of models under multiple criteria remains unidimensional in nature, on the one hand, and the "big picture" is not considered in that a single or a very limited number of criteria only are used, on the other hand. Taken all criteria and their measures into consideration, empirical results of unidimensional evaluations indicate that the rankings corresponding to different criteria or measures are often different, which lead to uncertainty for a decision maker in choosing the best prediction model.

Second, the existing comparative studies failed to have a comprehensive comparison between all types of prediction models; i.e., traditional statistical models, contingent claims analysis (CCA) models, and survival analysis (SA) models.

Third, the existing performance evaluation of competing prediction models is static in nature. In other words, the multi-period performance of competing models over time has never been considered in the literature.

Fourth, taking into account two-stage distress prediction models, no study provides a comprehensive comparison between two-stage prediction models; neither considering different DEA models at the first stage that are used to estimate company efficiency nor using different classifier models at the second stage.

Fifth, the survey of using DEA in corporate credit risk and failure indicates that the choice of input and output for DEA models to estimate efficiency measures of companies is restricted to accounting variables; and other useful sources of information such as market variables are overlooked.

Considering the previous literature, the main motivation of this research is to provide frameworks for both static and dynamic multi-criteria assessment to compare the performance of competing bankruptcy and distress prediction models. Further, inspired by the recent trend in failure prediction studies in developing two-stage prediction models, it is necessary to have a comparative analysis of two-stage models that use different DEA models, under alternative assumptions, e.g. Constant Return to Scale (CRS) or Variable Return to Scale (VRS), with different types of inputs and outputs, to estimate efficiency measures of companies in the first stage, and to use different classifiers in the second stage.

1.3 Objectives

Considering the above-mentioned gaps in the literature, this research has the following major objectives and aims to achieve;

1) To provide a comprehensive comparison between the most pioneer and applied statistical bankruptcy (Chapter two) and distress prediction models (Chapters three and four).

2) To propose a cross-sectional multi-criteria framework to compare the performance of competing bankruptcy (Chapter two) and distress prediction models (Chapter three).
3) To propose a dynamic multi-criteria framework to compare the performance of competing distress prediction models (Chapter four).
4) To provide a comprehensive comparison of the discriminatory power of different DEA efficiency measures (original vs. decomposed DEA scores and market vs. management DEA scores) incorporated as features in static and dynamic classifiers in the second stage of two-stage DPMs (Chapter five).

1.4 Contributions

This research makes several contributions to the current literature. First, it proposes an orientation free super-efficiency DEA framework (chapter two) and an orientation free slack-based context-dependent (SBM-CDEA) DEA framework (chapter three) – as methodological contributions - to assist both academics and practitioners with the rankings of a set of competing bankruptcy and distress prediction models under multiple criteria.

Second, this study uses Malmquist DEA as a dynamic framework for evaluating and monitoring the relative performance of distress prediction models over time and ranking them (chapter four).

Third, from two points of view, this research is the most comprehensive comparative analysis of competing prediction models. On the one hand, it provides a more in-depth classification of statistical distress prediction models and performs an exhaustive evaluation considering the most popular models of each class. On the other hand, to assist with the operationalization of the proposed evaluation frameworks, this study uses the most popular performance criteria for prediction models along with an exhaustive list of typical performance measures (chapter two, three, four and five).

Fourth, inspiring by existing two-stage prediction models, this study suggests to apply the decomposition of the Non-Radial Technical Efficiency score, i.e., Slack-Based Measure

(SBM) of efficiency (Tone, 2001) into Pure Technical Efficiency (PTE), which presents the ability to improve the effectiveness by prudently allocating resources and using new technology, Scale Efficiency (SE), which indicates capacity to attain better efficiency by adjusting to its optimal scale, and Mix Efficiency (ME), which shows capacity to improve the effectiveness by managing input- or output-slacks, and analyse how each of these measures individually contributes to developing distress prediction models (chapter five).

Sixth, in addition to the conventional approach of using accounting variables as inputs and outputs of DEA models to estimate the measure of management efficiency, this study is the first that suggests using market information variables as inputs and outputs of DEA models to calculate the measure of the market efficiency of companies that is retained to be used as a predictor in a classifier model (chapter five).

Seventh, this study provides a comprehensive analysis of two-stage distress prediction models that apply different DEA models – say, input-oriented vs. output-oriented, radial vs. non-radial, static vs. dynamic, to compute the measures of management and market efficiency of companies at the first stage of two-stage models and use dynamic and static classifier frameworks at the second stage of two-stage models (chapter five).

1.5 Importance

The methods and findings of this study are important for a range of academics and practitioners who are interested in the field of corporate risk management and credit scoring. Further, the results of this research are important for stakeholders because distress prediction models protect them from downside effects of failure such as the cost of lawyers and court, opportunity cost, higher rate of financing, and more restrictions in capital raising.

To be more specific, bankruptcy and distress prediction models can provide facility to creditors (e.g., commercial banks, saving and loan associations) in estimating the probability of default and failure of their customers. This research contributes to the decision of creditors in allocating funds to customers with lower risk. Further, managers of the company could use the findings of this research to select the best early warning system that could take proper action against failure and immune their business. Also, financial institutions and investors can employ this study to have a more informative evaluation and the decision about their investment. Moreover, job seekers can make advantage to choose job positions in companies with economic stability. Further, the findings of this research could be applied by auditors as users of prediction models in the going-concern evaluation of companies.

1.6 Research Outline

This thesis is the combination of four separate projects that are presented in individual chapters. Chapters two, three and four focus on multi-criteria performance evaluation of bankruptcy and distress prediction models, and chapter five focuses on the two-stage distress prediction models.

Chapter two is about the multi-criteria assessment of bankruptcy prediction model (BPMs). It provides a survey on bankruptcy prediction models and a classification of applied criteria and measures to assess the performance of competing prediction models. Also, this chapter proposes and discusses the slack-based super-efficiency DEA and explains how to use this multi-criteria assessment framework to compare the performance of bankruptcy prediction models. This chapter ends with a discussion on empirical findings and conclusions.

Chapter three provides a multi-criteria assessment of competing distress prediction models (DPMs). It provides a survey on comparative studies related to competing statistical distress prediction models. Also, it describes the proposed multi-criteria methodology, namely context-dependent DEA that is used to compare the relative performance of competing DPMs. Finally, the chapter presents the empirical results and discussions.

Chapter four is about a dynamic multi-criteria assessment of DPMs. It begins with a review of advances in DPMs and a survey of comparative studies on DPMs. This chapter describes the proposed dynamic multi-criteria framework; namely, an orientation-free super-efficiency Malmquist DEA, for the comparison of prediction models. Further, it provides details on the experimental designs and the applied DPMs for comparison. Finally, the chapter concludes with empirical results and discusses its findings.

Chapter five is about the comparative analysis of two-stage DPMs. It provides a detailed literature on using DEA in corporate credit risk and failure prediction. Also, it describes different DEA models, the decomposed DEA scores and the hybrid two-stage models to be assessed in comparative analysis. Finally, the chapter describes the empirical results and findings.

Last but not the least, chapter six concludes this research by providing the summary of main findings, and the limitations of this research. It also offers some suggestions for future research in this field.

References

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.

Basel Committee on Banking Supervision, 2000. *Principles for the management of credit risk*. Basel.

Bauer, J., Agarwal, V., 2014. Are hazard models superior to traditional bankruptcy prediction approaches? A comprehensive test. *Journal of Banking & Finance* 40(Mar), 432–442.

Beaver, W.H., 1968. Alternative accounting measures as predictors of failure. *Accounting review* 43(1), 113–122.

Beaver, W.H., 1966. Financial Ratios as Predictors of Failure. *Journal of Accounting Research* 4 (Jan), 71–111.

Bellotti, T., Crook, J., 2009. Credit scoring with macroeconomic variables using survival analysis. *Journal of the Operational Research Society* 60(12), 1699-1707.

Bharath, S.T., Shumway, T., 2008. Forecasting default with the Merton distance to default model. *The 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.

Campbell, J.Y., Hilscher, J., Szilagyi, J., 2008. In search of distress risk. *The Journal of Finance* 63(3), 2899–2939.

Darrat, A.F., Gray, S., Park, J.C., Wu, Y., 2016. Corporate Governance and Bankruptcy Risk. *Journal of Accounting, Auditing & Finance* 31(2), 163–202.

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.

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.

Li, Z., Crook, J., Andreeva, G., 2017. Dynamic prediction of financial distress using Malmquist DEA. *Expert Systems with Applications* 80(Sep), 94–106.

Li, Z., Crook, J., Andreeva, G., 2014. Chinese companies distress prediction: an application of data envelopment analysis. *Journal of the Operational Research Society* 65(3), 466–479.

Liang, D., Lu, C.-C., Tsai, C.-F., Shih, G.-A., 2016. Financial ratios and corporate governance indicators in bankruptcy prediction: A comprehensive study. *European Journal of Operational Research* 252(2), 561–572.

Mousavi, M.M., Ouenniche, J., Xu, B., 2015. Performance evaluation of bankruptcy prediction models: An orientation-free super-efficiency DEA-based framework. *International Review of Financial Analysis* 42(Dec), 64–75.

Nam, C.W., Kim, T.S., Park, N.J., Lee, H.K., 2008. Bankruptcy prediction using a discrete-time duration model incorporating temporal and macroeconomic dependencies. *Journal of Forecasting* 27(6), 493–506.

Ohlson, J.A., 1980. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of accounting research* 18(Apr), 109–131.

Piesse, J., Lee, C.-F., Kuo, H.-C., Lin, L., 2006. Corporate failure: Definitions, methods, and failure prediction models, in: *Encyclopedia of Finance*. Springer, pp. 477–490.

Psillaki, M., Tsolas, I.E., Margaritis, D., 2010. Evaluation of credit risk based on firm performance. *European Journal of Operational Research* 201(3), 873–881.

Shumway, T., 2001. Forecasting bankruptcy more accurately: A simple hazard model. *The Journal of Business* 74(1), 101–124.

Sueyoshi, T., Goto, M., Omi, Y., 2010. Corporate governance and firm performance: Evidence from Japanese manufacturing industries after the lost decade. *European Journal of Operational Research* 203(3), 724–736.

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 among listed companies using accounting, market and macroeconomic variables. *International Review of Financial Analysis* 30(Dec), 394–419.

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

Wilson, R.L., Sharda, R., 1994. Bankruptcy prediction using neural networks. *Decision support systems* 11(5), 545–557.

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.

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

Chapter Two

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

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. This research overcomes this methodological issue by proposing an orientation-free super-efficiency data envelopment analysis model as a multi-criteria assessment framework. Furthermore, it performs an exhaustive comparative analysis of the most popular bankruptcy modelling frameworks for UK data including new models. In addition, this study addresses 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 the findings.

Keywords: Bankruptcy Prediction; Performance Criteria; Performance Measures; Data Envelopment Analysis; Slacks-Based Measure

2.1 Introduction

Corporate failure often occurs when a firm experiences serious loss 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, the reader is referred 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 a 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).

The 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 (HL) 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 the 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. This research addresses this methodological issue and fills the 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 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 behavior. Note also that the DEA terminology is motivated by an analogy between DMUs and production systems according to the economic theory of production.

Since its early days, DEA witnessed many methodological developments as well as a large number of applications. In the bankruptcy prediction area, DEA has so far been used either to classify firms into healthy and non-healthy categories (e.g., 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., 2014). Unlike these uses of DEA in bankruptcy research, this study proposes to use DEA as a performance evaluation framework of competing bankruptcy prediction models.

In sum, the key contribution of this chapter 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, this study performs an exhaustive comparative analysis of the most popular bankruptcy modelling frameworks for UK data; namely, statistical and stochastic models including the ones that are 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 the findings. The 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 to 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 chapter is organised as follows. Section 2.2 provides and classifies the literature on bankruptcy prediction models. Section 2.3 presents the proposed multi-criteria methodology; namely, an orientation-free super-efficiency DEA framework to evaluate the relative performance of competing forecasting models of bankruptcy. Section 2.4 presents and discusses the empirical findings. Finally, section 2.5 concludes the chapter.

2.2 Bankruptcy Prediction Models

Bankruptcy prediction models can be divided into two main categories; namely, accounting-based models and market-based models. Accounting-based models can 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. This chapter focuses on the relative performance of accounting-based models, market-based models, and hybrids. Hereafter, this study provides 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.

2.2.1 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 maximise 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 summarised as follows:

$$z = f\left(\sum_{j=1}^{p} \beta_j x_j\right), \qquad \text{Eq. 2-1}$$

where z is commonly referred to as a score or a z-score, x_js are explanatory variables, β_js represent the coefficients of the explanatory variables in the model, and f denotes the mapping of $\beta_j x_j$ on the set of real numbers $\mathfrak{R}$- 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. This study follows Hillegeist et al. (2004) in using a logit transformation:

$$PD = \frac{e^z}{1 + e^z} \qquad \text{Eq. 2-2}$$

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 2-A, that are part of most comparative analysis exercises and the comparative analysis of this research is no exception.

2.2.2 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 someone to overcome some of the limitations of the discriminant analysis. For example, within a regression framework for discrete response variables, the normality and the homoscedasticity assumptions are relaxed, on the 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:

$$\begin{cases} PD = Prob(y = 1) \\ PD = F(\beta, x) \end{cases} \qquad \text{Eq. 2-3}$$

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:

$$PD = \Lambda(\beta^t x) = \frac{e^{\beta^t x}}{1 + e^{\beta^t x}}, \qquad \text{Eq. 2-4}$$

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

$$PD = \beta^t x. \qquad \text{Eq. 2-5}$$

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 2-A, that are part of most comparative analysis exercises.

2.2.3 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. This chapter is limited 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: $S(t) = P(T>t)$, 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:

$$H(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T \leq t + \Delta t | T \geq t)}{\Delta t} = -\frac{S'(t)}{S(t)}, \qquad \text{Eq. 2-6}$$

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:

$$S(t, x; \theta) = 1 - \sum_{j<t} f(j, x, \theta) \text{ and } H(t, x; \theta) = \frac{f(t,x,\theta)}{S(t,x;\theta)} \qquad \text{Eq. 2-7}$$

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

$$PD_t = \frac{e^{H_0(t)+\beta^t x_t}}{1 + e^{H_o(t)+\beta^t x_t}}, \qquad \text{Eq. 2-8}$$

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

2.2.4 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, $P(V_A < D)$:

$$PD = \Phi(-\frac{\ln\left(\frac{V_A}{D}\right) + (\mu - \delta - 0.5\sigma^2) \times T}{\sigma\sqrt{T}}, \qquad \text{Eq. 2-9}$$

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:

$$\begin{cases} V_E = V_A e^{-\delta T}\phi(d_1) - De^{-rT}\phi(d_2) + \left(1 - e^{-\delta T}\right)\phi(d_1)V_A \\ \sigma_E = \dfrac{V_A e^{-\delta T}\phi(d_1)\sigma}{V_E} \end{cases} \qquad \text{Eq. 2-10}$$

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:

$$d_1 = \frac{\ln\left(\frac{V_A}{D}\right) + (r - \delta - 0.5\sigma^2) \times T}{\sigma\sqrt{T}}; d_2 = d_1 - \sigma\sqrt{T} \qquad \text{Eq. 2-11}$$

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

$$\mu = \frac{V_{At} + Dividends - V_{A,t-1}}{V_{A,t-1}} \qquad \text{Eq. 2-12}$$

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:

$$V_A = V_E + D \, ; \sigma = \frac{V_E}{V_A}\sigma_E + \frac{D}{V_A}\sigma_D \qquad \text{Eq. 2-13}$$

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%.

The next section describes the proposed DEA framework for assessing the relative performance of bankruptcy prediction models based on these modelling frameworks.

2.3 A Slacks-based DEA Framework for Assessing Bankruptcy Prediction Models

This chapter proposes a DEA-based framework for assessing the relative performance of competing bankruptcy prediction models. Hereafter, the basic concepts and models of DEA (see, section 2.3.1) are presented. Then, the way that someone might adapt a DEA framework to assess the relative performance of competing bankruptcy prediction models is explained (see, section 2.3.2).

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: Maximise 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, Cooper, and Rhodes (1978) and often referred to as CCR models are presented in Table 2.1, 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).

Table 2.1: Basic DEA Multiplier Models

Input-Oriented	**Output-Oriented**
$Maximize \quad e_k^{input} = \sum_{r=1}^{s} u_r y_{r,k}$	$Minimize \quad e_k^{output} = \sum_{i=1}^{m} v_i x_{i,k}$
$s.t.: \quad \sum_{i=1}^{m} v_i x_{i,k} = 1$	$s.t.: \quad \sum_{r=1}^{s} u_r y_{r,k} = 1$
$\sum_{r=1}^{s} u_r y_{r,j} \le \sum_{i=1}^{m} v_i x_{i,j};\ j = 1,\ldots,n$	$\sum_{r=1}^{s} u_r y_{r,j} \le \sum_{i=1}^{m} v_i x_{i,j};\ j = 1,\ldots,n$
$v_i \ge 0;\ i = 1,\ldots,m$	$v_i \ge 0;\ i = 1,\ldots,m$
$u_r \ge 0;\ r = 1,\ldots,s$	$u_r \ge 0;\ r = 1,\ldots,s$

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, $e_k^{output} = 1$) and slacks =0 means that DMU_k is efficient in that its weighted sum of outputs is equal to its weighted sum of inputs, and $e_k^{input} < 1$ (respectively, $e_k^{output} > 1$) means that DMU_k is inefficient in that it produces less output than the input it requires. The set of efficient DMU_s is referred to as the efficient frontier and represents the empirical standard of excellence.

Table 2.2: Basic DEA Envelopment Models

Input-Oriented	**Output-Oriented**
$Minimize \quad \theta_k$	$Maximize \quad \phi_k$
$s.t.: \quad \sum_{j=1}^{n} \lambda_j x_{i,j} \le \theta_k \cdot x_{i,k}; \forall i$	$s.t.: \quad \sum_{j=1}^{n} \lambda_j x_{i,j} \le x_{i,k}; \forall i$
$\sum_{j=1}^{n} \lambda_j y_{r,j} \ge y_{r,k}; \forall r$	$\sum_{j=1}^{n} \lambda_j y_{r,j} \ge \phi_k \cdot y_{r,k}; \forall r$
$\lambda_j \ge 0; \forall j$	$\lambda_j \ge 0; \forall j$
θ_k *unrestricted*	ϕ_k *unrestricted*

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 2.2, 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.

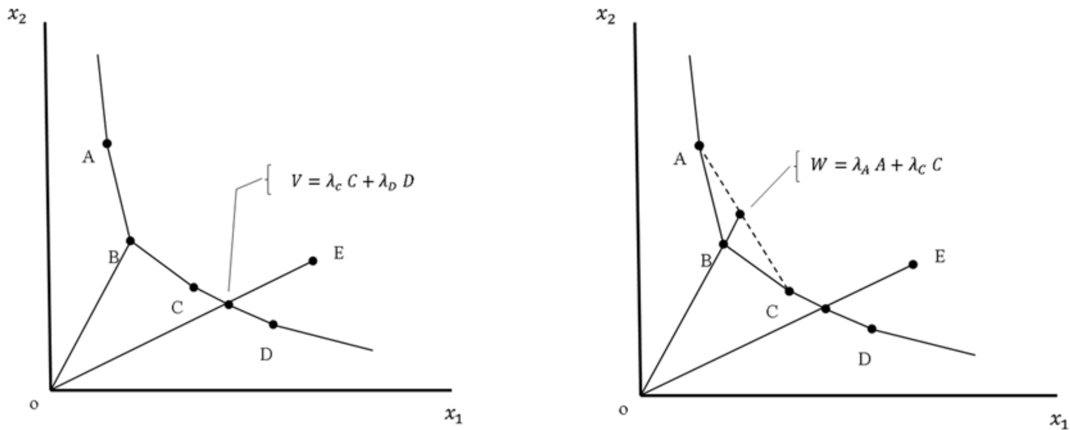
Last, but not least, DEA models – whether expressed in multiplier form or in envelopment form – allow one to identify the reference set or peer group used to benchmark each DMU in seeking improvements; for example, the reference set of a specific DMU, say DMU_k, is the set of DMU_s with positive dual variables λ_j. Note that if the optimal value of θ_k, say θ_k^*, (respectivelyϕ_k, say ϕ_k^*) is equal to 1 and slacks = 0 , then the DMU_k under evaluation is efficient; else, $\theta_k^* < 1$ (respectively, $\phi_k^* > 1$) indicates that DMU_k is inefficient and the current level of inputs (respectively, outputs) should be decreased (respectively, increased). As the objective is to provide a multidimensional ranking to get rid of the inconsistencies of unidimensional rankings, DEA efficiency scores are only used. 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).

2.3.2 Adaption of DEA framework

Under the basic DEA model discussed above, a considerable number of DMU_s are typically classified as efficient with a score of 1.0, which result in impossible differential analysis. Super-efficiency DEA introduced by Anderson and Peterson (1993) is a method to break the ties between efficient DMU_s. To demonstrate the application of super-efficiency DEA in breaking ties between efficient DMUs, we refer to Figure 2-1 showing a simple example with five DMU_s (A-E) that are described with two inputs and one output. To have a two-dimensional illustration, inputs are standardised on output.

From the Figure 2-1 the DMU_s of *A*, *B*, *C* and *D* are efficient with the assigned efficiency score of 1.0. As mentioned in the last section, the combination of efficient DMU_s constructs the efficient frontier. The DEA model compares the inefficient DMU_s with their reference points on the efficient frontier to indicate how they would have to improve.

Figure 2-1: Basic DEA vs. Supper-efficiency DEA model



Basic DEA model Super-efficiency DEA models

Source: Staat, M., & Hammerschmidt, M. (2005)

Under both basic and super-efficiency DEA, the results obtained for inefficient DMU_S are the same. In Figure 2-1, the inefficient DMU, *E,* is placed closest to *C* and *D*; thus, a virtual DMU, *V*, is made as a weighted average of *C* and *D* to represent as reference point for *E*; therefore*,* the efficiency score of E is $OV/OE < 1$. However, under basic DEA model, the reference point of an efficient DMU, for example B in Figure 2-1, is itself; therefore, the efficiency score of B is $OB/OB = 1$. Under super-efficiency DEA, on the other hand, the degree of super-efficiency of an efficient DMU, for example *B*, can be calculated by extracting that efficient DMU from the efficient frontier and comparing it with the new contracted efficient frontier by the remaining efficient DMU_S. Therefore, as it is shown in Figure 2-1, under super-efficiency DEA, the reference point of *B* is *W*, which is a linear combination of *A* and *C*; therefore, the super-efficiency score of B is $OW/OB > 1$ (Anderson and Peterson ,1993).

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* DMU_S*?* In this chapter, DMUs are thirty competing bankruptcy prediction models – see Appendix 2-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.

This chapter focuses 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 minimised (respectively, maximised) – see Appendix 2-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 – see for example, Paradi et al., 2004; Yeh et al., 2010; Li et al., 2014. 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 DMU_s evaluation and this research application is no exception. This chapter proposes 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. An orientation-free analysis has been deliberately chosen over input-oriented analysis or output-oriented analysis because, in the 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 my 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. The reason is that the super-efficiency DEA model was developed under (i) constant returns to scale (CRS) condition and (ii) the simultaneous and same proportion of change in all inputs (or outputs). Once any of these conditions is violated, it is high likely that infeasibility of the related DEA mode occurs (see, e.g., Seiford and Zhu, 1998a,b). 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 summarised 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):

$$Min \;\; \rho_k = \left(1 - \frac{1}{m}\sum_{i=1}^{m}\frac{s_{i,k}^-}{x_{i,k}}\right) \Bigg/ \left(1 + \frac{1}{s}\sum_{r=1}^{s}\frac{s_{r,k}^+}{y_{r,k}}\right) \qquad \text{Eq. 2-14}$$

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

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

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

where n denotes the number of DMU_s, 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, $s_{i,k}^-$ and $s_{r,k}^+$ are slack variables associated with the first and the second sets of constraints, respectively. If the optimal objective function value $\rho_k^* = 1$ and slacks = 0, then DMU_k is classified as efficient. If $\rho_k^* \neq 1$, then DMU_k is classified as inefficient. Note that model 2-14 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:

$$\sum_{j=1}^{n} \lambda_j = 1\,;\ \sum_{j=1}^{n} \lambda_j \geq 1\,;\ \sum_{j=1}^{n} \lambda_j \leq 1. \qquad \text{Eq. 2-15}$$

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):

$$Min\ \ \delta_k = \left(1 - \frac{1}{m}\sum_{i=1}^{m}\frac{t_{i,k}^{-}}{x_{i,k}}\right) \Big/ \left(1 - \frac{1}{s}\sum_{r=1}^{s}\frac{t_{r,k}^{+}}{y_{r,k}}\right) \qquad \text{Eq. 2-16}$$

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

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

$$\lambda_j \geq 0, \forall_j \neq k; t_{i,k}^{-} \geq 0, \forall_i; t_{r,k}^{+} \geq 0, \forall_r$$

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 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 (VRS) 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, the above-described methodology is used to rank order competing bankruptcy prediction models and discuss the empirical results obtained using UK data for the period of 1989-2006.

2.4 Empirical Investigation

In this section, we first describe the process of data gathering and sample selection (see section 2.4.1). Then, I present the list of models that I used in my comparative analysis (see section 2.4.2). Next, I provide the list of criteria and measures that I applied to evaluate bankruptcy prediction models (see, section 2.4.3). Finally, I discuss my empirical findings on the relative performance evaluation of bankruptcy prediction models under both a single criterion and multiple criteria (see, section 2.4.4).

2.4.1 Data and Sample Selection

In this chapter, I 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, I further reduced such dataset by excluding both financial and utility 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, my 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 my 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.

2.4.2 Bankruptcy Models to be Assessed

In this chapter, I 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 my 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 my 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. I refer to these models as the "original" models and represent them in Figure 2-2 and Figure 2-3 with white shapes, where the shapes differ from one modelling framework to another – see legends of Figure 2-2 and Figure 2-3.

I also include in my comparative analysis three additional categories of models that I 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 my sample data (i.e., Altman, 1968, Lis, 1972, Altman, 1983, Ohlson, 1980, Taffler, 1984, Zmijewski, 1984, Shumway, 2001) and represented in Figure 2-2 and Figure 2-3 with dotted shapes, where the shapes differ from one modelling framework to another – see legends of Figure 2-2 and Figure 2-3. 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 Figure 2-2 and Figure 2-3 with grey shapes, where the shapes represent the original modelling frameworks – see, legends of Figure 2-2 and Figure 2-3. 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 2-B) are represented in Figure 2-2 and Figure 2-3 with black shapes, where the shapes represent the original modelling frameworks – see, legends of Figure 2-2 and Figure 2-3.

In sum, a total of 30 models are assessed in my comparative analysis – see Appendix 2-A and Appendix 2-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, I 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, section 2.3.2).

2.4.3 Performance Criteria and Measures

With respect to performance criteria for evaluating bankruptcy prediction models, the focus of this chapter shall be on the most commonly used criteria and their; namely, the discriminatory power criterion, the calibration accuracy criterion, the information content criterion, and the correctness of categorical predictions criterion measures (Bauer and Agarwal, 2014). 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 my comparative analysis of models, I use Kolmogorov-Smirnov (KS) Statistic, Area under Receivable Operating Characteristic (AUROC) – also known as c-statistic, Gini Index, and Information Value (IV) to measure the discriminatory power criterion; I use Brier Score (BS) to measure the quality of fit under calibration accuracy criterion; I use log-likelihood statistic (LL) and pseudo-coefficient of determination (pseudo-R^2) to measure the information content under calibration accuracy criterion; and I 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 2-C for descriptions of these measures.

2.4.4 Performance Evaluation of Bankruptcy Prediction Models

In my empirical investigation, I first generated the unidimensional rankings of the 30 models under evaluation (see, Figure 2-2) 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 minimised (respectively, maximised). 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 my 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 (AUROC) 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 my dataset, 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.

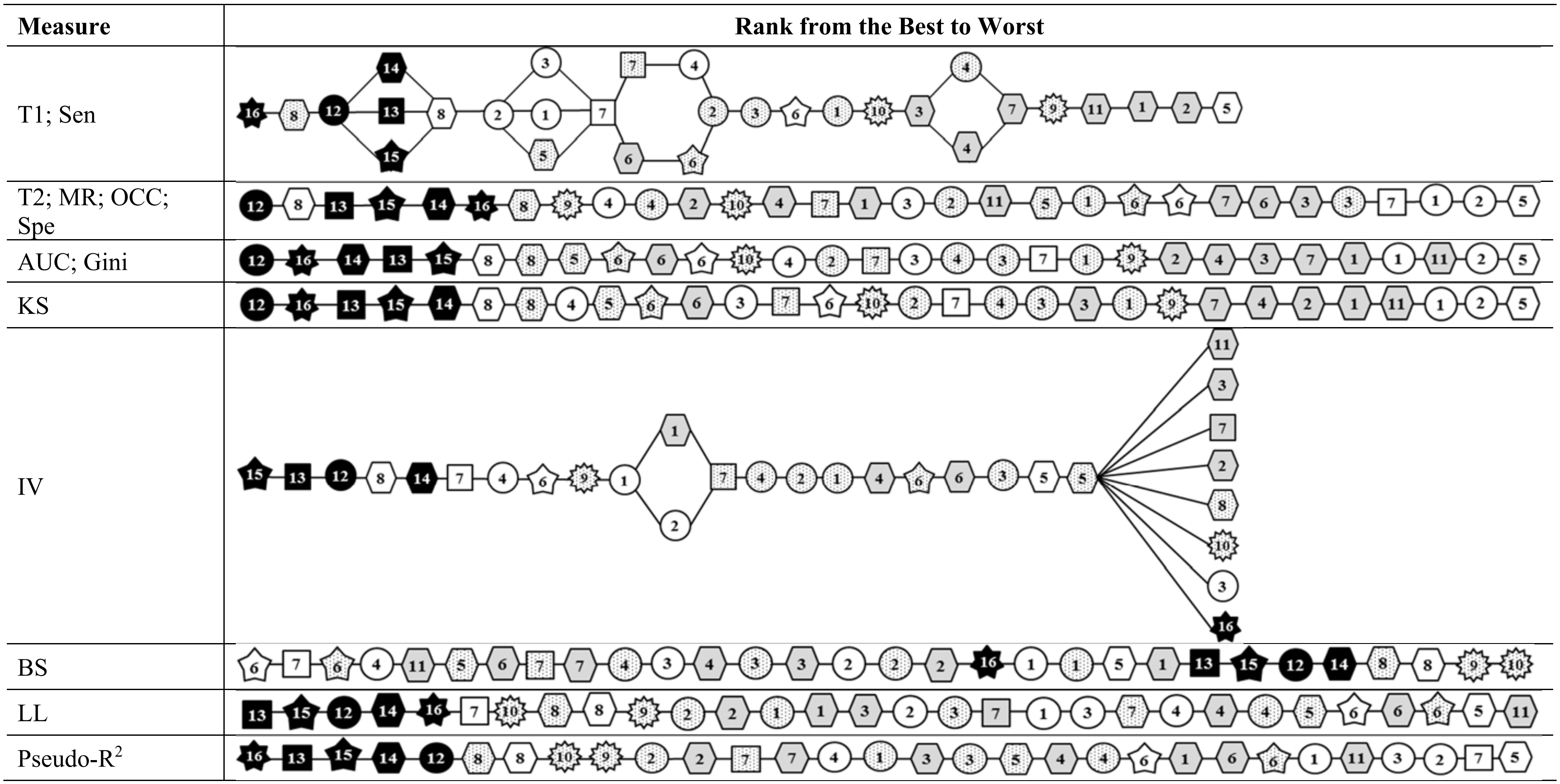
On the other hand, for most performance measures, reworking the original MDA, probit and linear probability models with the same explanatory variables in a logit framework seems to improve the ranks – with the exception of the MDA models of Lis (1972) and Taffler (1984) which were originally fitted to 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 in rankings; however, Hillegeist et al. (2004) model seems to always outperform Bharath and Shumway (2008) model.

At this stage, I 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 I provide using the proposed DEA framework.

Figure 2-2: Unidimensional Rankings of Bankruptcy Prediction Models

This figure presents the unidimensional rankings of 30 competing bankruptcy models, where models are ranked from best to worst using a single measure of a single criterion at a time. T1 (type I error), T2 (type II error), MR (misclassification rate), Sen (sensitivity), Spe (specificity) and OCC (overall correct classification) are used as measures of correctness of categorical prediction; AUC (area under receiver operating character), Gini coefficient, KS (Kolmogorov-Smirnov) and IV (information value) are used as measures of discriminatory power; BS (brier score) is used as measures of calibration accuracy; and log-likelihood (LL) and Pseudo-R^2 (R^2) are used as measures of information content. Different shapes represent different modelling frameworks; namely, multivariate Discriminant Analysis (MDA), Linear Probability (LPA), Logit Analysis (LA), Probit Analysis (PA), Survival Analysis (SA), and Black-Scholes-Merton-based Model (BSM-based). White, dotted white, grey, and black shapes represent the original models, the original models refitted, the reworked models with the same explanatory variables, and the new models, respectively.



[1]Altman (1968); [2]Altman (1983); [3]Lis (1972); [4]Taffler (1984); [5]Ohlson (1980); [6]Zmijewski (1984); [7]Theodossiou (1995); [8]Shumway (2001); [9]Bharath & Shumway (2008); [10]Hilligieski et.al (2004); [11]Bemmann (2005); [12]New MDA model; [13]New LPA model; [14]New LA model; [15]New PA model; [16]New SA model

= MDA Framework; = LP Framework; = LA Framework; = PA Framework; = SA Framework; = BSM-based Framework

The multi-criteria rankings of the above mentioned 30 models are provided in Figure 2-3 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, section 2.3.2). 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, I 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 my dataset 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 the multi-criteria setting, refitting models is not necessarily a mean for improvement.

Figure 2-3: SBM Super-Efficiency DEA Scores-based Multidimensional Rankings of Bankruptcy Prediction Models

This figure presents the multi-criteria rankings of 30 competing bankruptcy models using a DEA ranking framework, where models are ranked from best to worst using DEA scores. A multi-criteria ranking is produced for each combination of a variety of metrics of the performance criteria under consideration, where inputs (resp. outputs) are chosen according to the principle of the less (resp. more) the better. T1 (type I error), T2 (type II error), MR (misclassification rate), Sen (sensitivity), Spe (specificity) and OCC (overall correct classification) are used as measures of correctness of categorical prediction; AUC (area under receiver operating character), Gini coefficient, KS (Kolmogorov-Smirnov) and IV (information value) are used as measures of discriminatory power; BS (brier score) is used as measures of calibration accuracy; and log-likelihood (LL) and Pseudo-R^2 (R^2) are used as measures of information content. Different shapes represent different modelling frameworks; namely, multivariate Discriminant Analysis (MDA), Linear Probability (LPA), Logit Analysis (LA), Probit Analysis (PA), Survival Analysis (SA), and Black-Scholes-Merton-based Model (BSM-based). White, dotted white, grey, and black shapes represent the original models, the original models refitted, the reworked models with the same explanatory variables, and the new models, respectively.

Inputs	Outputs	Rank from the Best to Worst
T1; BS; LL	ROC	16-13-12-7-5-6-15-14-6-6-4-7-3-8-3-4-7-11-4-3-8-2-2-1-1-2-1-10-9-5
BS; LL	ROC; Sen	13-16-7-12-6-5-15-6-14-6-4-7-3-3-8-4-7-3-4-11-2-2-1-1-2-1-8-10-5-9
T2; BS; LL	ROC	13-12-16-7-4-6-5-7-6-6-15-14-4-11-7-3-4-3-3-8-2-2-1-8-1-2-1-10-9-5
MR; BS; LL	ROC	13-12-16-7-4-6-5-7-6-6-15-14-4-11-7-3-4-3-3-8-2-2-1-8-1-2-1-10-9-5
BS; LL	ROC; OCC	13-7-16-12-4-6-5-7-6-15-6-14-4-11-7-4-3-3-3-2-2-1-1-8-2-1-8-10-9-5
BS; LL	ROC; Spe	13-7-16-12-4-6-5-7-6-15-6-14-4-11-7-4-3-3-3-2-2-1-1-8-2-1-8-10-9-5
T1; BS; LL	KS	16-13-12-7-5-4-6-15-14-6-6-7-8-3-3-7-8-4-3-11-4-2-1-2-2-10-1-1-9-5
BS; LL	KS; Sen	13-16-7-12-4-5-6-15-6-14-6-7-3-8-3-7-4-3-4-11-2-8-1-2-1-2-1-10-9-5
T2; BS; LL	KS	12-13-16-4-7-6-5-6-15-7-14-6-3-7-4-11-3-8-4-3-8-2-2-1-1-10-1-2-9-5
MR; BS; LL	KS	12-13-16-4-7-6-5-6-15-7-14-6-3-7-4-11-3-8-4-3-8-2-2-1-1-10-1-2-9-5
BS; LL	KS; OCC	13-16-12-7-4-6-5-7-15-6-14-6-4-3-7-11-4-3-3-2-8-2-1-8-1-1-2-10-9-5
BS; LL	KS; Spe	13-16-12-7-4-6-5-7-15-6-14-6-4-7-3-11-4-3-3-2-8-2-1-8-1-1-2-10-9-5

To be continued …

Figure 2.3 continue

Inputs	Outputs	Rank from the Best to Worst
T1; BS	ROC; R^2	16 – 12 – 6 – 5 – 7 – 7 – 4 – 6 – 6 – 13 – 15 – 14 – 7 – 8 – 11 – 8 – 4 – 3 – 4 – 3 – 3 – 2 – 10 – 2 – 9 – 1 – 2 – 1 – 1 – 5
BS	ROC; Sen; R^2	16 – 12 – 6 – 5 – 7 – 7 – 13 – 4 – 15 – 14 – 6 – 6 – 7 – 8 – 8 – 11 – 4 – 4 – 3 – 3 – 3 – 2 – 10 – 2 – 1 – 9 – 1 – 2 – 1 – 5
T2; BS	ROC; R^2	16 – 12 – 6 – 4 – 13 – 15 – 14 – 7 – 8 – 5 – 6 – 7 – 6 – 7 – 8 – 11 – 4 – 4 – 3 – 3 – 3 – 10 – 2 – 9 – 2 – 1 – 1 – 2 – 1 – 5
MR; BS	ROC; R^2	16 – 12 – 6 – 4 – 13 – 15 – 14 – 7 – 8 – 5 – 6 – 7 – 6 – 7 – 8 – 11 – 4 – 4 – 3 – 3 – 3 – 10 – 2 – 9 – 2 – 1 – 1 – 2 – 1 – 5
BS	ROC; OCC; R^2	16 – 12 – 6 – 4 – 13 – 15 – 7 – 14 – 5 – 8 – 7 – 6 – 6 – 7 – 8 – 11 – 4 – 4 – 3 – 3 – 3 – 2 – 10 – 2 – 9 – 1 – 1 – 2 – 1 – 5
BS	ROC; Spe; R^2	16 – 12 – 6 – 4 – 13 – 15 – 7 – 14 – 5 – 8 – 7 – 6 – 6 – 7 – 8 – 11 – 4 – 4 – 3 – 3 – 3 – 2 – 10 – 2 – 9 – 1 – 1 – 2 – 1 – 5
T1; BS	KS; R^2	16 – 12 – 6 – 5 – 4 – 7 – 7 – 6 – 6 – 13 – 15 – 14 – 8 – 7 – 11 – 8 – 4 – 3 – 4 – 3 – 3 – 2 – 10 – 2 – 9 – 1 – 2 – 1 – 1 – 5
BS	KS; Sen; R^2	16 – 12 – 6 – 5 – 4 – 7 – 7 – 13 – 15 – 14 – 6 – 6 – 7 – 8 – 8 – 11 – 4 – 3 – 4 – 3 – 3 – 2 – 10 – 2 – 1 – 9 – 1 – 2 – 1 – 5
T2; BS	KS; R^2	16 – 12 – 6 – 4 – 13 – 15 – 14 – 7 – 5 – 8 – 6 – 6 – 7 – 7 – 8 – 11 – 4 – 4 – 3 – 3 – 3 – 10 – 2 – 9 – 2 – 1 – 1 – 2 – 1 – 5
MR; BS	KS; R^2	16 – 12 – 6 – 4 – 13 – 15 – 14 – 7 – 5 – 8 – 6 – 6 – 7 – 7 – 8 – 11 – 4 – 4 – 3 – 3 – 3 – 10 – 2 – 9 – 2 – 1 – 1 – 2 – 1 – 5
BS	KS; OCC; R^2	16 – 12 – 4 – 6 – 13 – 15 – 7 – 5 – 14 – 8 – 6 – 6 – 7 – 7 – 8 – 11 – 4 – 4 – 3 – 3 – 3 – 2 – 10 – 2 – 9 – 1 – 1 – 2 – 1 – 5
BS	KS; Spe; R^2	16 – 12 – 4 – 6 – 13 – 15 – 7 – 5 – 14 – 8 – 6 – 6 – 7 – 7 – 8 – 11 – 4 – 4 – 3 – 3 – 3 – 2 – 10 – 2 – 9 – 1 – 1 – 2 – 1 – 5

[1]Altman (1968); [2]Altman (1983); [3]Lis (1972); [4] Taffler (1984); [5]Ohlson (1980); [6] Zmijewski (1984); [7] Theodossiou (1995); [8] Shumway (2001); [9] Bharath & Shumway (2004); [10] Hilligieski et.al (2004); [11]Bemmann (2005); [12]New MDA model; [13]New LPA model; [14]New LA model; [15] New PA model; [16] New SA model

= MDA Framework; = LP Framework; = LA Framework; = PA Framework; = SA Framework; = BSM-based Framework;

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 – except for 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 the 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 multi-criteria rankings. Therefore, under the multi-criteria 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, multi-criteria 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 the changes in most combinations of performance metrics.

However, overall the multi-criteria 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, my findings suggest the following answers to my research questions. First, the survival analysis model tends to 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 accounting-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.

2.5 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 many 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, I 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, I performed an exhaustive comparative analysis of the most popular six bankruptcy modelling frameworks resulting in 30 prediction models for UK firms including my own models organised into four categories; namely, original models, original models refitted, reworking models in a logit framework with the same original explanatory variables, and new models. I 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. I have considered several measures for each criterion to find out about the robustness of multidimensional rankings with respect to different combinations of measures. Furthermore, I 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?

My main findings may 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 the survival analysis model tends to 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 accounting-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.

It must be borne in mind that the present chapter has some limitations. Firstly, this chapter does not take into account non-statistical bankruptcy prediction models. Also, this study considers only one dynamic modelling framework, i.e. discrete time hazard models. Secondly, for this research, I did not have access to non-listed UK companies, and therefore, in this study I used listed UK companies. Third, one of the issues related to the employed multi-criteria assessment framework in this chapter is that within a super-efficiency DEA, the reference benchmark changes from one efficient DMU to another one (refer to Figure 2-1), which in some contexts might be viewed as “unfair” benchmarking.

Future research could incorporate non-statistical frameworks and more variety of dynamic prediction models. Also, future research could use data of non-listed UK companies as well as the data of other countries. Further, to overcome the issue related to possible ‘unfair’ benchmarking of super-efficiency DEA, future research could incorporate the context-dependent DEA framework.

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.

Black, F., & Scholes, M. , 1973. The pricing of options and corporate liabilities. *Journal of political economy*, 81(3), 637-654.

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., Bhabar, 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.

Ravikumar, 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.

Staat, M. and Hammerschmidt, M., 2005. Product performance evaluation: a super-efficiency model. *International Journal of Business Performance Management*, *7*(3), 304-319.

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

Seiford, L.M. and Zhu, J., 1998. Stability regions for maintaining efficiency in data envelopment analysis. *European Journal of Operational Resea*rch, 108(1):127-139.

Seiford, L.M. and Zhu, J., 1998. Sensitivity analysis of DEA models for simultaneous changes in all the data. *Journal of the Operational Research Society*, 49(10):1060-1071.

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.

Appendix 2-A: Statistical Models of Bankruptcy Prediction

Framework	Type	Variables	
Discriminant Analysis	**Altman (1968)**	WCTA	= Working capital / Total assets,
		RETA	= Retained earnings / Total Assets,
		EBITTA	= Earnings before interest and taxes / Total assets,
		METD	= Market value of equity / Total debt,
		STA	= Sales / Total assets.
	Altman (1983)	WCTA	= Working capital / Total assets,
		RETA	= Retained earnings / Total assets,
		EBITTA	= Earnings before interest and taxes / Total assets,
		BETL	= Book value Equity / Total liabilities,
		STA	= Sales / Total assets.
	Lis (1972)	WCTA	= Working capital / Total assets,
		EBITTA	= Earnings before interest and taxes / Total assets,
		METL	= Market value of equity / Total liabilities,
		NWTA	= Net wealth / Total assets,
	Taffler (1984)	PBTCL	= Profit before tax /Current liabilities,
		CLTA	= Current liabilities/ Total assets,
		CATL	= Current assets/ Total liabilities,
		NCI	= No-credit intervals.
Regression Analyses	**Theodossiou**	WCTA	= Working capital / Total assets,
	Linear	NITA	= Net income / Total assets,
	Model	LTDTA	= Long term debt / Total assets,
		TDTA	= Total debt / Total assets,
		RETA	= Retained earnings / Total assets.
	Ohlson (1980)	TLTA	= Total liabilities / Total assets
	Logit Model	WCTA	= Working capital / Total assets
		CLCA	= Current liabilities / Current assets
		OENEG	= 0 if total liabilities exceed total assets, 1 otherwise
		NITA	= Net income / Total assets
		FUTL	= Funds from operations (operating income minus depreciation) / Total liabilities
		INTWO	= 1 if net income has been negative for the last 2 years, 0 otherwise,
		CHIN	$=NI_t - NI_{t-1}/(\lvert NI_t\rvert + \lvert NI_{t-1}\rvert)$, NI_t is the net income for the last period. The variable is a proxy for the relative change in net income.
		Ohlsonsize	= log (Total assets/GNP price-level index)
	Zmijewski	NITA	= Net income / Total assets
	(1984)	TLTA	= Total liabilities / Total assets
	Probit Model	CACL	= Current assets / Current liabilities
	Bemmann (2005) *Logit Model*	TLTA	=Total Liabilities / Total assets
Survival Analysis Model	**Shumway**	NITL	=Net income/total liabilities,
	(2001)	TLTA	= Total liabilities/total assets,
		RealSize	= Log (the number of outstanding shares multiplied by year-end share price divided by total market value),
		LagExRet	= Cumulative annual return in year $t-1$ minus the value-weighted FTSE index return in year $t-1$,
		LagSigma	= Standard deviation of residuals derived from regressing monthly stock return on market return in year $t-1$.

BSM-based Models	**Hillegeist et al. (2004) & Bharath and Shumway (2008)**	V_E	= Market value of equity,
		V_A	= Market value of assets,
		μ	= Continuously compounded expected return on assets,
		δ	= Continuous dividend rate expressed in terms of V_A,
		D	= Face value of debt maturing at time T,
		σ	= Asset volatility,
		T	= Time to debt maturity, considers as 1 year.

Appendix 2-B: New Designed Models

This table presents the explanatory variables and coefficients of the new models in 5 different frameworks; multivariate discriminant analysis (MDA), linear probability (LPA), logit analysis (LA), probit analysis (PA) and survival analysis (SA). A star and a dragger refer to 1% and 5% significance level, respectively.

Models / Explanatory Variables	MDA	LPA	LA	PA	SA
Intercept	9.709	0.047 *	7.970 *	3.971*	
Working capital / Total assets	2.447	0.041 *	1.522 *	0.817 †	1.216 *
Net income / Capital	-0.003				
Net income / Current assets	-0.176				
Net income / Equity	0.004				0.002 †
Total debt / Equity	0.150				
EBIT / Total assets	1.099				
Total liabilities / Total assets	-2.016	- 0.022 *			
Inventory / Working capital	0.148	0.001†			
Inventory / Sales	-0.567				
Quick asset / Sales	-0.056				
Current liabilities / Inventory	0.002				
Total liabilities / Working capital	-0.015				
Net worth / Total assets	-0.963				-1.629 *
Current liabilities / Total assets					-1.615 †
Real size	0.178	0.004 *	0.311 *	0.147 *	0.261 *
Lag Ex.Ret.	2.184	0.044 *	1.468 *	0.678 *	1.554 *
Sigma	-6.467	- 0.146 *	-2.089 †	-1.131†	
Long term debt / Total assets					-2.493*

Appendix 2-C: Performance Measures for Assessing Bankruptcy Prediction Models

Criteria	Measure	Formula	Definition
Discriminatory Power	IV	$IV = \sum_{i=1}^{I}(g_i/n_G - b_i/n_B)\ln\left(\frac{g_i/n_G}{b_i/n_B}\right)$ where g_i (respectively, b_i) denotes the number of goods (respectively, bads) in band I, $\sum_{i=1}^{I} g_i = n_G$ and $\sum_{i=1}^{I} b_i = n_B$.	The information value (IV) is nothing but the divergence statistic originally suggested by Kullback and Leibler (1951) as a measure of the relative distance between an empirical probability distribution and a theoretical one. The measure provided here is based on a discrete approximation of the density functions of the good cases and the bad cases.
	CIER	$CIER_S = \frac{H(P) - H_S}{H(P)}$ where $H(P) = -(P.\log(P) + (1-P).\log(1-P))$ is the unconditional entropy of a discrete random variable with say two categories; namely, non-default and default; and P denotes the probability of default.	In practice, one is also interested in the entropy of the probability of default given a specific rating score, say S; that is; $H(P(D\mid S)) = -(P(D\mid S)log\,P(D\mid S) + P(\bar{D}\mid S)logP(\bar{D}\mid S))$, which is a random variable with an expected value of $H_S = -E[P(D\mid S)logP(D\mid S) + P(\bar{D}\mid S)logP(\bar{D}\mid S)]$. that is known as the **conditional entropy** of the default event (with respect to the rating score S) and is upper bounded by the unconditional information entropy of the default event; i.e. $H_S \leq H(p)$. Note that the larger the difference between $H(p)$ and H_S, the more gain of information results from using rating scores.
	KS statistic	$KS = \max_s\left(F(s\mid B) - F(s\mid G)\right)$, where $F(.\mid G)$ & $F(.\mid B)$ denote the empirical cumulative distribution functions of the samples of good & bad cases, respectively.	The Kolmogorov-Smirnov (KS) statistic measures the distance between the empirical cumulative distribution functions of the samples of good cases and bad cases.
	AUC	$C - statistic = \frac{U}{N_G \cdot N_B}$ where U denotes the Mann-Whitney U statistic, N_G denotes the number of good cases in the sample, and N_B denotes the number of bad cases in the sample	The receiver operating characteristic (ROC) curve is a plot of the hit rate against the false alarm rate for all cut-off points. A good prediction model would generate a ROC curve well farther away from the diagonal, which suggests that the larger the area under the ROC curve, often referred to as AUC or AUROC and measured by a statistic called concordance or c-statistic, the better the prediction model performance.
	Gini Coefficient	$G = 2 \cdot AUC - 1$	The Gini coefficient, say G, is a measure of the area between the ROC curve and the diagonal – also referred to as Somer's D.

Criteria	Measure	Formula	Definition
Information Content	Log-likelihood statistic	Formulae are model-dependent and shall not be presented for space considerations	The Log-Likelihood statistic (LL) is a measure of goodness-of-fit and is computed as the natural logarithm of the maximum value of the likelihood function of a model, where the likelihood function is a function of the parameters of the model which are determined so that the model is in maximum "agreement" with the data. In my empirical investigation, I computed LL values as suggested by Hillegeist et al. (2004).
	Pseudo-R^2	$R^2 = 1 - \frac{LL(Logit\ model\ with\ intercept\ and\ predictors)}{LL(Logit\ model\ with\ intercept\ only)}$ where LL denotes the log-likelihood of a model	Pseudo-R^2 is a measure of the strength of association between the output of a logistic regression model and the set of explanatory variables and its range lies between 0 and 1, with higher values indicating better logit model with intercept and predictors likelihood; i.e., better "agreement" of the selected model with the observed data.
Correctness of Categorical Predictions	Sensitivity	$S_e = \frac{N_{B\|B}}{N_{B\|B}+N_{B\|G}}$ where $N_{B\|B}$ denotes the number of bad cases predicted as bad, $N_{B\|G}$ denotes the number of bad cases predicted as good	Given a specific cut-off score, say s_c, sensitivity (S_e) is defined as the fraction of the bad cases that would have scores below the cut-off and would therefore be rightly rejected; i.e., the proportion of bad cases who are predicted as bad. Sensitivity is also referred to as the hit rate.
	Specificity	$S_p = \frac{N_{G\|G}}{N_{G\|G}+N_{G\|B}}$ where $N_{G\|G}$ denotes the number of good cases predicted as good, $N_{G\|B}$ denotes the number of good cases predicted as bad	Given a specific cut-off score, say s_c, specificity (S_p) is defined as the fraction of the good cases that have scores above the cut-off and would therefore be rightly accepted; i.e., the proportion of good cases who are predicted as good.
	Type I Error	Type I error = 1- Specificity	Type I error is the proportion of bad cases being misclassified as good cases
	Type II Error	Type II error = 1- Sensitivity	Type II error is the proportion of good cases being misclassified as bad cases – also referred to as the false alarm rate.

Chapter Three

Multi-Criteria Ranking of Corporate Distress Prediction Models: An Orientation-Free Context-dependent DEA-based Framework

Abstract: Although many modelling and prediction frameworks for corporate bankruptcy and distress have been proposed, the relative performance evaluation of prediction models is criticised due to the assessment using a single measure of one criterion at a time, which lead to reporting conflicting results. Mousavi et al. (2015) proposed an orientation-free super-efficiency data envelopment analysis (DEA) -based framework to overcome this methodological issue. However, within a super-efficiency DEA framework, the reference benchmark changes from one prediction model evaluation to another one, which in some contexts might be viewed as "unfair" benchmarking. In this chapter, I overcome this issue by proposing a slacks-based context-dependent DEA (SBM-CDEA) framework to evaluate competing distress prediction models. Furthermore, using data on UK firms listed on London Stock Exchange (LSE), I exercise a comprehensive comparative analysis of the most popular corporate distress prediction models under both a mono criterion and multiple criteria frameworks considering several performance measures. Also, I propose new models using macroeconomic indicators as features.

Keywords: Corporate Distress Prediction; Performance Criteria; Performance Measures; Context-Dependent Data Envelopment Analysis; Slacks-Based Measure

3.1 Introduction

Business distress refers to a situation in which a company is unable to continue its operations because the revenue generated by the company is not enough to cover its expenses. Early detection of a company deteriorating condition or distress has such economic benefits, which motivate both academics and practitioners in finance and accounting to invest in developing a range of corporate distress prediction models. From a statistical point of view, a distress prediction model or DPM is a typical classification problem, which uses the selected features; say accounting, market, and macroeconomic-based information, to classify the firms into distressed or non-distressed categories or classes. During the last decades, numerous studies have employed different types of techniques from statistics, operational research, and artificial intelligence fields to design new DPMs. Initial studies on distress prediction use statistical techniques such as univariate discriminant analysis (e.g., Beaver, 1966, 1968), and multivariate discriminant analysis (e.g., Altman, 1968, 1973, 1983) as classification techniques. Later on, conditional probability models such as linear probability models (e.g., Meyer and Pifer, 1970; Maddala, 1986), logit models (e.g., Martin, 1977; Ohlson, 1980), and probit models (e.g., Zmijewski, 1984) are used to predict the probability of distress. The common characteristic of these models, however, is that they are time-independent (i.e., static) in nature and as such fail to take time-varying features of a firm into account. Dynamic models such as survival (hazard) models (e.g., Lane et al., 1986; Crapp and Stevenson, 1987; Luoma and Laitinen, 1991; Shumway, 2001; Bharath and Shumway, 2008; Chava and Jarrow, 2004), and contingent claims models (e.g., Bharath and Shumway, 2008; Hillegeist et al., 2004) are the next group of models, which by design could take account of changes in the condition of firms over time. Statistical techniques, however, are constrained by the potential severity of the underlying assumptions, i.e., linearity, multivariate normality, independence among predictor or input variables, and equal within-group variance-covariate matrices. The artificial intelligence and mathematical programming techniques are alternatives that overcome the methodological restrictions related to statistical techniques.

Considering the massive increase in the number of DPMs, a stream of the literature has focused on answering the question: which of these models are superior in performance? According to Zhou (2013), DPMs are data-fitting based empirical research consisting of a series of processes including sampling, features selection, modelling, and performance evaluation. Obviously, the performance of DPMs is not only dependent on the sample selection, modelling techniques and feature selection procedures but also reliant on the evaluation process and the chosen performance criteria. In practice, several studies have compared the performance of competing DPMs taking into account different modelling frameworks – see for example, Bauer and Agarwal (2014) , Mousavi et al.(2015) and Wu et al. (2010); alternative sampling techniques – see, for example, Neves and Vieira (2006), and Zhou (2013); and various features – see, for example, Tinoco and Wilson (2013), Trujillo-Ponce et al. (2014). Furthermore, several criteria, including, *discriminatory power*, *calibration accuracy*, *information content* and *correctness of categorical prediction* have been used for the performance evaluation of alternative models.

My survey of the existing studies concerned with the comparison of competing statistical DPMs supports Bauer and Agarwal (2014) and Mousavi et al. (2015) arguments in addressing two main drawbacks in the related literature. Firstly, most of the existing studies failed to have a comprehensive comparison between all types of statistical models, i.e. traditional statistical models, contingent claim analysis (CCA) models and survival analysis models. Secondly, the existing literature has used a restricted number of criteria to evaluate the performance of competing models. Thirdly, as mentioned by Mousavi et al. (2015), the nature of the performance evaluation of competing DPMs remains mono-criterion, as they use a single measure of a single criterion at a time. Therefore, under mono-criterion evaluation, the rankings corresponding to different criteria are mostly different, which lead to a situation that practitioners cannot make a well-informed decision as to which model performs best when taken all criteria into account (see, for example, Theodossiou, 1991; Bandyopadhyay, 2006; Tinoco and Wilson, 2013). To overcome this methodological drawback, Mousavi et al. (2015) proposed a multi-criteria assessment framework; namely, an orientation-free super-efficiency DEA. However, within a super-efficiency DEA framework, the reference benchmark may change from one efficient decision-making unit (DMU) evaluation to another one, which in some contexts might be viewed as "unfair" benchmarking (Ouenniche et al., 2014) - for more details about super-efficiency DEA framework and its related issue, the reader is referred to section 2.3.2. This study overcomes this issue by proposing a slacks-based context-dependent DEA (SBM-CDEA) framework (Seiford and Zhu, 2003; Morita et al., 2005; Zhu, 2014) for evaluating the relative performance of competing DPMs. I assess the performance of the proposed SBM-CDEA framework by performing a comparative analysis of the most commonly used and cited statistical corporate distress prediction modelling frameworks. I organised models into three categories; namely, original models, refitted models and new models. Last, but not least, I use different measures of four commonly used criteria in the literature; namely, calibration accuracy, information content, the correctness of categorical prediction, and discriminatory power, to evaluate the relative performance of models.

The remainder of this chapter is organised as follows. Section 3.2 provides a review of comparative studies related to competing statistical models. Section 3.3 explains the proposed multi-criteria methodology that used to compare the relative performance of competing distress prediction models. Section 3.4 presents the research methodology. Then, section 3.4 presents the empirical results and discussions. Finally, section 3.6 outlines the main conclusions of the chapter.

3.2 Existing Literature on Comparison of Competing Statistical Models

Since the existing literature on the comparative performance of distress prediction models is substantial, this section provides a review of the studies, which focus on comparisons of different types of statistical models; i.e., traditional statistical models, contingent claim analysis (CCA) models, and survival analysis (SA) models.

Panel I of Table 3.1 presents the comparison between traditional statistical models. From the introduction of univariate discriminant analysis (UDA) by Beaver (1966) through the early years of the 1980s, the multivariate discriminant analysis (MDA) was the superior method for predicting corporate failure. From the 1980s until 2001, the logit (introduced by Ohlson, 1980) and probit (introduced by Zmijewski, 1984) models dominated statistical techniques.

Panel II of Table 3.1 presents the comparison between traditional statistical models and SA model. Shumway (2001) proposed the breakthrough discrete-time hazard (DTH) model – using a multi-period logit framework – for distress prediction. In theory, SA models take advantage of their dynamic structure, and therefore outperform traditional statistical models, which are static in nature. However, in practice, the results of comparative studies indicate that the type of information that models fed with have a significant impact on the performance of models and could overcome the design shortcomings of static models (Shumway, 2001); therefore, static models should not be discarded entirely.

Panel III of Table 3.1 presents the comparison between statistical models and CCA models. Hilligeist et al. (2004) proposed a Black-Scholes-Merton (BSM) based model that outperforms two types of traditional statistical techniques; namely, logit and MDA.

Reisz and Perlich (2007) compared the performance of three CCA models; namely, BSM, KMV, and Down-and-Out Call option (DOC) based models. Further, Agarwal and Taffler (2008) compared the performance of two types of market-based models; namely, Hillegeist et al. (2004) and Bharath and Shumway (2008) and the MDA model of Taffler (1984). The comparison results indicate that CCA models outperform traditional statistical models under most measures of performance.

Panel IV of Table 3.1 shows the comparison between CCA and hazard models. Campbell et al. (2008) compared the performance of a CCA model; namely, KMV (Kealhofer, McQuown and Vasicek) and two types of hazard models; namely, Shumway (2001) and Campbell et al. (2008). The results indicate that their suggested hazard model outperforms both KMV and Shumway (2001) models.

Panel V of Table 3.1 presents the comparison between CCA, hazard and traditional statistical models. Wu et al. (2010) compared the performance of three frameworks of traditional statistical models; namely, MDA model of Altman (1968), logit model of Ohlson (1980), probit model of Zmijewski (1984) with DTH model of Shumway (2001) and BSM-based model of Hillegeist et al. (2004). Bauer and Agarwal (2004) compared the performance of traditional statistical, CCA and DTH models. The results of both studies suggest that DTH model outperforms other models. However, there are conflicts in the ranking of other models regarding different measures.

Table 3.1: Literature on Comparative Studies of Distress Prediction Models

Author	Models	Criteria (Measure)	Result
Panel I: Comparison between traditional statistical models			
Press and Wilson (1976)	LA and MDA models	Correctness of categorical prediction (T1 and T2 errors)	Two models unlikely will give significantly different results.
Collins and Green (1982)	LPA, MDA, and LA models	Correctness of categorical prediction (OCC, T1 and T2)	The models produce identical, uniformly results.
Lo (1986)	MDA and LA models	Power of models	There are no differences between models.
Theodossiou (1991)	LPA, LA, and PA models	Correctness of categorical prediction (T1 and T2 errors), Calibration (BS), Information content (pseudo-R^2)	logit model outperforms others; CONFLICT in the ranking of others on different measures
Lennox (1999)	LA, PA, and MDA models	Correctness of categorical prediction (T1 and T2)	A well-specified non-linear PA and LA are superior over DA
Bandyopadhyay (2006)	MDA models and logit models	Correctness of categorical prediction (OCC, T1 and T2) Discriminatory power (ROC), Information content (pseudo-R^2, LL)	CONFLICT in rankings using different criteria and measures
Tinoco and Wilson (2013)	logit models taking to account different categories of features	Discriminatory power (ROC, Gini, KS), Calibration accuracy (HL)	CONFLICT in rankings using different criteria and measures
Panel II: Comparison between traditional statistical models and survival analysis models			
Luoma and Laitinen (1991)	Cox-hazard, MDA and LA models	Correctness of categorical prediction (T1 and T2)	SA model is inferior to MDA and LA models
Shumway (2001)	Discrete-time SA model, MDA, LA and PA	Correctness of categorical prediction (OCC)	SA model, which encompasses both accounting and market information (respectively, only accounting information) outperforms (respectively, underperforms) other statistical techniques.

Panel III: Comparison between statistical models and contingent claim analysis (CCA) models			
Hilligeist et al. (2004)	BSM-based, LA and MDA models	Information content (LL and Pseudo-R^2)	BSM-based model outperforms both original and refitting version of LA and MDA models
Reisz and Perlich (2007)	BSM-based, KMV, DOC and MDA models	Discriminatory power (AUROC)	DOC and MDA outperforms others for 3-, 5- and 10-year ahead; MDA outperforms others for 1-year ahead distress prediction
Agarwal and Taffler (2008)	Contingent claim based models [HKCL (2004) and BHSH (2008)] and MDA model of Taffler (1984)	Discriminatory power (ROC), Information content (pseudo-R^2, LL), Correctness of categorical prediction (EV for different cost of misclassification)	MDA model outperforms HKCL (2004) on ROC and pseudo-R^2. CONVERSELY, HKCL (2004) outperforms BHSH (2008) and MDA model on LL.
Panel IV: Comparison between CCA models and survival analysis models			
Campbell et al. (2008)	A new duration dependent SA without time-variant baseline, SA model [Shumway (2001)] and KMV(Kealhofer, McQuown and Vasicek) model	Information content (pseudo-R^2, LL)	The suggested new SA model outperforms both Shumway (2001) and KMV models.
Panel V: Comparison between CCA, survival analysis and traditional statistical models			
Wu et al. (2010)	MDA [Altman (1968)], Logit model [Ohlson (1980)], probit model [Zmijewski (1984)] hazard model [Shumway (2001)] and BSM- model [HKCL (2004)]	Information content (pseudo-R^2, LL) Correctness of categorical prediction (OCC), Discriminatory power (ROC)	Shumway outperforms others on LL and Pseudo-R2. Logit model performs better that others on OCC. CONFLICT in rankings
Bauer and Agarwal (2014)	Traditional model, contingent claim based model and hazard model	Discriminatory power (ROC), Information content (LL, R^2) and Correctness of categorical prediction (OCC, T1, T2)	Hazard model outperforms others; CONFLICT in ranking of others on different measures

3.3 A Slacks-based CDEA Framework for Assessing Corporate Distress Predictions

In this chapter, I propose an orientation-free non-radial (slacks-based measure) context-dependent DEA (SBM-CDEA) framework for evaluating the relative performance of competing corporate DPMs. Hereafter, I first present the SBM-CDEA framework. Then, I discuss how one might adapt it to evaluate the relative performance of competing corporate DPMs.

Data envelopment analysis (DEA), proposed by Charnes, Cooper and Rhodes (1978), is a linear programming technique to assess the relative efficiency of a set of similar decision making units (DMUs), where each DMU is considered as a system, which uses multiple inputs to produce a number of outputs. The decision variables of these linear programming models are the weights allocated to inputs and outputs, and these models are referred to as multiplier models. The objective function value of the chosen DEA model – commonly referred to as a DEA score, allows one to classify a DMU as being efficient or not depending on whether its DEA score is equal to 1 or not, respectively. In DEA terminology, the set of efficient DMUs is referred to as the efficient frontier and represents the empirical standard of excellence against which benchmarking is done. Solutions to DEA models allow one to identify the reference set or peer group to use for benchmarking each DMU in seeking improvements. For detailed presentations of different DEA models, the reader is referred to Cooper et al. (2006).

Following the pioneering study by Mousavi et al. (2015) in using non-radial (slacks-based measure), orientation-free super-efficiency DEA to evaluate the performance of bankruptcy prediction models, I propose the non-radial (slacks-based measure), orientation-free context dependent DEA framework as a device for multi-criteria ranking of DPMs. I use an orientation-free evaluation because I intend to assess DPMs and thus the choice between input-oriented or output-oriented analysis is irrelevant. Further, input-oriented and output-oriented DEA studies may result in different scores and rankings of DMUs. On the other hand, I use a non-radial framework because the radial DEA models may be infeasible for some DMUs, which could result in having ties in rankings. The reason is that the super-efficiency DEA model was developed under (i) constant returns to scale (CRS) condition and (ii) the simultaneous and same proportion of change in all inputs (or outputs). Once any of these conditions is violated, it is high likely that infeasibility of the related DEA mode occurs (see, e.g., Seiford and Zhu, 1998a,b). As a result, we do not have a value associated with infeasibility to represent the super-efficiency, and the use of super-efficiency DEA is restricted. Furthermore, radial DEA models do not take account of possible excesses and shortfalls; namely, slacks, in inputs and outputs, respectively, which could result in over-estimating the efficiency scores due to ignoring mix efficiency. Finally, the reason to use context-dependent rather than super-efficiency scores to rank DMUs is that within the latter one, the scores are used to rank order the efficient DMUs; however, the efficient DMUs have different reference sets, which in some contexts could be considered as “unfair” benchmarking. On the other hand, within CDEA, a set of DMUs can be divided into different levels of efficient frontiers (evaluation context), and the attractiveness measure or the progress measure are used to rank those efficient DMUs belonging to the same specific evaluation context; that is, having the same level of efficiency or score. The proposed SBM-CDEA framework is summarised in the following stages:

Stage 1 – Returns-to-scale (RTS) Analysis: Perform RTS analysis to find out which type of RTS to include in DEA models; that is, constant-returns-to-scale (CRS), increasing returns-to-scale (IRS), decreasing return-to-scale (DRS), or variable returns-to-scale (VRS) – see Banker, Cooper, Thrall and Zhu (2004) for details. Note that depending on whether VRS, IRS or DRS conditions prevail, one must add $\sum_{j\in J^\lambda}\lambda_j = 1$, $\sum_{j\in J^\lambda}\lambda_j \geq 1$ or $\sum_{j\in J^\lambda}\lambda_j \leq 1$, respectively, as an additional constraint to the linear programming (LP) models 1, 2 and 3 below.

Stage 2 – Classification of DMUs: Use the following algorithm to identify several levels of efficiency or several efficient frontiers (evaluation contexts), say L:

- Step 1: Set the performance level counter, say λ, equal to 1. Let $\mathrm{J}^\lambda = \{DMU_k, k = 1, \dots, n\}$ be the set of all n DMUs at efficiency level λ. Evaluate the entire set of DMUs, J^λ, by solving the relevant DEA model to construct the λ-level efficient frontier, say E^λ, where $E^\lambda = \left\{k \in J^\lambda \middle| DEA\ score\ \rho_k^\lambda = 1\right\}$.
- Step 2: Drop the current efficient DMUs; that is, E^λ, from the next DEA

analysis run; that is, $J^{\lambda+1} = J^{\lambda} - E^{\lambda}$, and increase the counter λ by 1.

- Step 3: Evaluate the "globally inefficient" set of DMUs identified in the previous step; this is, $J^{\lambda+1}$, by solving the relevant DEA model and set the λ+1-level efficiency frontier to $E^{\lambda+1}$.

- Step 4: If $J^{\lambda+1} = \emptyset$, then stop; otherwise, set $\lambda = \lambda + 1$ and go to step 2.

where the relevant DEA model to determine the λ^{th} −level efficiency frontier is the slacks-based measure (SBM) model of Tone (2001):

$$Min \quad p_k^{\lambda} = \left(1 - \frac{1}{m}\sum_{i=1}^{m}\frac{s_{i,k}^{-}}{x_{i,k}}\right) \Big/ \left(1 + \frac{1}{s}\sum_{r=1}^{s}\frac{s_{r,k}^{+}}{y_{r,k}}\right) \qquad \text{Eq. 3-1}$$

$$s.t.: \quad \sum_{j\in J^{\lambda}} \lambda_j x_{i,j} + s_{i,k}^{-} = x_{i,k}; \forall_i$$

$$\sum_{j\in J^{\lambda}} \lambda_j y_{r,j} - s_{r,k}^{+} = y_{r,k}; \forall_r$$

$$\lambda_i \geq 0; \forall_j \in J^{\lambda}; s_{i,k}^{-} \geq 0, \forall_i; s_{r,k}^{+} \geq 0, \forall_r$$

where the $x_{i,j}$ $(i = 1, \dots, m)$ and $y_{r,j}$ $(r = 1, \dots, s)$ are the i^{th} input and the r^{th} output of DMU_k $(k = 1, \dots, n)$, respectively, λ_k is the weight allocated to DMU_k in constructing its ideal benchmark, $s_{i,k}^{-} \in \mathbb{R}^m$ and $s_{r,k}^{+} \in \mathbb{R}^s$ denote the slacks of the first and second constrains; that is, input excesses and output shortfalls, and ρ_k^{λ} is the SBM efficiency score of DMU_k with respect to evaluation context λ. In the case that the optimal value of $\rho_k^{\lambda} = 1$, then DMU_k is part of λ-level efficient frontier; otherwise DMU_k is inefficient and will be evaluated in future DEA runs. Obviously, DMUs are partitioned into L efficient frontiers, which indicate different performance levels. One could rank order DMUs considering the 1st-level efficient frontier DMUs as best and the L^{th} -level efficient frontier DMUs as worst, however, ties exist between DMUs on the same level efficient frontier and the next stage is designed to break those ties.

Stage 3 – Breaking of Efficiency Ties: Perform the following steps to break the ties between DMUs in the same level efficient frontier:

- Step 1: solve the LP (2) for all DMUs obtained at performance level λ, say $DMU_k \in E^{\lambda}$, where $\lambda = 2, 3, \dots, L$, to compute *relative progress scores*, $\delta_k^1 s$, with reference to the best evaluation context, E^1, and rank DMUs on efficient frontier E^{λ} based on the calculated scores:

$$Min\ \delta_k^1 = \left(1 - \frac{1}{m}\sum_{i=1}^{m}\frac{t_{i,k}^-}{x_{i,k}}\right) \Big/ \left(1 + \frac{1}{s}\sum_{r=1}^{s}\frac{t_{r,k}^+}{y_{r,k}}\right) \qquad \text{Eq. 3-2}$$

$$\begin{aligned} s.t.: \quad & \textstyle\sum_{j\in E^1}\lambda_j x_{i,j} \geq x_{i,k} - t_{i,k}^-; \forall_i \\ & \textstyle\sum_{j\in E^1}\lambda_j y_{r,j} \leq y_{r,k} + t_{r,k}^+; \forall_r \\ & \lambda_i \geq 0; \forall_j \in E^1; t_{i,k}^- \geq 0, \forall_i; t_{r,k}^+ \geq 0, \forall_r \end{aligned}$$

where $t_{i,k}^-$ (respectively, $t_{r,k}^-$) indicates the amount by which input i (respectively, output r) of DMU_k should be decreased (respectively, increased) to reach the evaluation context E^1 (see, Figure 3.1).

- Step 2: solve the LP (3) for all DMUs obtained at the best efficient frontier E^1; that is, $DMU_k \in E^1$, to compute *relative attractiveness scores*, $\gamma_k^2 s$, with reference to the second-best evaluation context, E^2, and rank DMUs on the best efficient frontier E^1, based on the calculated scores (see, Figure 3.1).

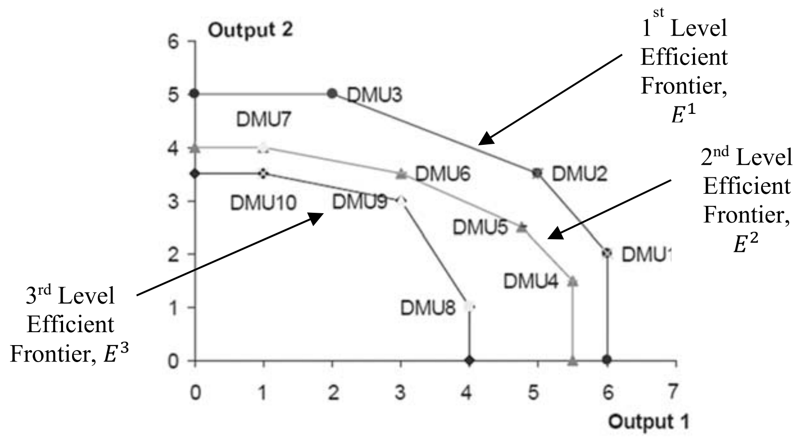
$$Max\ \ \gamma_k^2 = \left(1 - \frac{1}{m}\sum_{i=1}^{m}\frac{t_{i,k}^+}{x_{i,k}}\right) \Big/ \left(1 + \frac{1}{s}\sum_{r=1}^{s}\frac{t_{r,k}^-}{y_{r,k}}\right) \qquad \text{Eq. 3-3}$$

$$\begin{aligned} s.t.: \quad & \textstyle\sum_{j\in E^2}\lambda_j x_{i,j} \leq x_{i,k} + t_{i,k}^+;\ \forall_i \\ & \textstyle\sum_{j\in E^2}\lambda_j y_{r,j} \geq y_{r,k} - t_{r,k}^-;\ \forall_r \\ & \lambda_i \geq 0; \forall_j \in E^2; t_{i,k}^- \geq 0, \forall_i; t_{r,k}^+ \geq 0, \forall_r \end{aligned}$$

where $t_{i,k}^+$ (respectively, $t_{r,k}^-$) indicates the amount by which input i (respectively, output r) of $DMU_k \in E^1$ should be increased (respectively, decreased) to reach to evaluation context E^2 (see, Figure 3.1).

In the next section, I shall use the above-described methodology to rank order competing corporate distress prediction models and discuss the empirical results obtained using UK data on firms listed on the London Stock Exchange (LSE) for the period 2008-2014. In this chapter, DMUs are thirty competing corporate distress prediction models – see Appendix 3-A for a general description of these models. The inputs and outputs are the performance measures of the relevant criteria for assessing corporate prediction models. This study considers discriminatory power, calibration accuracy, information content, and correctness of categorical predictions criteria and their measures. Further, inputs (respectively, outputs) are selected based on the rule of the less (respectively, the more), the better; therefore, inputs (respectively, outputs) refer to the performance measures to be minimised (respectively, maximised).

Figure 3-1: Context Dependent DEA



3.4 Empirical Investigation

This section provides the details of my research methodology, where I compare the performance of competing DPMs using both mono-criterion and multi-criteria performance evaluation frameworks. Hereafter, I provide the details on my dataset (see, section 3.4.1), features selection (see, section 3.4.2), model evaluation criteria and measures (see, section 3.4.3), and distress prediction models (see, section 3.4.4).

3.4.1 Data

I took the following steps to select my dataset. First, I considered all non-financial and non-utility UK companies listed on the London Stock Exchange (LSE) at any time during an 8-year period from 2007 through 2014. Second, I excluded the firms which are listed less than two years in LSE, as historical information is the requirement for some modelling frameworks. Third, I excluded the firms with missing values for the main accounting information (e.g., sales, total assets) and market information (e.g., price), which are essential items for calculating many financial ratios (Lyandres and Zhdanov, 2013).

I replaced the remaining missing values with the recently observed ones for each firm (Zhou, 2013; Shumway, 2001). Fourth, for each variable, I winsorised the outliers by replacing the values higher (respectively, lower) than 99th (respectively, 1st) percentile with that 99th (respectively, 1st) percentile value (Shumway, 2001).

Table 3.2: Sample Sizes

Samples	Year	Healthy	Distressed	Total	Distress rate
Training sample (2007 - 2010)	2007	1,826	81	1,907	4.25%
	2008	1,704	106	1,810	5.86%
	2009	1,456	165	1,621	10.18%
	2010	1,409	61	1,470	4.15%
	Total	*6,395*	*413*	*6,808*	*6.07%*
Hold out sample (2011-2014)	2011	1,354	27	1,381	1.96%
	2012	1,255	69	1,324	5.21%
	2013	1,143	101	1,244	8.12%
	2014	1,120	66	1,186	5.56%
	Total	*4,872*	*263*	*5,135*	*5.12%*
Total		**11,267**	**676**	**11,943**	**5.66%**

With respect to the classification of firms into distressed and non-distressed, I followed the definition of Pindado et al. (2008) where a company is classified as distressed if it experiences both of the following conditions for two consecutive years: (1) its earnings before interest, taxes, depreciation and amortization (EBITDA) is lower than its interest expenses, and (2) it shows negative growth in market value. To be more specific, the distress variable, say y, equals 1 for financially distressed companies and equals 0 otherwise. In sum, my dataset consists of 2,096 firms and 11,943 firm-year observations. Among the total number of observations, 676 firm-year observations are classified as distressed, which result in the average distress rate of 5.66 percent per year. The models are developed using training sample period ranging from 2007 to 2010 and tested using holdout sample period ranging from 2011 to 2014. Table 3.2 presents my sample sizes.

3.4.2 Feature Selection

To select proper features for prediction models, I applied the following steps. First, I reviewed the literature to select the most commonly used features in other studies (e.g., Hebb, 2016; du Jardin, 2015; Zhou, 2015, 2013; Ravi Kumar and Ravi, 2007), where I end up with 83 accounting-based ratios and seven market-based items. Second, I used *t-tests* to choose features which show a significant difference between the means of two groups of distressed and healthy firms (Shin and Lee, 2002; Huang et al., 2004; Shin et al., 2005). Third, for further reduction of features, I applied factor analysis, and principal component analysis with VARIMAX technique (Chen, 2011, Mousavi et al., 2015). To be more specific, I used factors analysis to select the variables that both the absolute values of their loadings and communities are greater than 0.5 and 0.8, respectively. Finally, 34 variables, which presented high factor loadings and high communality values, were retained as input features into a stepwise procedure in each statistical framework.

3.4.3 Corporate Distress Models to be Assessed

In this chapter, I compare the performance of the most cited statistical, probability and stochastic models in the literature on distress prediction. To be more specific, I consider the MDA models proposed by Altman (1968), Altman (1983), Lis (1972), and Taffler (1984); the logit model proposed by Ohlson (1980); the probit model proposed by Zmijewski (1984); and the linear probability model proposed by Theodossiou (1991); the contingent claim analysis models proposed by Bharath and Shumway (2008), Hillegeist et al. (2004) and Jackson and Wood (2013); and the survival analysis model proposed by Shumway (2001).

I also included in my comparative analysis two additional categories of models that I refer to as refitted original models and new models. In the case of refitted models, I keep the explanatory variables of each original model and refit them with my new dataset. On the other hand, in the case of new models, I develop new distress prediction models using different static and dynamic frameworks and fit them with my new dataset. The static frameworks used to develop new models are MDA, logit, probit and linear probability analysis. The dynamic frameworks used to develop new models are duration-independent with (or without) time-independent baseline hazard rate, and different duration-dependent models, which contain a variety of time-varying baseline hazard rates.

Note that depending on the existence and specification of baseline hazard rate in dynamic or duration models, one could classify them into two subcategories; namely, duration-independent and duration-dependent frameworks (Nam et al., 2008). The duration independent models could be classified into duration-independent with time-independent baseline (DIWTIB) and duration independent without baseline (DIWOB). The difference between these two types of models is that the former one contains a constant baseline hazard rate, while the latter one does not contain baseline hazard rate (e.g., Shumway, 2001). On the other hand, the duration-dependent framework contains a time-dependent baseline hazard rate, as mentioned in Beck et al. (1998), who use time dummies to proxy the baseline hazard rate.

Table 3.3: List of Financial Ratios

Category	Ratio or item	Category	Ratio or item
Profitability (9)	Net income to total liabilities EBIT to total assets Return on assets Operating income after depreciation to total assets Retained earnings to total assets Expected return on assets Total liabilities exceed total assets Changes in net income in two consecutive years Negative net income for last two years	Liquidity (9)	Current asset turnover Current assets to total liabilities Current liabilities to current assets Inventory to current assets Inventory turnover Inventory to total assets Profit before tax to current liabilities Quick asset to total assets Quick asset to inventory
Asset utilisation (2)	Asset turnover ratio Quick assets to sales	Solvency (3)	Current liabilities to liabilities Equity to capital Long term and current liabilities to total assets
Cash flow (2)	Operating cash flow to liabilities Funds Provided by Operations to Total Liabilities	Market information (5)	Lag of excess return Lag sigma Ln (price) Real size Failure rate in last year
Mixed (2)	GDP $\times$ Sales Interest rate $\times$ Income	Firm characteristics (2)	Ln(age) Log (total assets to GNP price level index)

Since the use of time dummies as an indirect proxy for the baseline rate is less efficient, I follow Nam et al. (2008) and Gupta et al. (2015) in using time-varying features to proxy the time-dependent baseline rate. Therefore, taking into account the duration dependent (DD) framework, the models differ based on the type of baseline hazard rates, i.e., ln (age), 1/ln(age), last year probability of distress (LPD), and volatility of exchange rate (VEX) – see Appendix 3-A for more details on models.

Table 3.3 and Table 3.4 present the list of features used to develop new models, and the new models, respectively. In sum, 30 models are assessed in my comparative analysis. Note that all chosen models are tested out-of-sample and the training period ranges from 2007 to 2010 including 413 distressed and 6,395 non-distressed firm-year observations.

Table 3.4: New Designed Models

The table presents the features and coefficients of the new models, namely MDA (20), LPA (21), LA (22), PA (23), DIWOB (24), DIWTIB_ln(age) (25), DD_ln(age) (26), DD_VEX (27), DD_LPD (28), DDWTIB_1/ln(age) (29), DD_1/ln(age) (30). *** and ** refer to 1% and 5% significance level, respectively.

Models	Model 20	Model 21	Model 22	Model 23	Model 24	Model 25	Model 26	Model 27	Model 28	Model 29*	Model 30
Explanatory variables											
Intercept	17.11	1.18***	0.62	-0.54***	0.77	0.22+ln(age)$_i$	0.22	1.78	0.74	0.77+1/ln(age)$_i$	2.68
INTWO	1.86	0.09***	1.59***	0.78***	1.53***	1.62***	1.62***	1.53***	1.53***	1.53***	1.60***
Lag of excess return	-4.67	-0.22***	-3.36***	-1.84***	-3.31***	-3.77***	-3.77***	-3.35***	-3.34***	-3.31***	-3.95***
Current Assets to Total Liabilities	-16.69	-0.79***									
Current Liabilities to Liabilities	21.19	1.03***									
Net worth over total liabilities			-1.38***	-0.55***	-1.47***	-2.02***	-2.02***	-1.54***	-1.53***	-1.47***	-1.97***
Retained Earnings to Total Assets	-12.98	-0.74***									
Equity to Capital	-10.87	-0.54***									
CHIN	0.48	0.02***									
Real size	-1.81	-0.14***	-2.78***	-1.49***	-1.97***	-1.89***	-1.89***	-1.90***	-1.91***	-1.97***	-1.93***
Asset Turnover Ratio	-5.97	-0.29***	-14.82***	-8.61***	-14.16***	-20.54***	-20.54***	-13.99***	-14.03***	-14.16***	-20.69***
Inventory Turnover	5.14	0.25***	3.34***	1.77***	3.34***	3.30***	3.30***	3.31***	3.30***	3.34***	3.21***
Interest rate × Net Income			-12.47***		-12.58***	-15.78***	-15.78***	-13.19***	-13.15***	-12.58***	-16.66***
log (price)	0.14	0.007***	0.11***	0.05***							
GDP × Sales	1.44	0.05***									
Ohlson size	-1.26										
1/ln (age)											-6.17***
Last period distress rate (LPD)									1.9864		
Volatility of exchange rate (VEX)								-0.2981			
ln (age)						2.52***	2.52***				

* Note that in models 25 and 29, the ln(age) and 1/ln(age) of firm i is added to intercept as the baseline hazard rate.

In the next section, I 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 section 3.5.2).

3.4.4 Performance Criteria and Measures

The objective of this study is to evaluate the relative performance of distress prediction models using UK data. For this, I follow Mousavi et al. (2015) to assess the performance of different models under four commonly used criteria; namely the discriminatory power, the calibration accuracy, the information content, and the correctness of categorical prediction. On the discriminatory power criterion, I use Receivable Operating Characteristic (ROC), Kolmogorov-Smirnov (KS) statistics, Gini Index (GI), and Information Value (IV) to measure how much a model is capable of discriminating between the distressed firms and the healthy ones. On the calibration accuracy criterion, I use Brier Score (BS) to measure how much a model is qualified in estimating the probability of distress (PD). On the information content criterion, I follow Agarwal and Taffler (2008) and use a log-likelihood statistic (LL) and pseudo-R^2 to measure the extent to which the output of a model (e.g., PD, scores) carries enough information for prediction. Finally, with respect to the correctness of categorical prediction criterion, I use Type I errors (T1), Type II errors (T2), misclassification rate (MR), sensitivity (Sen), specificity (Spe), and overall correct classification (OCC) to measure how often a model can predict distressed firms (respectively, healthy firms) as distressed (respectively, healthy) ones.

3.5 Empirical Results

In this section, I organise the analyses into the mono-criterion analysis (see section 3.5.1) and multi-criteria analysis (see section 3.5.2) and summarise main findings. The original models, refitted original models, and new models are presented in Figure 3-2, Figure 3-3, and Figure 3-4 with white, grey, and black shapes, respectively, and static and dynamic models with a circle and non-circle shapes, respectively – see legends of Figure 3-2, Figure 3-3, and Figure 3-4.

3.5.1 Mono-Criterion Analysis

Figure 3-2 presents the mono-criterion (unidimensional) rankings of the 30 models using several measures under four commonly used criteria. For my dataset, mono-criterion rankings results could be summarised as follows. First, regarding the performance of all competing models in my study, the new developed models outperform original models and refitted models. This finding suggests that the change in trend of information during time, as someone would expect, tend to affect the performance of corporate distress prediction models; therefore, out-dated original models or refitted original models with new dataset do not seem to be as efficient as new models with respect to most of the performance measures. The only exceptions are original models 1, 2 and 6 (Altman, 1968; Lis, 1972; and Ohlson, 1990, respectively), which perform amongst the best regarding T1 and Sen as performance measures.

Second, relating to the comparison of new dynamic models and new static models in my study, for most of the performance measures, the new dynamic models outperform static ones. To be more specific, on most performance measures – see, for example, T1, Sen, AUC, Gini and KS, new dynamic models 27 and 28 (DD_VEX and DD_LPD, respectively) are superior to other models. However, considering the performance measures KS, IV, and BS, new static models 21 and 23 (New LPA and New PA, respectively) are amongst the best performers. In general, the density of new dynamic models amongst the top-ranking performers is an indicator of their superiority.

Third, contingent claim analysis (CCA) models (models 16, 17 and 18) are not amongst the best performers. The only exception is model 17 (Hillegeist et al., 2004), which is ranked second under T2, MR, OCC, and Spe; however, model 18 (Bharath and Shumway, 2008) seems to outperform other CCA models for most performance measures. Fourth, regarding the performance of the original MDA models refitted (i.e., models 9, 10, 11 and 12), for most performance measures, models 9 and 11 (Altman, 1968 and Altman, 1983, respectively) outperform others. Also, amongst the refitted regression models (i.e., models 13, 14 and 15), for most performance measures, the logit model 14 (refitted Ohlson, 1990) outperforms others.

In general, considering refitted models, the logit model 14 (Ohlson, 1990) and MDA model 10 (Altman, 1968) outperform other refitted models, for most performance measures. Last, but not least, regarding the out-of-sample performance of the original models, static models 6, 2 and 1 (Ohlson, 1990; Lis, 1972 and Altman, 1968, respectively) seem to outperform other original models including the dynamic model 8 (Shumway, 2001). This result suggests that the static models mentioned above have more stability over time. Note that the results of stability index confirm this finding – see, for example, stability index in Figure 3-2.

Much like typical outcomes in the existing literature, the rankings under mono-criterion are facing two main issues. Firstly, the rankings of models are different not only for measures under different criteria – see, for example, T1 under correctness of categorical prediction criterion and ROC under discriminatory power criterion, but also for measures under the same criterion; see, for example, OCC and MR under correctness of categorical prediction criterion or KS and ROC under discriminatory power criterion – as it is the case in Theodossiou (1991), Bandyopadhyay (2006), and Tinoco and Wilson (2013). Secondly, the models' rankings tend to have ties corresponding to some measures – see, for example, measures of T1 and Sen. Consequently, practitioners cannot make an informed decision about the best distress prediction model. To overcome these issues, I propose a multi-criteria ranking framework, namely SBM-CDEA, which not only provides a single ranking using multiple criteria at the same time but also breaks the possible ties in the ranking of competing models.

3.5.2 Multi-Criteria Analysis

Figure 3.4 presents the multi-criteria (multidimensional) rankings of the mentioned 30 models using SBM-CDEA. Further, Table 3.5 provides the efficient frontiers obtained with SBM-CDEA. Also, following Mousavi et al. (2015), I provide the rankings of models using SBM-super efficiency DEA, see, Figure 3.3, to compare the performance of two multi-criteria assessment frameworks.

In my empirical investigation, RTS analysis revealed that VRS conditions hold and therefore an additional constraint (i.e., $\sum_{j\in J^{\lambda}} \lambda_j = 1$) needs to be added to linear programming models 1, 2 and 3. In addition, for my dataset, multi-criteria rankings under SBM-super efficiency DEA and SBM-CDEA show considerable consistency in the rankings of top five models, for most of combinations of measures; however, they do not provide a general consistency in the rankings of all models.

Furthermore, under SBM-CDEA, the results could be summarised as follows. First, on the performance of all competing models in my study, the new developed models outperform the original models and the original models refitted. Contrary to the mono-criterion ranking, multi-criteria ranking indicates the superiority of dynamic models, for all combinations of performance measures.

Second, on the performance of dynamic and static models in my study, for most of the combinations of measures, the dynamic models outperform static ones. To be more accurate, regardless of the combinations of measures, the dynamic models 25, 30 (DIWTIB_ln (age) and DD_1/ln (age), respectively) followed by models 24 and 27 (DIWOB and DD_VEX, respectively) are always amongst the top five best performers. The exceptional performance of the dynamic models seems to suggest that taking account of the time-varying nature of predictors pays off. However, for all combinations of measures, static models 21 and 23 (new LPA and new PA, respectively) are superior to other models.

Third, with respect to CCA models – which are systematically amongst the worst ranked models, model 17 (Hillegeist et al., 2004) outperforms model 18 (Bharath and Shumway, 2008) and model 19 (Jackson and Wood, 2013), for combinations of measures, which include T1 and OCC; conversely, model 18 outperforms, for combinations of measures, which include T2.

Fourth, amongst the original MDA models refitted (i.e., models 9, 10, 11 and 12), for most of the combinations of measures, model 9 (refitted Altman, 1968) outperforms others. Further, amongst the refitted regression models (i.e., models 13, 14 and 15), for most performance measures, model 14 (refitted Ohlson, 1990) outperforms others. Further, the logit model 14 seems to be the best performer amongst all refitted models. Finally, considering the performance of original models, the static models 6, 2 and 1 (Ohlson, 1990; Lis, 1972 and Altman, 1968, respectively) followed by the dynamic model 8 (Shumway, 2001) are amongst the best performers.

The empirical findings reveal that the multi-criteria rankings differ from the mono-criterion ones. Further, the multi-criteria rankings have no ties as compared to the mono-criterion ones, which suggests that the choice of the SBM-CDEA framework is an effective one to overcome ties in the ranking of corporate distress prediction models. Furthermore, I have considered several combinations of measures of the performance criteria under consideration to find out about the robustness of the multi-criteria rankings on the choice of measures.

3.6 Conclusion

Prediction of corporate distress and bankruptcy is one of the most crucial inputs to decisions making processes related to financing and investing actives. During recent decades, academics and practitioners have developed a large number of distress prediction models, which raise the question that "which of these models performs better in predicting distress?" To answer this question, the unidimensional ranking of competing prediction models has been the dominant approach; however, it results in conflicting rankings of models once someone shifts from one performance criteria to another. Mousavi et al. (2015) proposed a multi-criteria evaluation framework; namely, an orientation free super-efficiency DEA-based framework, to evaluate the performance of different bankruptcy prediction models, which provides a single ranking based on multiple performance criteria; such a framework faces one main issue, which is referred to as "fair benchmarking" that was overcome in this chapter. This study proposes an orientation-free slack-based context dependent DEA framework to overcome the methodological issues of both super-efficiency DEA-based and unidimensional ranking. Furthermore, I performed an exhaustive comparative analysis of the most popular distress modelling frameworks resulting in 30 prediction models including new models organised into three categories; namely, original models, original models refitted, and new models. I used several measures under four commonly used criteria, which are often employed in the literature to compare the performance of prediction models using a UK dataset on firms listed on the London Stock Exchange.

The major conclusions could be summarised as follows. First, although the rankings of distress prediction models under orientation-free SBM-super efficiency and orientation-free SBM-CDEA are very similar, the latter one, however, does not suffer from the changes of reference benchmark from one DMU or prediction model to another. Second, the numerical results reveal that amongst the duration models, which are always superior in performance, duration dependent models DD_VEX and

DD_1/ln(age) that use volatility of exchange rate (VEX) and 1/ln(age) as time-varying baseline, respectively, followed by duration independent without baseline (DIWOB) tend to be superior. Third, numerical results seem to suggest that amongst the static models, LPA and PA models outperform others. Last, but not least, my empirical results suggest that developing new models using the most recent accounting, market, and macroeconomic information enhances the performance of distress prediction models.

This research has some limitations. First, because of time constraint, this study focuses on statistical distress prediction models. Second, because of limited access to data, in this research I have focused on listed UK companies. Third, due to time constraint, in this study I focused on financial distress and the main failure event. Future studies could take into account non-statistical prediction models. Also, future research could choose other failure events such as bankruptcy, debt restructure, etc. Further, using data from different countries, the extent to which these models are globalized could be tested.

Chapter 2 and 3 are restricted to static multi-criteria assessment frameworks. The next chapter employs dynamic multi-criteria assessment to evaluate the performance of models over time.

Figure 3-2: Mono-Criterion Rankings of Corporate Distress Prediction Models

This table presents the mono-criterion rankings of 30 competing corporate distress prediction models, where models are ranked from best to worst using a single measure of a single criterion at a time. T1 (type I error), T2 (type II error), MR (misclassification rate), Sen (sensitivity), Spe (specificity) and OCC (overall correct classification) are used as measures of correctness of categorical prediction; AUC (area under receiver operating character), Gini coefficient, KS (Kolmogorov Smirnov) and IV (information value) are used as measures of discriminatory power; BS (Brier score) is used as a measure of calibration accuracy; and log-likelihood (LL) and Pseudo-R^2 (R^2) are used as measures of information content. Circle shapes represent static models, namely Multivariate Discriminant Analysis (MDA), Linear Probability (LPA), Logit Analysis (LA), and Probit Analysis (PA). Non-circle shapes represent dynamic models, namely duration models, and Contingent Claim Analysis (CCA) models. White, grey, and black coloured shapes represent the original models, the original models refitted, and the new models, respectively.

Measure	Rank from the Best to Worst
T1; Sen	27, 1, 2, 6, 28, 23, 20, 9, 11, 22, 14, 10, 29, 3, 30, 26, 25, 24, 16, 21, 13, 5, 15, 8, 18, 7, 12, 19, 4, 17
T2; MR; OCC; Spe	24, 17, 21, 25, 26, 22, 29, 30, 28, 23, 20, 27, 16, 15, 8, 13, 5, 18, 11, 9, 14, 3, 1, 2, 19, 10, 4, 12, 6, 7
AUC; Gini	27, 28, 24, 25, 29, 23, 30, 22, 26, 21, 20, 16, 10, 2, 14, 9, 1, 11, 3, 6, 8, 13, 18, 15, 5, 12, 19, 4, 17, 7
KS	23, 27, 28, 25, 24, 21, 29, 30, 22, 26, 20, 10, 14, 16, 2, 9, 11, 1, 3, 6, 8, 15, 13, 5, 18, 17, 12, 19, 4, 7
IV	23, 27, 25, 29, 28, 24, 22, 30, 26, 21, 20, 14, 16, 18, 19, 9, 11, 10, 6, 15, 13, 5, 3, 2, 8, 1, 12, 4, 17, 7
BS	21, 23, 25, 27, 29, 28, 24, 30, 14, 26, 22, 20, 16, 6, 11, 9, 10, 8, 2, 3, 1, 4, 12, 13, 15, 7, 5, 18, 19, 17
LL, Pseudo-R^2	30, 26, 25, 22, 23, 27, 28, 24, 29, 20, 21, 16, 14, 18, 8, 17, 19, 6, 9, 11, 3, 1, 10, 5, 15, 2, 13, 12, 4, 7
Stability index	23, 16, 22, 29, 27, 28, 18, 24, 10, 20, 26, 11, 9, 25, 30, 6, 17, 21, 15, 2, 12, 19, 14, 5, 13, 3, 1, 8, 7, 4

[1,9] Altman (1968); [2,10] Lis (1972); [3,11] Altman (1983); [4,12] Taffler (1984); [5,13] Theodossiou (1991); [6,14] Ohlson (1990); [7,15] Zmijewski (1984); [8,16] Shumway (2001); [17] Hillegeist et al. (2004); [18] Bharath and Shumway (2008); [19] Jackson and Wood (2013); [20] New MDA; [21] New LPA; [22] New LA; [23] New PA; [24] DIWOB; [25] DIWTIB_ln(age); [26] DD_ln(age); [27] DD_VEX; [28] DD_LPD; [29] DIWTIB_1/ln(age); [30] DD_1/ln(age)

Figure 3-3: SBM-Super Efficiency DEA-based Multi-Criteria Rankings of Corporate Distress Prediction Models

This table presents the multi-criteria rankings of 30 competing corporate distress models using a DEA ranking framework, where models are ranked from best to worst using SBM-super efficiency scores. A multi-criteria ranking is produced for each combination of a variety of metrics of the performance criteria under consideration, where inputs (resp. outputs) are chosen according to the principle of the less (resp. more) the better. T1 (type I error), T2 (type II error), MR (misclassification rate), Sen (sensitivity), Spe (specificity) and OCC (overall correct classification) are used as measures of correctness of categorical prediction; ROC (area under receiver operating character), Gini coefficient, KS (Kolmogorov Smirnov) and IV (information value) are used as measures of discriminatory power; BS (Brier score) is used as a measure of calibration accuracy; and log-likelihood (LL) and Pseudo-R^2 (R^2) are used as measures of information content. Circle shapes represent static models, namely Multivariate Discriminant Analysis (MDA), Linear Probability (LPA), Logit Analysis (LA), and Probit Analysis (PA). Non-circle shapes represent dynamic models, namely duration models, and Contingent Claim Analysis (CCA) models. White, grey, and black coloured shapes represent the original models, the original models refitted, and the new models, respectively.

Inputs	Outputs	Rank from the Best to Worst
T1; BS; LL	ROC	27–30–23–25–21–26–28–22–20–29–24–14–16–10–2–1–9–11–6–3–8–13–15–5–18–12–19–4–17–7
T2; BS; LL	ROC	24–21–30–25–27–23–28–26–22–29–20–16–14–2–9–10–11–1–3–8–15–13–6–5–18–17–12–19–4–7
BS; LL	ROC; OCC	21–30–24–25–23–27–28–26–22–29–20–16–14–9–11–2–1–8–10–3–15–13–6–5–18–12–17–19–4–7
T1; BS; LL	KS	27–23–30–21–25–26–28–22–20–29–24–14–16–10–2–1–9–11–3–6–8–15–13–5–18–17–12–19–7–4
T2; BS; LL	KS	24–21–23–25–30–28–26–27–22–29–20–16–14–11–10–9–2–1–3–8–15–13–6–5–18–17–12–19–4–7
T1; BS	ROC; R^2	27–30–23–21–25–26–28–22–20–29–24–14–16–18–8–6–17–19–9–11–1–3–10–2–15–5–13–12–4–7
T2; BS	ROC; R^2	24–30–21–25–23–27–28–26–22–29–20–16–14–18–17–8–19–6–9–11–3–1–10–15–5–2–13–12–4–7
BS	ROC; OCC; R^2	21–30–25–24–23–27–28–26–22–29–20–16–14–18–8–17–6–19–9–11–3–1–10–15–5–2–13–12–4–7
T1; BS	KS; R^2	27–30–23–21–25–26–28–22–20–29–24–14–16–18–8–6–17–9–11–1–19–3–10–2–15–5–13–12–4–7
T2; BS	KS; R^2	24–30–21–25–23–28–26–22–27–29–20–16–14–18–17–8–6–19–9–11–3–1–10–15–5–2–13–12–4–7

[1,9] Altman (1968); [2,10] Lis (1972); [3,11] Altman (1983); [4,12] Taffler (1984); [5,13] Theodossiou (1991); [6,14] Ohlson (1990); [7,15] Zmijewski (1984); [8,16] Shumway (2001); [17] Hillegeist et al. (2004); [18] Bharath and Shumway (2008); [19] Jackson and Wood (2013); [20] New MDA; [21] New LPA; [22] New LA; [23] New PA; [24] DIWOB; [25] DIWTIB_ln(age); [26] DD_ln(age); [27] DD_VEX; [28] DD_LPD; [29] DIWTIB_1/ln(age); [30] DD_1/ln(age)

Figure 3-4: SBM-Context Dependent DEA-based Multi-Criteria Rankings of Corporate Distress Prediction Models

This table presents the multi-criteria rankings of 30 competing corporate distress models using a DEA ranking framework, where models are ranked from best to worst using SBM-CDEA scores. A multi-criteria ranking is produced for each combination of a variety of metrics of the performance criteria under consideration, where inputs (resp. outputs) are chosen according to the principle of the less (resp. more) the better. T1 (type I error), T2 (type II error), MR (misclassification rate), Sen (sensitivity), Spe (specificity) and OCC (overall correct classification) are used as measures of correctness of categorical prediction; ROC (area under receiver operating character), Gini coefficient, KS (Kolmogorov Smirnov) and IV (information value) are used as measures of discriminatory power; BS (Brier score) is used as a measure of calibration accuracy; and log-likelihood (LL) and Pseudo-R^2 (R^2) are used as measures of information content. Circle and non-circle shapes indicate static and dynamic frameworks, respectively. Black, grey and white shapes represent new models, BSM-based models, and original models refitted, respectively.

Inputs	Outputs	Rank from the Best to Worst
T1; BS; LL	ROC	27 – 21 – 23 – 30 – 25 – 26 – 28 – 29 – 22 – 20 – 24 – 14 – 1 – 6 – 2 – 16 – 10 – 9 – 11 – 8 – 18 – 3 – 19 – 17 – 13 – 15 – 12 – 4 – 5 – 7
T2; BS; LL	ROC	21 – 30 – 28 – 27 – 25 – 24 – 23 – 26 – 22 – 29 – 17 – 20 – 14 – 16 – 2 – 9 – 10 – 11 – 8 – 15 – 6 – 18 – 1 – 3 – 13 – 19 – 5 – 12 – 4 – 7
BS; LL	ROC; OCC	21 – 24 – 25 – 23 – 30 – 27 – 28 – 26 – 22 – 29 – 17 – 20 – 14 – 16 – 9 – 11 – 2 – 8 – 10 – 15 – 6 – 18 – 1 – 3 – 13 – 19 – 5 – 12 – 4 – 7
T1; BS; LL	KS	27 – 21 – 23 – 25 – 30 – 26 – 28 – 29 – 22 – 20 – 24 – 14 – 1 – 6 – 2 – 10 – 9 – 16 – 11 – 8 – 18 – 3 – 17 – 19 – 15 – 13 – 12 – 4 – 5 – 7
T2; BS; LL	KS	24 – 21 – 25 – 23 – 30 – 28 – 26 – 27 – 29 – 17 – 22 – 14 – 20 – 16 – 10 – 11 – 9 – 2 – 8 – 15 – 6 – 18 – 1 – 3 – 13 – 19 – 5 – 12 – 4 – 7
T1; BS	ROC; R^2	27 – 30 – 21 – 23 – 25 – 26 – 28 – 29 – 22 – 20 – 24 – 14 – 6 – 1 – 16 – 9 – 10 – 2 – 18 – 8 – 11 – 17 – 19 – 3 – 15 – 13 – 12 – 4 – 5 – 7
T2; BS	ROC; R^2	24 – 21 – 25 – 30 – 23 – 27 – 28 – 26 – 22 – 29 – 17 – 20 – 14 – 16 – 18 – 8 – 6 – 9 – 11 – 10 – 15 – 2 – 19 – 3 – 1 – 13 – 5 – 12 – 4 – 7
BS	ROC; OCC; R^2	21 – 23 – 25 – 30 – 24 – 27 – 28 – 26 – 22 – 29 – 17 – 20 – 14 – 16 – 18 – 8 – 6 – 10 – 9 – 2 – 11 – 19 – 15 – 3 – 1 – 13 – 5 – 12 – 4 – 7
T1; BS	KS; R^2	27 – 30 – 21 – 23 – 25 – 26 – 28 – 29 – 22 – 20 – 24 – 14 – 6 – 1 – 16 – 9 – 10 – 2 – 18 – 8 – 11 – 17 – 19 – 3 – 15 – 13 – 12 – 4 – 5 – 7
T2; BS	KS; R^2	21 – 23 – 24 – 25 – 30 – 17 – 26 – 27 – 28 – 29 – 14 – 22 – 20 – 10 – 16 – 2 – 6 – 8 – 9 – 11 – 15 – 18 – 1 – 3 – 13 – 19 – 4 – 5 – 12 – 7

[1,9] Altman (1968); [2,10] Lis (1972); [3,11] Altman (1983); [4,12] Taffler (1984); [5,13] Theodossiou (1991); [6,14] Ohlson (1990); [7,15] Zmijewski (1984); [8,16] Shumway (2001); [17] Hillegeist et al. (2004); [18] Bharath and Shumway (2008); [19] Jackson and Wood (2013); [20] New MDA; [21] New LPA; [22] New LA; [23] New PA; [24] DIWOB; [25] DIWTIB_ln(age); [26] DD_ln(age); [27] DD_VEX; [28] DD_LPD; [29] DIWTIB_1/ln(age); [30] DD_1/ln(age)

Table 3.5: Efficient Frontiers with Different Performance Levels

Efficient Frontiers	(I) T1; BS, LL (O) ROC	(I) T2; BS; LL (O) ROC	(I) BS; LL (O)ROC; OCC	(I) T2; BS; LL (O) KS	(I) T1; BS; LL (O) KS	(I) T1; BS (O) ROC; R^2	(I) T2; BS (O) ROC; R^2	(I) BS (O) ROC; OCC; R^2	(I) T1; BS (O) KS; R^2	(I) T2; BS (O) KS; R^2
E^1	{25,30,27,23}	{21,30,28,27,25,24,23}	{21,30,28,27,25,24,23}	{21,30,27,25,23}	{23,30,25,24,21}	{21,30,27,25,23}	{21,30,28,27,25,24,23}	{21,23,24,25,27,28,30}	{21,23,25,27,30}	{21,23,24,25,30}
E^2	{26,28,29}	{26,22,29,17}	{26,22,29,17}	{26,28,29}	{28,26,27,29,17}	{26,28,29}	{26,22,29,17}	{26,22,29,17}	{26,28,29}	{17,26,27,28,29}
E^3	{22,20,24,14,1,6}	{20,14}	{20,14}	{22,20,24,14,1,6}	{22,14}	{22,20,24,14,6,1}	{20,14}	{20,14}	{22,20,24,14,6,1}	{14,22}
E^4	{2,16,10,9}	{16}	{16}	{2,10,9,16}	{20}	{16,9,10,2}	{18,8,6,9,11,10,15,2}	{16}	{16,9,10,2}	{20}
E^5	{11,8,18}	{2,9,10,11,8,15,6,18}	{9,11,2,8,10,15,6,18}	{11,8,18,}	{16,10}	{18,8,11,17, 19,3}	{19,3,1,13}	{18,8,6,9,11,10,15,2}	{18,8,11}	{10,16}
E^6	{3,19,17}	{1,3,13,9}	{1,3,13,19}	{3,7,19}	{11,9,2,8,15,6,18}	{15,13,12,4}	{5,12,4}	{19,3,1,13}	{17,19,3}	{2,6,8,9,11,15,18}
E^7	{13,15,12,4}	{5,12,4}	{5,12,4}	{15,13,12,4}	{1,3,13,19}	{5,7}	{7}	{5,12,4}	{15,13,12,4}	{1,3,13,19}
E^8	{5,7}	{7}	{7}	{5,7}	{7}			{7}	{5,7}	{4,5,12}
E^9										{7}

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.

Altman, E., 1983. *Corporate financial distress: A complete guide to predicting, avoiding and dealing with bankruptcy*. John Wiley and Sons.

Altman, E.I., 1973. Predicting railroad bankruptcies in America. *The Bell Journal of Economics and Management Science* 4 (1), 184.

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

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., 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.

Bauer, J., Agarwal, V., 2014. Are hazard models superior to traditional bankruptcy prediction approaches? A comprehensive test. *Journal of Banking & Finance* 40(Mar), 432–442.

Beaver, W.H., 1968. Alternative accounting measures as predictors of failure. *Accounting review* 43(1), 113–122.

Beaver, W.H., 1966. Financial Ratios as Predictors of Failure. *Journal of Accounting Research* 4, 71–111.

Beck, N., Katz, J.N., Tucker, R., 1998. Taking time seriously: Time-series-cross-section analysis with a binary dependent variable. *American Journal of Political Science* 42, 1260–1288.

Bharath, S.T., Shumway, T., 2008. Forecasting default with the Merton distance to default model. *The Review of Financial Studies* 21(3), 1339–1369.

Campbell, J.Y., Hilscher, J., Szilagyi, J., 2008. In search of distress risk. *The Journal of Finance* 63(6), 2899–2939.

Charnes, A., Cooper, W.W., Rhodes, E., 1978. Measuring the efficiency of decision making units. *European journal of operational research* 2(6), 429–444.

Chava, S., Jarrow, R.A., 2004. Bankruptcy prediction with industry effects. *Review of Finance* 8(4), 537–569.

Cleary S and Hebb G, 2016. An efficient and functional model for predicting bank distress: In and out of sample evidence. *Journal of Banking & Finance* 64 (Mar): 101–111.

Chen, M. Y. (2011). Bankruptcy prediction in firms with statistical and intelligent techniques and a comparison of evolutionary computation approaches. *Computers & Mathematics with Applications* 62(12), 4514-4524.

Collins, R. A., & Green, R. D. (1982). Statistical methods for bankruptcy forecasting. *Journal OF Economics and business* 34(4), 349-354.

Cooper, W.W., Seiford, L.M., Tone, K., 2006. *Introduction to data envelopment analysis and its uses*. Springer.

Crapp, H.R., Stevenson, M., 1987. Development of a method to assess the relevant variables and the probability of financial distress. *Australian journal of management* 12(2), 221–236.

du Jardin P., 2015. Bankruptcy prediction using terminal failure processes. *European Journal of Operational Research* 242(1): 286–303.

Gupta, J., Gregoriou, A., Healy, J., 2015. *Using hazard models correctly: a comparison employing different definitions of SMEs financial distress*.

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.

Huang, Z., Chen, H., Hsu, C.-J., Chen, W.-H., Wu, S., 2004. Credit rating analysis with support vector machines and neural networks: a market comparative study. Decision Support Systems, *Data mining for financial decision making* 37(4) 543–558.

Jackson, R.H., Wood, A., 2013. The performance of insolvency prediction and credit risk models in the UK: A comparative study. *The British Accounting Review* 45(3), 183–202.

Lane, W.R., Looney, S.W., Wansley, J.W., 1986. An application of the Cox proportional hazards model to bank failure. *Journal of Banking & Finance* 10(4), 511–531.

Lennox, C. (1999). Identifying failing companies: a re-evaluation of the logit, probit and DA approaches. *Journal of economics and Business* 51(4), 347-364.

Lo, A. W., 1986, Logit versus discriminant analysis: A specification test and application to corporate bankruptcies. *Journal of econometrics* 31(2), 151-178.

Luoma, M., Laitinen, E.K., 1991. Survival analysis as a tool for company failure prediction. *Omega* 19(6), 673–678.

Lyandres, E., & Zhdanov, A., 2013. Investment opportunities and bankruptcy prediction. *Journal of Financial Markets* 16(3), 439-476.

Maddala, G.S., 1986. *Limited-dependent and qualitative variables in econometrics*. Cambridge university press.

Martin, D., 1977. Early warning of bank failure: A logit regression approach. *Journal of Banking & Finance* 1(3), 249–276.

Meyer, P.A., Pifer, H.W., 1970. Prediction of bank failures. *The Journal of Finance* 25(4), 853–868.

Morita, H., Hirokawa, K., Zhu, J., 2005. A slack-based measure of efficiency in context-dependent data envelopment analysis. *Omega* 33(4), 357–362.

Mousavi, M.M., Ouenniche, J., Xu, B., 2015. Performance evaluation of bankruptcy prediction models: An orientation-free super-efficiency DEA-based framework. *International Review of Financial Analysis* 42, 64–75.

Nam, C.W., Kim, T.S., Park, N.J., Lee, H.K., 2008. Bankruptcy prediction using a discrete-time duration model incorporating temporal and macroeconomic dependencies. *Journal of Forecasting* 27, 493–506.

Neves, J.C., Vieira, A., 2006. Improving bankruptcy prediction with Hidden Layer Learning Vector Quantization. *European Accounting Review* 15, 253–271.

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

Ouenniche, J., Xu, B., Tone, K., 2014. Relative performance evaluation of competing crude oil prices' volatility forecasting models: a slacks-based super-efficiency DEA model. *American Journal of Operations Research* 4(4), 235–245.

Pindado, J., Rodrigues, L., & de la Torre, C., 2008, Estimating financial distress likelihood. *Journal of Business Research* 61(9), 995-1003.

Press, S. J., & Wilson, S. ,1978. Choosing between logistic regression and discriminant analysis. *Journal of the American Statistical Association* 73(364), 699-705.

Kumar, P. R., & 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.

Reisz, A.S., Perlich, C., 2007. *A market-based framework for bankruptcy prediction. Journal of Financial Stability* 3(2), 85–131.

Seiford, L.M. and Zhu, J., 1998. Stability regions for maintaining efficiency in data envelopment analysis. *European Journal of Operational Resea*rch, 108(1):127-139.

Seiford, L.M. and Zhu, J., 1998. Sensitivity analysis of DEA models for simultaneous changes in all the data. *Journal of the Operational Research Society*, 49(10):1060-1071.

Seiford, L.M., Zhu, J., 2003. Context-dependent data envelopment analysis—measuring attractiveness and progress. *Omega* 31(5), 397–408.

Shin, K.-S., Lee, T.S., Kim, H., 2005. An application of support vector machines in bankruptcy prediction model. *Expert Systems with Applications* 28(1), 127–135.

Shin, K.-S., Lee, Y.-J., 2002. A genetic algorithm application in bankruptcy prediction modelling. *Expert Systems with Applications* 23(3), 321–328.

Shumway, T., 2001. Forecasting bankruptcy more accurately: A simple hazard model. *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 among listed companies using accounting, market and macroeconomic variables. *International Review of Financial Analysis* 30, 394–419.

Trujillo-Ponce, A., Samaniego-Medina, R., Cardone-Riportella, C., 2014. Examining what best explains corporate credit risk: accounting-based versus market-based models. *Journal of Business Economics and Management* 15, 253–276.

Wu, Y., Gaunt, C., Gray, S., 2010. A comparison of alternative bankruptcy prediction models. *Journal of Contemporary Accounting & Economics* 6(1), 34–45.

Zhou, L., 2013. Performance of corporate bankruptcy prediction models on imbalanced dataset: The effect of sampling methods. *Knowledge-Based Systems* 41, 16–25.

Zhou, L., 2014, A comparison of dynamic hazard models and static models for predicting the special treatment of stocks in China with comprehensive variables. *Journal of the Operational Research Society* 66(7), 1077-1090.

Zhu, J., 2014. Context-dependent Data Envelopment Analysis, in: Quantitative Models for performance evaluation and benchmarking, International Series in Operations Research & Management Science. *Springer International Publishing*, pp. 153–174.

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

Appendix 3-A: Statistical Models of Corporate Distress Prediction

Framework	Model	Explanation
Multiple discriminant analysis (MDA)	**Altman (1968)** $Z = 1.2\,WCTA + 1.4\,RETA + 3.3\,EBITTA + 0.6\,MVTL + 0.999\,STA$ WCTA: Working capital / Total Assets; RETA: Retained Earnings / Total Assets; EBITTA: Earnings before interest and taxes / Total assets; METL: Market value of equity / Total Liabilities; STA: Sales / Total assets	Assuming there are n groups, the generic form of DA model for the group k could be shown as follows; $$z_k = f\left(\sum_{j=1}^{p} \beta_{kj} x_j\right)$$ where x_j is the discriminant features j, β_{kj} is the discriminant coefficients of group k for discriminant feature j, z_k represents the score of group k, and f is the linear or non-linear classifier that maps the scores, say $\beta^t x$ onto a set of real numbers. To compare DA models to other statistical models, I need to estimate the probability of failure, which is used as an input for estimating many measures of performance. For this, I follow Hillegeist et al. (2004) in using a logit link to calculate the probability of failure for companies; $$P(distress)_i = \frac{e^z}{1 + e^z}$$
Multiple discriminant analysis (MDA)	**Altman (1983)** $Z = 0.717\,WCTA + 0.847\,RETA + 3.107\,EBITTA + 0.42\,BVTL + 0.998\,STA$ WCTA: Working capital / Total Assets; RETA: Retained Earnings / Total Assets; EBITTA: Earnings before interest and taxes / Total assets; BVETL: Book value of equity / Total Liabilities; STA: Sales / Total assets	
Multiple discriminant analysis (MDA)	**Lis (1972)** $Z = 0.063\,WCTA + 0.092\,RETA + 0.057\,EBITTA + 0.0014\,NWTL$ $WCTA$: Working capital/ Total assets; $EBITIA$: Earnings before interest and taxes/ Total assets; $METL$: Market value of equity /Total liabilities; $NWTA$: Net wealth / Total assets	
Multiple discriminant analysis (MDA)	**Taffler (1984)** $Z = 3.2 + 2.5\,CATL + 12.18\,PBTCL + 0.029\,NCI - 10.68\,CLTA$ $CLTA$: Current liabilities/ Total assets; $PBTCL$: Profit before tax/ Current liabilities; NCI: Number of credit intervals as (quick assets - current liabilities) / ((sales - PBT - depreciation)/365); $CATL$: Current assets / Total liabilities	
Linear probability model (LPA)	**Theodossiou (1991)** $Z = -0.075 + 0.51\,WCTA - 0.21\,TDTA + 0.449\,NITA + 0.663\,RETA - 0.446\,LTDTA$ $WCTA$: Working capital/Total assets; $TDTA$= Total debt/Total assets; $NITA$: Net income/Total assets; $RETA$= Retained earnings/ Total assets; $LTDTA$= Long term debt/Total assets	The generic linear probability model (LPA) is a particular case of OLS regression and results in an estimate of probability of distress, the formula for which is as follows; $$P(distress)_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}$$
Logit analysis (LA)	**Ohlson (1980)** $log\left[\frac{Pi}{1 - Pi}\right] = -1.32 - 1.43\,WCTA + 6.03\,TLTA - 2.37\,NITA - 0.407\,OSIZE - 1.83\,FUTL + 0.0757\,CLCA + 0.285\,INTWO - 1.72\,OENEG - 0.521CHIN$	The generic model for binary variables could be stated as follows: $$\begin{cases} P(distress)_i = P(Y = 1) \\ P(distress)_i = G(\beta, X) \end{cases}$$ where Y denotes the binary response variable, X denotes the vector of features, β denotes the vector of coefficients of X in the model, and

	$WCTA$: Working capital/Total assets; $TLTA$: Total liabilities/ Total assets; $NITA$: Net income/ Total assets; $OSIZE$ = log (Total assets/GNP price-level index); $FUTL$: Funds from operations (operating income minus depreciation) / Total liabilities; $CLCA$: Current liabilities/ Current assets; $INTWO$= 1 if net income has been negative for the last 2 years, 0 otherwise; $OENEG$ = 0 if total liabilities exceed total assets, 1 otherwise; $CHIN = (NI_t - NI_{t-1})/(\lvert NI_t\rvert + \lvert NI_{t-1}\rvert)$, where NI_t is the net income for the last period. The variable is thus a proxy for the relative change in net income.	$G(.)$ is a link function that maps the scores of $\beta^t x$, onto a probability. In practice, depending on the choice of link function, the type of probability model is determined. For example, the logit model (respectively, probit model) assumes that the link function is the cumulative logistic distribution, say Θ (respectively, cumulative standard normal distribution, say N) function.
Probit analysis (PA)	**Zmijewski (1984)** $\log[\mathrm{Pt}/(1-\mathrm{Pt})] = 4.336 - 5.769\,TLTA + 4.513\,NITA - 0.004\,CACL$ $NITA$: Net income/ Total assets; $TLTA$: total liabilities/ Total assets; $CACL$: Current assets/ Current liabilities	
Contingent claim analysis (CAA): Black-Scholes-Merton (BSM) Based Models	**Hillegeist et al. (2004), Bharath and Shumway (2008)** $P(distress_i) = N\left(-\frac{\ln\left(\frac{Va}{L}\right)+(\mu-\delta-0.5\sigma_a^2)\times T}{\sigma_a\sqrt{T}}\right)$ $N(.)$: the cumulative normal distribution function,V_a:the value of the company's assets; L: total liabilities; μ: the expected return of the firm; σ_a : volatility of the company's asset; δ is the divided rate; which is estimated by the ratio of dividends to the sum of L and V_e (market value of common equity); T is time to maturity for both of call option and liabilities.	The probability of failure is extracted as the probability that call option expires worthless at the end of maturity data - i.e. the value of the company's assets (Va) be less than the face value of its debt liabilites (L) at the end of the holding period [$P(Va < L)$]. In Hillegeist et al. (2004), V_a and σ_a are estimated by solving the systems of equations; i.e. the call option equation (1) and the optimal hedge equation (2). $\begin{cases} V_e = V_a e^{-\delta T} N(d_1) - L e^{-rT} N(d_2) + \left(1 - e^{\delta T}\right) N(d_1) V_a & (1) \\ \sigma_e = \frac{V_a e^{-\delta T} N(d_1)\sigma_a}{V_e} & (2) \end{cases}$ where V_e is the market value of common equity at the time of estimation, σ_e is the annualized standard deviation of daily stock returns over 12 months prior to estimation, r is the risk-free interest rate, and d_1 and d_2 are calculated as follows; $d_1 = \frac{\ln\left(\frac{Va}{L}\right)+(r-\delta-\frac{1}{2}.\sigma_e^2)\times T}{\sigma_e\sqrt{T}}; d_2 = d_1 - \sigma_e\sqrt{T}$ Where $V_{a,t}$ is the value of the company's assets in year t and $V_{a,t-1}$ is the value of the company's assets in year $t-1$. Bharath and Shumway (2008) proposed a naïve approach to estimate V_a and σ_a as follows; $V_a = V_e + D\,; \sigma = \frac{V_e}{V_a}\sigma_e + \frac{D}{V_a}\sigma_d$

		Where $\sigma_d = 0.05 + 0.25\sigma_e$. Further, the firm's expected return μ is peroxided by the risk-free rate, r or the stock return of previous year restricted to be between r and 100%.
Contingent claim analysis (CAA): Down-and-Out Call (DOC) Barrier Option Model	**Jackson and Wood (2013)** $$P(distress)_i = N\left[\frac{ln\left(\frac{L}{V_a}\right) - \left(\mu - \frac{1}{2}\sigma_e^2\right)T}{\sigma_e\sqrt{T}}\right] + \left(\frac{L}{V_a}\right)^{\frac{2(\mu)}{\sigma_e^2}-1} N\left[\frac{ln\left(\frac{L}{V_a}\right) - \left(\mu - \frac{1}{2}\sigma_e^2\right)T}{\sigma_e\sqrt{T}}\right]$$	A naïve DOC barrier option as an extension of BSM model, which assumes that debt holder's position in the firm is like holding a portfolio of risk-free debt and a DOC option with a strike price (or Barrier) equal to total liabilities (L). The model rests on the assumptions of no dividends, zero rebate, costless failure proceedings, and set return on asset equal to risk-free rate.
Discrete time hazard model (Duration dependent hazard model)	**Shumway (2001)** $$log\ [p_{i,t}/(1 - P_{i,t})] = -13.303\ -\ 1.983\ NITA\ +\ 3.593\ TLTA\ -\ 0.467\ R.size\ -\ 1.809\ LAGEXRET\ +\ 5.791\ SIGMA$$ $NITA$: Net income /Total assets; $TLTA$: Total liabilities / Total assets; $RSIZE$: Relative size; $LAGEXRET$: Lag of excess return ($r_{it-1} - r_{mt-1}$)	Shumway proposed a discrete time hazard model using an estimation procedure like the one used for estimating the parameters of a multi-period logit model. $$P(y_{i,t} = 1 \mid x_{i,t}) = h(t \mid x_{i,t}) = \frac{\exp^{\alpha(t)+X_{i,t}\beta}}{1 + \exp^{\alpha(t)+X_{i,t}\beta}}$$ where $h(t \mid x_{i,t})$ represent the individual hazard rate of firm i at time t, $X_{i,t}$ is the vector of covariates of each firm i at time t. Shumway employed a constant time invariant term, say $ln\ (age)$, as proxy of baseline rate.
Duration-independent hazard model	$$h(t \mid x_{i,t}) = h_0.\, e^{x_{i,t}.\beta}$$ $$p(y_{i,t} = 1) = \frac{1}{1 + e^{-x_{i,t}.\beta}}$$	where, α_t is the time-varying baseline hazard function related, which could be relate to firm, e.g. ln(age) or related to macroeconomic variables, e.g. foreign exchange rate.
Duration-dependent hazard model	$$h(t \mid x_{i,t}) = h_0(t).\, e^{x_{i,t}.\beta}$$ $$p(y_{i,t} = 1) = \frac{1}{1 + e^{-(\alpha_t + x_{i,t}.\beta)}}$$	

Chapter Four

Dynamic Ranking of Corporate Distress Prediction Models

Abstract: The design of reliable models to predict corporate distress is crucial as the likelihood of filing for bankruptcy increases with the level and persistence of distress. Although many corporate failure and distress prediction models exist in the literature, the relative performance evaluation of competing prediction models remains an exercise that is mono-criterion in nature, which leads to conflicting rankings of models. This methodological issue has been addressed by Mousavi et al. (2015) by proposing a static multi-criteria assessment framework based on data envelopment analysis (DEA). In this research, I propose a dynamic DEA framework to assess and monitor the relative performance of an exhaustive range of distress prediction models and rank them accordingly. Also, I address several research questions including, what is the effect of information on the performance of distress prediction models? How the out-of-sample performance of dynamic distress prediction models compare to the out-of-sample performance of static? What is the effect of the length of training sample on the performance of models? Which models perform better in forecasting distress over the years with high distress rate (HDR)?

Keywords: Corporate Distress Prediction; Bankruptcy; Performance Criteria; Performance Measures; Data Envelopment Analysis; Malmquist Index

4.1 Introduction

Predicting bankruptcy or corporate failure before it happens has such economic benefits for a range of stakeholders (e.g., managers, investors, auditors, regulators) that many prediction models have been designed. In practice, managers could use distress prediction models (DPMs) as early warning systems to take proper preventive actions against bankruptcy. From a conceptual point of view, failure and distress predictions are classification problems, which use several features – often extracted from accounting, market, or macroeconomic information – to classify firms into one out of two or more risk categories. During the last decades, numerous studies have employed different types of prediction models or methods from fields such as probability and statistics, operational research, and artificial intelligence – for a detailed classification of distress prediction models, the reader is referred to Aziz and Dar (2006), Bellovary et al. (2007) and Abdou and Pointon (2011).

With the increasing number of prediction models, a strand of the literature has focused on assessing the performance of these models and identifying the factors that drive performance such as modelling frameworks, features selection, estimation methods, sampling, and performance criteria and their measures (e.g., Zhou, 2013; Mousavi et al., 2015). As demonstrated by Mousavi et al. (2015), the performance of prediction models is not only dependent on the nature of the modelling frameworks and the type of features, but also is related to the performance evaluation process and the underlying performance assessment method (i.e., mono-criterion methods, multi-criteria methods) and the performance criteria and measures. In fact, recent comparative studies have assessed the performance of competing failure prediction models grounded into different modelling frameworks (e.g., Wu et al., 2010; Fedorova et al., 2013; Bauer and Agarwal, 2014; Mousavi et al., 2015) and using alternative sampling techniques (e.g., Gilbert et al., 1990; Neves and Vieira, 2006; Zhou, 2013), various features (e.g., Tinoco and Wilson, 2013; Trujillo-Ponce et al., 2014; Mousavi et al., 2015), different feature selection procedures (Tsai, 2009; Unler and Murat, 2010) and a range of performance criteria (e.g., discriminatory power, calibration accuracy, information content, correctness of categorical prediction) and their measures along with different performance evaluation methodologies (Mousavi et al., 2015).

The survey of the literature on comparative studies of failure prediction models revealed a variety of shortcomings that prevent practitioners from a proper ranking of models. As pointed out by Bauer and Agarwal (2014), the literature on comparative studies suffers from two main drawbacks. First, most of the existing studies failed to have a comprehensive comparison between all types of prediction models; i.e., traditional statistical models, contingent claims analysis (CCA) models, and survival analysis (SA) models. Second, the existing literature has used a restricted number of criteria to evaluate the performance of competing models. To have a more comprehensive comparative assessment, Bauer and Agarwal (2014) evaluated the performance of Taffler (1983), Bharath and Shumway (2008) and Shumway (2001) as representative of the traditional statistical models, CCA models, and SA models, respectively. Further, they used three types of criteria; namely, discriminatory power, information content, and correctness of categorical prediction to compare the performance of these models. On the other hand, Mousavi et al. (2015) emphasised a methodological shortcoming in comparative studies arguing that although some studies consider multiple criteria and related measures to compare competing models, the nature of the comparison exercise remains mono-criterion, as they use a single measure of a single criterion at a time. The drawback of this mono-criterion approach is that the rankings corresponding to different criteria are often different (e.g., Bandyopadhyay, 2006; Theodossiou, 1991; Tinoco and Wilson, 2013), which result in a situation where one cannot make an informed decision as to which model performs best when taken all criteria into consideration. To overcome this methodological drawback, Mousavi et al. (2015) proposed a multi-criteria assessment framework; namely, an orientation-free super-efficiency data envelopment analysis framework. Finally, Zavgren (1983) argued that most of the traditional failure and distress prediction models are based on the assumption that the relationship between the dependent variable (e.g., the probability of failure) and all independent variables (e.g., accounting and market information) is stable over time. Empirical studies, however, indicate that this stability is highly arguable (e.g. Jardin and Severin, 2012; Charitou et al., 2004) and that the performance of models is sensitive to changes in macroeconomic conditions (Menash, 1984; Platt et al., 1994). For example, the logit model of Ohlson (1980) performs better in the mid- to the late 1980s, whereas the SA model of Shumway (2001) outperforms other models in the 2000s. The changes in patterns of accounting- and market-based information during time suggest that prediction models need to be re-estimated frequently to encompass the most recent patterns of information (Grice and Ingram, 2001). In this research, I argue that another shortcoming of the existing literature lies in the use of static performance evaluation frameworks to compare prediction models, and I propose a dynamic multi-criteria performance assessment framework. An additional feature of a dynamic framework is its ability, by design, to monitor the performance of models.

Recent studies have substituted financial distress for corporate failure in the implementation of failure prediction models (e.g., Tinoco and Wilson, 2013; Geng et al., 2015; Wanke et al., 2015; Laitinen and Suvas, 2016). Financial distress refers to the inability of a company to pay its financial obligations as they mature (Beaver, 1966). Obviously, the financial situation of a distressed company differs from a healthy one suggesting that, while a company moves toward deterioration, its financial features shift towards the characteristics of failed firms. This movement towards failure is a process that could take several time periods (e.g., years) and manifest itself through a variety of signals, which could prevent failure if predicted with a reasonable level of accuracy. In this research, in addition to proposing new models to predict distress or detect its signals, I propose a dynamic multi-criteria framework for assessing and monitoring the performance of distress prediction models, which, as a by-product, allows someone to detect signals of distress. To the best of my knowledge, no previous research proposed a dynamic framework for the performance evaluation and monitoring of prediction models. In practice, such a framework for the early detection of signs of distress is both necessary and beneficial.

In this chapter, I contribute to the academic literature in several respects. First, following the lead of Xu and Ouenniche (2012) and Mousavi et al. (2015) who proposed static multi-criteria frameworks for assessing the relative performance of prediction models of continuous and discrete variables respectively, I propose a new dynamic multi-criteria framework for evaluating and monitoring the relative performance of prediction models over time and ranking them. Second, I consider a more in-depth classification of statistical distress prediction models and perform an exhaustive evaluation considering the most popular models of each class. In sum, I assess the performance of univariate discriminant analysis (UDA), multivariate discriminant analysis (MDA), linear probability analysis (LPA), probit analysis (PA) and logit analysis (LA) models as traditional techniques; Black-Scholes-Merton (BSM)-based models, naïve BSM-based models, and naïve down-and-out call (DOC) barrier option models as contingent claims analysis (CCA) models; and duration independent and duration dependent survival analysis (SA) models. To best of my knowledge, this study is the first to propose the Cox model with time-varying variables using UK data for distress prediction, or equivalently estimating distress probabilities. To date, this study provides the most comprehensive empirical comparative analysis of statistical, probabilistic and stochastic distress prediction models. Third, I provide answers to several important research questions using a rolling horizon sampling framework and a multi-period performance evaluation and monitoring framework; namely, what is the effect of information on the performance of distress models? How the out-of-sample performance of dynamic distress prediction models compare to the out-of-sample performance of static ones with respect to sample type and sample period length? What is the effect of the length of training sample on the performance of models? Which models perform better in forecasting distress over the years with high distress rate (HDR)?

The rest of the chapter unfolds as follows. Section 4.2 reviews the literature on advances in and comparative studies on distress prediction models. Section 4.3 describes the proposed dynamic multi-criteria framework; namely, an orientation-free super-efficiency Malmquist DEA, for the comparison of prediction models. Section 4.4 provides details on my experimental design including data, sample selection, and the variety of distress prediction models compared as part of this study. Section 4.5 summarises my empirical results and discusses my findings. Finally, section 4.6 concludes the chapter.

4.2 Comparative Studies on Distress Prediction Models

In this section, a concise account of advances in distress prediction modelling (see section 4.2.1) along with a detailed survey of comparative studies (see section 4.2.2) are provided

4.2.1 Advances in Distress Prediction Models

Failure and distress prediction models could be divided into several categories depending on the choice of the classification criteria. In this chapter, I focus on a variety of models except for the artificial intelligence and mathematical programming ones. In sum, I consider the first generation of models; namely, discriminant analysis (DA) models (e.g. Beaver, 1966, 1968; Deakin, 1972; Blum, 1974; Altman et al., 1977; Altman, 1968), the second generation of models; namely, probability models such as linear probability (LPA) models (e.g. Meyer and Pifer, 1970), logit analysis (LA) models (e.g., Martin, 1977; Ohlson, 1980), and probit analysis (PA) models (e.g., Zmijewski, 1984), and the third generation of models; namely, survival analysis (SA) models (e.g., Lane et al., 1986; Crapp and Stevenson, 1987; Luoma and Laitinen, 1991; Shumway, 2001) and contingent claims analysis (CCA) models (e.g., Hillegeist et al., 2004; Bharath and Shumway, 2008).

Beaver (1966,1968) is the pioneering study which proposed a univariate discriminant analysis model fed with financial ratios information to predict failure. However, the first multivariate study was undertaken by Altman (1968) who estimated a score, commonly referred to as a “Z-score,” as a proxy of the financial situation of a company using multivariate discriminant analysis (MDA). The later studies frequently have employed the suggested MDA technique (e.g., Deakin, 1972; Blum, 1974; Altman et al., 1977; Altman, 1983). The majority of subsequent studies have applied the second generation models; that is, linear probability models (e.g., Meyer and Pifer, 1970), logit models (e.g., Martin, 1977; Ohlson, 1980), and probit models (e.g., Zmijewski, 1984). This first and second generation of models could be viewed as empirical models in that they are driven by practical considerations such as an accurate prediction of the risk class or an exact estimate of the probability of belonging to a risk class; in sum, the choice of the explanatory variables is driven by the predictive performance of the models. These models and their application in some previous studies are not without limitations. In fact, some of the assumptions underlying the modelling frameworks may not be reasonably satisfied for some datasets, on the one hand, and earliest studies restricted the type of information to accounting-based one, on the other hand. Also, these models are static in nature and therefore fail to adequately account for changes over time in the profiles of companies. The third generation of models; namely, survival analysis (SA) models and contingent claims analysis (CCA) models overcome some of these issues. In fact, the underlying modelling frameworks of both SA models and CCA models are dynamic by design. In addition, most previous studies made use of additional sources of information to enhance the performance of these models; namely, market-based information (e.g., Hillegeist et al., 2004; Bharath and Shumway, 2008) and macroeconomic information (e.g., Tinoco and Wilson, 2013; Kim and Partington, 2014; Charalambakis and Garrett, 2016), although one might argue that the approximation process of unobservable variables (e.g., volatility, expected return, and market value of assets) is not free of potential measurement errors (Aktug, 2014). To be more specific, SA models are used to estimate time-varying probabilities of failure. Despite the application of SA models in failure prediction dates back to the mid-1980s (e.g., Lane et al., 1986; Crapp and Stevenson, 1987; Luoma and Laitinen, 1991), Shumway (2001) was the pioneering study which made its use popular by providing an attractive estimation methodology based on an equivalence between multi-period logit models and a discrete-time hazard model. Thereafter, the suggested discrete-time hazard model – also referred to as a discrete-time logit model – was frequently used in later studies (e.g., Chav and Jarrow, 2004; Wu et al, 2010, Tinoco and Wilson, 2013; Bauer and Agarwal, 2014; Mousavi et al., 2015) to estimate the coefficients of time-varying accounting and market-based covariates of SA models. Unlike, the first-generation models, the second-generation models and SA models, which are empirical models, CCA models – also referred to as Black-Scholes-Merton (BSM)-based models – are theoretically grounded. In fact, these models are based on option-pricing theory, as set out in Black and Scholes (1973) and Merton (1974) whereby the equity holders' position in a firm is assumed to be the long position in a call option. Therefore, as suggested by McDonald (2002), the probability of failure could be interpreted as the likelihood that the value of firm's assets will be less than the face value of company's liabilities at maturity; i.e., the call option expires worthless. These models make use of market-based information by incorporating company stock returns and their volatility in estimating the probability of failure (Hillegeist et al., 2004; Bharath and Shumway, 2008). Like any modelling framework, CCA models are not without their limitations. For example, CCA models implicitly assume that the liabilities of the firm have the same maturities, which in practice is a limitation (Saunders and Allen, 2002).

4.2.2 Comparative Studies of Distress Prediction Models

This section provides a survey of the studies, which focus on the comparison of different types of failure or distress prediction models; namely, the first generation of models, the second generation of models, and the third generation of models. My survey focus is on models and performance criteria and their measures, which have been applied to the existing literature on the evaluation of competing prediction models.

Comparison between first and second generation models: Before the breakthrough model of Shumway (2001), the first and second generations of models were the common techniques in classification. Since the implementation of DA in failure prediction by Beaver (1966) and Altman (1968) to the early 1980s, MDA was the superior method for predicting corporate failure. In fact, ease of use and interpretation were the main reasons for the popularity of DA. However, the validity of these models depends on the extent to which the underlying assumptions (i.e., multivariate normality, equal groups' variance-covariate matrices) hold in a dataset. From the 1980s to 2001, LA models (introduced by Ohlson, 1980) and PA models (introduced by Zmijewski, 1984) became the common techniques. Even though probability models are more attractive from a practical perspective in that the underlying assumptions are less restrictive, most comparative studies have indicated that the prediction powers of LA models and PA models are like those of DA models (e.g., Press and Wilson, 1978; Collins and Green, 1982; Lo, 1986). A notable exception is Lennox (1999) who suggested that well-specified probit and logit models outperform DA models.

Comparison between first and second-generation models and survival analysis models: From a conceptual perspective, SA models are superior to discriminant analysis models and probability models, because of their dynamic nature. However, empirical results across several comparative studies seem to report mixed findings. From an empirical perspective, the features of a modelling framework design that are not being adequately supported or exploited by the dataset under consideration nullify its conceptual advantage. In sum, the choice combination of a modelling framework and the features to feed into it has a more significant role in enhancing or downgrading prediction performance.

For example, Luoma and Laitinen (1991) compared the performance of a semiparametric Cox hazard model with a DA model and an LA model – all models fed with accounting based information – considering type I and type II errors as measures of correctness of categorical prediction. The results suggested that the proposed SA model was inferior to both DA and LA models regarding type I and type II errors. Further, their research was limited to the number of criteria, since they only used correctness of categorical prediction.

Shumway (2001) proposed a discrete-time SA model – using a multi-period logit estimation technique – for failure prediction and compared its performance with the performance of DA, LA, and PA using overall correct classification rate (OCC) as a measure of correctness of categorical prediction. The results indicate that an SA model which encompasses both accounting and market information (respectively, only accounting information) outperforms (respectively, underperforms) DA, LA and PA models. However, on the choice of performance criteria and their measures, this study is also restricted to the correctness of categorical prediction as a criterion and overall accuracy – also known as overall correct classification rate – as its measure.

Comparison between first and second generation models and contingent claims models; Hilligeist et al. (2004) compared the performance of a BSM-based model with two types of representative models of the first and second generation of models; namely, MDA (Altman, 1968) and LA models (Ohlson, 1980), respectively. They used Log-Likelihood and Pseudo-R^2 as measures of information content to evaluate the performance of these models. The results suggested that the BSM-based model outperforms both the original and the refitted versions of Altman (1968) and Ohlson (1980) models on information content. Furthermore, they found out that the original Altman (1968) with coefficients estimated with a small dataset from decades earlier outperformed the refitted one with updated coefficients using recent data suggesting that refitting models with more recent data not necessarily improve performance. However, this study is restricted to one criterion; i.e., information content, for comparing the performance of models.

Reisz and Perlich (2007) compared the performance of three contingent claims models; namely, a BSM model, a KMV model developed by KMV Corporation in 1993 and then acquired by Moody's Corporation in 2002, and a Down-and-Out Call option (DOC) model, with the MDA model of Altman (1968). Recall that the KMV model – also referred to as the Expected Default Frequency (EDF) model – is actually a four-step procedure based on Merton's framework, which determines a default point, estimates asset value and volatility, calculates distance to default (DD), and converts DD into EDF. They use ROC as a measure of discriminatory power and Log-Likelihood as a measure of information content. They found out that the DOC model outperforms the other types of contingent claims models as well as MDA for 3-, 5- and 10-year ahead failure prediction. Unexpectedly, Altman (1968) outperforms all contingent claims models for 1-year ahead failure prediction. Although this study encompassed two types of criteria; i.e., discriminatory power and information content, the comparison is somehow incomplete as the log-likelihood cannot be computed for Altman's model.

Agarwal and Taffler (2008) compared the performance of two types of BSM models; namely, Hillegeist et al. (2004) and Bharath and Shumway (2008), with the MDA model of Taffler (1983) with respect to ROC as a measure of discriminatory power, Log-likelihood and Pseudo-R^2 as measures of information content, and return on assets (ROA) and return on risk-weighted assets (RORWA) as measures of economic value, given different costs of misclassification. The empirical results showed that the MDA model outperforms Hillegeist et al. (2004) significantly on ROC as a measure of discriminatory power. Meanwhile, MDA model does not outperform Bharath and Shumway (2008) significantly on ROC. On the other hand, considering Log-likelihood as a measure of information content, Hillegeist et al. (2004) perform significantly better than Bharath and Shumway (2008) and MDA model, respectively. However, Pseudo-R^2 was higher for Taffler (1983) compared to BSM models, which suggests that these two information content measures carry different elements of information. Furthermore, considering differences in misclassification costs, they analysed the economic benefit of applying Bharath and Shumway (2008) or Taffler (1983) as classifiers using the approach proposed by Blochlinger and Leippold (2006). The results suggest that the MDA model of Taffler (1983) outperforms BSM-based models. It is worth mentioning that with respect to the number of criteria, this study was innovative in its era since three criteria; namely, the correctness of categorical prediction, discriminatory power, and information content, were used for evaluating models.

Comparison between contingent claims models and survival analysis models: Campbell et al. (2008) proposed a duration-dependent SA model and evaluated its performance with the performances of a KMV model and the duration-independent SA model of Shumway (2001) using log-likelihood and Pseudo-R^2 as measures of information content. The results indicate that their SA model outperforms both the SA model of Shumway (2001) and the KMV model. However, this study fails to incorporate more criteria for comparing the performance of models.

Comparison between first, second and third generations of models: Wu et al. (2010) compared the performance of the MDA model of Altman (1968), the LA model of Ohlson (1980), the PA model of Zmijewski (1984), the duration-independent SA model of Shumway (2001) and the BSM model of Hillegeist et al. (2004). The results indicate that considering Log-likelihood and Pseudo-R^2 as measures of information content, the discrete time SA model of Shumway outperforms LA, PA, BSM and MDA models, respectively. Unexpectedly, taking to account overall correct classification rate as a measure of correctness of categorical prediction, Ohlson’s model of LA outperforms MDA, PA, BSM, and SA models, respectively, under a rolling window implementation. For ROC as a measure of discriminatory power, the authors failed to take account of BSM model, the result suggests that the duration-independent SA model of Shumway performs better than LA, MDA, and PA, respectively. Regarding the number of criteria, this study puts comparison into effect with three types of criteria; namely, the correctness of categorical prediction, discriminatory power, and information content.

4.3 A Multi-Period Framework for Assessing Distress Prediction Models: Orientation-free Super-Efficiency Malmquist DEA

Malmquist productivity index (MPI) is a multi-criteria assessment framework for performing performance comparisons of DMUs over time. Fare et al. (1992, 1994) employed DEA to extend the original MPI proposed by Malmquist (1953) and constructed the DEA-based Malmquist productivity index as the product of two components, one measuring the efficiency change (EC) of DMU with respect to the efficiency possibilities defined by the frontier in each period (also referred to as caching-up to the frontier), and the other measuring the efficient frontier-shift (EFS) between the two time periods t and $t+1$ (also referred to as change in the technical efficiency evaluation).

Let x_{i0}^{t} denote the ith input and y_{r0}^{t} denote the rth output for DMU_0, both at period t. Figure 4-1 shows the change of efficiency of DMU_0 from point A (with respect to efficient frontier at period t) to point B (with respect to efficient frontier at period $t+1$) assuming to have one input and one output. The efficiency change, EC component in Figure 4-1, is measured by the following formula:

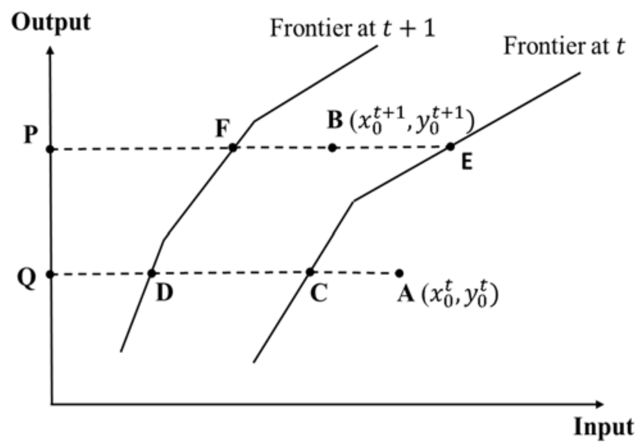
$$\mathrm{EC} = \frac{PF/_{PB}}{QC/_{QA}} = \frac{\text{Efficiency of } \mathrm{DMU}_0 \text{ with respect to the period } t+1}{\text{Efficiency of } \mathrm{DMU}_0 \text{ with respect to the period } t} \qquad \text{Eq. 4-1}$$

Let $\Delta^{t_2}((x_0, y_0)^{t_1})$ denote the efficiency score of DMU with x_0 input and y_0 output at period t_1 (say, $DMU(x_0, y_0)^{t_1}$) relative to frontier t_2 in Figure 4-1. Replacing t_1 and t_2 with t and $t+1$, respectively, the EC effect (say, α) can be presented as:

$$EC: \quad \alpha = \frac{\Delta^{t+1}((x_0, y_0)^{t+1})}{\Delta^{t}((x_0, y_0)^{t})} \qquad \text{Eq. 4-2}$$

Thus, $EC > 1$ shows an improvement in relative efficiency from period t to $t+1$, while $EC = 1$ and $EC < 1$ shows stability and deterioration in relative efficiency, respectively.

Figure 4-1: Efficiency Change and Efficient Frontier-Shift



Also, Figure 4-1 indicates that the reference point of (x_0^t, y_0^t) moved from C on the frontier of period t to D on the frontier of period $t+1$. Therefore, the efficient frontier-shift (EFS) effect at (x_0^t, y_0^t) is equivalent to:

$$EFS_t = \frac{QC}{QD} = \frac{QC/_{QA}}{QD/_{QA}} \qquad \text{Eq. 4-3}$$

$$= \frac{\text{Efficiency of } (x_0^t, y_0^t) \text{ with respect of the period t frontier}}{\text{Efficiency of } (x_0^t, y_0^t) \text{ with respect of the period t + 1 frontier}}$$

Similarly, the EFS effect at (x_0^{t+1}, y_0^{t+1}) is equivalent to:

$$EFS_{t+1} = \frac{PE}{PF} = \frac{PE/_{PB}}{PF/_{PB}} \qquad \text{Eq. 4-4}$$

$$= \frac{\text{Efficiency of } (x_0^{t+1}, y_0^{t+1}) \text{with respect of the period t frontier}}{\text{Efficiency of } (x_0^{t+1}, y_0^{t+1}) \text{with respect of the period t + 1 frontier}}$$

The EFS component is measured by the geometric mean of EFS effect at (x_0^t, y_0^t) (say, EFS_t) and EFS effect at (x_0^{t+1}, y_0^{t+1}) (say, EFS_{t+1});

$$EFS = [EFS_t \times EFS_{t+1}]^{1/2} \qquad \text{Eq. 4-5}$$

Using my notation, the EFS effect can be expressed as:

$$EFS: \quad \beta = \left[\frac{\Delta^t((x_0, y_0)^t)}{\Delta^{t+1}((x_0, y_0)^t)} \times \frac{\Delta^t((x_0, y_0)^{t+1})}{\Delta^{t+1}((x_0, y_0)^{t+1})}\right]^{1/2} \qquad \text{Eq. 4-6}$$

Therefore, the Malmquist Productivity Index (MPI) can be written as;

$$MPI = EC \times EFS \qquad \text{Eq. 4-7}$$

Using my notation, the MPI can be presented as:

$$\begin{aligned} MPI: \quad \gamma &= \alpha \times \beta \qquad \text{Eq. 4-8} \\ &= \frac{\Delta^{t+1}((x_0, y_0)^{t+1})}{\Delta^t((x_0, y_0)^t)} \\ &\times \left[\frac{\Delta^t((x_0, y_0)^t)}{\Delta^{t+1}((x_0, y_0)^t)} \times \frac{\Delta^t((x_0, y_0)^{t+1})}{\Delta^{t+1}((x_0, y_0)^{t+1})}\right]^{1/2} \end{aligned}$$

MPI could be rearranged as;

$$\gamma = \left[\frac{\Delta^t((x_0, y_0)^{t+1})}{\Delta^t((x_0, y_0)^t)} \times \frac{\Delta^{t+1}((x_0, y_0)^{t+1})}{\Delta^{t+1}((x_0, y_0)^t)}\right]^{1/2} \qquad \text{Eq. 4-9}$$

This explanation of MPI could be interpreted as the geometric mean of efficiency change measured by period t and $t + 1$ technology, respectively. $MPI > 1$ shows an improvement in the total factor productivity of DMU_0 from period t to $t + 1$, while $MPI = 1$ and $MPI < 1$ shows stability and deterioration in total factor productivity, respectively.

Comment 1: Caves et al. (1982) introduced a distance function, $\Delta(.)$, to measure technical efficiency in the basic CCR model (Charnes et al., 1978). Though, in the non-parametric framework, instead of using a distance function, DEA models are implemented. For example, Fare et al. (1994) used input (or output) oriented radial DEA to measure the MPI. However, the radial model faces a lack of attention to slacks, which could be overcome using non-radial (slacks-based measure) oriented (or orientation-free) DEA model (Tone, 2001, 2002).

In this study, I use the non-radial (slacks-based measure) orientation-free super- efficiency DEA (Tone, 2001, 2002) Malmquist index to evaluate the performance of competing distress prediction models. The reason to choose an orientation-free evaluation is that the aim is to evaluate distress prediction models, and thus, the choice between input-oriented or output-oriented analysis is irrelevant. In other words, I do not have any priority with respect to minimising inputs at the same level of outputs (input oriented) or maximising outputs at the same level of inputs (output oriented). Further, my study is under the assumption of the variable return to scale (VRS), where input-oriented and output-oriented analysis may result in different scores and rankings of DMUs. On the other hand, the reason to choose non-radial framework is that radial super-efficiency DEA models may be infeasible for some DMUs; therefore, ties would stay in rankings. The reason is that the super-efficiency DEA model was developed under (i) constant returns to scale (CRS) condition and (ii) the simultaneous and same proportion of change in all inputs (or outputs). Once any of these conditions is violated, it is high likely that infeasibility of the related DEA mode occurs (see, e.g., Seiford and Zhu, 1998a,b). Moreover, radial DEA models overlook possible slacks in inputs and outputs, and therefore, would possibly over-estimate the efficiency scores by ignoring mix efficiency.

Further, basic DEA techniques cannot distinguish between efficient DMUs (here, distress prediction models) because all their scores are equal to 1 (Anderson and Peterson, 1993). Therefore, I choose super-efficiency DEA framework, as I am interested in acquiring a complete ranking of distress prediction models. It should be acknowledged that within a super-efficiency DEA framework, the reference benchmark changes from one prediction model evaluation to another one, which in some contexts might be viewed as "unfair" benchmarking. However, super-efficiency DEA is a valuable method in distinguishing DMUs, that has been employed in many recent studies (Li et al, 2017; Huang et al, 2015; Mousavi et al, 2015).

Considering the production possibility set P defined by Cooper et al. (2006) as

$$P = \{(x, y) | x \geq X^{t_1}\lambda, y \leq Y^{t_1}\lambda, 1 \leq e\lambda \leq 1, \lambda \geq 0\}, \qquad \text{Eq. 4-10}$$

SBM-DEA (Tone, 2001) measures the efficiency of DMU $(x_0, y_0)^{t_2}$ $(t_2 = 1,2)$ with respect to the benchmark set $(X, Y)^{t_1}(t_1 = 1,2)$ using the following linear programing (LP):

$$\Delta^{t_1}((x_0, y_0)^{t_2}) = \min_{\lambda, s^-, s^+} \frac{1 - \frac{1}{m}\sum_{i=1}^{m} \frac{s_i^-}{x_{io}^{t_2}}}{1 + \frac{1}{r}\sum_{i=1}^{r} \frac{s_i^+}{y_{io}^{t_2}}} \qquad \text{Eq. 4-11}$$

$$\begin{aligned} \text{subject to} \quad & x_0^{t_2} = X^{t_1}\lambda + s^-, \\ & y_0^{t_2} = Y^{t_1}\lambda - s^+, \\ & 1 \leq e\lambda \leq 1, \\ & \lambda \geq 0, s^- \geq 0, s^+ \geq 0. \end{aligned}$$

where $\Delta^{t_1}((x_0, y_0)^{t_2})$ is the efficiency score of $DMU(x_0, y_0)^{t_1}$ relative to frontier t_2; $X^{t_1} = \left(x_1^{t_1}, \dots, x_n^{t_1}\right) \in \mathbb{R}^n$ and $Y^{t_1} = (y_1^{t_1}, \dots, y_n^{t_1}) \in \mathbb{R}^n$ are matrices of inputs and outputs at the period t_1, respectively; $s^- \geq 0$ and $s^+ \geq 0$ are the vectors of input surpluses and output shortages in $\mathbb{R}^n$, respectively, and are named *slacks;* e is a row vector with all items equal to one, and λ is a nonnegative vector in $\mathbb{R}^n$.

Equivalently;

$$\Delta^{t_1}((x_0, y_0)^{t_2}) = \min_{\theta, \eta, \lambda} \frac{\frac{1}{m}\sum_{i=1}^{m} \theta_i}{\frac{1}{r}\sum_{i=1}^{r} \eta_i} \qquad \text{Eq. 4-12}$$

$$\begin{aligned} \text{subject to} \quad & \theta_i x_{io}^{t_2} \geq \textstyle\sum_{j=1}^{n} x_{ij}^{t_1} \lambda_j \ (i = 1, \dots, m), \\ & \eta_i x_{io}^{t_2} \geq \textstyle\sum_{j=1}^{n} y_{ij}^{t_1} \lambda_j \ (i = 1, \dots, r), \\ & \theta_i \leq 1 (i = 1, \dots, m), \eta_i \geq 1 (i = 1, \dots, r), \\ & 1 \leq e\lambda \leq 1, \\ & \lambda \geq 0. \end{aligned}$$

where $\theta_i$and η_i are $\left(1 - \frac{s_i^-}{x_{io}^{t_2}}\right)$ and $\left(1 + \frac{s_i^+}{y_{io}^{t_2}}\right)$, respectively.

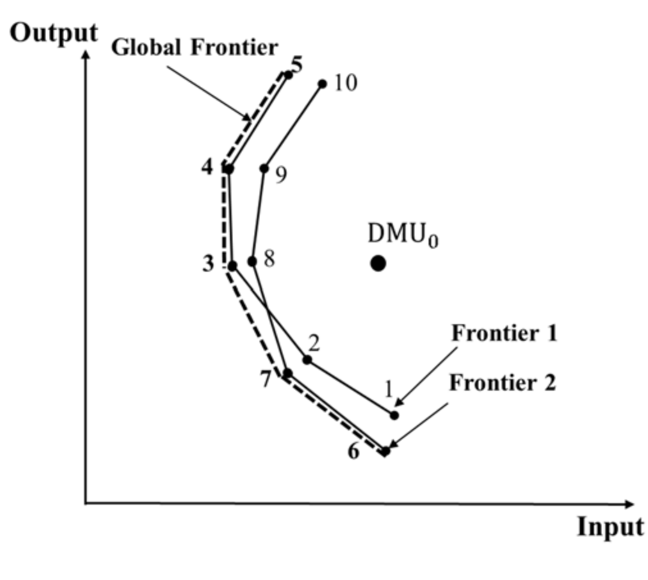
Referring to equation 4-9, one can use equation 4-12 to estimate $\Delta_0^t(x_0^t, y_0^t)$, $\Delta_0^{t+1}(x_0^{t+1}, y_0^{t+1})$, $\Delta_0^t(x_0^{t+1}, y_0^{t+1})$ and $\Delta_0^{t+1}(x_0^t, y_0^t)$ as four required terms for calculating MPI.

Comment 2: The main objective of this study is to estimate the relative efficiency of $DMUs$ in each period. However, the estimated Malmquist productive index, say, $MPI_0^{t,t+1}$, indicates the change of efficiency score between period t and $t+1$, and should be modified for my purpose. Further, according to Pastor and Lovell (2005), the contemporaneous MPI is not circular, its adjacent period components can give conflicting signals, and it is sensitive to LP infeasibility.

The adjacent reference index, proposed by Fare et al. (1992), suggests multiplying $MPI_0^{t,t+1}$ by $\Delta_0^t(x_0^t, y_0^t)$, which results in the relative efficiency of DMU_0 at period $t+1$ compared to period t. However, the main drawback of this index is that it cannot estimate the relative efficiency score of non-adjacent periods, e.g., period t and $t+2$ or $t+1$ and $t+3$.

To overcome this drawback, Berg et al. (1992) used a fixed reference index, which compares and refers the relative efficiencies of all periods (say, t $(t \geq 2)$) to the first period (say, $t=1$). Therefore, it is possible that the efficiency scores of the periods later than the first one are more than 1 since the technology develops over time. Although, fixed reference index acquires the circularity property with a base period dependence, it remains sensitive to LP infeasibility.

Figure 4-2: Global Frontier



More recently, Pastor and Lovell (2005) suggested a global MPI, which its components are circular, provides single measures of productivity change, and is not susceptible to LP infeasibility. Further, in situation where efficient frontiers of multiple periods cross each other, the global index can be measured by the best practices in all periods.

As Figure 4-2 presents, the relative efficiency of DMU_0 can be measured in terms of either the frontier of period 1 (consists of four DMUs of 1,2,3,4 and 5) or the frontier of period 2 (consist of four DMUs of 6,7,8,9 and 10). An alternative is the global frontier, which is the combination of the best DMUs in the history, i.e. five DMUs of 6,7,3,4 and 5.

It is argued that if the length of observation period is long enough, the current DMUs would be covered by the best historical DMUs, probably themselves. Thus, the relative efficiency to the global frontier could be considered as an absolute efficiency with the scores less than or equal to 1 (Pastor and Lovell, 2005).

4.4 Empirical Investigation

In this section, the details of the empirical investigation are provided, where the performance of both existing and new distress prediction models using both mono- and multi-criteria performance evaluation frameworks are compared. In the remainder of this section, the details on the dataset (see section 4.4.1), features selection (see section 4.4.2), sampling and fitting choices (see section 4.4.3), and distress prediction models (see section 4.4.4) used in this study are provided.

4.4.1 Data

The dataset used in my empirical analysis is chosen as follows. First, I considered all non-financial and non-utility UK companies listed on the London Stock Exchange (LSE) at any time during a 25-year period from 1990 through 2014. Second, since only post-listing information is used as input to my prediction models and these models have minimum historical data requirements, I excluded companies that have been listed for less than two years. In all databases, there are several companies with missing data. My dataset is no exception. Excluding those companies with missing data is a source of potential error in evaluating prediction models (Zmijewski, 1984; Platt and Platt, 2012). Therefore, to minimise any bias related to this aspect, I only excluded those companies with missing values for the main accounting book items (e.g., sales, total assets) and market information (e.g., price) which are required for computing many accounting and market-based ratios (Lyandres and Zhdanov, 2013). The remaining companies with missing values were dealt with by replacing the missing values for each company by its most recently observed ones (Zhou et al., 2012). As to outlier values amongst the observed variables, these variables are winsorized; that is, the values lower (respectively, greater) than the 1st (respectively, 99th) percentile of each variable are sit equal to that value (Shumway, 2001).

With respect to the definition of distress, I considered the proposed definition by Pindado et al. (2008). A binary variable, say D that equals 1 for financially distressed companies and 0 otherwise, represents the distress definition.

A company is considered financially distressed if it meets both following conditions: (1) its earnings before interest, taxes, depreciation and amortization (EBITDA) is lower than its interest expenses for two consecutive years, and (2) the company experience negative growth in market value for two consecutive years. In sum, my dataset consists of 3,389 companies and 36,984 company-year observations.

Among the total number of observation, there are 1,414 company-year observations classified as distressed resulting in a distress rate average of 3.82% per year. Details on the number of healthy and distress firms in the dataset and the proportion of distress rate in samples are provided in Table 4.1 and Table 4.2, respectively.

Table 4.1: Basic Sample Statistics

This table presents the total number of distressed companies versus healthy ones for the period of 1990 and 2014.

Observation (1990-2014)	***#***	***%***
Distressed company-year observations (D)	1414	3.82%
Healthy company-year observations	35,570	96.18%
Total company-year Observation	36,984	100%

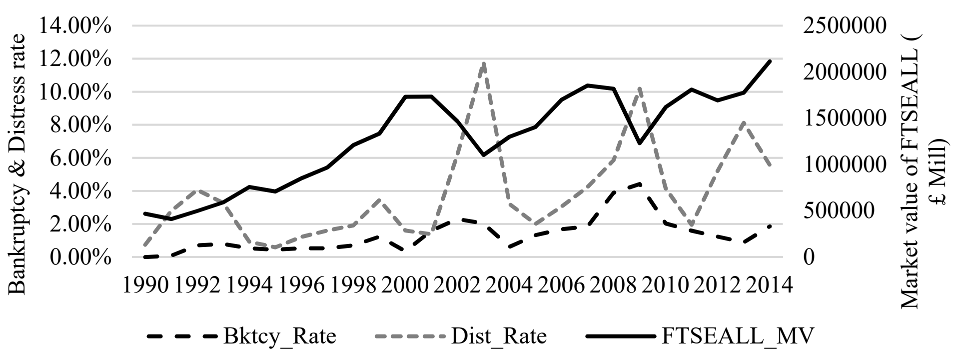
Table 4.2: The Proportion of Distress Firms (D) in Training and Holdout Samples

This table presents the yearly proportion of distress in my training and hold-out samples. The percentage of distress is presented based on the definition of distress (D) and three different length of training period.

Hold out sample		3-year training sample		5-year training sample		10-year training sample	
Year	**D %**	**Years**	**D %**	**Years**	**D %**	**Years**	**D %**
2000	1.60%	1997-1999	2.32%	1995-1999	1.79%	1990-1999	2.04%
2001	1.39%	1998-2000	2.32%	1996-2000	1.96%	1991-2000	2.11%
2002	6.22%	1999-2001	2.15%	1997-2001	1.99%	1992-2001	1.97%
2003	11.78%	2000-2002	3.04%	1998-2002	2.89%	1993-2002	2.23%
2004	3.21%	2001-2003	6.42%	1999-2003	4.82%	1994-2003	3.09%
2005	2.00%	2002-2004	6.97%	2000-2004	4.77%	1995-2004	3.29%
2006	3.06%	2003-2005	5.37%	2001-2005	4.76%	1996-2005	3.38%
2007	4.25%	2004-2006	2.75%	2002-2006	4.99%	1997-2006	3.54%
2008	5.86%	2005-2007	3.13%	2003-2007	4.62%	1998-2007	3.81%
2009	10.18%	2006-2008	4.37%	2004-2008	3.69%	1999-2008	4.21%
2010	4.15%	2007-2009	6.59%	2005-2009	4.94%	2000-2009	4.86%
2011	1.96%	2008-2010	6.77%	2006-2010	5.41%	2001-2010	5.10%
2012	5.21%	2009-2011	5.66%	2007-2011	5.37%	2002-2011	5.18%
2013	8.12%	2010-2012	3.76%	2008-2012	5.63%	2003-2012	5.09%
2014	5.56%	2011-2013	4.99%	2009-2013	6.01%	2004-2013	4.71%

Figure 4-3 displays the market value of LSE as measured by the FTSE-all index, the average of financial distress (Dist_Rate) and bankruptcy rate (Bktcy_Rate) during 25 years from 1990 through 2014. This graphical snapshot indicates that the percentage of bankrupt and distress companies expressed in percentage terms and the performance of the UK stock market are, as one would expect, inversely moving together in a consistent fashion, which suggests that the use of market information would in principle enhance distress prediction accuracy.

Figure 4-3: Financial Distress Rate and Market Value of LSE Trend



4.4.2 Feature Selection

There is a variety of strategies and methods for identifying the most effective group of features to feed failure prediction models with (Balcaen and Ooghe, 2006). Feature selection strategies could be theoretically grounded, empirically grounded, or both – see, for example, Laitinen and Suvara (2016). On the other hand, feature selection methods could be objective or subjective. Objective feature selection methods could be statistical (e.g., Tsai, 2009; Zhou et al., 2012) or non-statistical (e.g. Pacheco, 2007, 2009; Unler, 2010), but adopt a common approach; that is, optimising an effectiveness criterion. Whereas subjective feature selection methods make often use of a subjective decision rule including reviewing the literature and selecting the most commonly used features (e.g., Hebb, 2016; du Jardin, 2015; Zhou, 2014, 2013; Ravi Kumar and Ravi, 2007). In this research, I used a statistical objective feature selection method.

To be more specific, I reduced my very large initial set of accounting-based ratios (i.e., 83 accounting-based ratios) using factor analysis, where factors are selected so that both the absolute values of their loadings are greater than 0.5 and their communities are greater than 0.8, and the stopping criterion is either no improvement in the total explained variance or no more variables are excluded. This factor analysis was run using principal component analysis with VARIMAX as a factor extraction method (Chen, 2011, Mousavi et al., 2015).

Finally, 31 accounting-based ratios with high factor loadings and high communality values, 5 frequent used market-based information and 2 mixed ratios (interaction effect of macroeconomic indicators and accounting-based information) were retrained as input features into a stepwise procedure in each statistical framework. Table 4.3 represents the final features selected for my analysis. Please note that I fed the newly developed models with six different combinations of financial accounting (FA), Market variables (MV) and Macroeconomic indicators (MI); namely, FA, MV, FAMV, FAMI, FAMVMI, and MVMI. Also, note that for a better model fit, I considered the interaction effect of Macroeconomic indicators on financial accounting items on distress prediction models.

4.4.3 Sample Selection

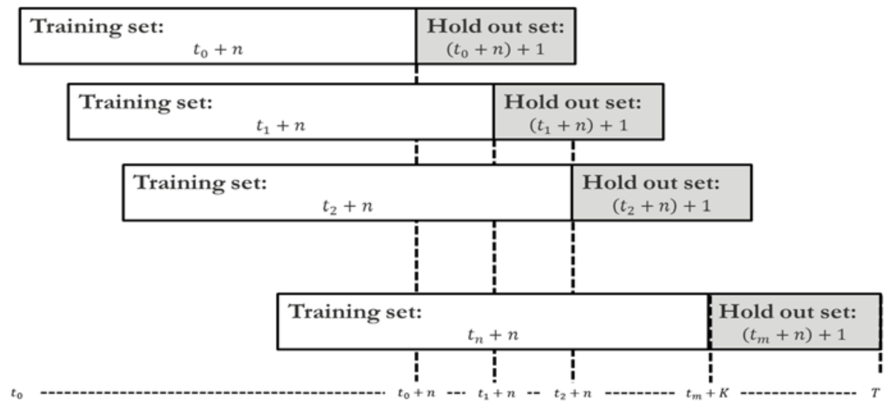
Following the lead of Mousavi et al. (2015), I test the performance of distress prediction models out-of-sample; however, in this chapter out-of-sample testing is implemented within a rolling horizon framework. The aim here is to find out how robust is the out-of-sample performance of dynamic distress prediction models relative to static ones with respect to sample period length. In my empirical investigation, I considered three sample period lengths; namely, 3, 5, and 10 years. In sum, I used firm-year observations from year $t-n+1$ to year t ($n=3,5,10$) as a training sample to fit models; that is, estimate their coefficient. Then, I used the fitted models to predict distress in year $t+1$.

Table 4.3: List of Financial Ratios

Category	Ratio or item	Category	Ratio or item
Liquidity (13)	Current assets to total assets Cash and equivalent to current liabilities Current assets to current liabilities Sales to inventory Current assets to sales Quick assets to current liabilities Current assets to total liabilities Quick assets to inventory Quick assets to assets Inventory to assets Inventory to current assets Net fixed assets to total assets Current liabilities to total assets	Solvency (7)	Liabilities to sales Long-term and current liabilities to assets Equity to capital Book value of equity to total liabilities Net worth to total debt Shareholders capital to total capital ABD=|1- (fixed assets to equity) |
Market information (5)	Lag of excess return Lag sigma Ln (price) Real size Failure rate in last year	Cash flow (3)	Operating cash flow to liabilities Cash and equivalent to sales Funds provided by operations to total liabilities
Asset utilisation (2)	Working capital to sales Quick assets to sales	Profitability (4)	Net income to capital Net income to long-term funding Net worth to total liabilities ROI × average payment period
Mixed (2)	GDP × Sales Interest rate × Income	Firm characteristics (2)	Ln(age) Log (total assets to GNP price level index)

For the sake of comparing the predictive ability of different models for different sample period lengths, I am concerned with predicting distress from 2000 onwards; that is, $t=1999$ to 2013. Therefore, I used 45 different training samples for developing models. The reader is referred to Figure 4-4 for a graphical representation of this process. Also, the details about the proportion of distressed firms for 45 training samples and 15 holdout samples are presented in Table 4-2.

Figure 4-4: Rolling Window Periodic Sampling



4.4.4 Distress Prediction Models for Comparative Study

In this research, I develop a variety of distress prediction models in different static and dynamic frameworks. To be more specific, I use MDA, logit, probit and linear probability as static frameworks; Bharath and Shumway (2008) (BhSh_2008), Hillegeist et al. (2004) (HKCL_2004) and Jackson and Wood (2013) (JW_2013) as contingent claim analysis (CCA) frameworks; and discrete time hazard model (Shumway, 2001) and Cox hazard model (Kim and Partington, 2014) as the most recent and frequent cited dynamic or duration frameworks. Regarding discrete time hazard model, I followed Nam et al. (2008) in classifying dynamic models into two subcategories; namely, duration-independent and duration-dependent models. Based on containing a constant (time-independent) baseline hazard rate, the duration-independent models could be further classified into two subcategories, namely duration-independent with time-independent baseline (DIWTIB) and duration independent without baseline (DIWOB). The corresponding DIWTIB, which use ln(age) and 1/ln(age) as time-independent baseline rates are named DIWTIB_ln(age) and DIWTIB_1/ln(age), respectively. On the other hand, based on the type of time-dependent baseline hazard rate, a variety of duration-dependent (DD) models could be developed. Beck et al. (1998) used time dummies to proxy the baseline hazard rate. Since the use of time dummies as an indirect proxy for the baseline rate is less efficient, I follow

Nam et al. (2008) and Gupta et al. (2015) in using time-varying features to proxy the time-dependent baseline rate. For this, I use ln (age), 1/ln(age), last year probability of distress (LPD), and volatility of exchange rate (VEX) as alternative features to proxy baseline hazard rates in duration-dependent models. Therefore, the duration dependent (DD) models in my analysis are named, DDWTD_ln(age), DDWTD_1/ln(age), DDWTD_LPD and DDWTD_VEX. Further, Considering Cox hazard model, I followed Kim and Partington (2014) in estimating the time-dependent baseline rate using the historical information of the firm. I refer to this model as duration dependent with firm's specific baseline rate (DDWFSB). Considering 15 static and dynamic frameworks, 45 training samples, and 6 combinations of features, I ended up with 3,375 new developed models. Table 4-4 presents the new models fed with 3-year training sample from 2011 to 2013 using FAMVMI as features. See Appendix 4-A for more details on models.

4.5 Performance Evaluation of Distress Prediction Models

In this section, firstly, the criteria and measures employed to evaluate the performance of models are explained (see section 4.5.1). Then, the mono-criteria evaluation of prediction models (see section 0) is provided. Finally, the suggested multi-criteria evaluation approach to evaluate the performance of models is implemented (see section 4.5.3).

4.5.1 Criteria and Measures for Performance Evaluation

In this chapter, I have focused on the most frequently used criteria and their measures for performance evaluation of prediction models. The first criterion is the discriminatory power, which is defined as the power of a prediction model to discriminate between the healthy firms and the unhealthy firms. In the comparative evaluation, I use Hand measure (H) (Hand and Anagnostopoulos, 2014), Kolmogorov-Smirnov (KS), Area under Receivable Operating Characterise (AUROC), Gini index (GI) and Information Value (IV) to measure this criterion. The second criterion is the calibration accuracy, which is defined as the quality of estimation of the probability of failure (or distress). I use Brier Score (BS) to measure this criterion.

Table 4.4: New Developed Models Fed with 3-year Training Samples Using FAMVMI

The table presents the features and coefficients of the new models, namely MDA (1), LPA (2), LA (3), PA (4), DIWOB (5), DIWTIB_ln(age) (6), DIWTIB_1/ln(age) (7), DDWTD_ln(age) (8), DDWTD_VEX (9), DDWTD_LPD (10), DDWTD_1/ln(age) (11) and DDWFSB (12). *** and ** refer to 1% and 5% significance level, respectively.

Models Explanatory variables	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6*	Model 7*	Model 8	Model 9	Model 10	Model 11	Model 12
Intercept	-2.94	0.073	-3.44	-3.24	-1.53	-1.53 $+ln(age)_i$	-1.53 $+1/ln(age)_i$	-1.9749	-0.2996	-2.171	-0.591	
Current Assets to Total Liabilities	-0.002	-0.0001*										
Net income to long term funding	-0.014	-0.0006***										0.0022
Current Assets to Sales	-.0002	-0.0001***										
Total liabilities to Total Assets	0.066	0.0028***										-0.0079
Cash and equivalent to Sales	0.006	0.0002***										0.0003
Inventory to Assets												-0.9997
Equity to Sales	-0.0003	-0.0001***										
Lag of Excess Return	-1.289	-0.059***	-0.832***	-0.9322***	-0.987***	-0.987***	-0.987***	-1.001***	-1.041***	-1.033***	-1.016***	-0.985***
Lag of Sigma	2.865	0.117***										
Ln (price)	-3.842	-0.015***	-0.281***	-0.211***	-0.217***	-0.217***	-0.217***	-0.218***	-0.209***	-0.213***	-0.214***	-0.226***
Equity to Capital		0.003	0.045	0.0204								
Current Liabilities to Total Assets		0.0067	0.038	0.032								
Real size			-0.2562***	-0.1524**								-0.111
Ohlson size					-0.197***	-0.197***	-0.197***	-0.201***	-0.193***	-0.193***	-0.207***	
Interest rate × Net Income	-0.0001	3.23	-0.0006***	-0.00006***	-0.00007***	-0.00007***	-0.00007***	-0.00007***	-0.00006***	-0.00006***	-0.00007***	-0.00004***
GDP × Sales	-0.0001	-3.34	-0.0004***	-0.0001***	-0.0001***	-0.0001***	-0.0001***	-0.0001***	-0.0001***	-0.0001***	-0.0001***	-0.0001***
Ln (age)								0.185				
1/ ln (age)											-2.124***	
Volatility of Exchange Rate (VEX)									-0.711***			
Last year distress rate										15.319***		

* Note that in models 25 and 29, the ln(age) and 1/ln(age) of firm i is added to intercept as the baseline hazard rate.

The third criterion is the information content which is defined as the extent to which the outcome of a prediction model (e.g. score or probability of failure) carries enough information for failure (or distress) prediction. I employ log-likelihood statistic (LL) and Pseudo-coefficient of determination (Pseudo-R^2) to measure this criterion. The last criterion is the correctness of categorical prediction, which is defined as the capability of the failure (or distress) model to correctly classify firms into healthy or non-healthy categories considering the optimal cut-off point. I use Type I errors (T1), Type II errors (T2), misclassification rate (MR), sensitivity (Sen), specificity (Spe), and overall correct classification (OCC) to measure this criterion. – See, Appendix 2-C for the descriptions of these measures.

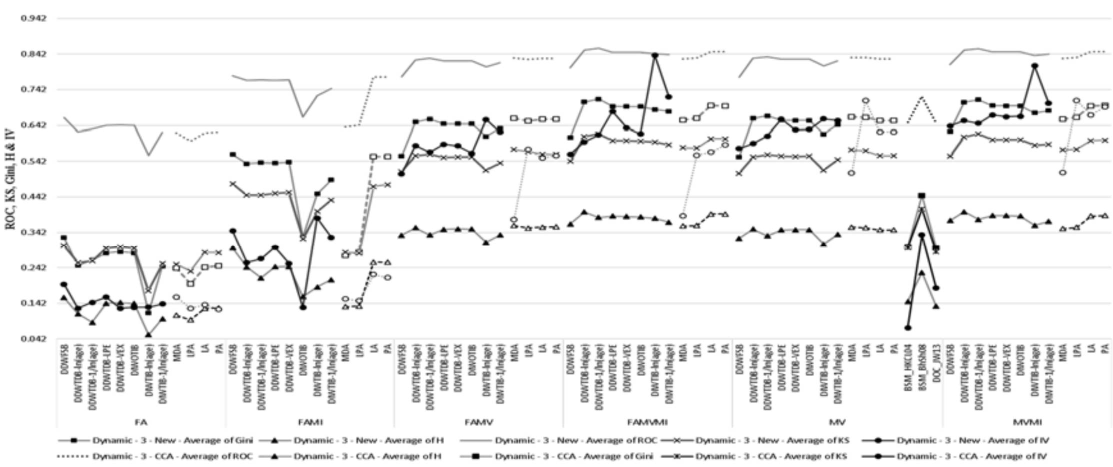
4.5.2 Mono-criterion Performance Evaluation of Distress Prediction Models

In order to answer the first question about the effect of information on the performance of distress models, I employ different combinations of information such as financial accounting (FA), financial accounting and market variables (FAMV), financial accounting and macroeconomic indicators (FAMI), financial accounting, market variables and macroeconomic indicators (FAMVMI), market variables (MV), and market variables and macroeconomic indicators (MVMI) to fed models.

When the availability of information is limited to accounting information (e.g., situations, where firms under evaluation are not listed on stock exchanges and macroeconomic information is not available or not reliable), the empirical results demonstrate that accounting information on its own can predict distressed firms. As one would expect, additional information enhances the ability of all models, whether static or dynamic, to discriminate between firms.

In fact, regardless of the selected performance criterion (i.e., Discriminatory Power, Correctness of Categorical Prediction, Calibration Accuracy) and its measures, empirical results demonstrate that most static and dynamic models perform better when fed with information beyond accounting ones – see, for example, Figure 4-5, and this enhancement in performance is statistically significant as demonstrated by a substantially large number of one-tailed t-tests of hypotheses involving all combinations of 15 modelling frameworks, 6 categories of information, and 15 measures of 3 performance criteria, where the Null hypothesis H_0 is: Average performance of modelling framework X fed with information category $Y \leq$ Average performance of modelling framework X' fed with information category Y' – see Appendix 4-A for an illustrative example of the typical outcome of these hypothesis tests. In addition, market information (e.g., (log) stock prices, (log) excess returns, volatility of stock returns (unsystematic risk), firm size as proxied by log (number of outstanding shares × year end share price / total market value), its market value, or market value of assets to total liabilities) on its own informs models better than accounting information on its own. However, market and macroeconomic information combined slightly enhance the performance of distress prediction models whether static or dynamic. Furthermore, the feature selection procedure and empirical results suggest that the choice of how a specific piece of information is modelled affects its relevance in adding value to a prediction model. In fact, for example, with respect to the market information category, log (price) is a better modelling choice compared to the price itself and excess return is generally better than log (price).

Figure 4-5: Measures of Discriminatory Power (ROC, H, Gini, KS, IV) of New Models Designed in Different Static and Dynamic Frameworks and Fed with 3-Year Information



In regard to the second question, which considers the out-of-sample performance of dynamic distress prediction models compare to the out-of-sample performance of static ones on sample period length, empirical evidence suggests that the out-of-sample implementation of static models within a rolling horizon framework overcomes the priori limitation of their static nature. In fact, under several combinations of categories of information (e.g., FAMV, FAMVMI, MV), the performance of static models is comparable to the performance of the dynamic ones across all measures of all criteria. This finding suggests that static models are not to be discarded and explains why static models are popular amongst practitioners – see, for example, Figure 4-5, Figure 4-7, Figure 4-9, and Figure 4-10. Also, the performance of four static models is consistent across different combinations of information categories for all measures of all criteria except for information value (IV) and Type I error – see, for example, Figure 4-6 and Figure 4-9. With respect to ROC, H, Gini, and KS, as measures of discriminatory power, LA and PA models fed with 3-year MVMI information outperform other static models. For example, the ROC for PA and LA is 0.8486 and 0.8484, respectively. However, considering IV as a measure of discriminatory power, LPA models fed with 3-year MV and MVMI seem to deliver the best performance, i.e., 0.711 whereas PA fed with 3-year FA is the worst performer. However, when static modelling frameworks are fed with both financial accounting and macroeconomic information, there is a clear difference in discriminatory power which suggests that macroeconomic information enhances the performance of LA and PA for discriminatory power measures– see, for example, Figure 4-6, and Appendix 4-D.

Figure 4-6: Measures of Discriminatory Power (ROC, H, Gini, KS, IV) of New Static Models Designed in Different Frameworks

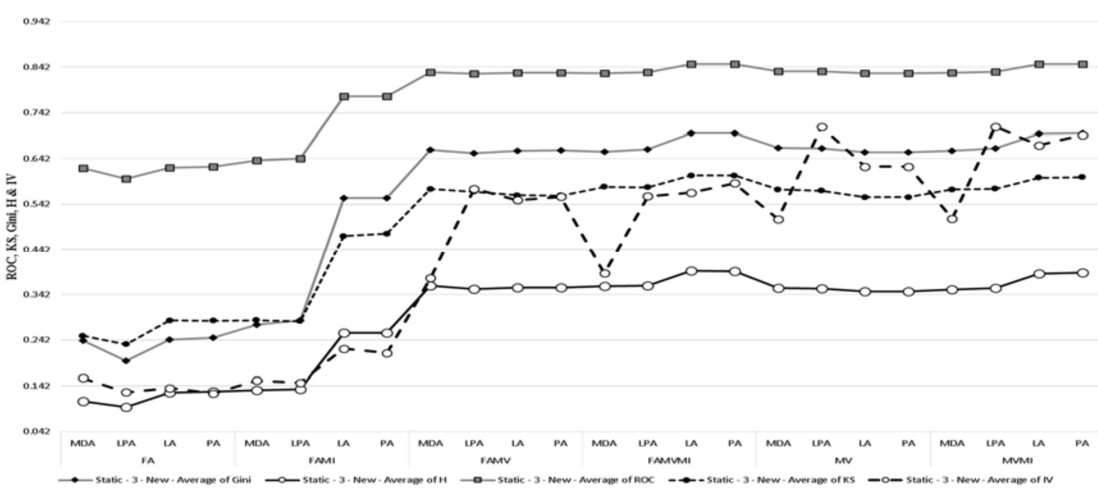
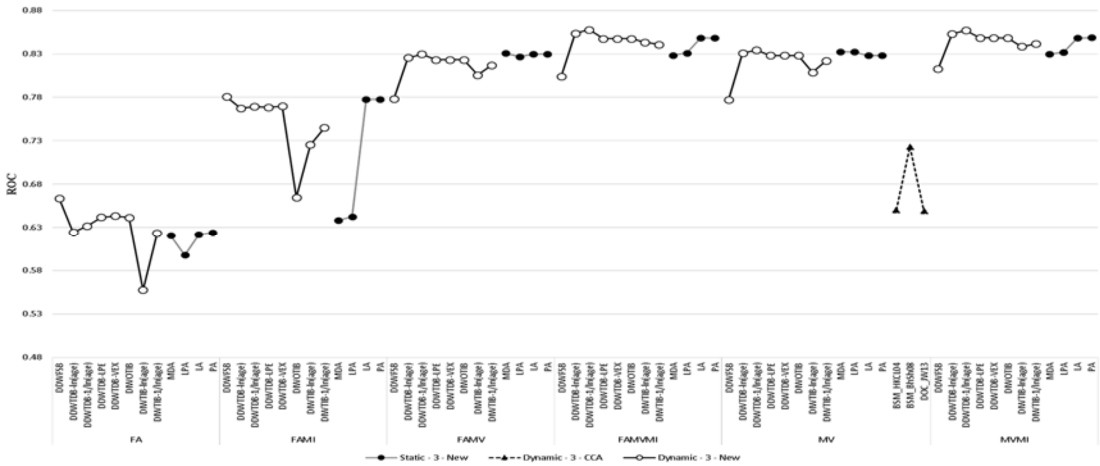


Figure 4-7: ROC of New Models Designed in Different Static and Dynamic Frameworks Fed with 3-year Information



Considering static models and with respect to T1 error, as a measure of correctness of categorical prediction criterion, MDA and LPA models seem to deliver the best performance, whereas PA is the worst performer. Also, PA performance suggests that this modelling framework is good at properly classifying healthy firms (i.e., it has the smallest T2 error), but relatively speaking, it poorly classifies the distressed ones (i.e., it has the largest Type I error) – see, for example, Figure 4-8. Amongst all new static models, MDA and LPA models fed with 10-year MV information have the lowest T2 error, 16.45% - see, Figure 4-13 and Appendix 4-C.

Figure 4-8: Measures of Correctness of Categorical Prediction of New Static Models Designed in Different Frameworks

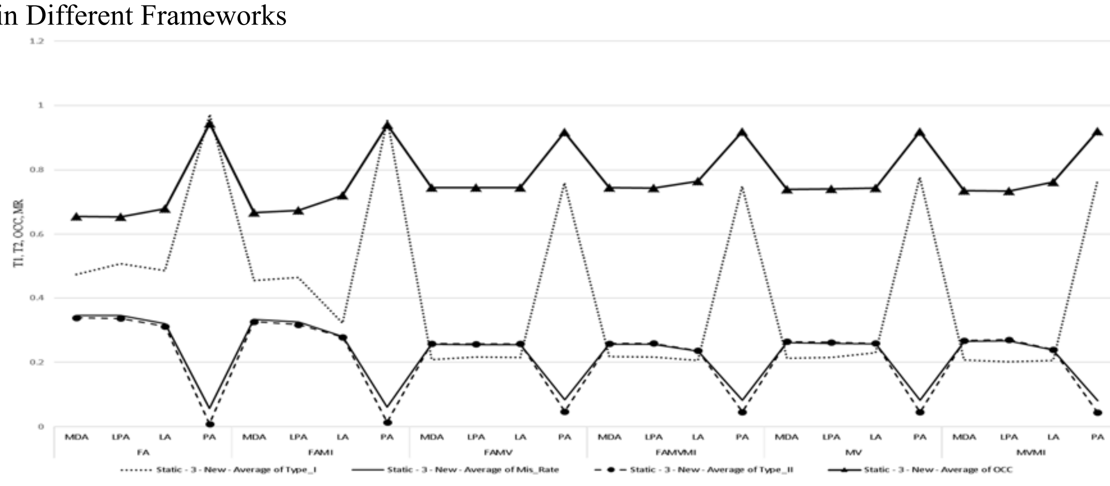
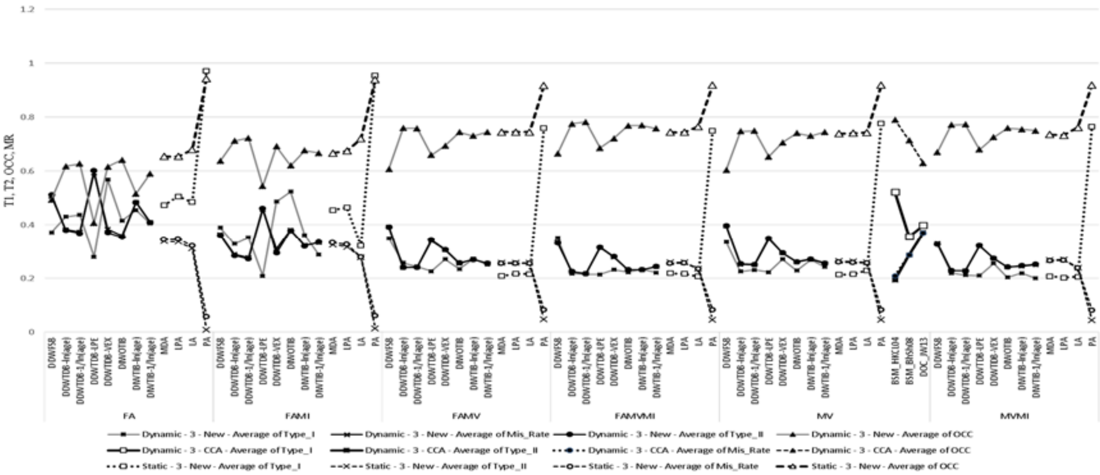


Figure 4-9: Correctness of Categorical Prediction of New Models Designed in Different Static and Dynamic Frameworks



With respect to Pseudo-R^2 and Log Likelihood, as measures of information content, LA and PA outperform other static models when fed with accounting and macroeconomic information – see, for example, Figure 4-10, however, MDA stands out as the best model when MVMI information is used. On the other hand, with respect to measures of calibration accuracy, such as Brier score, LPA outperforms other static models when fed with market information - see, for example, Figure 4-11.

Much like static models, empirical results suggest that dynamic models perform better when fed with information beyond accounting one; in fact, the performance of dynamic models across most measures of the three criteria under consideration is not only further enhanced when market information is taken on board, but it is consistent across all combinations of categories of information that include market variables – see, for example, Figure 4-5.

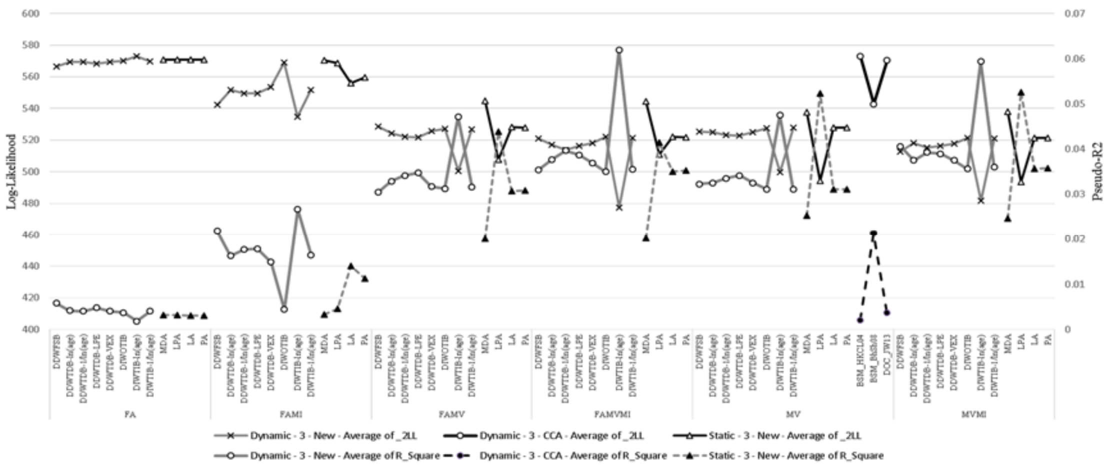
With respect to OCC, T2 and MR, as measures of correctness of categorical prediction, DDWTDB_1/ln(age) and DDWTDB_ln(age) (baseline is ignored or equal to 1, and 1/ln(age) and ln(age) are explanatory variables) are the best and second best performers, followed by DIWTIB_1/ln(age) and DIWTIB_ln(age) (1/ln(age) and ln(age) are baselines or intercepts) as average performers, and DDWFSB and DDWTDB_LPE being the worst ones – see, for example, Figure 4-9. Note however that, with respect to T1, DDWTDB_LPE is the best performer or amongst the best performers regardless of the information categories considered. On the other hand, DDWFSB and DDWTDB_VEX are, as expected, being the worst for any combination of information categories that includes market information; however, when market information is not considered, DDWFSB's performance improves while DDWTDB_VEX's performance remains weak – see Figure 4-9. Considering CCA models, under OCC, T2 and MR, BSM_HKCL_2004 outperforms BSM_BhSh_2008 and DOC_JW_2013. Also, under T1 error, BhSh_2008 performs better than other CCA models.

With respect to ROC, H, Gini and KS, as measures of discriminatory power, DDWTDB_1/ln(age) fed with FAMVMI and MVMI are the best performers amongst dynamic models, whereas DIWTIB-ln(age) and DDWFSB are the worst performers – see Figure 4-5. The findings suggest that majority of dynamic modes that fed with FA perform weakly on discriminatory power. Also, dynamic models perform better when fed with information beyond accounting ones. Note that the performance of CCA models is the worst between all dynamic models. The performance of BSM_BhSh_2008 is better than other CCA models considering all measures of discriminatory power.

Regarding the information content, as measured by Pseudo-R^2 and Log likelihood, DIWTIB_ln(age) (respectively, DDWFSB) outperform (respectively, underperform) other dynamic models. Findings suggest that using market variable information enhances the performance of models on information content– see, for example, Figure 4-10. Also, considering CCA models, BSM_HKCL_2004 outperforms BSM_BhSh_2008 and DOC_JW_2013.

With respect to the calibration accuracy, under its BS measure, the performance of DIWTIB_1/ln(age) (respectively, DIWTIB_ln(age)) models fed with market variables outperform (respectively, underperform) other dynamic models - see, for example, Figure 4-11.

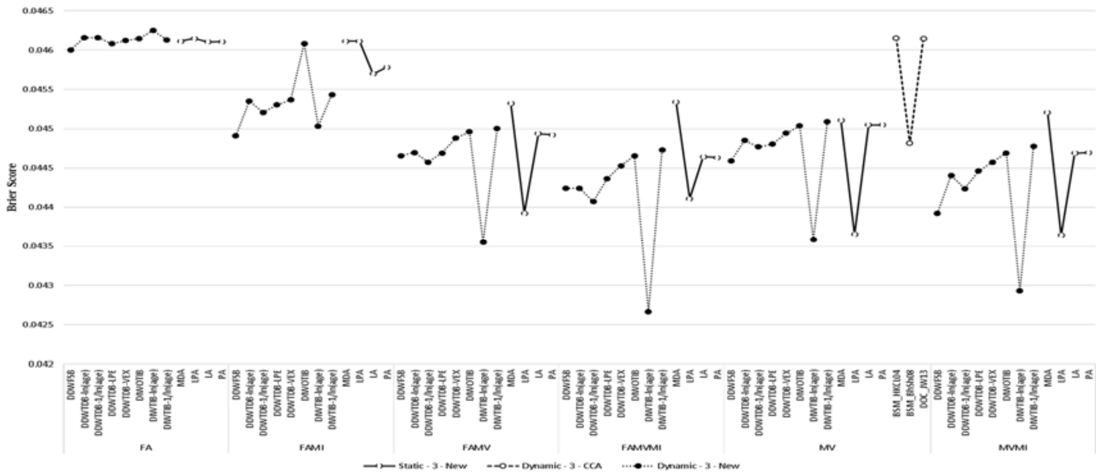
Figure 4-10: Log-likelihood and Pseudo-R^2 of New Dynamic and Static Models Fed with Different Type of Information



However, feeding dynamic frameworks with information beyond accounting one enhances their calibration accuracy, which suggests that macroeconomic and market information improve the performance of models. Moreover, considering CCA models, the calibration accuracy of BSM_BhSh_2008 is the best.

With respect to both static and dynamic models, under the correctness of categorical prediction criterion, the performance profiles of both static and dynamic models are consistent across different combinations of information categories, however they deliver different performances on different performance measures except for T2 and MR for which both static and dynamic models deliver the same average performance figures – see Figure 4-9 and Appendix 4-C. This latter empirical finding is explained by the fact that MR is a weighted combination of T1 and T2 errors and healthy firms count for most firms in my sample. However, although T1 and OCC are consistent in the way they drive performance, they deliver different figures as expected. One notable behaviour in performance is that of PA fed with all combinations of information being the best performer amongst all static and dynamic models with respect to T2 error, MR, and OCC; whereas its performance is consistently the worst under T1 errors, on the one hand, and DDWFSB fed with all combinations of information being the worst performer across T2 error, MR, and OCC; whereas the best performer under T1 error, on the other hand.

Figure 4-11: Brier Score of New Dynamic and Static Models Fed with Different Type of Information



Under the discriminatory power criterion, the performance profiles of both static and dynamic models are also consistent across different combinations of information categories – see, Figure 4-5 and Appendix 4-D. Furthermore, their performance is similar for all measures of discriminatory power except for information value (IV). DDWTDB_1/ln(age) fed with MVMI is the best performer between all dynamic and static models. Also, taking to account MVMI in addition to FA enhance the performance of PA model to become the best amongst static models.

As to the calibration accuracy criterion, under measures of both information content and quality of fit, the dynamic model DIWTIB_ln(age) fed with FAMVMI and MVMI has the highest Pseudo-R^2, the lowest Log Likelihood, and the lowest Brier score; therefore, it outperforms all other models whether static or dynamic – see for example; Figure 4-10, Figure 4-11 and Appendix 4-E. However, market information boosts LPA models performance to become the best amongst static models. In sum, although accounting, market and macroeconomic information are correlated to varying degrees over time, the market information proved to be the most informative prediction-wise.

To conclude the comparative analysis of static and dynamic models, I would like to stress out that, in general, static modelling frameworks are as good performers as dynamic ones when implemented under a dynamic scheme. This conclusion suggests that in case of dynamic frameworks, the design of models along with the type of information they are fed with requiring more attention from the academic community to perform to the standard it is expected, on the one hand, and to become a real contender for practitioners, on the other hand.

The third research question is about the effect of the length of training sample on the performance of models. Under the discriminatory power criterion, a comparison of models under different lengths of the training sample revealed that their empirical performance when market information is taken account of is not significantly affected, except for DDWFSB. In fact, the performance of DDWFSB deteriorates with a longer time window of the training sample – see, Figure 4-12. This result could be linked to the efficient market hypothesis which claims that existing stock prices contain and reflect all relevant information. Therefore, taking to account market information, a short length of training sample is enough to predict failure. However, when market information is not considered, the performance of models depends on to varying extents on the length of the training sample and thus their historical information needs might become lower or higher; e.g., dynamic models fed with 5-year training sample tend to outperform 3-year and 10-year models.

Under the discriminatory power measures, both static and dynamic models that use 5-year training sample outperform the models that use 3-year and 10-year training samples, respectively. This difference of performance is more significant for dynamic models – see, Figure 4-12.

Under the correctness of categorical prediction criterion, a longer time window of the training sample improves the performance of both static and dynamic models under T1 – see, Figure 4-13. However, under T2, MR, and OCC, a shorter time window of the training sample improves the performance of both static and dynamic models – see, Figure 4-14. In sum, under T1, both static and dynamic modelling frameworks require more historical information than what is required under T2, MR and OCC for a good performance.

Figure 4-12: ROC of New Models Fed with Different Length and Type of Information

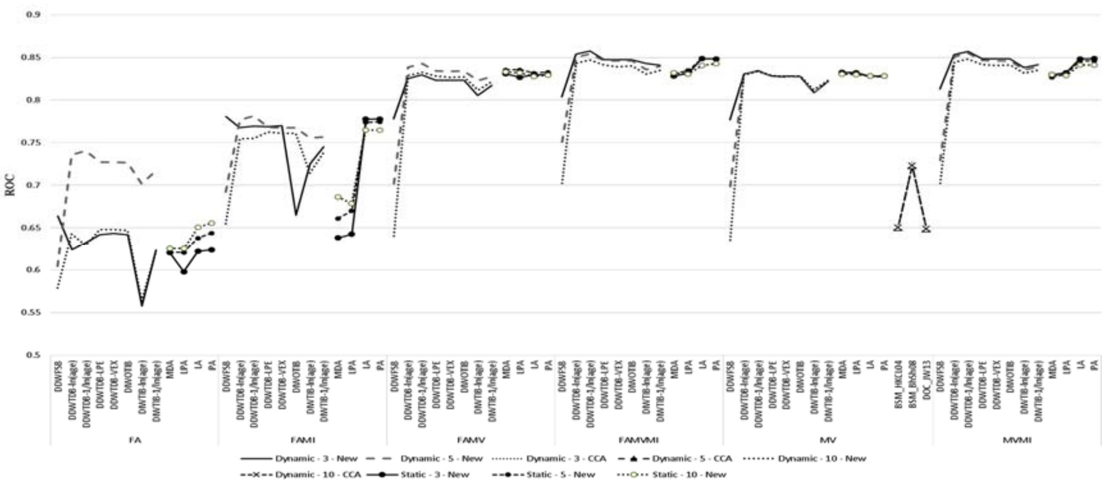
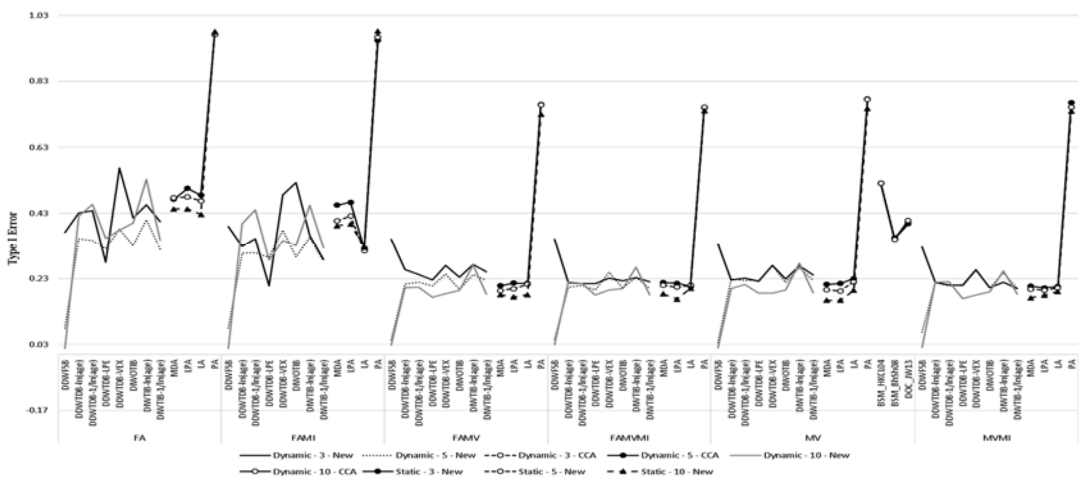


Figure 4-13: T1 Error of New Models Fed with Different Length and Type of Information



Under the information content criterion and its measures; namely, Pseudo-R^2 and Log likelihood, most dynamic models fed with 5-year training sample outperform 3-year and 10-year trained models, respectively. Also, most of static models fed with 3-year training sample outperform 5-year and 10-year models, respectively– see, for example, Figure 4-15.

.

Figure 4-14: T2 Error of New Models Fed with Different Length and Type of Information

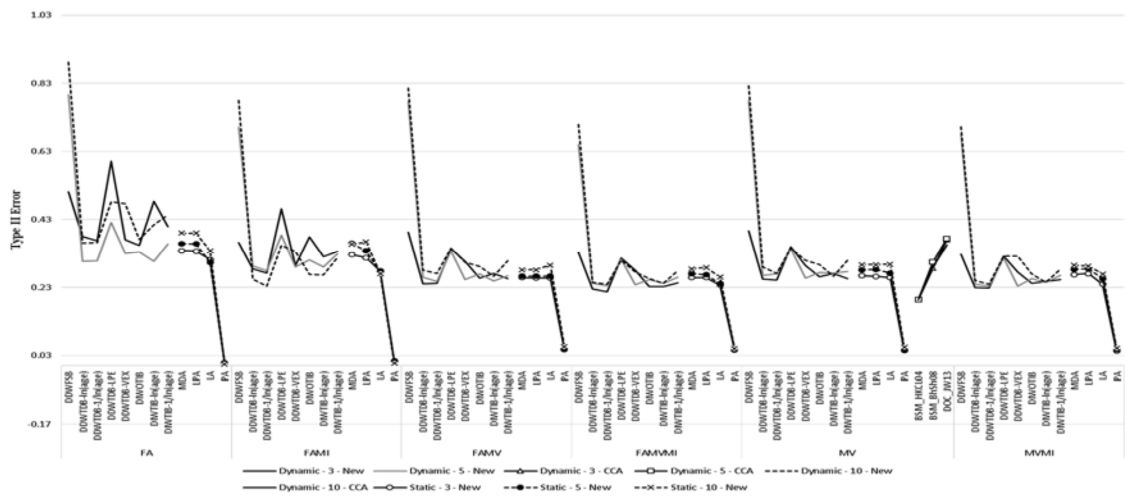
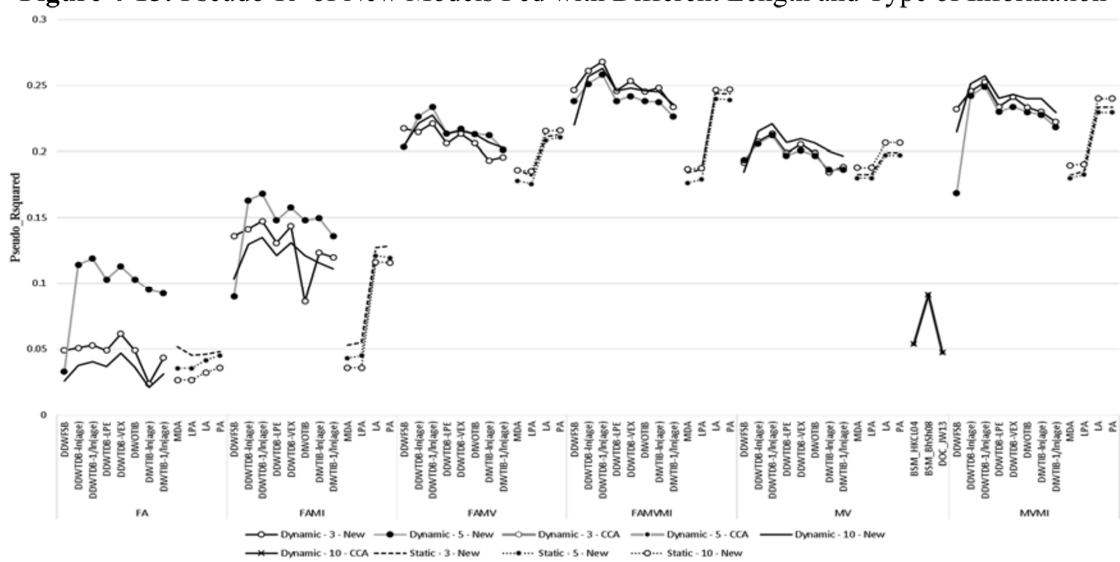


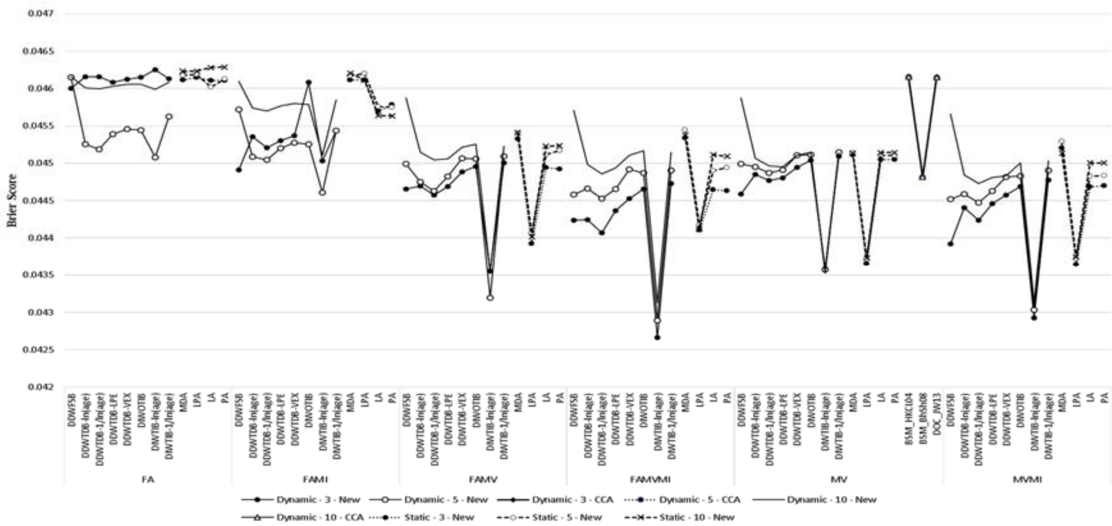
Figure 4-15: Pseudo-R^2 of New Models Fed with Different Length and Type of Information



Under Brier score, as a calibration accuracy measure, the dynamic models fed with 5-year training sample perform better in average than those models fed with 3-year and 10-year training samples. Also, in average, the static models fed with 3-year information outperform those static models that are fed with 5-year and 10-year training samples. However, when static and dynamic models are fed with market information, a shorter time window of the training sample improves their performance – see, Figure 4-16.

As suggested by the unidimensional ranking of distress prediction models, considering different performance criteria and measures, there are considerable conflicts and ties in the ranking of models. Therefore, considering multiple criteria, one cannot make an informed decision as to which model performs best. Although I insist that unidimensional rankings are not to be discarded, I would like to propose a dynamic multi-criteria assessment, which provides a single ranking under multiple criteria.

Figure 4-16: Brier Score of New Models Fed with Different Length and Type of Information



4.5.3 Multi-criteria Performance Evaluation of Distress Prediction Models

In this study, I developed 216 new models using 12 forecasting frameworks (i.e., MDA, LPA, LA, PA, DDWFSB, DDWTDB_ln(age), DDWTDB_1/ln(age), DDWTDB_LPE, DDWTDB_VEX, DDWOTIB, DDWTIB_ln(age), and DDWTIB_1/ln(age)) that are fed with 6 groups of information (i.e., FA, FAMI, FAMV, FAMVMI, MV, and MVMI) using 3 different training periods (i.e., 3, 5 and 10-year training samples). Also, I used 3 contingent claims analysis models (i.e., BSM_HKCL_2004, BSM_BhSh_2008 and DOC_JW_2013) to find out the probability of distress of firms. As mentioned above, the advantage of the multi-criteria framework is that it facilitates taking multiple performance criteria into account, which results in a comprehensive performance evaluation. Also, someone can present and monitor the performance of models over time. Depending on the preference of practitioners, alternative measures could be selected for each criterion. Here, I present the results of two rounds of multi-criteria evaluation.

4.5.3.1 Round one - Inputs T1, BS and outputs Pseudo-R^2, ROC

In the first round of multi-criteria assessment using Malmquist DEA, I used T1 error (a measure of correctness of categorical prediction) and Brier score (a measure of the quality of fit) as inputs, and Pseudo-R^2 (a measure of information content) and ROC (a measure of discriminatory power) as outputs. Table 4.5 presents the ranking of the best 20 and the worst 10 models out of 216 new models. For easier comparison, I rank the overall efficiency of models during a 15-year period using a total rank score. The results suggest that models developed in dynamic frameworks and fed with FAMV and FAMVMI features outperform other models. Also, models developed in static frameworks and fed with FA and FAMI features are amongst the worst performers. To save space, I only present the multi-criteria performance evaluation of models fed with FAMVMI, since models developed using all data available present better performance.

Table 4.6 provides the rankings of models based on the estimated efficiency scores using multi-criteria framework during period 2000 to 2014. With respect to the performance of models fed with 3-year training sample, DDWTDB_1/ln(age) outperforms other models, following by DDWTDB_ln(age) and DIWTIB_ln(age) which are the second and third best performers. In regards to the models fed with either 5 or 10-year training samples, DDWFSB, DDWTDB_1/ln(age), and DDWTDB_ln(age) are the best performers.

More, comparing all models fed with 3, 5 and 10-year training sample, dynamic models perform better than static models. DDWFSB fed with 5-year information has the best performance over the 15-year period, following by other duration dependent models that use ln(age) or 1/ln(age) as baseline rate. In addition, amongst the static models, LA outperforms others.

Table 4.5: The Yearly Rank of Top 20 Models Using Multi-Criteria Evaluation Framework

Framework	Feature	Training Period	Type	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	Point	Overall rank
Panel A: The best 20 models																				
DDWFSB	MVMI	10	Dynamic	15	15	38	5	17	9	41	39	2	7	11	4	5	2	6	216	**1**
DDWFSB	FAMVMI	5	Dynamic	8	6	25	29	34	12	6	3	30	6	51	1	6	27	1	245	**2**
DDWTDB-1/ln(age)	FAMVMI	3	Dynamic	2	45	55	40	37	2	4	9	13	12	35	10	10	5	23	302	**3**
DDWTDB-1/ln(age)	FAMVMI	10	Dynamic	32	4	7	12	1	5	29	14	23	20	32	24	22	21	68	314	**4**
DDWTDB-1/ln(age)	FAMVMI	5	Dynamic	11	1	12	23	24	1	26	15	6	18	34	39	21	45	43	319	**5**
DDWFSB	FAMVMI	10	Dynamic	14	142	20	1	16	8	35	60	5	9	9	3	3	4	5	334	**6**
DDWTDB-ln(age)	FAMVMI	3	Dynamic	12	51	58	48	45	4	5	19	18	13	44	19	13	6	29	384	**7**
DDWTDB-1/ln(age)	MVMI	10	Dynamic	33	27	8	15	18	18	27	17	38	24	36	53	20	26	66	426	**8**
DDWTDB-1/ln(age)	MVMI	3	Dynamic	7	24	55	40	43	13	9	23	27	17	60	62	25	7	16	428	**9**
DDWTDB-1/ln(age)	MVMI	5	Dynamic	20	13	24	31	31	28	43	13	9	21	41	59	29	41	35	438	**10**
DDWTDB-ln(age)	FAMVMI	10	Dynamic	34	33	10	17	6	16	38	25	28	27	37	40	35	28	71	445	**11**
DIWTIB-ln(age)	FAMVMI	3	Dynamic	3	62	42	33	66	6	8	10	16	16	52	58	16	14	49	451	**12**
DDWTDB-VEX	FAMVMI	3	Dynamic	25	59	39	42	79	22	1	48	26	26	13	23	12	16	32	463	**13**
DDWTDB-ln(age)	FAMVMI	5	Dynamic	16	36	14	24	30	17	44	18	7	22	42	64	30	65	45	474	**14**
DDWTDB-LPE	FAMVMI	3	Dynamic	30	53	79	105	8	38	25	41	39	11	12	12	17	9	14	493	**15**
DDWTDB-LPE	FAMVMI	10	Dynamic	45	54	19	13	2	27	54	40	25	31	20	17	61	32	60	500	**16**
DDWTDB-ln(age)	MVMI	10	Dynamic	26	34	11	19	25	21	30	29	47	29	39	67	27	31	79	514	**17**
DDWTDB-VEX	MVMI	10	Dynamic	51	37	4	14	28	47	56	4	32	38	1	35	51	56	78	532	**18**
DDWFSB	MV	5	Dynamic	37	10	13	70	21	15	14	45	93	23	111	6	4	79	4	545	**19**
DDWTDB-LPE	FAMVMI	5	Dynamic	1	23	27	67	5	31	78	42	17	44	2	31	60	75	44	547	**20**
To save the space other models are not presented																				
Panel B: the worst 10 models																				
LPA	FAMI	10	Static	203	190	194	213	200	204	201	199	196	200	206	201	207	207	206	3027	**206**
PA	FA	10	Static	216	198	190	203	193	201	215	204	209	196	210	204	199	201	202	3041	**207**
DIWOTIB	FA	10	Dynamic	210	202	200	187	197	207	202	210	206	203	208	207	200	204	203	3046	**208**
MDA	FA	5	Static	195	188	196	206	204	209	204	202	202	214	200	214	214	209	210	3067	**209**
LA	FA	10	Static	215	211	200	211	205	200	214	208	200	201	209	202	197	196	203	3072	**210**
LPA	FA	5	Static	196	191	195	207	207	208	205	201	201	214	201	215	213	210	211	3075	**211**
DIWTIB-1/ln(age)	FA	10	Dynamic	213	210	205	190	201	210	210	214	214	207	212	210	208	208	207	3119	**212**
LPA	FA	10	Static	212	199	203	216	209	212	213	207	210	212	214	212	212	212	213	3156	**213**
MDA	FA	10	Static	211	207	204	215	208	211	212	206	211	211	213	211	211	213	213	3157	**214**
DIWTIB-ln(age)	FA	3	Dynamic	214	213	212	204	211	213	209	212	212	206	211	213	216	214	215	3175	**215**
DIWTIB-ln(age)	FA	10	Dynamic	208	214	207	189	210	214	216	215	216	216	216	216	215	216	216	3184	**216**

Table 4.6: 1st Round of Multi-Criteria Ranking of Models Fed with FAMVMI Information

Panel A

Models	Training Period	Framework	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	Total Rank
DDWFSB	5	Dynamic	5	1	14	14	13	9	4	1	17	1	19	1	2	12	1	**1**
DDWTDB_1/ln(age)	3	Dynamic	1	15	25	17	14	1	2	3	7	4	11	3	4	2	7	**2**
DDWFSB	10	Dynamic	9	36	13	1	7	8	14	24	1	2	2	2	1	1	2	**3**
DDWTDB_1/ln(age)	5	Dynamic	7	1	6	9	8	1	12	7	2	7	10	20	11	20	13	**4**
DDWTDB_1/ln(age)	10	Dynamic	15	1	4	3	1	5	13	6	13	8	9	14	12	10	23	**5**
DDWTDB_ln(age)	3	Dynamic	8	17	26	21	18	4	3	9	10	5	15	11	6	3	9	**6**
DDWTDB_ln(age)	10	Dynamic	16	10	5	6	5	10	16	11	16	12	12	21	16	13	24	**7**
DDWTDB_ln(age)	5	Dynamic	10	12	7	10	11	11	19	8	3	9	14	27	14	23	15	**7**
DIWTIB_ln(age)	3	Dynamic	1	25	21	16	25	6	6	4	8	6	20	25	8	6	19	**9**
DDWTDB_LPE	3	Dynamic	14	19	29	29	6	23	11	17	21	3	3	5	9	5	3	**10**
DDWTDB_VEX	3	Dynamic	12	23	20	18	28	13	1	22	15	11	4	13	5	7	10	**11**
DDWTDB_LPE	10	Dynamic	20	20	12	4	1	16	20	16	14	13	5	9	26	15	22	**12**
DIWTIB_ln(age)	5	Dynamic	6	8	3	12	20	1	17	5	5	10	26	31	18	31	26	**13**
DIWTIB_ln(age)	10	Dynamic	16	1	1	2	10	7	9	10	20	17	28	29	15	21	34	**14**
DDWTDB_VEX	10	Dynamic	21	13	2	5	9	18	23	2	12	15	25	8	24	19	27	**15**
DDWTDB_LPE	5	Dynamic	1	7	16	24	4	17	30	18	9	16	1	16	25	26	14	**16**
LA	10	Static	19	18	10	7	16	25	18	15	19	19	6	7	19	17	28	**17**
DIWOTIB	10	Dynamic	22	16	11	8	12	20	21	21	22	18	8	6	21	16	25	**18**
DDWFSB	3	Dynamic	1	27	22	34	1	12	5	36	33	20	36	4	3	14	5	**19**
LA	5	Static	23	14	17	20	21	21	22	14	4	21	13	15	20	24	12	**20**
DIWOTIB	3	Dynamic	11	24	28	26	29	24	10	19	24	23	17	10	7	4	8	**21**
DDWTDB_VEX	5	Dynamic	27	21	8	15	17	14	28	12	11	14	18	17	22	30	16	**22**
LA	3	Static	13	22	24	23	23	19	7	13	26	22	27	28	13	9	4	**23**
DIWOTIB	5	Dynamic	28	6	15	13	22	15	29	20	6	24	16	18	23	27	17	**24**
DIWTIB_1/ln(age)	10	Dynamic	26	11	9	11	19	28	27	23	23	25	7	12	27	22	29	**25**
DIWTIB_1/ln(age)	3	Dynamic	18	26	30	27	36	30	15	28	29	27	21	23	10	8	6	**26**
PA	3	Static	24	31	27	28	26	22	8	25	31	32	31	30	17	11	11	**27**
DIWTIB_1/ln(age)	5	Dynamic	30	9	19	19	30	26	32	27	18	26	24	26	29	29	20	**28**
PA	10	Static	25	30	18	22	15	29	24	26	27	28	22	19	28	25	30	**29**
PA	5	Static	29	29	23	25	24	27	31	29	25	31	23	22	30	28	18	**30**
LPA	10	Static	36	5	31	30	27	34	35	30	32	29	30	24	35	32	33	**31**
LPA	3	Static	33	32	33	33	34	31	26	34	35	35	34	35	31	18	21	**32**
LPA	5	Static	34	28	35	32	31	33	33	32	30	33	29	33	32	34	32	**33**
MDA	10	Static	35	34	32	31	32	35	34	31	34	30	32	32	36	35	35	**34**
MDA	3	Static	32	33	34	35	35	32	25	35	36	36	35	36	33	33	31	**35**
MDA	5	Static	31	35	36	36	33	36	36	33	28	34	33	34	34	36	36	**36**

Panel B

		2003				2009				2013			
		All	Dynamic	Static	z	All	Dynamic	Static	z	All	Dynamic	Static	z
Training Period	3	26	24	30	1.52	19	12	31	2.37**	10	6	18	2.37**
	5	19	15	28	2.54**	19	13	30	2.37**	27	25	31	1.35
	10	11	5	23	2.38**	18	14	27	2.54**	19	15	27	2.20**
	x^2	11.82**	17.80***	1.5		0.98	0.86	3.23		15.03***	14.59***	3.23	

Regarding the third research question that investigates the effect of the length of training sample on the performance of model, numerical results in Figure 4-6 indicate that the dynamic model developed with 3-year training sample outperform the 5-year and 10-year dynamic model in the same framework. Contrary to the dynamic models, most of the static models fed with 10-year training information perform better than the 5-year and 3-year trained models in the same framework. The reason is that dynamic models by nature are capable to incorporate time-varying covariates, and therefore adopt to the changes in economic and firm operational environment. This characteristic could be a drawback for dynamic models, since taking into account a long period training sample could be misleading because of changes in economic environment and financial ratios patterns. On the other hand, static models are not capable to incorporate time-varying covariates, and as such employing more firm-year observations as training sample would improve their performance.

With regards to the fourth research question, which considers the performance of models in predicting distress during financial crises, I took into account the rankings of models in 2003, 2009 and 2013, say high distress rate (HDR) years, since the distress rates in these years are higher than other years because of financial crises in the years before – see, Figure 4-3. The panel B of Table 4-6 compares the average performance rankings of all dynamic and static prediction models during HDR years. The results suggest that first, based on z*-statistics* of Wilcoxon rank-sum test, in average, the dynamic models outperform the static ones in predicting distress during HDR years. Second, numerical results indicate that in 2003, the 10-year trained distress models outperform 5- and 3-year trained models, respectively. Also, in 2013, the 3-year trained distress models outperform 5, and 10-year trained models. However, in 2009, the average ranking of models with different length of training sample is almost the same. The results of x^2-statistics of Kruskal-Wallis test indicate that there are significant (respectively, no significant) differences between the performance of dynamic (respectively, static) models fed with 3,5 or 10-year information in HDR years of 2003 and 2013. However, in 2009, there is no significant difference between the performance of 3,5 or 10-year trained models.

4.5.3.2 Round two – Inputs T2, BS and Outputs Pseudo-R^2, ROC

In the second round of multi-criteria assessment using Malmquist, I used T2 error (as a measure of correctness of categorical prediction) and Brier score (as a measure of the quality of fit) as inputs, and Pseudo-R^2 (as a measure of information content) and ROC (as a measure of discriminatory power) as outputs.

Table 4-7 presents the rankings of models based on the estimated efficiency scores using multi-criteria framework during period 2000 to 2014. Opposite of the last round of multi-criteria assessment where the static models underperform dynamic ones, the second round of assessment indicates that PA outperforms other models. Regarding the third research question that investigate the effect of length of training samples on the performance of models, the PA models with all length of training sample outperform other models. However, the dynamic model of DDWTDB_1/ln(age) is ranked second followed by DDWTDB_ln(age) with respect to all length of training samples. Regarding all models fed with FAMVMI, the static model of PA fed with 5,3 and 10-year information outperform other models over a 15-year period. DDWTDB_1/ln(age) fed with 3,5-year training sample is the second-best performer, see, Table 4-7. Further, most of dynamic models developed with 3-year training sample outperform the 5-year and 10-year dynamic model in the same framework.

As regards the fourth research question that considers the performance of models in predicting distress during HDR years, panel B of Table 4-7 compares the average performance rankings of all dynamic and static prediction models in HDR years. The average ranking of models suggests that the dynamic models have a lower average ranking that indicates a better performance in compare to the static models during HDR years. However, *z-statistics* of Wilcoxon rank-sum test does not show a significant difference between the rankings of static and dynamic models during HDR years.

Table 4.7: 2nd Round of Multi-Criteria Ranking of Models Fed with FAMVMI Information

Panel A

Models	Training Period	Framework	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	Total Rank
PA	5	Static	6	1	3	4	2	1	4	3	3	2	1	1	3	2	1	**1**
PA	3	Static	1	7	1	2	4	4	5	2	1	1	3	3	1	1	2	**2**
PA	10	Static	1	1	2	3	3	5	3	4	2	3	2	2	2	3	3	**3**
DDWTDB_1/ln(age)	3	Dynamic	5	16	21	16	13	1	1	6	4	4	5	6	5	6	5	**4**
DDWTDB_1/ln(age)	5	Dynamic	15	10	8	9	7	1	16	5	5	7	8	20	13	15	9	**5**
DDWTDB_ln(age)	3	Dynamic	9	24	23	13	17	7	2	8	6	5	10	13	7	8	7	**6**
DDWTDB_1/ln(age)	10	Dynamic	11	1	6	7	5	12	13	11	10	8	13	18	14	13	19	**7**
DIWTIB_ln(age)	3	Dynamic	1	8	20	25	20	8	6	10	7	6	19	24	8	12	17	**8**
DDWTDB_ln(age)	5	Dynamic	18	13	10	14	9	6	18	9	8	10	12	25	16	18	15	**9**
DDWTDB_ln(age)	10	Dynamic	13	9	7	11	6	13	17	13	12	11	15	23	15	16	22	**10**
DIWTIB_ln(age)	10	Dynamic	13	1	5	8	8	10	10	12	11	12	23	27	20	25	31	**11**
DIWTIB_ln(age)	5	Dynamic	8	1	4	12	12	9	12	7	9	13	25	28	23	31	30	**12**
DDWTDB_VEX	3	Dynamic	7	22	27	20	31	18	8	16	25	14	24	4	4	5	4	**13**
DDWFSB	3	Dynamic	1	29	26	5	1	11	7	1	32	28	4	22	18	28	21	**14**
DIWOTIB	3	Dynamic	12	23	25	26	29	17	15	20	23	9	6	7	9	10	8	**15**
DDWTDB_VEX	5	Dynamic	24	18	19	10	15	20	29	27	14	21	9	5	12	4	12	**15**
LA	3	Static	17	19	22	27	11	14	11	14	21	17	21	26	10	7	6	**17**
LA	5	Static	22	17	12	18	10	16	25	21	15	15	7	11	21	24	11	**18**
DDWTDB_LPE	3	Dynamic	10	28	24	1	33	21	14	18	24	25	27	12	6	11	16	**19**
LA	10	Static	16	11	18	17	19	25	19	15	17	18	18	16	26	17	24	**20**
DIWOTIB	10	Dynamic	21	15	11	19	16	22	23	17	18	20	14	14	27	21	23	**21**
DDWTDB_LPE	5	Dynamic	28	21	13	6	32	15	24	22	13	22	28	9	17	19	14	**22**
DDWTDB_VEX	10	Dynamic	20	12	17	15	14	24	22	25	22	23	29	8	24	14	20	**23**
DDWTDB_LPE	10	Dynamic	23	14	9	23	28	19	20	19	20	24	22	10	19	20	25	**24**
DIWOTIB	5	Dynamic	25	20	14	24	18	23	31	24	16	19	11	15	22	22	13	**25**
DIWTIB_1/ln(age)	3	Dynamic	19	27	28	28	34	28	21	23	27	16	16	17	11	9	10	**26**
DIWTIB_1/ln(age)	5	Dynamic	27	26	15	22	30	27	32	28	19	27	17	21	28	26	18	**27**
DIWTIB_1/ln(age)	10	Dynamic	26	25	16	21	22	26	30	26	26	26	20	19	29	27	26	**28**
LPA	10	Static	36	6	31	29	25	31	34	29	30	29	32	30	34	30	33	**29**
DDWFSB	5	Dynamic	29	31	29	35	35	36	9	35	31	30	35	29	25	34	29	**30**
LPA	3	Static	31	35	30	32	24	29	27	33	34	35	30	35	30	23	27	**31**
LPA	5	Static	33	32	34	31	21	35	35	30	28	32	26	32	32	29	28	**32**
MDA	10	Static	34	30	33	30	26	34	33	31	33	31	33	31	35	32	34	**33**
MDA	3	Static	30	36	32	34	23	30	26	34	35	36	34	36	31	33	32	**34**
MDA	5	Static	32	34	35	33	27	33	36	32	29	33	31	33	33	35	35	**35**
DDWFSB	10	Dynamic	35	33	36	36	36	32	28	36	36	34	36	34	36	36	36	**36**

Panel B

		2003				2007				2013			
		All	Dynamic	Static	z	All	Dynamic	Static	z	All	Dynamic	Static	z
Training Period	3	19	16.75	24	1.36	16	13	22.25	1.02	13	11	16	0.17
	5	18.18	16.50	23	0.51	20	19	22.33	0.51	21	21	22	0.51
	10	18.25	18	20	0.34	19	20	20	0.17	20	22	21	0.17
	x^2	0.06	0.10	0.73		0.78	2.20	0.26		5.37*	7.28**	0.50	

Also, considering panel A of Table 4-7, most of the dynamic models outperform static ones in predicting distress. The only exception is PA models trained with 3,5 and 10-year information, which outperform other dynamic models.

Also, the numerical results indicate that there are no differences between the performance of static models fed with 3,5 or 10-year information in HDR years. Further, the numerical results of dynamic models suggest that during HDR years, the 3-year trained distress models outperform 5 and 10-year trained models. However, the x^2 statistics of Kruskal-Wallis test indicate that there are no significant differences between the performance of 3,5 or 10-year trained dynamic models in 2003 and 2009.

4.6 Conclusion

Prediction of corporate distress is crucial for many stakeholders and decision makers in finance and investment. Although many models have been designed to predict bankruptcy and distress, the relative performance evaluation of competing distress models remains an exercise that (1) is unidimensional in nature, which results in conflicting rankings of models from one performance criterion to another, and (2) is static in practice and, therefore, ignore monitoring the performance of models over time. In this study, I proposed a dynamic framework based on an orientation-free super-efficiency Malmquist DEA index, which provides a single ranking based on multiple performance criteria. This dynamic framework makes it possible for practitioners and academics to use one measure under each criterion to evaluate the performance of distress prediction models. In addition, I performed a comprehensive comparative analysis of the most cited static and dynamic distress prediction models. For this, I used several measures under four commonly employed criteria (i.e., the discriminatory power, the information content, the calibration accuracy, and the correctness of categorical prediction) in the literature. Furthermore, I addressed the following important questions: what is the effect of information on the performance of distress models? How does the out-of-sample performance of dynamic distress prediction model compare to the out-of-sample performance of static ones?

What is the effect of the length of training sample on the performance of models? Which models perform better in forecasting distress during the years with higher distress rate (HDR)?

My main findings suggest that firstly, the proposed multi-criteria dynamic framework provides a useful tool in evaluating the relative performance of distress prediction models over time. Secondly, in defiance of the unidimensional ranking, the multidimensional ranking of models provides more consistent results. However, regarding the inconsistency between rankings of models with respect to T1 and T2 errors (i.e. PA model), multi-criteria rankings of models using each of these two measures would also present inconsistency. Third, with respect to the main research questions, the empirical results suggest that most static and dynamic models perform better when fed with information beyond accounting ones. Also, dynamic models, specifically DDWTDB_1/ln(age) and DDWTDB_ln(age) are always amongst the best distress prediction models and show consistency in multi-criteria ranking using different combinations of measures. Further, regarding the effect of length of training sample on the performance of models, most of dynamic frameworks show that the model with 3-year training period in a framework outperforms the models with longer training periods (e.g., 5-year or 10-year) in that framework. However, the static models do not show straightforward trend with respect to the effect of length of training sample on their performance. Also, the empirical results suggest that, in average, dynamic models outperform static ones during the years with HDR.

One of the limitations of this research is space constraint and as such I restricted this study to financial distress as an event. Further, same as chapters 2 and 3, this chapter is restricted with respect to data, i.e. listed companies in LSE, and type of models, i.e. statistical models. Future research could incorporate other definitions of failure such as bankruptcy, debt restructuring, etc. Moreover, future studies could analyse the extent to which failure prediction models are generalised by taking into account the data from other countries.

Chapters 2,3 and 4 are mutual with respect to developing and analysing one-stage failure prediction models. The recent trend in the literature has focused on developing two-stage models where the management efficiency of firms is estimated in the first stage and retained as a feature in developing models in the second stage. Next chapter is allocated to this model developing technique.

References

Abdou H A and Pointon J (2011). Credit scoring, statistical techniques, and evaluation criteria: A review of the literature. *Intelligent Systems in Accounting, Finance and Management* 18 (2-3): 59–88.

Agarwal V and Taffler R (2008). Comparing the performance of market-based and accounting-based bankruptcy prediction models. *Journal of Banking & Finance* 32(8): 1541–1551.

Aktug R E (2014). A critique of the contingent claims approach to sovereign risk analysis. *Emerging Markets Finance and Trade* 50(1): 294–308.

Allison P D (1982). Discrete-time methods for the analysis of event histories. *Sociological methodology* 13(1): 61–98.

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

Edward A I (1993). *Corporate Financial Distress and Bankruptcy–A Complete Guide to Predicting & Avoiding Distress and Profiting from Bankruptcy*. Edit. John Wiley and Sons.

Altman E I (1982). Accounting implications of failure prediction models. *Journal of Accounting, Auditing, and Finance* 6(1): 4–19.

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

Altman E I, Haldeman R G and Narayanan, P. (1977). ZETA analysis: A new model to identify bankruptcy risk of corporations. *Journal of banking & finance* 1(1): 29–54.

Andersen P K (1992). Repeated assessment of risk factors in survival analysis. *Statistical Methods in Medical Research* 1(3): 297–315.

Anderson R (2007). *The credit scoring toolkit, Theory, and practice for retail credit risk management and decision automation*. OXFORD university press: New York.

Aziz M A and Dar H A (2006). Predicting corporate bankruptcy: where we stand? *Corporate governance* 6(1): 18–33.

Balcaen S and 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.

Bauer J and Agarwal V (2014). Are hazard models superior to traditional bankruptcy prediction approaches? A comprehensive test. *Journal of Banking & Finance* 40 (Mar): 432–442.

Beaver W H (1968). Alternative accounting measures as predictors of failure. *Accounting review* 43(1): 113–122.

Beaver W H (1966). Financial ratios as predictors of failure. *Journal of Accounting Research* 4: 71-111.

Beck N, Katz J N and Tucker R (1998). Taking time seriously: Time-series-cross-section analysis with a binary dependent variable. *American Journal of Political Science* 42(4): 1260–1288.

Bellotti T and Crook J (2009). Forecasting and stress testing credit card default using dynamic models. *International Journal of Forecasting* 29(4): 563-574.

Bellovary J L, Giacomino D E and Akers M D (2007). A review of bankruptcy prediction studies: 1930 to present. *Journal of Financial Education* 33(Dec): 1–42.

Berg, S. A., Førsund, F. R., & Jansen, E. S. (1992). Malmquist indices of productivity growth during the deregulation of Norwegian banking, 1980-89. *The Scandinavian Journal of Economics*, S211-S228.

Bharath S T and Shumway T (2008). Forecasting default with the Merton distance to default model. *Review of Financial Studies* 21(3): 1339–1369.

Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of political economy* 81(3): 637-654.

Blöchlinger, A., & Leippold, M. (2006). Economic benefit of powerful credit scoring. *Journal of Banking & Finance* 30(3): 851-873.

Blum M (1974). Failing company discriminant analysis. *Journal of Accounting Research* 12(1): 1–25.

Campbell J Y, Hilscher J and Szilagyi J (2008). In search of distress risk. *The Journal of Finance* 63(6): 2899–2939.

Charalambakis, E. C., & Garrett, I. (2016). On the prediction of financial distress in developed and emerging markets: Does the choice of accounting and market information matter? A comparison of UK and Indian Firms. *Review of Quantitative Finance and Accounting* 47(1):1-28.

Chava S and Jarrow R A (2004). Bankruptcy prediction with industry effects. *Review of Finance* 8(4): 537–569.

Charitou, A., Neophytou, E., & Charalambous, C. (2004). Predicting corporate failure: empirical evidence for the UK. *European Accounting Review* 13(3): 465-497.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European journal of operational research* 2(6),: 429-444.

Caves, D. W., Christensen, L. R., & Diewert, W. E. (1982). The economic theory of index numbers and the measurement of input, output, and productivity. *Econometrica: Journal of the Econometric Society*: 1393-1414.

Chen L S, Yen M F, Wu H M, Liao C S, Liou D M, Kuo H S and Chen T H H (2005). Predictive survival model with time-dependent prognostic factors: development of computer-aided SAS Macro program. *Journal of evaluation in clinical practice* 11(2): 181–193.

Chen M Y., 2011. Predicting corporate financial distress based on integration of decision tree classification and logistic regression. *Expert Systems with Applications* 38(9): 11261–11272.

Cleary S and Hebb G, 2016. An efficient and functional model for predicting bank distress: In and out of sample evidence. *Journal of Banking & Finance* 64 (Mar): 101–111.

Collins R A, Green R D (1982). Statistical methods for bankruptcy forecasting. *Journal of Economics and Business* 34(4): 349–354.

Cooper, W. W., Seiford, L. M., & Tone, K. (2006). Introduction to data envelopment analysis and its uses: with DEA-solver software and references. *Springer Science & Business Media*.

Cox D R (1972). Regression models and life-tables. *Journal of Royal Statistical Society: Methodological* 34(2): 187–220.

Crapp H R, Stevenson M (1987). Development of a method to assess the relevant variables and the probability of financial distress. *Australian journal of management* 12(2): 221–236.

Deakin E B (1972). A discriminant analysis of predictors of business failure. *Journal of Accounting Research* 10(2): 167–179.

du Jardin P (2015). Bankruptcy prediction using terminal failure processes. *European Journal of Operational Research* 242(1): 286–303.

du Jardin P and Séverin E (2012). Forecasting financial failure using a Kohonen map: A comparative study to improve model stability over time. *European Journal of Operational Research* 221(2): 378–396.

Eisenbeis R A (1977). Pitfalls in the application of discriminant analysis in business, finance, and economics. *The Journal of Finance* 32(3): 875-900.

Färe, R., Grosskopf, S., Lindgren, B., & Roos, P. (1992). Productivity changes in Swedish pharamacies 1980–1989: A non-parametric Malmquist approach. In *International Applications of Productivity and Efficiency Analysis* (pp. 81-97). Springer Netherlands.

Färe, R., Grosskopf, S., Lindgren, B., & Roos, P. (1994). Productivity developments in Swedish hospitals: a Malmquist output index approach. In *Data envelopment analysis: theory, methodology, and applications* (pp. 253-272). Springer Netherlands.

Fedorova E, Gilenko E and Dovzhenko S (2013). Bankruptcy prediction for Russian companies: Application of combined classifiers. *Expert Systems with Applications* 40(18): 7285–7293.

Geng R, Bose I and Chen X (2015). Prediction of financial distress: An empirical study of listed Chinese companies using data mining. *European Journal of Operational Research* 241(1): 236–247.

Gilbert L R, Menon K and Schwartz K B (1990). Predicting bankruptcy for firms in financial distress. *Journal of Business Finance & Accounting* 17(1): 161–171.

Grice, J. S., & Ingram, R. W. (2001). Tests of the generalizability of Altman's bankruptcy prediction model. *Journal of Business Research*, 54(1): 53-61.

Hamer M M (1983). Failure prediction: Sensitivity of classification accuracy to alternative statistical methods and variable sets. *Journal of Accounting and Public Policy* 2(4): 289–307.

Hand, D.J. and Anagnostopoulos, C. (2013). When is the area under the receiver operating characteristic curve an appropriate measure of classifier performance? *Pattern Recognition Letters* 34(5): 492-495.

Hillegeist S A, Keating E K, Cram D P and Lundstedt K G (2004). Assessing the probability of bankruptcy. *Review of Accounting Studies* 9(1): 5–34.

Huang, C., Dai, C. and Guo, M., 2015. A hybrid approach using two-level DEA for financial failure prediction and integrated SE-DEA and GCA for indicators selection. *Applied Mathematics and Computation*, 251: 431-441.

Jackson, R H and Wood A (2013). The performance of insolvency prediction and credit risk models in the UK: A comparative study. *The British Accounting Review* 45(3): 183–202.

Kim, M H and Partington G (2014). Dynamic forecasts of financial distress of Australian firms. *Australian Journal of Management* 40(1):135-160.

Laitinen, E K and Suvas A (2016). Financial distress prediction in an international context: Moderating effects of Hofstede’s original cultural dimensions. *Journal of Behavioural and Experimental Finance* 9 (Mar): 98–118.

Lane W R, Looney S W and Wansley J W (1986). An application of the Cox proportional hazards model to bank failure. *Journal of Banking & Finance* 10 (4): 511–531.

LeClere M J (2000). The occurrence and timing of events: Survival analysis applied to the study of financial distress. *Journal of Accounting Literature* 19 (Jan): 158-189.

Lennox C (1999). Identifying failing companies: a re-evaluation of the logit, probit and DA approaches. *Journal of Economics and Business* 51(4): 347–364.

Lo A W (1986). Logit versus discriminant analysis: A specification test and application to corporate bankruptcies. *Journal of Econometrics* 31(2): 151–178.

Luoma M and Laitinen E K (1991). *Survival analysis as a tool for company failure prediction*. Omega 19(6): 673–678.

Lyandres E and Zhdanov A (2013). Investment opportunities and bankruptcy prediction. *Journal of Financial Markets* 16(3): 439–476.

Malmquist, S. (1953). Index numbers and indifference surfaces. *Trabajos de estadística* 4(2): 209-242.

Martin D (1977). Early warning of bank failure: A logit regression approach. *Journal of Banking & Finance* 1(3): 249–276.

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

McLeay S (1986). Student's t and the distribution of financial ratios. *Journal of Business Finance & Accounting* 13(2): 209-222.

Mensah, Y. M. (1984). An examination of the stationarity of multivariate bankruptcy prediction models: A methodological study. *Journal of Accounting Research* 22(1) :380-395.

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

Meyer P A and Pifer H W (1970). Prediction of bank failures. *The Journal of Finance* 25(4): 853–868.

Mousavi M M, Ouenniche J, Xu B (2015). Performance evaluation of bankruptcy prediction models: An orientation-free super-efficiency DEA-based framework. *International Review of Financial Analysis* 42 (Dec): 64–75.

Neves J C, Vieira A (2006). Improving bankruptcy prediction with Hidden Layer Learning Vector Quantization. *European Accounting Review* 15(2): 253–271.

Ohlson J A (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of accounting research* 18(1): 109–131.

Pacheco J, Casado S, Núñez L (2009). A variable selection method based on Tabu search for logistic regression models. *European Journal of Operational Research* 199(2): 506–511.

Pacheco J, Casado S, Nuñez L (2007). Use of VNS and TS in classification: variable selection and determination of the linear discrimination function coefficients. *IMA Journal of Management Mathematics* 18(2): 191–206.

Pastor, J. T., & Lovell, C. K. (2005). A global Malmquist productivity index. *Economics Letters* 88(2): 266-271.

Platt, H. D., Platt, M. B., & Pedersen, J. G. (1994). Bankruptcy discrimination with real variables. *Journal of Business Finance & Accounting* 21(4): 491-510.

Platt H, Platt M (2012). Corporate board attributes and bankruptcy. *Journal of Business Research* 65(8): 1139–1143.

Press S J, Wilson S (1978). Choosing between Logistic regression and discriminant analysis. *Journal of the American Statistical Association* 73 (364): 699–705.

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.

Reisz A S, Perlich C (2007). A market-based framework for bankruptcy prediction. *Journal of Financial Stability* 3(2): 85–131.

Shumway T (2001). Forecasting bankruptcy more accurately: A simple hazard model. *Journal of Business* 74(1): 101–124.

Seiford, L.M. and Zhu, J., 1998. Stability regions for maintaining efficiency in data envelopment analysis. *European Journal of Operational Resea*rch, 108(1):127-139.

Seiford, L.M. and Zhu, J., 1998. Sensitivity analysis of DEA models for simultaneous changes in all the data. *Journal of the Operational Research Society*, 49(10):1060-1071.

Siegel J G (1981). Warning signs of impending business failure and means to counteract such prospective failure. *National Public Accountant* 26(4): 9–13.

Taffler R J (1983). The assessment of company solvency and performance using a statistical model. *Accounting and Business Research* 13(52): 295–308.

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 and Wilson N (2013). Financial distress and bankruptcy prediction among listed companies using accounting, market, and macroeconomic variables. *International Review of Financial Analysis* 30 (Dec): 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 strange case of the cost and allocative efficiencies in DEA. *Journal of the Operational Research Society* 53(11): 1225-1231.

Trujillo-Ponce A, Samaniego-Medina R, Cardone-Riportella C (2014). Examining what best explains corporate credit risk: accounting-based versus market-based models. *Journal of Business Economics and Management* 15(2): 253–276.

Tsai C F (2009). Feature selection in bankruptcy prediction. *Knowledge-Based Systems* 22(2): 120–127.

Unler A, Murat A (2010). A discrete particle swarm optimization method for feature selection in binary classification problems. *European Journal of Operational Research* 206(3): 528–539.

Vassalou M, Xing Y (2004). Default risk in equity returns. *The Journal of Finance* 59(2): 831–868.

Wanke P, Barros C P, Faria J R (2015). Financial distress drivers in Brazilian banks: A dynamic slacks approach. *European Journal of Operational Research* 240(1): 258–268.

Wu Y, Gaunt C, Gray S (2010). A comparison of alternative bankruptcy prediction models. *Journal of Contemporary Accounting & Economics* 6(1): 34–45.

Xu, B., & Ouenniche, J. (2012). A data envelopment analysis-based framework for the relative performance evaluation of competing crude oil prices' volatility forecasting models. *Energy Economics* 34(2): 576-583.

Zavgren C V (1983). *Corporate Failure Prediction:* The State of the Art. Institute for Research in the Behavioral, Economic, and Management Sciences, Krannert Graduate School of Management, Purdue University.

Zhou L (2015). A comparison of dynamic hazard models and static models for predicting the special treatment of stocks in China with comprehensive variables. *Journal of the Operational Research Society* 66(7): 1077-1090.

Zhou L (2013). Performance of corporate bankruptcy prediction models on imbalanced dataset: The effect of sampling methods. *Knowledge-Based Systems* 41(Mar): 16–25.

Zhou L, Lai K K, Yen J (2012). Empirical models based on features ranking techniques for corporate financial distress prediction. *Computers & Mathematics with Applications* 64(8): 2484–2496.

Zmijewski M E (1984). Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research* 22(Jan): 59–82.

Appendix 4-A: The P-value of t-tests to Compare the Average Performance of Models Using ROC as a Measure of Discriminatory Power

This table presents the *p*-value of t-tests to compare the performance of models using ROC. Because of the lack of space, I show some models. The Null hypothesis (H_0) is: Average performance of modelling framework X fed with information category $Y \leq$ Average performance of modelling framework X' fed with information category Y'.

Models	DDWTDB_1/ln(age)_FA	DDWFSB_FAL1MI	DDWFSB_FAMI	DDWFSB_FAMV	DDWFSB_FAMVL1MI	DDWFSB_FAMVMI	DDWFSB_MV	DDWFSB_MVL1MI	DDWFSB_MVMI	DDWTDB_1/ln(age)_FA	DDWTDB_1/ln(age)_FAL1MI	DDWTDB_1/ln(age)_FAMI	DDWTDB_1/ln(age)_FAMV	DDWTDB_1/ln(age)_FAMVL1MI	DDWTDB_1/ln(age)_FAMVMI	DDWTDB_1/ln(age)_MV	DDWTDB_1/ln(age)_MVL1MI	DDWTDB_1/ln(age)_MVMI
DDWFSB_FA		0.007	0.973	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
DDWFSB_FAL1MI			1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
DDWFSB_FAMI				0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
DDWFSB_FAMV					0.145	0.000	0.000	0.000	0.000	0.000	0.000	0.915	0.885	0.000	0.000	0.000	0.000	0.000
DDWFSB_FAMVL1MI						0.000	0.000	0.000	0.000	0.000	0.000	0.942	0.929	0.000	0.000	0.000	0.000	0.000
DDWFSB_FAMVMI							0.001	0.000	0.711	0.004	0.002	1.000	1.000	0.667	0.009	0.006	0.627	0.001
DDWFSB_MV								0.034	0.999	0.978	0.690	1.000	1.000	0.997	0.845	0.663	0.999	0.386
DDWFSB_MVL1MI									0.999	0.997	0.978	1.000	1.000	0.999	0.946	0.857	1.000	0.780
DDWFSB_MVMI										0.001	0.001	1.000	1.000	0.465	0.007	0.005	0.344	0.001
DDWTDB_1/ln(age)_FA											0.036	1.000	1.000	0.995	0.534	0.323	0.998	0.032
DDWTDB_1/ln(age)_FAL1MI												1.000	1.000	0.997	0.795	0.583	0.999	0.233
DDWTDB_1/ln(age)_FAMI													0.359	0.000	0.000	0.000	0.000	0.000
DDWTDB_1/ln(age)_FAMV														0.000	0.000	0.000	0.000	0.000
DDWTDB_1/ln(age)_FAMVL1MI															0.002	0.002	0.422	0.001
DDWTDB_1/ln(age)_FAMVMI																0.071	0.994	0.066
DDWTDB_1/ln(age)_MV																	0.996	0.233
DDWTDB_1/ln(age)_MVL1MI																		0.000
DDWTDB_1/ln(age)_MVMI																		

Appendix 4-B: Models Explanation

Framework	Explanation
Multiple discriminant analysis (MDA)	Assuming there are n groups, the generic form of DA model for the group k could be shown as follows; $$z_k = f\left(\sum_{j=1}^{p} \beta_{kj} x_j\right) \qquad \text{Eq. 4-13}$$ where x_j is the discriminant features j, β_{kj} is the discriminant coefficients of group k for discriminant feature j, z_k represents the score of group k, and f is the linear or non-linear classifier that maps the scores, say $\beta^t x$ onto a set of real numbers. To compare DA models to other statistical models, I need to estimate the probability of failure, which is used as an input for estimating many measures of performance. For this, I follow Hillegeist et al. (2004) in using a logit link to calculate the probability of failure for companies; $$P(distress)_i = \frac{e^z}{1+e^z} \qquad \text{Eq. 4-14}$$
Linear probability model (LPA)	The generic linear probability model (LPA) is a special case of OLS regression and results in an estimate of probability of distress, the formula for which is as follows; $$P(distress)_i = \beta_o + \sum_{j=1}^{p} \beta_j x_{ij} \qquad \text{Eq. 4-15}$$
Logit analysis (LA)	The generic model for binary variables could be stated as follows: $$\begin{cases} P(distress)_i = P(Y=1) \\ P(distress)_i = G(\beta, X) \end{cases} \qquad \text{Eq. 4-16}$$ where Y denotes the binary response variable, X denotes the vector of features, β denotes the vector of coefficients of X in the model, and $G(.)$ is a link function that maps the scores of $\beta^t x$, onto a probability. In practice, depending on choice of link function, the type of probability model is determined. For example, the logit model (respectively, probit model) assumes that the link function is the cumulative logistic distribution, say Θ (respectively, cumulative standard normal distribution, say N) function.
Probit analysis (PA)	
Contingent claim analysis (CAA): Black-Scholes-Merton (BSM) Based Models	**Hillegeist et al. (2004), Bharath and Shumway (2008)** $$P(distress_i) = N\left(-\frac{\ln\left(\frac{V_a}{L}\right) + (\mu - \delta - 0.5\sigma_a^2) \times T}{\sigma_a \sqrt{T}}\right) \qquad \text{Eq. 4-17}$$

	$N(.)$: the cumulative normal distribution function, V_a: the value of the company's assets; L: total liabilities; μ: the expected return of the firm; σ_a : volatility of the company's asset; δ is the divided rate; which is estimated by the ratio of dividends to the sum of L and V_e (market value of common equity); T is time to maturity for both of call option and liabilities. The probability of failure is extracted as the probability that call option expires worthless at the end of maturity data - i.e. the value of the company's assets (Va) be less than the face value of its debt liabilites (L) at the end of the holding period [$P(Va < L)$]. In Hillegeist et al. (2004), V_a and σ_a are estimated by solving the systems of equations; i.e. the call option equation (4.18.1) and the optimal hedge equation (4.18.2). $\begin{cases} V_e = V_a e^{-\delta T} N(d_1) - L e^{-rT} N(d_2) + \left(1 - e^{\delta T}\right) N(d_1) V_a & (4.18.1) \\ \sigma_e = \frac{V_a e^{-\delta T} N(d_1) \sigma_a}{V_e} & (4.18.2) \end{cases}$ Eq. 4-18 where V_e is the market value of common equity at the time of estimation, σ_e is the annualised standard deviation of daily stock returns over 12 months prior to estimation, r is the risk-free interest rate, and d_1 and d_2 are calculated as follows; $d_1 = \frac{\ln\left(\frac{V_a}{L}\right) + (r - \delta - \frac{1}{2}.\sigma_e^2) \times T}{\sigma_e \sqrt{T}}; d_2 = d_1 - \sigma_e \sqrt{T}$ Eq. 4-19 Where $V_{a,t}$ is the value of the company's assets in year t and $V_{a,t-1}$ is the value of the company's assets in year $t-1$. Bharath and Shumway (2008) proposed a naïve approach to estimate V_a and σ_a as follows; $V_a = V_e + D \,; \sigma = \frac{V_e}{V_a}\sigma_e + \frac{D}{V_a}\sigma_d$ Eq. 4-20 Where $\sigma_d = 0.05 + 0.25\sigma_e$. Further, the firm's expected return μ is proxied by the risk-free rate, r or the stock return of previous year restricted to be between r and 100%.
Contingent claim analysis (CAA): Down-and-Out Call (DOC) Barrier Option Model	A naïve DOC barrier option as an extension of BSM model, which assumes that debt holder's position in the firm is like holding a portfolio of risk-free debt and a DOC option with a strike price (or Barrier) equal to total liabilities (L). The model rests on the assumptions of no dividends, zero rebate, costless failure proceedings, and set return on asset equal to risk-free rate. (Jackson and Wood, 2013) $P(distress)_i = N\left[\frac{ln\left(\frac{L}{V_a}\right) - \left(\mu - \frac{1}{2}\sigma_e^2\right)T}{\sigma_e\sqrt{T}}\right] + \left(\frac{L}{V_a}\right)^{\frac{2(\mu)}{\sigma_e^2}-1} N\left[\frac{ln\left(\frac{L}{V_a}\right) - \left(\mu - \frac{1}{2}\sigma_e^2\right)T}{\sigma_e\sqrt{T}}\right]$ Eq. 4-21

	$$P(distress)_i = N\left[\frac{ln\left(\frac{L}{V_a}\right) - \left(\mu - \frac{1}{2}\sigma_e^2\right)T}{\sigma_e\sqrt{T}}\right] + \left(\frac{L}{V_a}\right)^{\frac{2(\mu)}{\sigma_e^2}-1} N\left[\frac{ln\left(\frac{L}{V_a}\right) - \left(\mu - \frac{1}{2}\sigma_e^2\right)T}{\sigma_e\sqrt{T}}\right]$$ Eq. 4-22
Discrete time hazard framework: Duration-dependent hazard model (DD)	Shumway (2001) proposed a discrete time hazard model using an estimation procedure like the one used for estimating the parameters of a multi-period (dynamic) logit model. $$P(y_{i,t} = 1 \vert x_{i,t}) = h(t \vert x_{i,t}) = \frac{e^{(\alpha_t + x_{i,t}\beta)}}{1 + e^{(\alpha_t + x_{i,t}\beta)}}$$ Eq. 4-23 where $h(t \vert x_{i,t})$ represent the individual hazard rate of firm i at time t, $X_{i,t}$ is the vector of covariates of each firm i at time t. α_t is the time-variant baseline hazard function related, which could be relate to firm, e.g. ln(age) or related to macroeconomic variables, e.g. volatility of exchange rate (Nam et al, 2011). Shumway employed a constant time variant term; say $ln\ (age)$, as proxy of baseline rate. General notation of duration-dependent hazard model could be presented as follows: $$h(t \vert x_{i,t}) = h_0(t).\, e^{x_{i,t} \cdot \beta}$$ Eq. 4-24 $$p(y_{i,t} = 1) = \frac{1}{1 + e^{-(\alpha_t + x_{i,t} \cdot \beta)}}$$ Eq. 4-25
Discrete time hazard framework: Duration-independent model with time-invariant baseline (DIWTIB) & Duration-independent model without baseline hazard rate (DIWOB)	Duration independent hazard model uses the multi-period logit framework to estimate the coefficients of the features. However, conversely to duration dependent (DD) models, the baseline hazard rate is invariant to time. The time-invariant baseline hazard rate could be represented by firm related features, e.g. ln(age), 1/ln(age) or macroeconomic features, e.g., volatility of exchange rate. General notation of duration-independent hazard model could be presented as follows: $$h(t \vert x_{i,t}) = h_0.\, e^{x_{i,t} \cdot \beta}$$ Eq. 4-26 $$p(y_{i,t} = 1) = \frac{1}{1 + e^{-x_{i,t} \cdot \beta}}$$ Eq. 4-27 Conversely to the last two categories of discrete time hazard model, DIWOB does not use any baseline hazard rate.

Cox hazard framework	The cox hazard model with time-varying (TV) covariates could be presented as equation 4-28; $$h(t\|x_{i,t}) = h_0(t).\, e^{x_{i,t}.\beta} \quad \text{Eq. 4-28}$$ However, a partial likelihood function on the training sample is used to estimate the coefficients β ; $$PL\,(\beta) = \prod_{i=1}^{m} \left[\frac{exp\left(\sum_{j=1}^{p} \beta_j x_j^i(t)\right)}{\sum_{k \in R_t(t)} exp\left(\sum_{j=1}^{p} \beta_j x_j^k(t)\right)} \right] \quad \text{Eq. 4-29}$$ where i is the firm in the event of distress, k is the firm in the risk set at time t, and p is the number of features. This equation estimates β without requiring to consider the baseline hazard rate (Hosmer and Lemesho, 1999, Section 7.3). However, to use the developed model for estimation of distress probabilities, the baseline hazard rate is required. I follow Chen et al. (2005) in estimating the integrated baseline hazard function with time-varying covariates base on Anderson (1992) as follow: $$\widehat{H}_0(t) = \sum_{\tilde{T}_i \le t} \frac{D_i}{\sum_{j \in (\tilde{T}_i)} \exp\left(\hat{B}'.\, x_j(\tilde{T}_i)\right)} \quad \text{Eq. 4-30}$$ Where D_i is a dummy variable for whether the firm i faces the distress, i.e. 0 for survivors and 1 for distressed; $\hat{T}_i$ is the distress time for the ith firm; $\hat{\beta}$ is the vector of estimated coefficients; and $\tilde{T}_i$ is the distress time for the ith firm. Using Equations (4-29) and (4-30), I estimate the probability of distress for individual firms in Equation (4-28).

Appendix 4-C: Measures of Correctness of Categorical Prediction

The table presents the average measures of correctness of categorical prediction for all dynamic and static frameworks. The best performers are shown in block cells where darker represents better performance. The average performance of each group is bolded.

	T1			T2			MR			OCC		
Models	**3-year**	**5-year**	**10-year**	**3-year**	**5-year**	**10-year**	**3-year**	**5-year**	**10-year**	**3-year**	**5-year**	**10-year**
FA Information	**0.48**	**0.41**	**0.43**	**0.37**	**0.35**	**0.41**	**0.38**	**0.35**	**0.41**	**0.62**	**0.65**	**0.59**
DDWFSB	0.371	0.081	0.0183	0.512	0.797	0.892	0.506	0.761	0.848	0.494	0.239	0.152
DDWTDB-ln(age)	0.43	0.352	0.422	0.38	0.307	0.361	0.383	0.31	0.364	0.617	0.69	0.636
DDWTDB-1/ln(age)	0.437	0.347	0.457	0.368	0.308	0.362	0.373	0.311	0.367	0.627	0.689	0.633
DDWTDB-LPE	0.28	0.321	0.353	0.602	0.422	0.482	0.592	0.426	0.477	0.408	0.574	0.523
DDWTDB-VEX	0.567	0.381	0.378	0.371	0.331	0.477	0.384	0.334	0.467	0.616	0.666	0.533
DIWOTIB	0.415	0.331	0.4	0.354	0.334	0.371	0.359	0.334	0.372	0.641	0.666	0.628
DIWTIB-ln(age)	0.455	0.411	0.533	0.484	0.307	0.414	0.484	0.312	0.42	0.516	0.688	0.58
DIWTIB-1/ln(age)	0.405	0.32	0.348	0.409	0.358	0.444	0.409	0.357	0.438	0.591	0.643	0.562
MDA	0.473	0.478	0.443	0.338	0.359	0.39	0.346	0.367	0.394	0.654	0.633	0.606
LPA	0.506	0.48	0.443	0.337	0.358	0.39	0.346	0.367	0.394	0.654	0.633	0.606
LA	0.485	0.468	0.426	0.312	0.304	0.338	0.321	0.313	0.342	0.679	0.687	0.658
PA	0.972	0.972	0.981	0.0085	0.007	0.0047	0.0567	0.0553	0.0533	0.9433	0.9447	0.9467
FAMI Information	**0.43**	**0.37**	**0.39**	**0.3**	**0.32**	**0.32**	**0.31**	**0.33**	**0.32**	**0.69**	**0.67**	**0.68**
DDWFSB	0.389	0.081	0.0193	0.362	0.701	0.78	0.363	0.668	0.742	0.637	0.332	0.258
DDWTDB-ln(age)	0.328	0.308	0.397	0.285	0.293	0.254	0.289	0.294	0.262	0.711	0.706	0.738
DDWTDB-1/ln(age)	0.352	0.309	0.441	0.272	0.279	0.234	0.277	0.281	0.244	0.723	0.719	0.756
DDWTDB-LPE	0.209	0.296	0.287	0.462	0.384	0.354	0.457	0.388	0.351	0.543	0.612	0.649
DDWTDB-VEX	0.486	0.379	0.346	0.295	0.29	0.335	0.309	0.296	0.334	0.691	0.704	0.666
DIWOTIB	0.524	0.297	0.332	0.378	0.31	0.269	0.38	0.31	0.272	0.62	0.69	0.728
DIWTIB-ln(age)	0.361	0.356	0.454	0.321	0.29	0.267	0.324	0.294	0.276	0.676	0.706	0.724
DIWTIB-1/ln(age)	0.288	0.282	0.324	0.335	0.329	0.318	0.333	0.328	0.319	0.667	0.672	0.681
MDA	0.454	0.407	0.392	0.326	0.359	0.358	0.334	0.364	0.359	0.666	0.636	0.641
LPA	0.464	0.421	0.397	0.317	0.337	0.364	0.326	0.343	0.364	0.674	0.657	0.636
LA	0.322	0.316	0.325	0.278	0.275	0.271	0.28	0.277	0.275	0.72	0.723	0.725
PA	0.955	0.964	0.983	0.0137	0.0124	0.0079	0.0608	0.06	0.0564	0.9392	0.94	0.9436

FAMV Information	**0.29**	**0.25**	**0.23**	**0.26**	**0.29**	**0.32**	**0.26**	**0.29**	**0.31**	**0.74**	**0.71**	**0.69**
DDWFSB	0.35	0.0445	0.028	0.392	0.776	0.816	0.392	0.739	0.775	0.608	0.261	0.225
DDWTDB-ln(age)	0.257	0.214	0.203	0.24	0.259	0.281	0.241	0.259	0.277	0.759	0.741	0.723
DDWTDB-1/ln(age)	0.243	0.22	0.204	0.241	0.248	0.271	0.241	0.248	0.268	0.759	0.752	0.732
DDWTDB-LPE	0.226	0.208	0.173	0.344	0.337	0.346	0.341	0.334	0.338	0.659	0.666	0.662
DDWTDB-VEX	0.271	0.244	0.184	0.308	0.253	0.304	0.306	0.255	0.298	0.694	0.745	0.702
DIWOTIB	0.235	0.196	0.194	0.257	0.269	0.292	0.256	0.267	0.287	0.744	0.733	0.713
DIWTIB-ln(age)	0.273	0.242	0.272	0.27	0.249	0.263	0.27	0.25	0.263	0.73	0.75	0.737
DIWTIB-1/ln(age)	0.25	0.224	0.184	0.256	0.263	0.309	0.255	0.261	0.302	0.745	0.739	0.698
MDA	0.208	0.194	0.181	0.259	0.263	0.282	0.256	0.26	0.277	0.744	0.74	0.723
LPA	0.217	0.2	0.174	0.257	0.263	0.282	0.255	0.259	0.277	0.745	0.741	0.723
LA	0.215	0.212	0.182	0.258	0.263	0.296	0.255	0.261	0.29	0.745	0.739	0.71
PA	0.759	0.758	0.728	0.047	0.049	0.056	0.083	0.084	0.089	0.917	0.916	0.911
FAMVMI Information	**0.27**	**0.24**	**0.23**	**0.24**	**0.27**	**0.29**	**0.24**	**0.27**	**0.28**	**0.76**	**0.73**	**0.72**
DDWFSB	0.351	0.0464	0.03	0.333	0.65	0.709	0.336	0.621	0.674	0.664	0.379	0.326
DDWTDB-ln(age)	0.218	0.204	0.22	0.225	0.241	0.245	0.225	0.24	0.244	0.775	0.76	0.756
DDWTDB-1/ln(age)	0.215	0.209	0.215	0.217	0.236	0.239	0.218	0.235	0.239	0.782	0.765	0.761
DDWTDB-LPE	0.214	0.196	0.18	0.316	0.313	0.309	0.315	0.312	0.303	0.685	0.688	0.697
DDWTDB-VEX	0.232	0.249	0.196	0.281	0.237	0.274	0.279	0.239	0.271	0.721	0.761	0.729
DIWOTIB	0.223	0.202	0.199	0.232	0.253	0.256	0.232	0.251	0.254	0.768	0.749	0.746
DIWTIB-ln(age)	0.233	0.234	0.264	0.232	0.242	0.242	0.231	0.242	0.244	0.769	0.758	0.756
DIWTIB-1/ln(age)	0.221	0.2	0.18	0.244	0.26	0.279	0.243	0.258	0.274	0.757	0.742	0.726
MDA	0.218	0.21	0.184	0.259	0.271	0.285	0.256	0.268	0.28	0.744	0.732	0.72
LPA	0.216	0.204	0.168	0.259	0.267	0.288	0.256	0.264	0.283	0.744	0.736	0.717
LA	0.207	0.21	0.201	0.236	0.241	0.26	0.235	0.24	0.258	0.765	0.76	0.742
PA	0.749	0.75	0.74	0.046	0.048	0.0501	0.082	0.083	0.0845	0.918	0.917	0.9155
MV Information	**0.32**	**0.29**	**0.27**	**0.27**	**0.3**	**0.31**	**0.27**	**0.3**	**0.31**	**0.73**	**0.7**	**0.69**
DDWFSB	0.335	0.0365	0.0195	0.397	0.778	0.822	0.396	0.742	0.781	0.604	0.258	0.219
DDWTDB-ln(age)	0.226	0.229	0.199	0.254	0.266	0.29	0.252	0.264	0.286	0.748	0.736	0.714
DDWTDB-1/ln(age)	0.231	0.225	0.212	0.251	0.27	0.271	0.251	0.268	0.267	0.749	0.732	0.733

DDWTDB-LPE	0.222	0.224	0.186	0.35	0.344	0.346	0.347	0.342	0.338	0.653	0.658	0.662
DDWTDB-VEX	0.271	0.27	0.186	0.295	0.256	0.31	0.294	0.258	0.303	0.706	0.742	0.697
DIWOTIB	0.228	0.218	0.195	0.261	0.273	0.297	0.26	0.271	0.291	0.74	0.729	0.709
DIWTIB-ln(age)	0.269	0.264	0.277	0.271	0.272	0.263	0.27	0.271	0.264	0.73	0.729	0.736
DIWTIB-1/ln(age)	0.242	0.225	0.188	0.256	0.276	0.31	0.255	0.274	0.303	0.745	0.726	0.697
MDA	0.213	0.196	0.165	0.264	0.282	0.297	0.261	0.278	0.29	0.739	0.722	0.71
LPA	0.215	0.192	0.165	0.262	0.282	0.297	0.259	0.278	0.29	0.741	0.722	0.71
LA	0.23	0.22	0.194	0.259	0.272	0.298	0.257	0.27	0.292	0.743	0.73	0.708
PA	0.776	0.775	0.747	0.045	0.046	0.054	0.082	0.0821	0.088	0.918	0.9179	0.912
BSM_HKCL04	0.522	0.521	0.521	0.194	0.194	0.193	0.209	0.209	0.208	0.791	0.791	0.792
BSM_BhSh08	0.355	0.352	0.35	0.287	0.304	0.29	0.287	0.305	0.292	0.713	0.695	0.708
DOC_JW13	0.398	0.398	0.409	0.371	0.374	0.355	0.37	0.373	0.356	0.63	0.627	0.644
MVMI Information	**0.27**	**0.25**	**0.23**	**0.25**	**0.28**	**0.3**	**0.25**	**0.28**	**0.29**	**0.75**	**0.72**	**0.71**
DDWFSB	0.327	0.069	0.021	0.327	0.681	0.702	0.33	0.648	0.668	0.67	0.352	0.332
DDWTDB-ln(age)	0.219	0.217	0.219	0.229	0.239	0.25	0.229	0.239	0.249	0.771	0.761	0.751
DDWTDB-1/ln(age)	0.211	0.209	0.22	0.228	0.234	0.24	0.227	0.234	0.239	0.773	0.766	0.761
DDWTDB-LPE	0.21	0.212	0.169	0.322	0.317	0.322	0.321	0.316	0.315	0.679	0.684	0.685
DDWTDB-VEX	0.256	0.26	0.181	0.275	0.233	0.322	0.275	0.236	0.315	0.725	0.764	0.685
DIWOTIB	0.2032	0.202	0.19	0.242	0.255	0.27	0.241	0.253	0.266	0.759	0.747	0.734
DIWTIB-ln(age)	0.219	0.248	0.254	0.247	0.245	0.245	0.245	0.245	0.245	0.755	0.755	0.755
DIWTIB-1/ln(age)	0.1995	0.199	0.184	0.253	0.265	0.284	0.25	0.262	0.279	0.75	0.738	0.721
MDA	0.207	0.198	0.172	0.268	0.283	0.295	0.265	0.279	0.289	0.735	0.721	0.711
LPA	0.2023	0.196	0.179	0.27	0.285	0.292	0.267	0.28	0.287	0.733	0.72	0.713
LA	0.206	0.202	0.191	0.24	0.255	0.269	0.239	0.253	0.265	0.761	0.747	0.735
PA	0.765	0.751	0.739	0.044	0.047	0.051	0.0808	0.083	0.085	0.9192	0.917	0.915
Total	**0.34**	**0.3**	**0.29**	**0.28**	**0.3**	**0.32**	**0.28**	**0.3**	**0.32**	**0.72**	**0.7**	**0.68**

Appendix 4-D: Measures of Discriminatory Power

The table presents the average measures of discriminatory power for all dynamic and static frameworks. The best performers are shown in block cells where darker represents better performance. The average performance of each group is bolded.

	ROC			H			Gini			KS			IV		
Models	**3-year**	**5-year**	**10-year**	**3-year**	**5-year**	**10-year**	**3-year**	**5-year**	**10-year**	**3-year**	**5-year**	**10-year**	**3-year**	**5-year**	**10-year**
FA Information	**0.62**	**0.68**	**0.63**	**0.12**	**0.16**	**0.11**	**0.25**	**0.37**	**0.24**	**0.27**	**0.33**	**0.26**	**0.142**	**0.241**	**0.115**
DDWFSB	0.664	0.604	0.579	0.158	0.12	0.074	0.327	0.208	0.052	0.305	0.252	0.245	0.196	0.155	0.1149
DDWTDB-ln(age)	0.624	0.736	0.642	0.113	0.197	0.122	0.248	0.472	0.285	0.257	0.392	0.274	0.128	0.312	0.1099
DDWTDB-1/ln(age)	0.632	0.741	0.629	0.089	0.178	0.107	0.264	0.481	0.258	0.261	0.402	0.245	0.144	0.330	0.1046
DDWTDB-LPE	0.642	0.727	0.648	0.142	0.182	0.126	0.283	0.454	0.296	0.297	0.385	0.281	0.159	0.311	0.1236
DDWTDB-VEX	0.643	0.727	0.648	0.144	0.182	0.127	0.287	0.454	0.296	0.3	0.385	0.283	0.128	0.302	0.0955
DIWOTIB	0.642	0.727	0.647	0.142	0.182	0.126	0.283	0.453	0.294	0.296	0.386	0.283	0.131	0.329	0.0988
DIWTIB-ln(age)	0.558	0.702	0.564	0.054	0.156	0.05	0.115	0.404	0.083	0.177	0.335	0.179	0.131	0.296	0.143
DIWTIB-1/ln(age)	0.623	0.717	0.621	0.099	0.169	0.092	0.246	0.434	0.242	0.253	0.375	0.247	0.140	0.310	0.1297
MDA	0.621	0.62	0.625	0.109	0.114	0.102	0.241	0.241	0.25	0.251	0.251	0.232	0.159	0.139	0.1144
LPA	0.598	0.62	0.625	0.096	0.114	0.102	0.196	0.241	0.25	0.233	0.251	0.232	0.128	0.145	0.1369
LA	0.622	0.638	0.65	0.128	0.143	0.137	0.243	0.276	0.301	0.285	0.299	0.297	0.138	0.125	0.0964
PA	0.624	0.644	0.656	0.129	0.13	0.136	0.248	0.287	0.311	0.285	0.281	0.294	0.126	0.139	0.1095
FAMI Information	**0.74**	**0.75**	**0.73**	**0.22**	**0.22**	**0.2**	**0.47**	**0.49**	**0.47**	**0.41**	**0.42**	**0.4**	**0.250**	**0.331**	**0.22**
DDWFSB	0.781	0.692	0.655	0.298	0.19	0.144	0.561	0.383	0.309	0.478	0.358	0.338	0.346	0.198	0.1251
DDWTDB-ln(age)	0.767	0.777	0.755	0.244	0.241	0.224	0.535	0.554	0.509	0.448	0.455	0.431	0.256	0.408	0.2381
DDWTDB-1/ln(age)	0.769	0.781	0.755	0.213	0.228	0.216	0.538	0.563	0.51	0.447	0.462	0.43	0.267	0.425	0.2465
DDWTDB-LPE	0.768	0.768	0.762	0.244	0.23	0.235	0.537	0.536	0.525	0.452	0.448	0.445	0.300	0.429	0.2372
DDWTDB-VEX	0.77	0.767	0.761	0.245	0.23	0.234	0.539	0.535	0.522	0.455	0.449	0.442	0.254	0.443	0.2796
DIWOTIB	0.665	0.767	0.761	0.161	0.23	0.234	0.33	0.534	0.522	0.324	0.448	0.443	0.131	0.425	0.241
DIWTIB-ln(age)	0.725	0.755	0.714	0.188	0.214	0.17	0.451	0.511	0.429	0.401	0.427	0.38	0.382	0.480	0.3818
DIWTIB-1/ln(age)	0.745	0.756	0.739	0.208	0.216	0.2	0.49	0.513	0.478	0.433	0.441	0.418	0.328	0.415	0.291
MDA	0.638	0.661	0.686	0.133	0.14	0.149	0.276	0.322	0.372	0.285	0.292	0.309	0.154	0.133	0.1231
LPA	0.642	0.67	0.679	0.135	0.149	0.145	0.285	0.339	0.357	0.283	0.299	0.302	0.149	0.142	0.1698
LA	0.778	0.774	0.765	0.258	0.257	0.239	0.555	0.547	0.529	0.472	0.463	0.454	0.223	0.233	0.1577
PA	0.778	0.775	0.765	0.257	0.257	0.237	0.555	0.549	0.529	0.477	0.468	0.451	0.214	0.237	0.151
FAMV Information	**0.82**	**0.82**	**0.81**	**0.35**	**0.35**	**0.33**	**0.64**	**0.64**	**0.62**	**0.55**	**0.55**	**0.53**	**0.562**	**0.536**	**0.521**

DDWFSB	0.778	0.702	0.641	0.334	0.241	0.134	0.556	0.404	0.275	0.513	0.381	0.277	0.506	0.361	0.2002
DDWTDB-ln(age)	0.826	0.839	0.829	0.355	0.367	0.351	0.651	0.677	0.657	0.558	0.574	0.557	0.586	0.512	0.5045
DDWTDB-1/ln(age)	0.83	0.843	0.833	0.335	0.355	0.341	0.659	0.686	0.665	0.561	0.58	0.566	0.567	0.537	0.5044
DDWTDB-LPE	0.823	0.834	0.828	0.351	0.359	0.349	0.646	0.668	0.656	0.552	0.571	0.558	0.589	0.560	0.5724
DDWTDB-VEX	0.823	0.834	0.827	0.352	0.358	0.346	0.646	0.667	0.654	0.554	0.569	0.553	0.585	0.534	0.5496
DIWOTIB	0.823	0.834	0.827	0.351	0.358	0.347	0.646	0.668	0.654	0.554	0.568	0.554	0.564	0.546	0.5381
DIWTIB-ln(age)	0.806	0.823	0.811	0.314	0.338	0.317	0.611	0.646	0.623	0.516	0.54	0.526	0.658	0.7356	0.6856
DIWTIB-1/ln(age)	0.817	0.827	0.822	0.335	0.345	0.335	0.634	0.655	0.644	0.537	0.56	0.547	0.623	0.610	0.6147
MDA	0.831	0.835	0.833	0.361	0.366	0.357	0.661	0.671	0.666	0.575	0.585	0.577	0.379	0.390	0.3542
LPA	0.827	0.835	0.832	0.355	0.366	0.356	0.654	0.67	0.664	0.569	0.582	0.575	0.576	0.610	0.6238
LA	0.83	0.832	0.827	0.357	0.362	0.348	0.659	0.663	0.655	0.561	0.566	0.558	0.551	0.539	0.5562
PA	0.83	0.833	0.829	0.357	0.36	0.352	0.66	0.665	0.658	0.561	0.566	0.564	0.558	0.498	0.5431
FAMVMI Information	**0.84**	**0.84**	**0.83**	**0.38**	**0.37**	**0.35**	**0.68**	**0.67**	**0.65**	**0.59**	**0.58**	**0.57**	**0.614**	**0.579**	**0.537**
DDWFSB	0.804	0.751	0.702	0.366	0.28	0.181	0.608	0.501	0.405	0.543	0.432	0.375	0.562	0.404	0.1842
DDWTDB-ln(age)	0.8536	0.85	0.843	0.3997	0.3954	0.3807	0.7072	0.700	0.686	0.6109	0.597	0.5941	0.596	0.584	0.4979
DDWTDB-1/ln(age)	0.8578	0.8541	0.8471	0.384	0.385	0.366	0.7156	0.7082	0.6941	0.6188	0.6044	0.5999	0.616	0.592	0.5202
DDWTDB-LPE	0.847	0.846	0.841	0.388	0.3902	0.371	0.695	0.693	0.683	0.599	0.595	0.584	0.681	0.600	0.5721
DDWTDB-VEX	0.847	0.846	0.839	0.387	0.388	0.365	0.695	0.691	0.678	0.6	0.594	0.579	0.635	0.581	0.5748
DIWOTIB	0.847	0.846	0.84	0.386	0.389	0.368	0.695	0.692	0.68	0.598	0.595	0.582	0.618	0.598	0.5581
DIWTIB-ln(age)	0.843	0.836	0.83	0.382	0.37	0.353	0.686	0.672	0.66	0.595	0.576	0.566	0.8386	0.81813	0.78155
DIWTIB-1/ln(age)	0.841	0.84	0.835	0.371	0.37	0.358	0.681	0.679	0.669	0.588	0.581	0.574	0.721	0.700	0.638
MDA	0.828	0.831	0.832	0.36	0.358	0.356	0.657	0.662	0.664	0.58	0.579	0.58	0.389	0.373	0.3654
LPA	0.831	0.835	0.83	0.361	0.366	0.353	0.662	0.669	0.66	0.579	0.587	0.577	0.559	0.546	0.6617
LA	0.848	0.848	0.84	0.394	0.389	0.367	0.697	0.697	0.681	0.605	0.595	0.583	0.567	0.567	0.5414
PA	0.848	0.848	0.843	0.394	0.388	0.3733	0.697	0.696	0.685	0.605	0.594	0.589	0.588	0.582	0.5428
MV Information	**0.79**	**0.79**	**0.78**	**0.31**	**0.3**	**0.3**	**0.59**	**0.58**	**0.57**	**0.51**	**0.5**	**0.49**	**0.539**	**0.510**	**0.473**
DDWFSB	0.777	0.698	0.636	0.325	0.239	0.137	0.554	0.397	0.264	0.508	0.374	0.267	0.578	0.358	0.2092
DDWTDB-ln(age)	0.831	0.83	0.83	0.351	0.351	0.35	0.662	0.661	0.659	0.554	0.552	0.551	0.591	0.564	0.5435
DDWTDB-1/ln(age)	0.834	0.834	0.833	0.332	0.336	0.342	0.669	0.668	0.667	0.56	0.562	0.559	0.613	0.570	0.5251
DDWTDB-LPE	0.828	0.828	0.829	0.349	0.349	0.349	0.656	0.655	0.657	0.556	0.552	0.555	0.660	0.626	0.5996
DDWTDB-VEX	0.828	0.827	0.828	0.349	0.348	0.348	0.656	0.654	0.655	0.555	0.551	0.552	0.630	0.620	0.5729
DIWOTIB	0.828	0.828	0.828	0.349	0.348	0.349	0.656	0.655	0.656	0.556	0.55	0.553	0.631	0.615	0.5621

DIWTIB-ln(age)	0.809	0.808	0.812	0.308	0.309	0.316	0.618	0.616	0.624	0.517	0.519	0.527	0.660	0.669	0.677
DIWTIB-1/ln(age)	0.822	0.822	0.823	0.336	0.336	0.339	0.644	0.644	0.646	0.547	0.545	0.546	0.656	0.661	0.6027
MDA	0.833	0.832	0.83	0.357	0.356	0.348	0.665	0.664	0.66	0.574	0.575	0.568	0.509	0.467	0.3869
LPA	0.832	0.832	0.83	0.355	0.356	0.348	0.664	0.663	0.66	0.571	0.575	0.568	0.711	0.686	0.68915
LA	0.828	0.828	0.828	0.349	0.349	0.349	0.656	0.655	0.657	0.557	0.554	0.553	0.624	0.607	0.5664
PA	0.828	0.828	0.828	0.349	0.349	0.35	0.656	0.655	0.657	0.557	0.554	0.554	0.624	0.607	0.5699
BSM_HKCL04	0.65	0.65	0.65	0.147	0.147	0.147	0.301	0.301	0.301	0.299	0.299	0.299	0.073	0.073	0.0732
BSM_BhSh08	0.723	0.723	0.723	0.228	0.228	0.228	0.446	0.446	0.446	0.408	0.408	0.408	0.335	0.335	0.3348
DOC_JW13	0.649	0.649	0.649	0.134	0.134	0.134	0.298	0.298	0.298	0.287	0.287	0.287	0.184	0.184	0.1844
MVMI Information	**0.84**	**0.83**	**0.83**	**0.38**	**0.36**	**0.35**	**0.68**	**0.67**	**0.65**	**0.59**	**0.57**	**0.56**	**0.671**	**0.624**	**0.558**
DDWFSB	0.813	0.729	0.702	0.375	0.246	0.187	0.625	0.458	0.405	0.557	0.423	0.374	0.641	0.370	0.2013
DDWTDB-ln(age)	0.853	0.8509	0.8441	0.4005	0.3938	0.3781	0.706	0.7019	0.6883	0.61	0.6035	0.591	0.655	0.606	0.5306
DDWTDB-1/ln(age)	0.8569	0.8549	0.8481	0.379	0.378	0.366	0.7139	0.7099	0.6963	0.618	0.6085	0.6002	0.648	0.623	0.5211
DDWTDB-LPE	0.848	0.846	0.842	0.39	0.383	0.372	0.697	0.693	0.683	0.602	0.593	0.586	0.671	0.661	0.6079
DDWTDB-VEX	0.848	0.846	0.841	0.389	0.382	0.368	0.697	0.692	0.681	0.603	0.591	0.583	0.666	0.631	0.596
DIWOTIB	0.848	0.846	0.841	0.388	0.383	0.368	0.696	0.692	0.682	0.601	0.591	0.584	0.667	0.642	0.5813
DIWTIB-ln(age)	0.838	0.835	0.831	0.362	0.356	0.351	0.677	0.67	0.663	0.586	0.573	0.57	0.8093	0.81452	0.78404
DIWTIB-1/ln(age)	0.842	0.84	0.835	0.373	0.369	0.357	0.683	0.679	0.67	0.589	0.582	0.571	0.704	0.722	0.6392
MDA	0.83	0.827	0.83	0.353	0.347	0.347	0.659	0.654	0.659	0.574	0.571	0.569	0.510	0.448	0.397
LPA	0.832	0.831	0.829	0.356	0.355	0.345	0.664	0.661	0.657	0.576	0.576	0.569	0.711	0.690	0.6809
LA	0.848	0.846	0.841	0.388	0.383	0.368	0.696	0.692	0.682	0.6	0.593	0.584	0.671	0.642	0.5767
PA	0.849	0.846	0.841	0.39	0.382	0.368	0.697	0.692	0.682	0.601	0.592	0.582	0.693	0.642	0.5798
Total	**0.78**	**0.78**	**0.77**	**0.29**	**0.29**	**0.27**	**0.55**	**0.57**	**0.54**	**0.49**	**0.49**	**0.47**	**0.466**	**0.472**	**0.407**

Appendix 4-E: Information Content and Quality of Fit under Calibration Accuracy

The table presents the average measures of calibration accuracy for all dynamic and static frameworks. The best performers are shown in block cells where darker represents better performance. The average performance of each group is bolded.

Models	LL			R2			BS		
	3-year	5-year	10-year	3-year	5-year	10-year	3-year	5-year	10-year
FA Information	**570**	**555**	**570**	**0.004**	**0.013**	**0.004**	**0.0461**	**0.046**	**0.046**
DDWFSB	566	572	572	0.006	0.003	0.003	0.046	0.0462	0.0462
DDWTDB-ln(age)	569	544	568	0.004	0.02	0.005	0.04616	0.0453	0.046
DDWTDB-1/ln(age)	569	542	568	0.004	0.021	0.005	0.04616	0.0452	0.046
DDWTDB-LPE	568	544	568	0.005	0.02	0.005	0.04608	0.0454	0.046
DDWTDB-VEX	569	546	569	0.004	0.019	0.005	0.04612	0.0455	0.0461
DIWOTIB	570	546	569	0.004	0.019	0.005	0.04615	0.0454	0.0461
DIWTIB-ln(age)	573	540	570	0.002	0.022	0.004	0.04625	0.0451	0.046
DIWTIB-1/ln(age)	570	549	569	0.004	0.017	0.005	0.04613	0.0456	0.0461
MDA	571	572	573	0.003	0.002	0.002	0.04611	0.0462	0.0462
LPA	571	570	571	0.003	0.003	0.003	0.04615	0.0462	0.0462
LA	571	569	573	0.003	0.005	0.002	0.0461	0.046	0.0463
PA	571	570	573	0.003	0.004	0.002	0.04611	0.0461	0.0463
FAMI Information	**555**	**548**	**558**	**0.014**	**0.02**	**0.01**	**0.0455**	**0.045**	**0.046**
DDWFSB	542	560	570	0.022	0.01	0.004	0.04491	0.0457	0.0461
DDWTDB-ln(age)	552	538	555	0.016	0.024	0.013	0.04535	0.0451	0.0457
DDWTDB-1/ln(age)	549	536	554	0.018	0.025	0.014	0.04521	0.045	0.0457
DDWTDB-LPE	549	538	554	0.018	0.025	0.014	0.04531	0.0452	0.0458
DDWTDB-VEX	553	540	555	0.015	0.023	0.014	0.04537	0.0453	0.0458
DIWOTIB	569	540	556	0.004	0.023	0.013	0.04608	0.0453	0.0458
DIWTIB-ln(age)	535	522	539	0.027	0.033	0.023	0.04503	0.0446	0.0451
DIWTIB-1/ln(age)	552	543	555	0.016	0.021	0.014	0.04544	0.0454	0.0458
MDA	571	571	572	0.003	0.003	0.002	0.04611	0.0462	0.0462
LPA	569	568	568	0.005	0.005	0.005	0.04611	0.0462	0.0461
LA	556	557	560	0.014	0.013	0.011	0.0457	0.0458	0.0456
PA	560	557	560	0.011	0.013	0.011	0.04578	0.0457	0.0456
FAMV Information	**524**	**525**	**530**	**0.033**	**0.032**	**0.029**	**0.0447**	**0.045**	**0.045**
DDWFSB	528	542	564	0.03	0.022	0.007	0.04466	0.045	0.0459
DDWTDB-ln(age)	524	525	531	0.033	0.032	0.028	0.0447	0.0447	0.0451
DDWTDB-1/ln(age)	522	523	529	0.034	0.033	0.03	0.04457	0.0446	0.045
DDWTDB-LPE	522	525	528	0.035	0.033	0.031	0.04469	0.0448	0.0451
DDWTDB-VEX	526	529	531	0.032	0.03	0.029	0.04488	0.0451	0.0452
DIWOTIB	527	529	533	0.031	0.03	0.028	0.04496	0.0451	0.0453
DIWTIB-ln(age)	501	490.2	499	0.047	0.0533	0.048	0.043554	0.0432	0.0435
DIWTIB-1/ln(age)	527	528	531	0.032	0.03	0.029	0.04501	0.0451	0.0452
MDA	545	544	545	0.02	0.021	0.02	0.04532	0.0454	0.0454
LPA	508	509	503	0.044	0.043	0.047	0.04392	0.0441	0.044
LA	528	530	532	0.031	0.03	0.028	0.04494	0.0451	0.0452
PA	528	531	532	0.031	0.028	0.028	0.04492	0.0452	0.0452

FAMVMI Information	**517**	**521**	**525**	**0.038**	**0.035**	**0.032**	**0.0444**	**0.045**	**0.045**
DDWFSB	521	534	561	0.035	0.027	0.009	0.04424	0.0446	0.0457
DDWTDB-ln(age)	517	522	526	0.038	0.035	0.032	0.04424	0.0447	0.045
DDWTDB-1/ln(age)	514	519	524	0.04	0.036	0.033	0.04407	0.0445	0.0449
DDWTDB-LPE	516	521	523	0.039	0.036	0.034	0.04437	0.0447	0.0449
DDWTDB-VEX	518	526	526	0.037	0.032	0.032	0.04453	0.0449	0.0451
DIWOTIB	522	525	528	0.035	0.033	0.031	0.04466	0.0449	0.0452
DIWTIB-ln(age)	477.2	480.8	487	0.0619	0.0592	0.0554	0.042666	0.04289	0.04313
DIWTIB-1/ln(age)	521	523	526	0.036	0.034	0.033	0.04473	0.0449	0.0451
MDA	545	544	544	0.02	0.02	0.021	0.04534	0.0454	0.0454
LPA	511	510	506	0.042	0.042	0.045	0.04411	0.0442	0.0441
LA	522	527	528	0.035	0.032	0.031	0.04464	0.0449	0.0451
PA	522	527	528	0.035	0.032	0.031	0.04463	0.0449	0.0451
MV Information	**530**	**532**	**535**	**0.029**	**0.028**	**0.026**	**0.0449**	**0.045**	**0.045**
DDWFSB	525	542	564	0.032	0.022	0.008	0.04459	0.045	0.0459
DDWTDB-ln(age)	525	526	530	0.033	0.032	0.029	0.04485	0.045	0.0451
DDWTDB-1/ln(age)	523	525	528	0.034	0.033	0.03	0.04477	0.0449	0.045
DDWTDB-LPE	523	525	527	0.034	0.033	0.031	0.0448	0.0449	0.045
DDWTDB-VEX	525	528	530	0.033	0.031	0.03	0.04494	0.0451	0.0451
DIWOTIB	528	529	531	0.031	0.03	0.029	0.04504	0.0451	0.0451
DIWTIB-ln(age)	500	499	499	0.047	0.048	0.048	0.04359	0.0436	0.04353
DIWTIB-1/ln(age)	528	528	530	0.031	0.031	0.03	0.04509	0.0451	0.0451
MDA	537	538	541	0.025	0.025	0.023	0.04511	0.0451	0.0451
LPA	494	495	497	0.052	0.052	0.0507	0.04366	0.0437	0.0437
LA	528	529	531	0.031	0.03	0.029	0.04505	0.0451	0.0451
PA	528	529	531	0.031	0.03	0.029	0.04505	0.0451	0.0451
BSM_HKCL04	573	573	573	0.002	0.002	0.002	0.04616	0.0462	0.0462
BSM_BhSh08	543	543	543	0.021	0.021	0.021	0.04482	0.0448	0.0448
DOC_JW13	570	570	570	0.004	0.004	0.004	0.04615	0.0461	0.0461
MVMI Information	**515**	**519**	**523**	**0.039**	**0.037**	**0.034**	**0.0444**	**0.045**	**0.045**
DDWFSB	513	538	560	0.041	0.025	0.01	0.04392	0.0445	0.0457
DDWTDB-ln(age)	518	520	525	0.038	0.036	0.033	0.04441	0.0446	0.0448
DDWTDB-1/ln(age)	515	518	522	0.039	0.037	0.034	0.04424	0.0445	0.0447
DDWTDB-LPE	516	519	522	0.039	0.037	0.035	0.04446	0.0446	0.0448
DDWTDB-VEX	518	522	522	0.038	0.035	0.035	0.04457	0.0448	0.0448
DIWOTIB	522	523	526	0.036	0.035	0.032	0.04469	0.0448	0.045
DIWTIB-ln(age)	481.6	482.4	487	0.0594	0.0582	0.0553	0.042927	0.04303	0.04313
DIWTIB-1/ln(age)	521	522	525	0.036	0.035	0.033	0.04478	0.0449	0.045
MDA	538	540	540	0.025	0.023	0.023	0.04521	0.0453	0.0451
LPA	493.7	495	497	0.0526	0.052	0.051	0.04364	0.0437	0.0437
LA	521	523	526	0.036	0.035	0.032	0.04469	0.0448	0.045
PA	521	523	526	0.036	0.035	0.032	0.0447	0.0448	0.045
Total	**535**	**533**	**540**	**0.026**	**0.027**	**0.023**	**0.045**	**0.045**	**0.045**

Chapter Five

A Comparative Analysis of Two-Stage Distress Prediction Models

Abstract: On feature selection, in addition to different types of information, i.e., accounting, market and macroeconomic variables, two-stage distress prediction studies used management performance to distinguish distress companies from those healthy ones. DEA is the programming algorithm that is used to estimate the cross-sectional and dynamic efficiency of companies. In addition to the conventional approach of considering management efficiency, this study uses the market efficiency as a measure of company efficiency. Also, it proposes the decomposition of Slack-Based Measure (SBM) into Pure Technical Efficiency (PTE), Scale Efficiency (SE) and Mix Efficiency (ME), and analyses how each of these measures contributes individually in developing distress prediction models. Further, this study provides a comprehensive comparison between static and dynamic two-stage distress prediction models that apply different types of DEA models to compute alternative DEA scores and two different efficiency measures. The results indicate that the measure of market efficiency is not superior to the managerial efficiency of a company, yet, it improves the performance of distress prediction models on some criteria. Moreover, feeding prediction models by decomposed efficiency measures enhances the performance of prediction models.

Keywords: Corporate Two-stage Distress Prediction; Efficiency; Data Envelopment Analysis; Malmquist Index

5.1 Introduction

Financial distress refers to a situation where a firm's cash flows are not enough to fulfill its contractual payments (Piesse et al., 2006, p. 478). Financially distressed companies face detrimental situations, which could adversely affect the value of the company, the welfare of stockholders and creditors, and finally, lead to bankruptcy. Corporate bankruptcy triggers significant losses to both business community and the society as a whole - for details about the costs of bankruptcy, the reader is referred to Davydenko et al. (2012), Elkamhi et al. (2012) and Branch (2002). Therefore, corporate distress prediction has received considerable attention and became a major subject of extensive research in the literature, especially after worldwide financial crises in 2007 and European recession in 2009.

Distress prediction models (DPMs) aim to use broadly recognised sources and indicators of financial distress such as difficulties in operating and financing activities, and poor performance in management and leadership of the company in developing an early warning system to take a proper action against bankruptcy and immune the firm (Altman et al., 2016; Bauer and Agarwal, 2014; Bauweraerts, 2016; Laitinen and Suvas, 2016; Liang et al., 2016; Wu et al., 2010; Yeh et al., 2010).

The related literature on corporate bankruptcy and distress prediction use different techniques from a variety of fields such as statistics and probability, artificial intelligence, and operations research to design new prediction models. Initial studies on failure prediction use statistical techniques such as univariate discriminant analysis (e.g., Beaver, 1966, 1968), and multivariate discriminant analysis (e.g., Altman, 1968, 1973, 1983) as classification techniques. Later on, conditional probability models such as linear probability analysis (e.g., Meyer and Pifer, 1970; Maddala, 1986), logit analysis (e.g., Martin, 1977; Ohlson, 1980) and probit analysis (e.g., Zmijewski, 1984) are used to predict the probability of distress. The common characteristic of these models, however, is that they are static in nature since they only account for a single-period data of studied firms. Dynamic models such as survival analysis (e.g., Lane et al., 1986; Crapp and Stevenson, 1987; Luoma and Laitinen, 1991; Shumway, 2001; Bharath and Shumway, 2008; Chava and Jarrow, 2004) and contingent claims analysis (CCA) (e.g., Bharath and Shumway, 2008; Hillegeist et al., 2004) are the next generation of models which could accommodate changes in the profile of firms over time. Statistical models, however, are restricted to the changes in underlying assumptions, i.e., linearity, multivariate normality, independence among predictor or input variables, and equal within-group variance-covariate matrices.

The less vulnerable techniques to the underlying statistical assumptions are the ones from the field of artificially intelligent expert systems (AIES) such as recursively partitioned decision trees (e.g., Frydman et al., 1985), case-based reasoning models (e.g., Li and Sun, 2009, 2011), neural networks (e.g., Kim and Kang, 2010; du Jardin and Séverin, 2012), rough set theory (e.g., McKee and Lensberg, 2002; Yeh et al., 2010), genetic programming (e.g., Back et al., 1996; Alfaro-Cid et al., 2007; Etemadi et al., 2009), as well as ones from field of operations research (OR), such as multi-criteria decision making analysis (MCDA) (e.g., Zopounidis and Doumpos, 2002) and data envelopment analysis (DEA) (e.g., Sueyoshi and Goto, 2009; Sueyoshi et al., 2010; Li et al., 2014) – for a detailed classification of failure prediction models, the reader is referred to Balcaen and Ooghe (2006), Aziz and Dar (2006), Bellovary et al. (2007), Baharammirzaee (2010), Abdou and Pointon (2011), and Chen et al. (2016).

According to Zhou (2013), bankruptcy and distress prediction models are data-fitting based empirical research containing four steps of sampling, features selection, modelling, and performance evaluation. Several studies have compared the performance of competing DPMs considering different sampling approaches – e.g., Neves and Vieira (2006), and Zhou (2013); alternative modelling frameworks – e.g., Wu et al. (2010), Bauer and Agarwal (2014), and Mousavi et al. (2015); various features – e.g., Tinoco and Wilson (2013), and Trujillo-Ponce et al. (2014); and different performance evaluation approach – e.g., Mousavi et al. (2015).

On feature selection, studies have used various types of information including accounting, market and macroeconomic variables in developing bankruptcy and distress prediction models. However, it is commonly acknowledged that one of the main reasons for corporate distress is the poor performance of management (Seballos et al., 1990; Gestel et al., 2006; Yeh et al., 2010). Recent studies have included the relative efficiency of the business operations as a capable reflection of the management efficiency of a company in the bankruptcy and distress prediction models (Xu and

Wang, 2009; Yeh et al., 2010; Li et al., 2014). The management efficiency refers to the company's ratio of weighted outputs (e.g., sales, profit, and net income) to weighted inputs (e.g., equity, asset, and employees) with regards to the performance of other companies.

A direct estimation of the company efficiency (i.e., technical, operating and productivity efficiency) using financial statements is difficult. One of the widely applied techniques is data envelopment analysis (DEA), which could incorporate multiple inputs and outputs to estimate the efficiency measure of a corporation relative to the most efficient ones (see next section). In this chapter, I develop two-stage distress prediction models through using different DEA techniques to compute various measures of company efficiency in the first stage and using the computed company efficiency measure as an input explanatory variable of the classifier model at the second stage.

My survey of the literature reveals that, to the best of my knowledge, no study provides a comprehensive comparison between two-stage distress prediction models; neither considering different DEA models that are used to estimate company efficiency nor using different classifier models at the second stage. Further, my survey indicates that the choice of input and output to estimate company efficiency measures using DEA models is restricted to accounting variables rather than market variables (see, Table 5.1).

This study adds to the current literature of two-stage distress prediction models in several respects. First, considering the pioneer of Li et al. (2014) in analysing the contribution of different DEA measures in predicting financial distress, I propose to study the decomposition of the Non-Radial Technical Efficiency score, i.e., Slack-Based Measure (SBM) of efficiency (Tone, 2001) into Pure Technical Efficiency (PTE), which presents the ability to improve the effectiveness by prudently allocating resources and using new technology, Scale Efficiency (SE), which indicates capacity to attain better efficiency by adjusting to its optimal scale, and Mix Efficiency (ME), which shows capacity to improve the effectiveness by managing input- or output-slacks, and analyse how each of these measures individually contributes to developing distress prediction models.

Second, in addition to the conventional approach of using accounting variables and firm's characteristics as inputs and outputs of DEA models to estimate the measure of management efficiency, this study is the first to use market information variables to calculate the measure of the market efficiency of companies as a predictor in a two-stage prediction model. Third, this study provides a comprehensive analysis of two-stage distress prediction models that apply different DEA models – say, input-oriented vs. output-oriented, radial vs. non-radial, static vs. dynamic, to compute the measures of management efficiency and market efficiency of companies at the first stage of two-stage models and use dynamic and static classifier frameworks at the second stage of two-stage models.

The remainder of the chapter is organised as follows. Section 5.2 reviews the literature background of DEA in distress prediction. Section 5.3 describes details on my experimental design including data, sampling, and hybrid two-stage models of distress prediction to be assessed and the proposed evaluation technique. Section 5.4 describes the empirical results and the findings. Finally, section 5.5 provides the conclusion of the chapter.

5.2 Literature review

DEA is a non-parametric technique which is introduced to measure the relative efficiency of a group of decision-making units (DMUs), e.g. firms, hospitals, products, prediction models, cities, and others, based on their respective inputs and outputs (Charnes et al., 1978). DEA has been one of the most successfully used techniques in the research activities related to performance evaluation of banking and other financial institutions – for a comprehensive survey on DEA in banking; the reader is referred to Emrouznejad and Yang (2017), Paradi and Zhu (2013) and Fethi and Pasiouras (2010). The rational association between the company's efficiency (as a proxy of management efficiency) and the probability of distress is commonly recognised in recent distress prediction studies (Seballos et al., 1990; Gestel et al., 2006; Yeh et al., 2010).

Table 5.1: **The Inputs and Outputs of DEA Models where DEA Score Used as a Predictor**

Reference	First stage DEA model	Second stage model	Inputs of DMUs	Outputs of DMUs
Barr et al. (1994)	CCR-IO	Probit	(1) Full-time equivalent employees, (2) Salary expenses, (3) Premises and fixed assets, (4) Other noninterest expenses, (5) Total interest expense, (6) Purchased funds	(1) Core deposit, (2) Earning assets, (3) Total interest income
Barr and Siems (1997)	BCC-IO	Probit	(1) Full-time equivalent employees, (2) Salary expenses, (3) Premises and fixed assets, (4) Other noninterest expenses, (5) Total interest expense, (6) Purchased funds	(1) Core deposit (2) Earning assets (3) Total interest income
Xu and Wang (2009)	CCR-IO	Logit, MDA, SVM	(1) Total assets, (2) Total liabilities, (3) Cost of sales	(1) Income from sales
Psillaki et al. (2010)	DD-VRS-OO	Logistic regression	(1) Capital stock, (2) Labor	(1) Value-added
Sueyoshi et al. (2010)	RAM-DEA	Tobit regression	(1) Cost of goods sold, (2) the total number of employees, (3) the book value of plant and equipment	(1) Total revenue
Yeh et al. (2010)	CCR-OO	RTS-SVM	(1) R&D expense, (2) R&D designers, (3) number of patents and trademarks	(1) Gross profit, (2) market share
Li et al. (2014)	SBM-VRS-IO SBM-CRS-IO	Logistic regression	(1) Number of employees, (2) share capital, (3) total cost, (4) total assets, (5) total liabilities	(1) Total sales, (2) total profit, (3) cash accrued
Li et al. (2017)	Malmquist SBM-VRS-IO Malmquist Super-SBM-VRS-IO	Multi-logit regression	(1) Number of employees, (2) share capital, (3) total cost, (4) total assets, (5) total liabilities	(1) Total sales, (2) total profit, (3) cash accrued

More relevant to this research, distress prediction studies applied DEA in two different ways. First, DEA is used as a classifier to discriminate between distress and healthy groups (e.g., Cielen et al., 2004; Paradi et al., 2004; Sueyoshi, 2006; Premachandra et al., 2011; Ouenniche and Tone, 2017). Second, in the hybrid two-stage prediction models, DEA is used to measure the relative efficiency of companies at the first stage.

Then, the estimated DEA efficiency score is used as an input explanatory variable of the model at the second stage (e.g., Xu and Wang, 2009; Sueyoshi et al., 2010; Psillaki et al., 2010; Yeh et al., 2010; Li et al., 2014, 2017). The next two sections provide a concise review of the application of DEA in distress prediction as a classifier (see section 5.2.1) and as a predictor (see section 5.2.2). Table 5.2 provides a brief explanation of these studies in the literature.

5.2.1 DEA as a Classifier

Compared with the conventional statistical models, DEA as a non-parametric classifier has some methodological advantages. For example, DEA is a distribution-free framework and does not require specifying the distribution of features. Also, DEA relaxes the assumption of equality of variance-covariance matrices among all groups. Further, it does not incorporate a *priori* probabilities to account for the relative occurrence of observations in different populations and does not require *a priori* specification of a functional form for the input-output correspondences (Paradi et al., 2004; Premachandra et al., 2009).

The application of DEA score as a classifier in the literature has been twofold; first, some studies used DEA score to discriminate between two groups of *Goods* (e.g. bankrupt) and *Bads* (e.g. non-bankrupt) using a cut-off point or statistical test (Simak, 1997; Pille and Paradi, 2002; Paradi et al., 2004; Cielen et al., 2004; Tsai et al., 2009; Shetty et al., 2012),and second, some studies used DEA score as dependent variable while other features of the firm as independent variables to establish the extent to which DEA results coincide with those of regression analysis, discriminant analysis, etc (Bowlin, 2004; Emel et al., 2003; Min and Lee, 2008) (see Table 5.2, for more details).

Table 5.2: DEA application in Distress Prediction

Reference	DEA incorporated as	Assumed technology	Main objectives
Barr et al. (1993)	Benchmark	CCR-IO	To apply DEA as a tool to quantify a bank's managerial efficiency for predicting bankruptcy.
Barr et al. (1994)	Efficiency score as predictor	CCR-IO	To apply DEA as a tool to quantify a bank's managerial efficiency as a predictor in developing bankruptcy prediction model.
Simak (1997)	Classifier	BCC-IO	To apply DEA as a classifier for predicting corporate distress.
Barr and Siems (1997)	Efficiency score as predictor	BCC-IO	To apply DEA as a tool to measure management efficiency as a predictor in a logistic regression model.
Sueyoshi (2001)	Classifier	DEA-DA	To propose a new type of DEA-Discriminant Analysis (DA), or "Extended DEA-DA," that can overcome some methodological shortcomings of its original formulation (Sueyoshi, 1999). The new extended framework is used as a classifier.
Pille and Paradi (2002)	Classifier	BCC-IO	To examine the effectiveness of various DEA scores computed using different inputs and outputs in detecting financial weakness of individual Credit Unions in Ontario, in the years before failure.
Emel et al. (2003)	Benchmark	CCR-IO	To apply DEA as a tool to measure creditability scores represent the dependent variable in a regression model with six independent variables (financial ratios)
Cielen et al. (2004)	Classifier	CCR-OO	To apply DEA as a classifier for predicting corporate distress and compare the prediction performance with two other bankruptcy prediction models (i.e. Linear programming and a rule induction (C5.0)).
Paradi et al. (2004)	Classifier	BCC-IO	To propose the concept of worse practice DEA in combination with a layering technique as a new classifier approach.
Sueyoshi (2004)	Classifier	DEA-DA	To apply DEA in a new proposed DEA-DA approach to suggest a new type of mixed integer programming formulation that estimates weights of a linear discrimination function by minimizing the total number of misclassified observations.
Bowlin (2004)	Benchmark	BCC-IO	To apply DEA in assessing the financial stability of CRAF participants.
Sueyoshi (2006)	Classifier	DEA-DA	To apply DEA in two DEA-DA approaches (i.e., Standard MIP & Two-stage MIP models) as classifiers and compare it with the prediction performance of six other bankruptcy prediction models (i.e. Logit, Probit, Fisher's linear DA, Smith's quadratic DA, Neural network & Decision tree)
Min and Lee (2008)	Benchmark	CCR-IO	To apply DEA to measure efficiency scores represent a dependent variable in two types of econometric model (i.e. regression analysis and discriminant analysis) to predict bankruptcy.

Reference	DEA incorporated as	Assumed technology	Main objectives
Premachandra et al. (2009)	Classifier	Additive-DEA	To apply DEA (additive model) as a classifier for assessing corporate failures compared to the Logistic technique.
Sueyoshi and Goto (2009a)	Classifier	DEA-DA	To apply DEA in a DEA-DA approach as a classifier and compare the prediction performance with Altman's Z-score (1968) model. Also, DEA-DA is applied to see whether R&D expenditure is effective on the financial performance of firms.
Sueyoshi and Goto (2009b)	Classifier	DEA-DA	To apply DEA in a DEA-DA approach in combination with PCA to reduce the computation burden and then alter DEA-DA weights to address both the sample imbalance problem and the location problem.
Tsai et al. (2009)	Classifier	DEA-DA	To apply DEA in combination with DA (DEA-DA) (Sueyoshi et al., 2004) as a classifier and compare the accuracy with another type of bankruptcy models.
Xu and Wang (2009)	Efficiency score as predictor	CCR-IO	To apply DEA to measure efficiency as a predictor in three main failure prediction models, i.e. MDA, Logit and SVM and compare the predictive performance of models.
Sueyoshi et al. (2010)	Benchmark	RAM-DEA	To apply RAM (Range Adjusted Measure; Cooper et al., 1999; Aida et al., 1998) as a DEA model to measure efficiency scores of firms, which are then represented as the dependent variable in a Tobit-regression model to investigate if the reform of the corporate governance influences the performance of companies.
Psillaki et al. (2010)	Efficiency score as predictor	DD-VRS-OO	To apply DEA as a tool to measure a productive inefficiency - the distance from the industry's best practice frontier- as a predictor in developing bankruptcy prediction model.
Yeh et al. (2010)	Efficiency score as predictor	CCR-OO	To apply DEA to measure efficiency as a predictor in a new model namely RTS-SVM.
Premachandra et al. (2011)	Classifier	Additive-DEA	To propose an approach based on the additive super-efficiency DEA to overcome drawbacks of the proposed model by Premachandra et al. (2009).
Mukhopadhyay et al. (2012)	Classifier	CCR-IO	To apply DEA in combination with (Multilayer Perception) MLP in two sequential stages to predict bankruptcy and compare the prediction performance of an MLP model.
Shetty et al. (2012)	Classifier	BCC-NO	To propose an orientation-free, not-radial directional distance DEA model to measure worst relative efficiency within the range of zero to one.
Khalili and Makvandi (2013)	Classifier	Additive-DEA	To apply DEA as a classifier to predict the bankruptcy probability of firms and compare the prediction performance with three other bankruptcy prediction models i.e. Logit, Probit, and MDA.

Reference	DEA incorporated as	Assumed technology	Main objectives
Li et al. (2014)	Efficiency score as predictor	SBM-VRS-IO SBM-CRS-IO	To propose a new application of DEA in bankruptcy prediction through assuming VRS and decomposing Technical Efficiency (TE) into Pure Technical Efficiency (PTE) and Scale Efficiency (SE).
Avkiran and Cai (2014)	Classifier	Super-SBM-CRS-NO	To apply super efficiency SBM-CRS-NO DEA model (Tone,2002) as a classifier to explore whether an ex-post sample of financially distressed bank holding companies from the USA can be identified as inefficient using pre-global financial crises data.
Paradi et al. (2014)	Classifier	SBM-CRS-IO	To apply DEA to fix appropriate cut-off points to classify healthy and non-healthy firms.
Li et al. (2017)	Efficiency score as predictor	Malmquist SBM-VRS-IO	To extend the cross-sectional DEA models to time- varying Malmquist DEA.

Barr et al. (1993) used DEA under CCR (Charnes, Cooper and Rhods, 1978) condition to measure the management efficiency of US banks. They found that the gap in the efficiency score between non-failed and failed banks is both significant and increasing as the failure date approaches.

Pille and Paradi (2002) developed four input-oriented DEA models (with different combinations of inputs and outputs) under BCC (Banker, Charnes, and Cooper) (1984) to predict financial failure of Credit Unions. The performance of DEA efficiency scores was statistically compared with a government modified “Z-score” model and “equity to asset” ratio. Overall, inconsistent with Barr et al. (1993), they found that failure unions, especially at one year before the failure time, have lower scores than healthy ones.

Paradi et al. (2004) proposed the worse practice DEA analysis under BCC _ aimed at finding the companies that are efficient at being bad_ in combination with a layering technique rather a fixed cut-off point to classify manufacturing firms into bankrupt and non-bankrupt. Also, they employed a different combination of inputs (the drivers of bad performance like current liabilities, interest expense, and bad debt) and outputs (the drivers of good performance like total asset, sales, profit) to find out the best set of inputs/outputs. The result suggested that combining three layers of the best of the worst practice improves the classification accuracy of identifying bankrupt and non-bankrupt firms up to 100 and 67 percent, respectively.

Cielen et al. (2004) applied DEA under CCR for bankruptcy prediction in comparison with a linear programming model (minimised sum of deviations (MSD)) and a rule induction (C5.0) model. They suggested using financial ratios with a positive correlation as inputs and those with a negative correlation as outputs. Regarding prediction accuracy, the result indicated that DEA outperforms both C5.0 and MSD models. However, the main methodological issue is that CCR cannot deal with negative values of financial ratios.

Bowlin (2004) analysed cross-sectional and longitudinal differences in DEA scores under BCC, over a 10-year period, 1988 -1997, to compare the financial stability of different groups of firms, using the presented statistical approach by Banker (1993).

Emel et al. (2003) and Min and Lee (2008) applied an input-oriented CCR-DEA to measure financial performance, namely, creditability scores, which then used to classify firms with scores equal to one and less than one as companies with a good and relatively worse financial performance, respectively. To validate the discriminatory power of DEA, they used DEA score as the dependent variable and financial ratios as independent variables in regression and discriminant analysis. The results suggested that DEA is a valid method for estimating the creditworthiness of companies. Premachandra et al. (2009) employed the additive DEA model of Charnes et al. (1982) for bankruptcy prediction to take advantage of its specific features. First, the additive DEA allows for negative values of inputs and outputs (namely, translation invariance property). Second, in contrast to radial models (CCR and BCC), which require examination of both a DEA efficiency score and slacks to estimate the efficiency of a DMU, the additive model requires the consideration of slacks only. Third, while the radial DEA model based on an input-oriented or an output-oriented measurement results in different efficiency scores, the additive model includes both input and output slacks in the efficiency analysis. The comparison of the additive DEA model with Logistic regression (LR) indicates that the DEA model (respectively LR) outperforms (respectively underperforms) in predicting non-bankrupt (respectively bankrupt) firms.

However, additive DEA has some drawbacks. First, the additive DEA and conventional DEA select reverse order of input and output variables, which lead to different results (Shetty et al., 2012). Second, the additive model does not provide an efficiency score in-between [0, 1]. In other words, although the estimated measure can discriminate bankrupt and non-bankrupt firms, it fails to evaluate the depth of bankruptcy (Premachandra et al., 2011; Shetty et al., 2012).

To overcome the above drawbacks, Premachandra et al. (2011) applied super-efficient additive DEA model (Du et al., 2010) to develop a discriminant index based upon two frontiers, namely failure and success. Switching input-output classification identifies these two frontiers. Therefore, for determining failure (respectively success) frontier, the smaller (respectively larger) values in the financial ratios are considered as input (respectively output), and the larger (respectively lower) values in those ratios are considered as output (respectively input). The result indicates that the super-efficiency

DEA model is relatively weaker in predicting failure firms compared to non-failure companies. However, the discriminant index based on two frontiers improves this weakness by giving the practitioners the option to choose different accuracy level of failure, non-failure, and total prediction.

Also, to overcome the shortcomings of the application of the additive DEA by Premachandra et al. (2009) in bankruptcy prediction, Shetty et al. (2012) proposed a modified efficiency measure using orientation-free non-radial directional distance formulation of DEA. In contrary to the additive DEA, this approach measures the worst relative efficiency within the range of zero to one. Further, in contrary to the conventional DEA methods, this method identifies the worst performers and locates an inefficient frontier.

Sueyoshi (1999) proposed a new type of discriminant analysis, namely "DEA-Discriminant Analysis (DEA-DA)" that incorporates the methodological advantages of DEA (for example, nonparametric and distribution-free features) into the discriminant analysis. This two-stage approach is designed to identify the existence of an overlap between the two groups at the first stage and to determine a group classification function for new observation samples at the second stage. Sueyoshi (2001) proposed the "extended DEA-DA" approach, which has two important features; 1) it can deal with negative values, and 2) it can estimate the weights of a DA function by minimising the total distance of misclassified observations. However, the drawback of the "extended DEA-DA" model is that it does not reduce the number of misclassified observations (as explained in accuracy performance evaluation), but the total distance of misclassified observations. To overcome this methodological issue, Sueyoshi (2004) proposed a mixed integer programming (MIP) version of DEA-DA to estimate the weights of the linear discrimination function by minimising the total number of misclassified observation. Furthermore, Sueyoshi (2006) compared the performance of two advanced versions of DEA-DA classifiers, namely standard MIP and two-stage MIP models with six other bankruptcy prediction models; logit, probit, Fisher's linear DA, Smith's quadratic DA, neural network and decision tree. Tsai et al. (2009) also used the MIP version of DEA-DA (Sueyoshi et al. 2004) as a predictor of loan default and compared its accuracy with DA, LR and NN models. The result suggests that DEA-DA and NN have the better-classifying capability.

Further, in a proposed two-stage model, Sueyoshi et al. (2010) applied RAM (range-adjusted measure: Cooper et al., 1999; Aida et al., 1998) as a DEA model to measure the operational efficiency scores of Japanese companies, in the first step. In the second step, the efficiency score is used as the dependent variable in a Tobit regression to investigate if the corporate governance variables influence the operational efficiency of firms.

Mukhopadhyay et al. (2012) proposed a combination of DEA and Multi-Layer Perceptron (MLP) to predict failure. For this, at the first step, they used super efficiency negative DEA to identify the worst performers amongst the non-failed firms (i.e. companies with an efficiency score more than 1). The recognised worst non-failed firms in combination with failed firms are labeled as failed group and then used to train the MLP at the second step. The developed MLP then used for failure prediction. The proposed technique, therefore, recognises firms that have a high likelihood of facing failure along with those that have filed for bankruptcy.

Avkiran and Cai (2014) applied a super-SBM DEA model (Tone, 2001, 2002) as a forward-looking approach to predict the distressed bank holding companies. Results suggested that DEA could identify distressed banks up to 2 years ahead.

More recently, Ouenniche and Tone (2017) applied BCC and SBM DEA models to estimate efficiency scores of LSE listed companies and proposed a customized k-Nearest Neighbour (K-NN) algorithm to determine an optimum DEA score-based cut-off point, which then is used to classify firms into bankrupt and non-bankrupt ones.

5.2.2 DEA Score as a Predictor

In the recent trend of DEA application in distress prediction, the use of DEA efficiency score as a feature in developing models is becoming more prevalent. In the earliest study, Barr and Siems (1997) used DEA under CCR condition to measure the managerial efficiency of US banks at the first stage and then applied the efficiency score as a predictor in a probit model at the second stage. Their findings suggest that removing management efficiency variable from the full model decreases the model's fit and classification accuracy.

Xu and Wang (2009) used DEA under BCC condition to estimate the efficiency score of Chinese firms at the first step. The second step compares the prediction accuracy of three failure prediction models, namely, SVM, MDA and logistic regression, with and without DEA efficiency. The results indicate that using efficiency score improves the performance of prediction models effectively.

Yeh et al. (2010) applied DEA under CCR condition to measure the efficiency of Taiwanese information and electronic manufacturing firms. The estimated DEA efficiency score and a list of frequently used financial ratios are employed as inputs of the second stage, namely rough set theory (RTS), to select the most significant features. Finally, the selected features from RTS are used as inputs of support vector machines (SVM) to predict business failures.

Psillaki et al. (2010) proposed a two-stage model of credit risk prediction. In the first stage, they applied DEA under BCC to estimate the directional distance function, which is used to determine the efficiency scores of a sample of French manufacturing firms. The firm efficiency score measures the company's distance from the industry's best practice frontier. In the second stage, they used logistic regression to evaluate the effect of company's efficiency in predicting failure over and above that explained by financial features.

Li et al. (2014) proposed a new application of DEA in bankruptcy prediction through using SBM-VRS to estimate Technical Efficiency (TE) and decomposing TE into Pure Technical Efficiency (PTE) and Scale Efficiency (SE) for a sample of Chinese companies, at the first stage. In the second stage, these efficiency measures along with other financial ratios are used in a Logistic analysis regression to predict the probability of failure. Further, they allowed the impact of a variety of efficiency scores across all industries on the probability of failure through introducing an interaction term into the model.

Most of the applications of DEA in corporate failure and distress prediction used cross-sectional or static DEA models that fail to consider the changes in efficiency over time. The only exception, to the best of my knowledge, is Li et al. (2017) that applied time-varying Malmquist DEA to estimate dynamic efficiency scores and conduct them in a dynamic prediction model. Further, most of the studies used DEA under constant returns-to-scale (CRS) condition (see for example, Paradi et al., 2004; Xu and Wang, 2009; Yeh et al., 2010; Avkiran and Cai, 2014; Mukhopadhyay et al., 2012) rather than VRS condition (see for example, Psillaki et al., 2010; Li et al., 2014). Also, on DEA orientation, three studies (Cielen et al., 2004; Psillaki et al., 2010; Yeh et al., 2010) are output oriented, though the majority are input oriented. Finally, while a large number of studies have estimated the DEA efficiency scores of firms using financial accounting variables (e.g. total assets, total liabilities, total sales, employees, cash flow, etc.), as far as I am aware only one (Avkiran and Cai, 2014) has estimated efficiency using market variables (e.g., market capitalisation, annual stock return, liquid asset,etc.) as inputs and outputs of DEA models.

5.3 Research Methodology

This section provides the details of my research methodology, where I compare the performance of two-stage distress prediction models. For this, I provide the details on my dataset (see section 5.3.1), the static and dynamic DEA models used in the first stage of two-stage DPMs (see section 5.3.2), and the static and dynamic models specification in the second stage of two-stage DPMs (see section 5.3.3).

5.3.1 Data

I took the following steps to select my dataset. First, I considered all non-financial and non-utility UK companies listed on the London Stock Exchange (LSE) at any time during an 8-year period from 2007 through 2014 - Financial and utility companies are excluded because they are regulated. Second, I excluded the firms which are listed less than two years in LSE as historical information is a requirement for some modelling frameworks. Third, I excluded the firms with missing values for the principal accounting items (e.g., sales, total assets) and market information (e.g., price), which are necessary elements for calculating many financial ratios (Lyandres and Zhdanov, 2013). I replaced the remaining missing values with the recently observed ones for each firm (Zhou et al., 2012). Fourth, I winsorized the outlier values through replacing the values higher (respectively, lower) than 99th (respectively, 1st) percentile of each variable with that 99th (respectively, 1st) percentile value (Shumway, 2001).

Regarding the classification of firms into distress and non-distress, I followed the proposed definition of financial distress by Pindado et al. (2008), where a company is classified as distressed if it experiences both of the following conditions for two consecutive years. First, the company's earnings before interest, taxes, depreciation and amortization (EBITDA) is lower than its interest expenses, and second, the company shows negative growth in market value. To be more specific, the distress variable, say y, equals 1 for financially distressed companies and equals 0 otherwise. In sum, my dataset consists of 2,096 firms and 11,943 firm-year observations. Among the total number of observations, there are 676 firm-year observations classified as distressed resulting in a distress rate average of 5.66 percent per year. The models are developed using training sample period ranges from 2007 to 2011, and tested using holdout sample period ranges from 2012 to 2014. Table 5.3 presents the sample sizes.

Table 5.3: Sample Sizes

Samples	Year	Healthy	Distressed	Total	Distress rate
Training sample (2007 - 2011)	2007	1,826	81	1,907	4.25%
	2008	1,704	106	1,810	5.86%
	2009	1,456	165	1,621	10.18%
	2010	1,409	61	1,470	4.15%
	2011	1,354	27	1,381	1.96%
	Total	*7,749*	*440*	*8,189*	*5.27%*
Training sample (2012-2014)	2012	1,255	69	1,324	5.21%
	2013	1,143	101	1,244	8.12%
	2014	1,120	66	1,186	5.56%
	Total	*3,518*	*236*	*3,754*	*6.29%*
Total		**11,267**	**676**	**11,943**	**5.66%**

5.3.2 Stage One: Estimating Efficiency Measures Using DEA Models

In this section, I explain cross-sectional (static) DEA models (see section 5.3.2.1), and Malmquist DEA model (see section 5.3.2.2) applied in the first stage of two-stage DPMs. Then, I describe the choice of inputs and outputs for DEA models (see section 5.3.2.3).

5.3.2.1 Static DEA Models

Several types of DEA models can be used depending on the conditions of the problem. Further, types of DEA model can be identified based on scale and orientation of the model. In this study, to compute the cross-sectional efficiency measures of companies, I apply CCR (Charnes, Cooper, Rhodes, 1978), BCC (Banker, Charnes, and Cooper, 1984), and Slack-Based Measure (SBM) (Tone, 2001) DEA models considering both input-oriented (IO) and output-oriented (OO) analyses – See Table 5.4 and Table 5.5 for details about DEA models. Also, I apply SBM-DEA model under assumptions of constant returns-to-scale (CRS) and variable returns-to-scale (VRS), separately.

Table 5.4: CCR and BCC Models

Formulation	Description
θ_k	Objective. This is Technical Efficiency (TE) score in CCR model and Pure Technical Efficiency (PTE) score in BCC model. The objective is to minimise the efficiency score (θ) in the input-oriented version of the model and to maximise the efficiency score (θ) in the output-oriented version of model.
$\sum_{j=1}^{n} \lambda_j x_{i,j} \leq \theta_k . x_{i,k};\ \forall_i$ (1) or $\sum_{j=1}^{n} \lambda_j x_{i,j} \leq x_{i,k};\ \forall_i$ (2)	For each input i $(i = 1, \dots, m)$, the amount used by DMU_k's ideal benchmark; i.e., its projection on the efficient frontier $(\sum_{j=1}^{n} \lambda_j x_{i,j})$, should be at most be equal to the amount used by DMU_k whether revised (i.e., amount of input i adjusted for the degree of technical efficiency of DMU_k) or not depending on whether the model is input-oriented (1) or output-oriented (2).
$\sum_{j=1}^{n} \lambda_j y_{r,j} \geq y_{r,k};\ \forall_r$ (1) Or $\sum_{j=1}^{n} \lambda_j y_{r,j} \geq \theta_k . y_{r,k};\ \forall_r$ (2)	For each output r $(r = 1, \dots, s)$, the amount used by DMU_k's ideal benchmark; i.e., its projection on the efficient frontier $(\sum_{j=1}^{n} \lambda_j y_{r,j})$, should be at least as large as the amount produced by DMU_k whether revised (i.e., amount of output r adjusted for the degree of technical efficiency of DMU_k) or not depending on whether the model is output-oriented (2) or input-oriented (1).
$\sum_{j=1}^{n} \lambda_j = 1$	BCC model requires that the technology is convex. CCR model does not need this restriction.
$\lambda_j \geq 0;\ \forall_j$	Non-negativity requirements

Source: Ouenniche and Tone (2017)

Note that the CCR and BCC scores are called the (global) technical efficiency (TE) and the (local) pure technical efficiency (PTE), respectively. The BCC model estimates the efficiency of DMUs when returns-to-scale (RTS) is not necessarily constant, i.e., it takes account of scale effect and postulates that convex combinations of the observed DMUs from the production possibility set (William W. Cooper et al., 2006, p. 153). If a DMU has full BCC efficiency but a low CCR efficiency, then it is operating locally efficient but not globally efficient because of the scale size of the DMU. Considering these concepts and denoting CCR and BCC as θ_{CCR}^* and θ_{BCC}^*, respectively, the *scale efficiency* (SE) is defined as (Charnes et al., 1978)

$$SE = \frac{\theta^*_{CCR}}{\theta^*_{BCC}} \qquad \text{Eq. 5-1}$$

Therefore, the technical efficiency could be decomposed as

$$TE = PTE \times SE \qquad \text{Eq. 5-2}$$

The advantage of this decomposition is that it determines the sources of inefficiency, i.e., whether it is due to inefficient operation (PTE) or due to detrimental conditions displayed by the scale efficiency (SE) or by both.

Table 5.5: SBM Model

Formulation	Description
$\rho_k = 1 - \frac{1}{m}\left(\sum_{i=1}^{m} \frac{s^-_{i,k}}{x_{i,k}}\right)$	Objective. That is, input-oriented SBM measure
$\rho_k = \frac{1}{1 + \frac{1}{s}\left(\sum_{r=1}^{s} \frac{s^+_{r,k}}{y_{r,k}}\right)}$	Objective. That is, output-oriented SBM measure
$\sum_{j=1}^{n} \lambda_j x_{i,j} + s^-_{i,k} = x_{i,k};\ \forall_i$	For each input $i\ (i = 1, \dots, m)$, the amount used by $DMU'_k s$ "ideal" benchmark; i.e., its projection on the efficient frontier, should be at most equal to the amount used by DMU_k; that is; $\sum_{j=1}^{n} \lambda_j x_{i,j} \le x_{i,k};\ \forall_i$
$\sum_{j=1}^{n} \lambda_j y_{r,j} - s^+_{r,k} = y_{r,k};\ \forall_r$	For each output $r\ (r = 1, \dots, s)$, the amount produced by $DMU_k{}'s$ "ideal" benchmark; i.e., its projection on the efficient frontier, should be at least as large as the amount produced by DMU_k; that is $\sum_{j=1}^{n} \lambda_j y_{r,j} \ge y_{r,k};\ \forall_r$
$\sum_{j=1}^{n} \lambda_j = 1$	BCC model requires that the technology is convex. CCR model does not need this restriction.
$\lambda_j \ge 0;\ \forall_j;\ s^-_{i,k};\ \forall_i;\ s^+_{r,k};\ \forall_r$	Non-negativity requirement

Source: Ouenniche and Tone (2017)

Moreover, radial DEA models, i.e., CCR and BCC, overlook possible slacks in inputs and outputs, and therefore, would possibly over-estimate the efficiency scores by ignoring mix efficiency. The SBM model is a non-radial model that considers slacks in inputs and outputs. Note that the equality of optimal input-oriented (respectively, output-oriented) SBM measure, i.e., p^*_{in} (respectively, ρ^*_{out}), and optimal input-oriented (respectively, output-oriented) CCR measure , i.e., θ^*_{CCR-in} (respectively, $\theta^*_{CCR-out}$) holds, i.e., $\rho^* = \theta^*_{CCR}$, if the input-oriented (respectively, output-oriented) CCR model has zero input-slacks (respectively, output-slacks) for every optimal solution. In other words, the strict inequality, i.e., $\rho^*_{in} < \theta^*_{CCR-in}$ (respectively, $\rho^*_{out} < \theta^*_{CCR-out}$) holds if and only if the CCR measure indicates an input (respectively, output) mix inefficiency. Considering these concepts, the input and output "mix efficiency" (ME) are defined by Cooper et al.(2006, p. 154)

$$ME_{in} = \frac{\rho_{in}^{*}}{\theta_{CCR-in}^{*}} \quad \text{and} \quad ME_{out} = \frac{\rho_{out}^{*}}{\theta_{CCR-out}^{*}} \qquad \text{Eq. 5-3}$$

Considering the equation 5-1 (the decomposition of TE), the non-radial input- or output- oriented technical efficiency (SBM) could be decomposed into mixed efficiency (ME), pure technical efficiency (PTE) and scale efficiency (SE) as:

$$SBM = PTE \times SE \times ME \qquad \text{Eq. 5-4}$$

In this study, I use CCR-IO, CCR-OO, BCC-IO, BCC-OO, SBM-CRS-IO, SBM-CRS-OO, SBM-VRS-IO and SBM-VRS-OO models to measure the cross-sectional managerial efficiency and market efficiency of companies. Also, I decompose SBM measure of each company into ME, PTE and SE, and incorporate them in developing distress prediction models in the second stage.

5.3.2.2 Dynamic DEA Model

To estimate the efficiency measures of companies over time, I use Malmquist DEA productivity index (Fare et al., 1992, 1994). Malmquist productivity index (MPI) is a multi-criteria assessment framework for comparing the performance of DMUs over time. Fare et al. (1992, 1994) used DEA to extend the original Malmquist Index proposed by Malmquist (1953) and constructed the DEA-based Malmquist productivity index as the product of two components; (1) caching-up to the frontier, which refers to the efficiency change (EC) of DMU with respect to the efficiency possibilities defined by the frontier in each period, and (2) efficient frontier-shift (EFS), which refers to the shift of efficient frontier between the two time periods t and $t+1$ (see, Table 5.6 for details about Malmquist productivity index).

Table 5.6: Malmquist DEA Model

Formulation	**Description**
$EC = \frac{PF/_{PB}}{QC/_{QA}}$ Equivalently, $EC = \frac{\Delta^{t+1}((x_0,y_0)^{t+1})}{\Delta^{t}((x_0,y_0)^{t})}$	The efficiency change (EC) component; Referring to Figure 5.1, $PF\ /\ PB$ and $QC\ /\ QA$ represent the efficiency of DMU_o at period $+1$, say $\Delta^{t+1}((x_0, y_0)^{t+1})$ and at period t, $\Delta^{t}((x_0, y_0)^{t})$, respectively. Also, x_{i0}^{t} denote the ith input and y_{r0}^{t} denote the rth output for DMU_0, both at period t. Figure 5-1shows the change of efficiency of DUM_0 from point A (with respect to efficient frontier at period t) to point B (with respect to efficient frontier at period $t+1$). Thus, $EC > 1$ shows an improvement in the relative efficiency from period t to $t+1$, while $EC = 1$ and $EC < 1$ shows stability and deterioration in the relative efficiency, respectively.
$EFS = [EFS_t \times EFS_{t+1}]^{1/2}$ $EFS_t = \frac{QC}{QD} = \frac{QC/_{QA}}{QD/_{QA}}$ and $EFS_{t+1} = \frac{PE}{PF} = \frac{PE/_{PB}}{PF/_{PB}}$ Equivalently, $EFS = \left[\frac{\Delta^{t}((x_0,y_0)^{t})}{\Delta^{t+1}((x_0,y_0)^{t})} \times \frac{\Delta^{t}((x_0,y_0)^{t+1})}{\Delta^{t+1}((x_0,y_0)^{t+1})}\right]^{1/2}$	The efficiency frontier-shift (EFS) component is the product of the Geometric mean of EFS_t and EFS_{t+1}; Referring to Figure 5.1, $QC\ /\ QA$ and $QD\ /\ QA$ ratios denote the efficiency of DMU_o at period t with respect to period t frontier, say $\Delta^{t}((x_0, y_0)^{t})$ and period $t+1$ frontier, say $\Delta^{t+1}((x_0, y_0)^{t})$, respectively. Also, $PE\ /\ PB$ and $PF\ /\ PB$ ratios represent the efficiency of DMU_0 at period $t+1$ with respect to period t frontier, say $\Delta^{t}((x_0, y_0)^{t+1})$ and period $t+1$ frontier, say $\Delta^{t+1}((x_0, y_0)^{t+1}$.
$MPI = EC \times EFS$ Equivalently, $MPI = \frac{\Delta^{t+1}((x_0,y)^{t+1})}{\Delta^{t}((x_0,y_0)^{t})} \times \left[\frac{\Delta^{t}((x_0,y_0)^{t})}{\Delta^{t+1}((x_0,y_0)^{t})} \times \frac{\Delta^{t}((x_0,y_0)^{t+1})}{\Delta^{t+1}((x_0,y_0)^{t+1})}\right]^{1/2}$	The Malmquist Productivity Index (MPI) is the product of the efficiency change (EC) and the efficiency frontier-shift (EFS). This explanation of MPI could be interpreted as the geometric mean of efficiency change measured by period t and $t+1$ technology, respectively. $MPI > 1$ shows an improvement in the total factor productivity of DMU_0 from period t to - $t+1$, while $MPI = 1$ and $MPI < 1$ shows stability and deterioration in total factor productivity, respectively.

Figure 5-1: Efficiency Change and Efficient Frontier-Shift

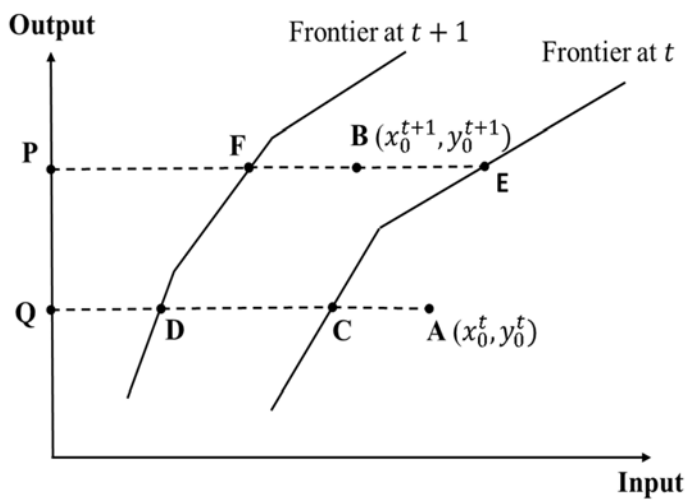
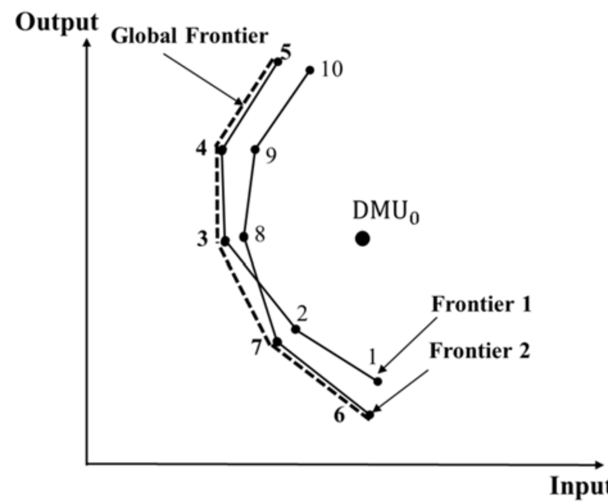


Figure 5-2: Global Frontier



Caves et al. (1982) introduced a distance function, $\Delta(.)$, to measure technical efficiency in the basic CCR model (Charnes et al., 1978). Though, in the non-parametric framework, instead of using a distance function, DEA models are implemented. For example, Fare et al. (1994) used input (or output) oriented radial DEA to measure the MPI. However, the radial model faces a lack of attention to slack that could be overcome using Slacks-based non-radial oriented (or orientation-free) DEA model (Tone, 2001, 2002). Along with measuring cross-sectional DEA scores (section 5.3.2.1), I incorporate CCR-IO, CCR-OO, BCC-IO, BCC-OO, SBM-CRS-IO, SBM-CRS-OO, SBM-VRS-IO and SBM-VRS-OO models to measure the MPI. Also, I decompose SBM measure of each company into ME, PTE and SE, and incorporate them in developing dynamic distress prediction models in the second stage.

Global Malmquist Productivity Index

The primary objective of this study is to estimate the relative efficiency of $DMUs$ (companies) in each period. However, the estimated Malmquist productive index (MPI), say, $MPI_0^{t,t+1}$, indicates the change of efficiency score between period t and $t+1$, and should be modified for my purpose. Further, referring to Pastor and Lovell (2005), the contemporaneous MPI is not circular, its adjacent period components can give conflicting signals, and it is sensitive to LP infeasibility.

The adjacent reference index, proposed by Fare et al. (1982), suggests multiplying $MPI_0^{t,t+1}$ by $\Delta_0^t(x_0^t, y_0^t)$, which results in the relative efficiency of DMU_0 at period $t+1$ compared to period t. However, the main drawback of this index is that it cannot estimate the relative efficiency score of non-adjacent periods, e.g., period t and $t+2$ or $t+1$ and $t+3$. To overcome this drawback, Berg et al., (1992) used a fixed reference index, which compares and refers the relative efficiencies of all periods (say, t $(t \geq 2)$) to the first period (say, $t = 1$). Therefore, it is possible that the efficiency scores of the periods later than the first one are more than 1 since the technology develops over time. Although, fixed reference index acquires the circularity property with a base period dependence, it remains sensitive to LP infeasibility. More recently, Pastor and Lovell (2005) suggested a global MPI, which its components are circular, it provides a single measure of productivity change, and it is not susceptible to LP infeasibility. Further, in a situation, where efficient frontiers of multiple periods cross each other, the global index can be measured by the best practices in all periods. As Figure 5.2 presents, the relative efficiency of DMU_0 can be measured in terms of either the frontier of period 1 (consists of four DMUs of 1,2,3,4 and 5) or the frontier of period 2 (consist of four DMUs of 6,7,8,9 and 10). An alternative is the global frontier, which is the combination of the best DMUs in the history, i.e. five DMUs of 6,7,3,4 and 5.

It is argued that if the length of observation period is long enough, the current DMUs would be covered by the best historical DMUs, probably themselves. Thus, the relative efficiency to the global frontier could be considered as an absolute efficiency with the scores less than or equal to 1 (Pastor and Lovell, 2005).

5.3.2.3 Choice of Inputs and Outputs for the First Stage

To select suitable inputs and outputs for DEA models, I considered the following issues. First, the survey on the application of DEA in bankruptcy and distress prediction indicates that there is no approved procedure for the choice of inputs and outputs of DEA – see Table 5.1 above. In practice, different DEA applications use different inputs and outputs, which is one of the drawbacks of DEA application (Premachandra et al., 2009). However, the choice of inputs and outputs should be related to the competitive environment (Oral and Yolanda, 1990).

Second, regarding most of two-stage prediction models, since financial ratios are used as features in the second stage, the monetary items of financial statements are used as inputs and outputs of DEA models in the first stage (Li et al., 2014, see, for example, 2017; Psillaki et al., 2010; Xu and Wang, 2009). Third, to the best of my knowledge, two-stage studies only used accounting items as inputs and outputs of DEA models to compute managerial efficiency of companies (see, Table 5.8 for details). Fourth, to deal with negative values in inputs and outputs of DEA, the following popular approaches have been proposed: The Range Directional Measure introduced by Portela et al., 2004), the Modified Slack-Based Measure introduced by Sharp et al., 2006, the Semi-Oriented Radial Measure introduced by Emrouznejad et al., 2010 and Variant of Radial Measure introduced by Cheng et al., 2013.

In this study, for estimating management efficiency, I selected three inputs (Total Liabilities, Total Shareholders' Equity and Number of Employees) and one output (Total Sales). Also, for estimating market efficiency, I selected one input (Lag of return volatility) and two outputs (Lag of excess return and Market Value).

To estimate efficiency measures using different DEA models, I employed MaxDEA that deals with negative values in inputs (such as shareholders' equity) and outputs (such as lag of excess return) using the variant radial measure approach (Cheng et al., 2013). Table 5.7 represents the descriptive statistics of inputs and outputs.

Table 5.7: Descriptive Statistics of Inputs and Outputs of DEA Models
This table presents the descriptive statistics of inputs and outputs of DEA models. The numbers of Total liabilities, Shareholder's equity, sales and market value are in thousand USD.

		Total liabilities	Shareholders' Equity	Employees	Sales	Lag (Sigma)	Lag (Excess Return)	Market Value
Healthy (N=11276)	mean	435881.2	287003.7	4762	595548.9	0.157	-0.108	531480.5
	sd	1501620.8	916985.6	22305	1897213.1	0.114	0.613	1755247.6
	skewness	5.0	4.8	14	4.6	1.672	-0.408	5.0
	kurtosis	28.5	26.2	282	24.8	6.266	4.469	29.1
	min	59.0	-23630.0	3	0.0	0.014	-1.967	590.0
	max	9476000.0	5592000.0	537784	11368000.0	0.601	1.580	11453960.0
Distress (N=676)	mean	45302.9	39396.6	472	26386.4	0.220	-0.706	24442.7
	sd	279600.1	206334.7	1444	136668.1	0.129	0.604	98950.0
	skewness	12.5	12.7	6	13.3	0.867	-0.010	12.8
	kurtosis	187.4	205.3	52	217.1	3.739	3.614	190.6
	min	59.0	-23630.0	2	0.0	0.014	-1.967	590.0
	max	5049200.0	3905187.0	18457	2595600.0	0.601	1.580	1597430.0
Total (N=11943)	mean	413773.6	272988.6	4519	563333.1	0.160	-0.142	502781.0
	sd	1462801.9	893836.9	21690	1847707.0	0.116	0.628	1709028.0
	skewness	5.1	4.9	15	4.8	1.603	-0.395	5.2
	kurtosis	30.2	27.8	298	26.4	5.947	4.235	30.8
	min	59.0	-23630.0	2	0.0	0.014	-1.967	590.0
	max	9476000.0	5592000.0	537784	11368000.0	0.601	1.580	11453960.0

5.3.3 Stage Two: Developing Distress Prediction Model

In this stage, I fed logistic regression with static DEA scores and selected features to develop static DPMs. Also, I fed multi-period logistic regression with dynamic DEA scores and selected features to develop dynamic DPMs.

5.3.3.1 Static Logit Model

Since the seminal work of Ohlson (1980), the logit analysis has become a frequently used static model in the distress and bankruptcy prediction (see, for example, Martin, 1977; Ohlson, 1980; Back et al., 1996; Duda et al., 2010). In the field of financial distress prediction, the dependent variable is a binary variable, takes on two values, zero or one. The generic model for binary variables could be stated as follows:

$$\begin{cases} P(distress) = P(y_i = 1|x_i) \\ P(distress) = G(\beta, X) \end{cases} \qquad \text{Eq. 5-5}$$

where Y denotes the binary response variable, X denotes the vector of covariates, β denotes the vector of coefficients of covariates in the model, and $G(.)$ is a link function that maps the scores of βx, onto a probability. In practice, depending on choice of link function, the type of probability model is determined. As for the logit regression model, the link function is the cumulative logistic distribution function, say Θ.

$$G(\beta, X) = \Theta^{-1}(\beta^t X) \qquad \text{Eq. 5-6}$$

which is between zero and one for all real numbers $\beta^t X$. For my analysis, I specified logistic regression to be

$$logit(p_i) = \alpha + \sum_{i=1}^{n}\sum_{r=1}^{m} \theta_r S_{ri} + \sum_{i=1}^{n}\sum_{j=1}^{l} \beta_j \, x_{ji} \qquad \text{Eq. 5-7}$$

where p_i denotes the probability of facing distress for company i; S_{ri} denotes the static efficiency score r for company i; θ_r denotes a parameter for the static efficiency score r to be estimated; x_{ji} denotes the feature j for company i, and β_j is a parameter for feature j to be estimated.

5.3.3.2 Dynamic Discrete-Time Hazard Model

Shumway (2001) proposed a discrete time hazard model using an estimation procedure similar to the one used for determining the parameters of a multi-period (dynamic) logit model. Many studies have applied this approach for computing the probability of a hazard occurrence (see, for example, Cheng et al., 2010; Shumway, 2001; Nam et al., 2008; El Kalak and Hudson, 2016). General notation of discrete time hazard model could be presented as follows:

$$P(y_{i,t} = 1|x_{i,t}) = h(t|x_{i,t}) = \frac{e^{(\alpha_t + x_{i,t}\beta)}}{1 + e^{(\alpha_t + x_{i,t}\beta)}} = h_0(t).\, e^{x_{i,t}\cdot\beta} \qquad \text{Eq. 5-8}$$

where $h(t|x_{i,t})$ represent the individual hazard rate of firm i at time t, $x_{i,t}$ is the vector of covariates of each firm i at time t; β denotes the vector of coefficients; α_t is the time-variant baseline hazard function related, which could be relate to firm, e.g. ln(age) or related to macroeconomic variables, e.g. volatility of exchange rate (Nam et al., 2011). Shumway (2001) used a constant time variant term; say $ln\ (age)$, as proxy of baseline rate. For my analysis, I modified the discrete-time hazard model to be

$$logit\big(h_{i,d=1}(t)\big) = \alpha + \beta_0 h_0(t) + \sum_{i=1}^{n}\sum_{r=1}^{m} \theta_r D_{rit} + \sum_{i=1}^{n}\sum_{j=1}^{l} \beta_j\, x_{jit} \qquad \text{Eq. 5-9}$$

where $h_{i,d=1}$ denotes the probability of facing distress for company i at time t; $h_0(t)$ denotes the baseline hazard function; D_{rit} denotes the dynamic efficiency score r for company i at time t; x_{jit} denotes the feature j for company i at time t; β_0 is the coefficient of the baseline hazard rate to be estimated; θ_r is a parameter for the dynamic efficiency score r at time t to be estimated; and β_j is a parameter for feature j at time t to be estimated.

5.3.3.3 Choice of Features for the Second Stage

To select suitable features for prediction models, I applied the following steps. First, I reviewed the literature to select the most commonly used features in other studies (e.g., Hebb, 2016; du Jardin, 2015; Zhou, 2014, 2013; Ravi Kumar and Ravi, 2007), including 83 accounting-based ratios and 7 market-based information. Second, I used *t-test* method to choose features which show a significant difference between two group's means (Shin and Lee, 2002; Huang et al., 2004; Shin et al., 2005).

Third, for further reduction of features, I applied factor analysis, and principal component analysis with VARIMAX technique (Chen, 2011, Mousavi et al., 2015). To be more specific, I used factors analysis to select the variables that both the absolute values of their loadings and communities are greater than 0.5 and 0.8, respectively. Fourth, 34 variables which presented high factor loadings and high communality values, were retained as input features into the stepwise procedure in the second stage of two-stage distress prediction models (see, Table 5.8), where a stepwise procedure for each framework is used to select the most significant features.

Table 5.8: List of Financial Ratios

Category	Ratio or item	Category	Ratio or item
Profitability (9)	Net income to total liabilities EBIT to total assets Return on assets Operating income after depreciation to total assets Retained earnings to total assets Expected return on assets Total Liabilities Exceed Total Assets Changes in net income in two consecutive years Negative net income for last two years	Liquidity (9)	Current asset turnover Current assets to total liabilities Current liabilities to current assets Inventory to current assets Inventory turnover Inventory to total assets Profit before tax to current liabilities Quick asset to total assets Quick asset to inventory
Asset utilization (2)	Asset turnover ratio Quick assets to sales	Solvency (3)	Current Liabilities to Liabilities Equity to capital Long term and current liabilities to total assets
Cash flow (2)	Operating cash flow to liabilities Funds Provided by Operations to Total Liabilities	Market information (5)	Lag of excess return Lag of sigma [1] Ln (price) Real size Failure rate in last year
Mixed (2)	GDP × Sales Interest rate × Income	Firm characteristics (2)	Ln (age) Log (total assets to GNP price level index)

5.3.3.4 Choice of Efficiency Scores for the Second Stage

Table 5.9 presents the descriptive statistics of static and dynamic managerial efficiency measures for two groups of distress and healthy companies. The result of *F-test* suggests that in most cases, input-oriented DEA scores discriminate better between two groups of distress and healthy firms. Therefore, I select input oriented managerial efficiency measures, i.e., CCR-IO, BCC-IO, SBM-CRS-IO and SBM-VRS-IO, and use equations 5-1 and 5-3 to compute SE-IO and ME-IO for the second stage. Table 5.10 shows the descriptive statistics of static and dynamic market efficiency measures for two groups of distress and healthy companies. The result of *F-test* indicates that in most cases output-oriented DEA models discriminate better between two groups of distress and healthy firms. Then, I chose output oriented market efficiency measures, i.e., CCR-OO, BCC-OO, SBM-CRS-OO, SBM-VRS-OO, SE-OO and ME-OO for the second stage. I retain the selected static and dynamic scores and relate them to the probability of distress using equations 5-4 and 5-6, respectively.

[1] The explanation about Lag of Sigma and Lag of excess return is provided in Table 2-A.

Table 5.9: Descriptive Statistics of Managerial Efficiency Scores

		Static										Dynamic									
		CCR (TE)	**BCC (PTE)**		**SE**		**SBM-CRS**		**ME**	**SBM-VRS**		**CCR (TE)**	**BCC (PTE)**		**SE**		**SBM-CRS**		**ME**	**SBM-VRS**	
		IO & OO	**IO**	**OO**	**IO**	**OO**	**IO**	**OO**	**IO**	**IO**	**OO**	**IO & OO**	**IO**	**OO**	**IO**	**OO**	**IO**	**OO**	**IO**	**IO**	**OO**
Healthy (N=11276)	mean	0.13	0.55	0.21	0.35	0.70	0.10	0.13	0.76	0.39	0.21	0.08	0.48	0.14	0.30	0.71	0.06	0.08	0.79	0.34	0.14
	sd	0.15	0.29	0.24	0.34	0.26	0.13	0.15	0.11	0.22	0.24	0.09	0.31	0.20	0.32	0.28	0.08	0.09	0.09	0.22	0.20
	skewness	2.75	0.10	1.90	0.64	-0.91	3.63	2.75	0.03	0.78	1.90	4.67	0.26	2.83	0.92	-0.77	5.74	4.67	-0.28	0.70	2.83
	kurtosis	13.52	1.70	6.32	1.91	2.89	21.63	13.52	2.30	3.22	6.32	36.81	1.67	11.42	2.47	2.41	55.39	36.81	3.36	2.88	11.42
	min	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.40	0.01	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.43	0.01	0.00
	max	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Distress (N=676)	mean	0.05	0.74	0.09	0.10	0.70	0.04	0.05	0.73	0.52	0.09	0.03	0.71	0.05	0.07	0.72	0.02	0.03	0.78	0.50	0.05
	sd	0.09	0.24	0.18	0.18	0.29	0.07	0.09	0.11	0.21	0.18	0.04	0.25	0.08	0.15	0.27	0.03	0.04	0.09	0.20	0.08
	skewness	5.33	-1.01	3.91	3.01	-0.92	7.03	5.33	0.18	0.04	3.91	3.01	-0.91	6.22	3.63	-1.00	2.58	3.01	-0.15	-0.14	6.22
	kurtosis	46.58	3.03	18.97	12.35	2.67	79.64	46.58	2.38	2.78	18.97	15.57	2.81	58.32	17.96	2.92	12.35	15.57	3.04	2.71	58.32
	min	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.42	0.02	0.00	0.00	0.02	0.00	0.00	0.01	0.00	0.00	0.45	0.02	0.00
	max	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.31	1.00	1.00	1.00	1.00	0.24	0.31	0.99	1.00	1.00
Total (N=11943)	mean	0.12	0.56	0.20	0.34	0.70	0.10	0.12	0.76	0.39	0.20	0.07	0.50	0.14	0.29	0.71	0.06	0.07	0.79	0.35	0.14
	sd	0.15	0.29	0.24	0.34	0.26	0.13	0.15	0.11	0.22	0.24	0.09	0.31	0.20	0.32	0.28	0.08	0.09	0.09	0.22	0.20
	skewness	2.80	0.04	1.95	0.70	-0.91	3.68	2.80	0.04	0.73	1.95	4.71	0.20	2.90	0.98	-0.79	5.79	4.71	-0.27	0.63	2.90
	kurtosis	13.91	1.68	6.53	2.00	2.88	22.28	13.91	2.30	3.09	6.53	37.70	1.63	11.95	2.61	2.44	56.89	37.70	3.34	2.76	11.95
	min	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.40	0.01	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.43	0.01	0.00
	max	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F-test	*182.67*	*312.23*	*157.30*	*363.37*	*0.02*	*159.71*	*182.67*	*34.21*	*237.73*	*157.30*	*156.07*	*370.11*	*150.50*	*335.84*	*2.44*	*148.50*	*156.07*	*13.86*	*325.24*	*150.72*
	p-value	0.00	0.00	0.00	0.00	0.89	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.12	0.00	0.00	0.00	0.00	0.00

Table 5.10: Descriptive Statistics of Market Efficiency Scores

		Static										Dynamic									
		CCR (TE)	**BCC (PTE)**		**SE**		**SBM-CRS**		**ME**	**SBM-VRS**		**CCR (TE)**	**BCC (PTE)**		**SE**		**SBM-CRS**		**ME**	**SBM-VRS**	
		IO & OO	**IO**	**OO**	**IO**	**OO**	**IO**	**OO**	**OO**	**IO**	**OO**	**IO & OO**	**IO**	**OO**	**IO**	**OO**	**IO**	**OO**	**OO**	**IO**	**OO**
Healthy (N=11276)	mean	0.14	0.20	0.60	0.73	0.21	0.14	0.03	0.11	0.20	0.07	0.16	0.17	0.98	0.97	0.16	0.16	0.02	0.08	0.17	0.06
	sd	0.17	0.21	0.21	0.26	0.18	0.17	0.10	0.21	0.21	0.18	0.15	0.16	0.01	0.13	0.16	0.15	0.07	0.19	0.16	0.16
	skewness	2.73	2.08	-0.46	-0.99	2.16	2.73	6.08	2.60	2.08	3.59	3.20	3.04	-0.33	-6.42	3.20	3.20	6.16	3.49	3.04	3.91
	kurtosis	11.92	7.56	3.27	3.13	9.01	11.92	46.10	9.15	7.56	16.01	15.47	14.02	4.03	43.32	15.50	15.47	51.09	15.10	14.02	18.95
	min	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.02	0.97	0.06	0.02	0.02	0.00	0.00	0.02	0.00
	max	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Distress (N=676)	mean	0.08	0.12	0.39	0.58	0.15	0.08	0.00	0.06	0.12	0.01	0.12	0.13	0.98	0.98	0.13	0.12	0.00	0.01	0.13	0.00
	sd	0.16	0.18	0.21	0.28	0.18	0.16	0.03	0.17	0.18	0.06	0.18	0.19	0.01	0.07	0.18	0.18	0.00	0.02	0.19	0.01
	skewness	3.71	3.49	0.48	-0.39	3.13	3.71	25.10	3.79	3.49	13.24	3.49	3.42	0.07	-12.05	3.49	3.49	11.93	12.07	3.42	11.89
	kurtosis	16.70	15.09	3.51	2.25	13.07	16.70	643.51	17.23	15.09	193.18	14.83	14.32	3.28	150.26	14.82	14.83	171.65	172.19	14.32	172.22
	min	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.02	0.97	0.09	0.02	0.02	0.00	0.00	0.02	0.00
	max	0.98	1.00	1.00	1.00	1.00	0.98	0.66	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	0.99	0.05	0.33	1.00	0.23
Total (N=11943)	mean	0.14	0.19	0.59	0.72	0.20	0.14	0.03	0.11	0.19	0.07	0.16	0.16	0.98	0.97	0.16	0.16	0.02	0.08	0.16	0.06
	sd	0.17	0.21	0.22	0.26	0.18	0.17	0.09	0.21	0.21	0.18	0.16	0.16	0.01	0.12	0.16	0.16	0.07	0.19	0.16	0.15
	skewness	2.75	2.13	-0.42	-0.94	2.20	2.75	6.25	2.64	2.13	3.70	3.21	3.05	-0.31	-6.56	3.21	3.21	6.35	3.61	3.05	4.04
	kurtosis	11.99	7.74	3.11	3.03	9.12	11.99	48.64	9.40	7.74	16.93	15.36	13.96	3.83	45.18	15.39	15.36	54.08	16.04	13.96	20.11
	min	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.02	0.97	0.06	0.02	0.02	0.00	0.00	0.02	0.00
	max	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	F-test	*83.43*	*94.02*	*618.33*	*217.48*	*57.11*	*83.43*	*44.30*	*35.92*	*94.02*	*85.72*	*27.66*	*34.15*	*610.93*	*2.66*	*26.99*	*27.66*	*52.74*	*98.67*	*34.15*	*89.45*
	p-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.00	0.00	0.00	0.00	0.00	0.00

5.4 Empirical Results

The objective of this study is to evaluate the relative performance of two-stage distress prediction models using UK data. Section 5.4.1provides the assessment of models using the conventional unidimensional-ranking framework. Section 5.4.2 assesses the models using a multi-criteria evaluation framework.

5.4.1 Unidimensional Ranking of Distress Prediction Models

For unidimensional ranking of different models, I use commonly applied performance criteria in the literature; i.e., the discriminatory power, the calibration accuracy, the information content, and the correctness of categorical prediction. Regarding the discriminatory power criterion that measures how much a prediction model is capable of discriminating between distressed firms and healthy ones, I use Receivable Operating Characteristic (ROC), Kolmogorov-Smirnov (KS) statistics, Gini Index (GI), and Information Value (IV) as measures. Regarding the calibration accuracy criterion that measures how much a model is qualified in estimating the probability of distress (PD), I use Brier Score (BS) as measure. Regarding the information content criterion that measures the extent to which the output of a model (e.g., PD, scores) carries enough information for prediction, I follow Agarwal and Taffler (2008) and use a log-likelihood statistic (LL) and pseudo-R^2 as measures. Finally, with respect to the correctness of categorical prediction criterion that measures how often a model can predict distressed firms (respectively, healthy firms) as distressed (respectively, healthy) ones, I use Type I errors (T1), Type II errors (T2), misclassification rate (MR), sensitivity (Sen), specificity (Spe), and overall correct classification (OCC) as measures (for more details about performance criteria and measures, the reader is referred to Mousavi et al. (2015)). Table 5.11, Table 5.12, Table 5.13 and Table 5.14 present the estimated distress prediction models. The χ^2 tests indicate that all 34 models explain significant amount of variation in the probability of distress. Tables 5.11 and 5.12 present the estimated static models using stepwise procedure in a logit framework using managerial efficiency scores and market efficiency scores, respectively.

Table 5.11: Static Model Results Using Managerial Efficiency Scores of Companies

This table presents the static models using managerial efficiency scores developed in a logit framework. Model 1 does not contain any measure of efficiency. Models 2, 3 and 4 use TE (CCR), PTE (BCC) and SE to develop prediction models. Model 5 uses decomposed measures of TE (i.e., BCC and SE) to develop prediction model. Models 6 and 7 use SBM-CRS and ME-CRS efficiency scores to develop models. Model 8 uses the decomposed measures of SBM-CRS (i.e., BCC-CRS, SE-CRS and ME-CRS). Model 9 uses SBM-VRS to develop prediction model. *** and ** refer to 1% and 5% significance level, respectively.

Covariates	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**	**Model 8**	**Model 9**
Intercept	-1.333***	-1.291***	-2.022***	6.598***	-1.723***	-1.291***	-0.848*	-1.449**	-1.722***
Retained Earnings to Total Assets				-6.488***					
Negative Net Income for Last Two Years	1.783***	1.709***	1.738***	1.608***	1.649***	1.703***	1.772***	1.647***	1.775***
Lag of Excess Return	-3.603***	-3.560***	-3.608***	-3.479***	-3.583***	-3.553***	-3.576***	-3.570***	-3.618***
Log (Total asset to GNP index)	-1.537***	-0.992*	-0.782		-0.2956	-0.946*	-1.500***	-0.3213	-1.061*
Real Size	-2.43***	-2.648***	-2.488***	-3.074***	-2.618***	-2.663***	-2.483***	-2.641***	-2.454***
Current liabilities over Current assets				-5.924*					
log (price)	0.103***	0.104***	0.103***	0.100***	0.1013***	0.103***	0.103***	0.101***	0.1049***
Inventory Turnover				2.399***					
IR × NI				-12.891***					
CCR-IO		-1.8769**							
BCC-IO			0.6936**		0.3923			0.3747	
SE-IO				-0.6797**	-0.9955***			-0.965***	
SBM-CRS-IO						-2.626***			
ME-CRS-IO							-0.647	-0.3424	
SBM-VRS-IO									0.495
Number of observations	8189	8189	8189	8189	8189	8189	8189	8189	8189
Log likelihood	2629	2621	2624	2594	2615	2620	2626	2614	2627
$Prob. > \chi^2$	0	0	0	0	0	0	0	0	0
Pseudo R^2	0.2720	0.2745	0.2735	0.2834	0.2767	0.2751	0.2728	0.2770	0.2725

Table 5.12: Static Model Results Using Market Efficiency Scores of Companies

This table presents the static models using market efficiency scores developed in a logit framework. Models 10,11 and 12 use TE (CCR), PTE (BCC) and SE to develop prediction models. Model 13 uses decomposed measures of TE (i.e., BCC and SE) to develop prediction model. Models 14 and 15 use SBM-CRS and ME-CRS efficiency scores to develop models. Model 16 uses the decomposed measures of SBM-CRS (i.e., BCC-CRS, SE-CRS and ME-CRS). Model 17 uses SBM-VRS to develop prediction model. *** and ** refer to 1% and 5% significance level, respectively.

Covariates	**Model 10**	**Model 11**	**Model 12**	**Model 13**	**Model 14**	**Model 15**	**Model 16**	**Model 17**
Intercept	1.006*	1.369**	1.054*	-1.323***	-1.327***	1.109*	-1.105***	-1.351***
Asset Turnover Ratio	-12.376***	-12.583***	-12.322***			-13.043***		
Quick Assets to Inventory	3.391***	3.336***	3.396***			3.524***		
Negative Net Income for Last Two Years	1.618***	1.631***	1.62***	1.758***	1.776***	1.636***	1.758***	1.776***
Lag of Excess Return	-3.300***		-3.419***	-0.457	-3.566***	-3.775***	-1.1058	-3.544***
Sigma		-0.5805**						
Log (Total asset to GNP index)	-1.599***	-1.575***	-1.592***	-1.580***	-1.522***	-1.308***	-1.349***	-1.519***
Real Size	-2.935***	-3.086***	-2.887***	-2.491***	-2.448***	-3.169***	-2.766***	-2.367***
log (price)	0.107***	0.109***	0.1084***	0.108***	0.102***	0.105***	0.107***	0.102***
IR × NI	-15.366***	-15.876***	-15.338***			-14.765***		
CCR-OO	-0.914**							
BCC-OO		-3.375**		-2.940**			-2.563**	
SE-OO			-0.648*	-0.3576			-0.1245	
SBM-CRS-OO					-2.596			
ME-CRS-OO						-0.120***	-1.193***	
SBM-VRS-OO								-1.615
Number of observations	8189	8189	8189	8189	8189	8189	8189	8189
Log likelihood	2596	2588	2597	2620	2626	2575	2599	2626
$Prob. > \chi^2$	0	0	0	0	0	0	0	0
Pseudo R^2	0.2828	0.2854	0.2824	0.2748	0.2731	0.2896	0.2816	0.2730

Also, Tables 5.13 and 5.14 indicate the estimated dynamic models using stepwise procedure in a multi-period logit framework using managerial efficiency scores and market efficiency scores, respectively. Retained earnings to total assets, negative net income for last two years, lag of excess return, log (total asset to GNP index), real size, current liabilities over current assets, log (price) and inventory turnover are amongst the selected variables using stepwise procedure.

Table 5.15 presents the performance measures of 34 developed distress prediction models. The results could be summarised as follows. First, considering the performance of models without efficiency measures, i.e., the one-stage static model 1 and the one-stage dynamic model 18, and models fed with efficiency measures, i.e., two-stage models, the results suggest that incorporating efficiency measures improve the performance of models.

Second, comparing the performance of dynamic models with static models in my study, for most of the performance measures, the dynamic models outperform static ones. To be more specific, on most performance measures – see, for example, T1, ROC, Gini, KS, IV, CIER, BS, LL and R^2 the two-stage dynamic models are superior to static ones. However, considering T2, MR and OCC as performance measures of correctness of categorical prediction, static models 16, 11 and 10 are amongst the best performers. In general, the density of dynamic models amongst the top-ranking performers suggests their superiority in performance. The superiority of dynamic to static models could be related to their competence in incorporating time-varying features of the firms. This finding indicates that taking to account the multi-period performance of companies over time as an explanatory variable in a dynamic framework is an appropriate technique to improve the performance of prediction models.

Third, considering the performance of two-stage models with different types of company efficiency measures, i.e., market efficiency and managerial efficiency, for most of the performance measures, the models with management efficiency outperform the models with market efficiency. The reason is that, as the *F*-tests in Table 5.9 and Table 5.10 indicate, the discriminatory power of the management efficiency scores are more than market efficiency measures.

Table 5.13: Dynamic Model Results Using Managerial Efficiency Scores of Companies

This table presents the dynamic models using managerial efficiency scores developed in a multi-period logit framework. Model 18 does not contain any measure of efficiency. Models 19,20 and 21 use TE (CCR), PTE (BCC) and SE to develop prediction models. Model 22 uses decomposed measures of TE (i.e., BCC and SE) to develop prediction model. Models 23 and 24 use SBM-CRS and ME-CRS efficiency scores to develop models. Model 25 uses the decomposed measures of SBM-CRS (i.e., BCC-CRS, SE-CRS and ME-CRS). Model 26 uses SBM-VRS to develop prediction model. *** and ** refer to 1% and 5% significance level, respectively.

Covariates	**Model 18**	**Model 19**	**Model 20**	**Model 21**	**Model 22**	**Model 23**	**Model 24**	**Model 25**	**Model 26**
Intercept	0.604***	0.554*	-0.063	-0.416	0.1052	-0.03315	0.9161	0.6454	0.5362
Asset Turnover Ratio	-18.169***	-11.175**	-16.189***		-10.725**		-17.654***	-9.931**	-17.81***
Negative Net Income for Last Two Years	1.636***	1.604***	1.613***	1.608***	1.564***	1.626***	1.635***	1.562***	1.636***
Lag of Excess Return	-3.843***	-3.851***	-3.864***	-3.816***	-3.926***	-3.625***	-3.847***	-3.930***	-3.847***
Log (Total asset to GNP index)	-2.119***	-1.774***	-1.568**		-1.254**		-2.120***	-1.300**	-2.054***
Real Size	-1.954***	-1.951***	-1.955***	-2.314***	-1.828***	-2.930***	-1.947***	-1.817***	-1.946***
Inventory Turnover	2.821***	2.746***	2.735***	1.992***	2.526***	2.511***	2.774***	2.464***	2.794***
IR × NI	-17.992***	-18.185***	-16.189***	-14.64***	-16.603***	-16.07***	-17.921***	-16.656***	-17.878***
Log (Age)	2.231***	2.307***	2.277***	2.364***	2.463***	2.248***	2.242***	2.488***	2.225***
CCR-IO		-3.517*							
BCC-IO			0.470		0.075			0.00968	
SE-IO				-2.539***	-1.549***			-1.618***	
SBM-CRS-IO						-11.264***			
ME-CRS-IO							-0.4247	-0.6247	
SBM-VRS-IO									0.0639
Number of observations	8189	8189	8189	8189	8189	8189	8189	8189	8189
Log likelihood	2565	2561	2563	2564	2552	2571	2564	2553	2565
$Prob. > \chi^2$	0	0	0	0	0	0	0	0	0
Pseudo R^2	0.2927	0.2938	0.2934	0.2931	0.2969	0.2908	0.2929	0.2966	0.2927

Table 5.14: Dynamic Model Results Using Market Efficiency Scores of Companies

This table presents the dynamic models using market efficiency scores developed in a multi-period logit framework. Models 27,28 and 29 use TE (CCR), PTE (BCC) and SE to develop prediction models. Model 30 uses decomposed measures of TE (i.e., BCC and SE) to develop prediction model. Models 31 and 32 use SBM-CRS and ME-CRS efficiency scores to develop models. Model 33 uses the decomposed measures of SBM-CRS (i.e., BCC-CRS, SE-CRS and ME-CRS). Model 34 uses SBM-VRS to develop prediction model. *** and ** refer to 1% and 5% significance level, respectively.

Covariates	**Model 27**	**Model 28**	**Model 29**	**Model 30**	**Model 31**	**Model 32**	**Model 33**	**Model 34**
Intercept	0.620	23.572***	0.619*	-14.355	0.685	0.644	-85.840	0.79***
Asset Turnover Ratio	-18.19***	-18.195***		-18.194***	-17.976***		-17.961***	-18.131***
Negative Net Income for Last Two Years	1.633***	1.633***	1.633***	1.633***	1.601***	1.710***	1.609***	1.577***
Lag of Excess Return	-3.8141***	-3.007***	-3.814***	-4.3461	-3.711***	-3.794***	-6.855	-3.720***
Log (Total asset to GNP index)	-2.130***	-2.127***	-2.130***	-2.130***	-2.070***	-2.283***	-2.015***	-2.333***
Real Size	-1.9671***	-1.961***	-1.966***	-1.968***	-1.221**		-0.546	
Inventory Turnover	2.807***		2.807***	2.813***	2.791***		2.828***	2.783***
Change in Net income in two years		2.805***						0.285
IR × NI	-18.028***	-18.021***	-18.01***	-18.014***	-20.044***	-22.099***	-21.777***	-23.134***
Log (Age)	2.235***	2.235***	2.235***	2.235***	2.303***	2.042***	2.331***	
CCR-OO	-0.129							
BCC-OO		-23.775		-15.509			-89.708	
SE-OO			-0.127	-0.188			-0.260	
SBM-CRS-OO					-118.14***			
ME-CRS-OO						-28.958***	-28.201***	
SBM-VRS-OO								-45.112***
Number of observations	8189	8189	8189	8189	8189	8189	8189	8189
Log likelihood	2565	2564	2565	2547	2551	2571	2565	2545
$Prob. > \chi^2$	0	0	0	0	0	0	0	0
Pseudo R^2	0.2928	0.2929	0.2928	0.2983	0.2972	0.2906	0.2927	0.2989

Table 5.15: Unidimensional Ranking of Distress Prediction Models

The bold figures represent the best one in each static or dynamic framework. The grey cells represent the top-three best models.

				Correctness of categorical prediction				Discriminatory Power					Calibration Accuracy		Information Content	
Model	**DEA Score**	**Efficiency Score**	**Framework**	**Type I**	**Type II**	**MR**	**OCC**	**AUROC**	**Gini**	**KS**	**IV**	**CIER**	**HL**	**BS**	**LL**	**R2**
Model 01	NA	NA	Static	0.314	0.202	0.209	0.791	0.8351	0.670	0.578	0.505	-0.395	128.78	0.05696	1587	4.56%
Model 02	CCR_IO	Managerial	Static	0.275	0.199	0.204	0.796	0.8421	0.684	0.573	0.551	-0.387	119.85	0.05679	1576	4.84%
Model 03	BCC_IO	Managerial	Static	0.305	0.203	0.209	0.791	0.8354	0.671	0.583	0.547	-0.385	**148.20**	0.05690	1584	4.64%
Model 04	SE_IO	Managerial	Static	**0.258**	0.209	0.212	0.788	**0.8466**	**0.693**	**0.593**	0.366	-0.387	**140.09**	0.05678	1584	4.65%
Model 05	BCC_IO & SE_IO	Managerial	Static	0.280	0.201	0.206	0.794	**0.8440**	**0.688**	**0.591**	**0.570**	**-0.359**	128.45	**0.05666**	**1568**	**5.06%**
Model 06	SBM_CRS_IO	Managerial	Static	0.271	0.198	0.203	0.797	0.8426	0.685	0.574	**0.562**	**-0.377**	120.72	**0.05674**	**1574**	**4.91%**
Model 07	ME-IO	Managerial	Static	0.267	0.208	0.212	0.788	0.8381	0.676	0.578	0.544	-0.401	132.62	0.05683	1581	4.72%
Model 08	BCC_IO & SE_IO & ME_IO	Managerial	Static	**0.263**	0.212	0.215	0.785	**0.8453**	**0.691**	0.589	**0.582**	**-0.378**	126.70	**0.05660**	**1564**	**5.08%**
Model 09	SBM_VRS_IO	Managerial	Static	0.271	0.211	0.215	0.785	0.8341	0.668	0.579	0.524	-0.401	137.45	0.05699	1590	4.50%
Model 10	CCR_OO	Market	Static	0.301	**0.193**	**0.200**	**0.800**	0.8407	0.681	0.580	0.513	-0.382	136.62	0.05715	1599	4.28%
Model 11	BCC_OO	Market	Static	0.335	**0.184**	**0.194**	**0.806**	0.8373	0.675	0.558	0.468	-0.391	134.82	0.05725	1601	4.21%
Model 12	SE_OO	Market	Static	0.271	0.205	0.209	0.791	0.8414	0.683	0.583	0.533	-0.385	135.43	0.05714	1597	4.31%
Model 13	BCC_IO & SE_IO	Market	Static	0.318	0.203	0.210	0.790	0.8326	0.665	0.570	0.452	-0.403	127.54	0.05691	1593	4.43%
Model 14	SBM_CRS_IO	Market	Static	0.309	0.202	0.209	0.791	0.8354	0.671	0.577	0.528	-0.413	127.74	0.05694	1586	4.60%
Model 15	ME_OO	Market	Static	**0.258**	0.204	0.208	0.792	0.8425	0.685	**0.592**	0.560	-0.387	**137.82**	0.05709	1586	4.60%
Model 16	BCC_OO & SE_OO & ME_OO	Market	Static	0.347	**0.183**	**0.194**	**0.806**	0.8334	0.667	0.577	0.464	-0.402	120.17	0.05683	1588	4.56%
Model 17	SBM_VRS_OO	Market	Static	0.292	0.210	0.215	0.785	0.8348	0.670	0.574	0.533	-0.413	128.60	0.05691	1585	4.63%
Model 18	NA	NA	Dynamic	0.288	**0.199**	**0.205**	**0.795**	0.8431	0.686	0.585	0.562	-0.366	155.22	0.05688	1586	4.60%
Model 19	CCR_IO	Managerial	Dynamic	**0.220**	0.220	0.220	0.780	0.8453	0.691	0.578	0.606	-0.370	155.10	0.05679	1581	4.72%
Model 20	BCC_IO	Managerial	Dynamic	0.233	0.221	0.221	0.779	0.8434	0.687	**0.593**	0.551	-0.357	159.50	0.05683	1583	4.67%
Model 21	SE_IO	Managerial	Dynamic	**0.212**	0.216	0.216	0.784	**0.8509**	**0.702**	**0.597**	**0.659**	-0.362	135.95	**0.05652**	**1559**	**5.27%**
Model 22	BCC_IO & SE_IO	Managerial	Dynamic	0.242	**0.205**	**0.208**	**0.792**	**0.8485**	**0.697**	0.589	0.622	**-0.341**	158.36	0.05674	1574	4.91%
Model 23	SBM_CRS_IO	Managerial	Dynamic	0.225	0.215	0.216	0.784	0.8481	0.696	**0.593**	**0.629**	**-0.345**	137.54	0.05668	1572	4.96%
Model 24	ME_IO	Managerial	Dynamic	0.263	0.211	0.215	0.785	0.8433	0.687	0.584	0.546	-0.382	**165.21**	0.05692	1588	4.56%
Model 25	BCC_IO & SE_IO & ME_IO	Managerial	Dynamic	**0.199**	0.228	0.226	0.774	**0.8513**	**0.703**	0.586	**0.650**	**-0.342**	**170.26**	0.05667	**1567**	**5.09%**
Model 26	SMB_VRS_IO	Managerial	Dynamic	0.263	0.211	0.214	0.786	0.8429	0.686	0.584	0.552	-0.356	159.44	0.05687	1586	4.60%
Model 27	CCR_OO	Market	Dynamic	0.267	0.208	0.212	0.788	0.8431	0.686	0.587	0.551	-0.373	159.70	0.05687	1587	4.58%
Model 28	BCC_OO	Market	Dynamic	0.267	0.209	0.213	0.787	0.8432	0.686	0.586	0.551	-0.361	156.81	0.05687	1587	4.58%
Model 29	SE_OO	Market	Dynamic	0.267	0.208	0.212	0.788	0.8431	0.686	0.587	0.551	-0.373	**160.29**	0.05687	1587	4.58%
Model 30	BCC_OO & SE_OO	Market	Dynamic	0.267	0.209	0.212	0.788	0.8363	0.673	0.572	0.570	-0.371	136.24	0.05656	1573	4.92%
Model 31	SBM_CRS_OO	Market	Dynamic	0.229	0.218	0.218	0.782	0.8388	0.678	0.568	0.591	-0.373	133.51	0.05668	1578	4.80%
Model 32	ME_OO	Market	Dynamic	0.301	**0.208**	**0.213**	**0.787**	0.8296	0.659	0.579	0.543	-0.408	124.97	**0.05635**	**1565**	**5.13%**
Model 33	BCC_OO & SE_OO & ME_OO	Market	Dynamic	0.267	0.208	0.212	0.788	0.8431	0.686	0.587	0.553	-0.363	160.29	0.05687	1587	4.58%
Model 34	SBM_VRS_OO	Market	Dynamic	0.254	0.217	0.219	0.781	0.8334	0.667	0.567	0.590	-0.402	131.24	**0.05648**	1574	4.89%

However, the findings suggest that taking to account T2, MR and OCC as measures of correctness of categorical prediction and BS as a measure of calibration accuracy, the models with market efficiency score outperform others.

Fourth, on the type of DEA scores that models are fed with, i.e., decomposed DEA scores and original DEA scores, the following findings are notable. On static models, models 8,5 and 4 that use decomposed managerial DEA scores, i.e., PTE, SE and ME, outperform the models that use original DEA scores, i.e., TE and SBM, considering most of the performance criteria. Also, model 16 that use decomposed market DEA scores, i.e., PTE, SE and ME outperform all models on T2, MR, and OCC. On dynamic models, models with decomposed managerial DEA scores, i.e., models 25, 22 and 21 are superior regarding most of the performance criteria. Further, model 32 with market ME score is the superior performer considering BS, LL, R2, T2, MR, and OCC. This finding suggests that using decomposed efficiency DEA scores improve the performance of prediction models. These results are consistent with Li et al. (Li et al., 2014, 2017) that suggest models with decomposed measures are superior.

5.4.2 Multi-criteria Ranking of Distress Prediction Models

For multi-criteria evaluation of DPMs, I followed Mousavi et al. (2015) in using super efficiency orientation-free SBM-DEA framework. I exercise two rounds of evaluation using four different measures. In the first round, I use T1 error (as the measure under correctness of categorical prediction), BS (as the measure of calibration accuracy) as inputs and ROC (as the measure of discriminatory power) and R^2 (as the measure of information content) as outputs of DEA model. Also, in the second round, I replace T1 with T2 error as the measure of correctness of categorical prediction.

From Table 5.16 the following results of multi-criteria assessment of DPMs are noteworthy. First, comparing the performance of models without efficiency measures with models fed with efficiency measures as predictor, the numerical results indicate that using efficiency measures improve the performance of models.

Second, comparing the performance of dynamic models with static models in my study, taking to account T1 error (Panel A of Table 5.16) as the measure of correctness of categorical prediction, the numerical results show that the dynamic models outperform static ones. However, in respect to T2 error (Panel B of Table 5.16) the results suggest that the static models are compatible with dynamic ones.

Third, comparing the performance of two-stage models with different types of company efficiency measures, i.e., market efficiency and managerial efficiency, under the choice of T1 error as the measure of correctness of categorical prediction (Panel A of Table 5.16), the results suggest that the model 21with managerial efficiency is the best model; though, the model 32 that uses market efficiency is the third in ranking. However, choosing T2 error (Panel B of Table 5.16), models 16, 31 and 11that employ market efficiency are among top five models. This result is consistent with unidimensional ranking of models that suggest models with market efficiency scores outperform others under T2, MR and OCC as measures of correctness of categorical prediction. These results could be linked to the efficient market hypothesis theory that claims in an efficient market the price of shares contains all available information, i.e. past, present and insider, about the company.

Fourth, considering the type of DEA scores that models are fed with, i.e., decomposed DEA scores and original DEA scores, the following findings are notable. In the panel A (respectively, the panel B) of multi-criteria assessment, the dynamic model 25 that uses decomposed dynamic managerial DEA score, i.e., PTE, SE and ME (respectively, the static model 16 that uses decomposed static market DEA score, i.e., PTE, SE and ME) are the best performers. Also, model 32 with market ME score is one of the best performers in both Panel A and Panel B of multi-criteria assessment. These results are consistent with unidimensional assessment and suggest that the models with decomposed measures are superior in performance. In practice, the decomposition of efficiency scores and employing them in the models control more effective variables on failure and therefore would improve the performance of failure prediction models.

Table 5.16: Multi-Criteria Performance Evaluation of Distress Prediction Models

Panel A: Super Efficiency Orientation-free SBM DEA
Input: T1, BS & **Output:** ROC, R^2

Model	DEA Score	Efficiency Score	Framework	Score	Rank
Model 25	BCC_IO & SE_IO & ME_IO	Managerial	Dynamic	1.032	1
Model 21	SE_IO	Managerial	Dynamic	1.018	2
Model 32	ME_OO	Market	Dynamic	1.001	3
Model 23	SBM_CRS_IO	Managerial	Dynamic	0.929	4
Model 34	SBM_VRS_OO	Market	Dynamic	0.920	5
Model 19	CCR_IO	Managerial	Dynamic	0.912	6
Model 31	SBM_CRS_OO	Market	Dynamic	0.901	7
Model 22	BCC_IO & SE_IO	Managerial	Dynamic	0.893	8
Model 20	BCC_IO	Managerial	Dynamic	0.882	9
Model 08	BCC_IO & SE_IO & ME_IO	Managerial	Static	0.878	10
Model 30	BCC_OO & SE_OO	Market	Dynamic	0.857	11
Model 05	BCC_IO & SE_IO	Managerial	Static	0.850	12
Model 06	SBM_CRS_IO	Managerial	Static	0.847	13
Model 04	SE_IO	Managerial	Static	0.842	14
Model 02	CCR_IO	Managerial	Static	0.834	15
Model 15	ME_OO	Market	Static	0.833	16
Model 07	ME-IO	Managerial	Static	0.832	17
Model 26	SMB_VRS_IO	Managerial	Dynamic	0.828	18
Model 24	ME_IO	Managerial	Dynamic	0.824	19
Model 28	BCC_OO	Market	Dynamic	0.822	20
Model 33	BCC_OO & SE_OO & ME_OO	Market	Dynamic	0.821	21
Model 27	CCR_OO	Market	Dynamic	0.820	22
Model 29	SE_OO	Market	Dynamic	0.819	23
Model 09	SBM_VRS_IO	Managerial	Static	0.804	24
Model 18	NA	NA	Dynamic	0.797	25
Model 17	SBM_VRS_OO	Market	Static	0.791	26
Model 12	SE_OO	Market	Static	0.787	27
Model 03	BCC_IO	Managerial	Static	0.779	28
Model 14	SBM_CRS_IO	Market	Static	0.771	29
Model 01	NA	NA	Static	0.763	30
Model 10	CCR_OO	Market	Static	0.751	31
Model 13	BCC_IO & SE_IO	Market	Static	0.747	32
Model 16	BCC_OO & SE_OO & ME_OO	Market	Static	0.734	33
Model 11	BCC_OO	Market	Static	0.712	34

Panel B: Super Efficiency Orientation-free SBM DEA
Input: T2, BS & **Output:** ROC, R^2

Model	DEA Score	Efficiency Score	Framework	Score	Rank
Model 16	BCC_OO & SE_OO & ME_OO	Market	Static	1.023	1
Model 21	SE_IO	Managerial	Dynamic	1.022	2
Model 05	BCC_IO & SE_IO	Managerial	Static	1.010	3
Model 32	ME_OO	Market	Dynamic	1.003	4
Model 11	BCC_OO	Market	Static	1.002	5
Model 22	BCC_IO & SE_IO	Managerial	Dynamic	1.001	6
Model 25	BCC_IO & SE_IO & ME_IO	Managerial	Dynamic	1.000	7
Model 06	SBM_CRS_IO	Managerial	Static	0.994	8
Model 02	CCR_IO	Managerial	Static	0.982	9
Model 08	BCC_IO & SE_IO & ME_IO	Managerial	Static	0.981	10
Model 30	BCC_OO & SE_OO	Market	Dynamic	0.969	11
Model 23	SBM_CRS_IO	Managerial	Dynamic	0.964	12
Model 18	NA	NA	Dynamic	0.960	13
Model 10	CCR_OO	Market	Static	0.957	14
Model 34	SBM_VRS_OO	Market	Dynamic	0.951	15
Model 03	BCC_IO	Managerial	Static	0.946	16
Model 04	SE_IO	Managerial	Static	0.944	17
Model 07	ME-IO	Managerial	Static	0.943	18
Model 14	SBM_CRS_IO	Market	Static	0.942	19
Model 15	ME_OO	Market	Static	0.939	20
Model 01	NA	NA	Static	0.938	21
Model 31	SBM_CRS_OO	Market	Dynamic	0.933	22
Model 33	BCC_OO & SE_OO & ME_OO	Market	Dynamic	0.932	23
Model 27	CCR_OO	Market	Dynamic	0.931	24
Model 29	SE_OO	Market	Dynamic	0.930	25
Model 28	BCC_OO	Market	Dynamic	0.929	26
Model 17	SBM_VRS_OO	Market	Static	0.928	27
Model 26	SMB_VRS_IO	Managerial	Dynamic	0.927	28
Model 19	CCR_IO	Managerial	Dynamic	0.924	29
Model 24	ME_IO	Managerial	Dynamic	0.922	30
Model 13	BCC_IO & SE_IO	Market	Static	0.921	31
Model 20	BCC_IO	Managerial	Dynamic	0.915	32
Model 09	SBM_VRS_IO	Managerial	Static	0.911	33
Model 12	SE_OO	Market	Static	0.905	34

5.5 Conclusion

The application of DEA in credit scoring and distress prediction has extended recently. This study has a comparative analysis between various static and dynamic two-stage distress prediction models to compare the contribution of different efficiency scores in estimating the probability of distress. This study estimates the efficiency of companies regarding market (using market information as input and output) and managerial (using accounting information as input and output) perspective. It uses CCR, BCC and SBM DEA models to estimate cross-sectional efficiency measures and Malmquist-DEA models to estimate dynamic efficiency measures. Also, it decomposes overall static and dynamic SBM efficiency scores into PTE, SE and ME scores, and overall static and dynamic TE efficiency score into PTE and SE scores and integrate them with accounting, market, and macroeconomic ratios to develop distress prediction models.

The results suggest that incorporating managerial efficiency measures have more contribution in predicting distress. The lower contribution of market efficiency measures of firms rather than managerial efficiency is because of choices of inputs and outputs of DEA models. However, market efficiency measures, especially using ME score, are valuable information in predicting distress. Also, the findings indicate that incorporating dynamic scores in a dynamic framework is the best approach to improve the accuracy of distress prediction models. This is because dynamic models by design could take account of changes in the condition of firms over time.

Moreover, the results show that decomposition of TE (respectively, SBM efficiency scores) into PTE and SE (respectively, PTE, SE and ME) improves the performance of prediction models. This is because incorporating decomposed measures of efficiency in the model would control more effective variables on the distress.

The main limitations of this research are time and space constraints and as such this study is restricted to specific DEA models in the first stage. Future studies could incorporate more DEA models to evaluate the managerial and market efficiency of firms. Further, because of the same constraints, this study is focused on Logit and multi-period Logit analysis models in the second stage. Future studies could apply more statistical and non-statistical prediction models.

References

Abdou, H.A., Pointon, J., 2011. Credit scoring, statistical techniques and evaluation criteria: A review of the literature. *Intelligent Systems in Accounting, Finance and Management* 18, 59–88.

Agarwal, V., Taffler, R., 2008. Comparing the performance of market-based and accounting-based bankruptcy prediction models. *Journal of Banking & Finance 32* 1541–1551.

Aida, K., Cooper, W.W., Pastor, J.T., Sueyoshi, T., 1998. Evaluating Water Supply Services in Japan with RAM: a Range-adjusted Measure of Inefficiency. *Omega* 26, 207–232.

Alfaro-Cid, E., Sharman, K., Esparcia-Alcázar, A.I., 2007. A Genetic Programming Approach for Bankruptcy Prediction Using a Highly-Unbalanced Database, in: *Giacobini, M. (Ed.), Applications of Evolutionary Computing, Lecture Notes in Computer Science. Springer Berlin Heidelberg*, pp. 169–178.

Altman, E., 1983. Corporate Financial Distress: A Complete Guide to Predicting, Avoiding and Dealing with Bankruptcy. Edit. John Wiley and Sons.

Altman, E.I., 1973. Predicting Railroad Bankruptcies in America. *The Bell Journal of Economics and Management Science* 4, 184-211.

Altman, E.I., 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The journal of finance* 23, 589–609.

Altman, E.I., Iwanicz-Drozdowska, M., Laitinen, E.K., Suvas, A., 2016. Financial Distress Prediction in an International Context: A Review and Empirical Analysis of Altman's Z-Score Model. *Journals of International Financial, Management & Accounting, March,* 1-40.

Araghi M.K., Makvandi S., 2013. Evaluating the Technique of Data Envelopment Analysis in Predicting Bankruptcy, *Journal of Basic and Applied Scientific Research* 3(3), 519-527.

Avkiran, N.K., Cai, L., 2014. Identifying distress among banks prior to a major crisis using non-oriented super-SBM. *Annals of Operations Research* 217, 31–53.

Aziz, M.A., Dar, H.A., 2006. Predicting corporate bankruptcy: where we stand? *Corporate governance* 6, 18–33.

Back, B., Laitinen, T., Sere, K., van Wezel, M., 1996. Choosing bankruptcy predictors using discriminant analysis, logit analysis, and genetic algorithms. *Turku Centre for Computer Science Technical Report* 40, 1–18.

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

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, 63–93.

Banker, R. D., Charnes, A., & Cooper, W. W. ,1984. Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management science* 30(9), 1078-1092.

Barr, R.S., Seiford, L.M., Siems, T.F., 1993. An envelopment-analysis approach to measuring the managerial efficiency of banks. *Annals of Operations Research* 45, 1–19.

Barr, R. S., Seiford, L. M., & Siems, T. F., 1994. Forecasting bank failure: a non-parametric frontier estimation approach. *Recherches Économiques de Louvain/Louvain Economic Review* 60(04), 417-429.

Barr, R.S., Siems, T.F., 1997. Bank failure prediction using DEA to measure management quality, in: *Interfaces in Computer Science and Operations Research. Springer,* pp. 341–365.

Bauer, J., Agarwal, V., 2014. Are hazard models superior to traditional bankruptcy prediction approaches? A comprehensive test. *Journal of Banking & Finance* 40, 432–442.

Bauweraerts, J., 2016. Predicting Bankruptcy in Private Firms: Towards a Stepwise Regression Procedure. *International Journal of Financial Research* 7, 147.

Beaver, W.H., 1968. Alternative accounting measures as predictors of failure. *Accounting review* 43, 113–122.

Beaver, W.H., 1966. Financial Ratios as Predictors of Failure. *Journal of Accounting Research* 4, 71–111.

Bellovary, J.L., Giacomino, D.E., Akers, M.D., 2007. A review of bankruptcy prediction studies: 1930 to present. *Journal of Financial Education* 33, 1–42.

Bharath, S.T., Shumway, T., 2008. Forecasting default with the Merton distance to default model. *The Review of Financial Studies* 21, 1339–1369.

Bowlin, W. F., 2004. Financial analysis of civil reserve air fleet participants using data envelopment analysis. *European Journal of Operational Research* 154(3), 691-709.

Branch, B., 2002. The costs of bankruptcy: A review. *International Review of Financial Analysis* 11, 39–57.

Caves, D.W., Christensen, L.R., Diewert, W.E., 1982. The Economic Theory of Index Numbers and the Measurement of Input, Output, and Productivity. *Econometrica* 50, 1393–1414.

Charnes, A., Cooper, W.W., Rhodes, E., 1978. Measuring the efficiency of decision making units. *European journal of operational research* 2, 429–444.

Chava, S., Jarrow, R.A., 2004. Bankruptcy prediction with industry effects. *Review of Finance* 8, 537–569.

Chen, N., Ribeiro, B., Chen, A., 2016. Financial credit risk assessment: a recent review. *Artificial Intelligence Review* 45, 1–23.

Cheng, K.F., Chu, C.K., Hwang, R.-C., 2010. Predicting bankruptcy using the discrete-time semiparametric hazard model. *Quantitative Finance* 10, 1055–1066.

Cheng, G., Zervopoulos, P., & Qian, Z., 2013. A variant of radial measure capable of dealing with negative inputs and outputs in data envelopment analysis. *European Journal of Operational Research* 225(1), 100-105.

Cielen, A., Peeters, L., Vanhoof, K., 2004. Bankruptcy prediction using a data envelopment analysis. *European Journal of Operational Research* 154, 526–532.

Cleary S and Hebb G, 2016. An efficient and functional model for predicting bank distress: In and out of sample evidence. *Journal of Banking & Finance* 64 (Mar): 101–111.

Cooper, W.W., Park, K.S., Pastor, J.T., 1999. RAM: A Range Adjusted Measure of Inefficiency for Use with Additive Models, and Relations to Other Models and Measures in DEA. *Journal of Productivity Analysis* 11, 5–42.

Cooper, W.W., Seiford, L.M., Tone, K., 2006. *Data envelopment analysis: A comprehensive text with models, applications, references and DEA-Solver Software*. Second editions. Springer.

Crapp, H.R., Stevenson, M., 1987. Development of a method to assess the relevant variables and the probability of financial distress. *Australian journal of management* 12, 221–236.

Davydenko, S.A., Strebulaev, I.A., Zhao, X., 2012. A market-based study of the cost of default. *Review of Financial Studies* 25, 2955–2999.

du Jardin, P., Séverin, E., 2012. Forecasting financial failure using a Kohonen map: A comparative study to improve model stability over time. *European Journal of Operational Research* 221, 378–396.

du Jardin P., 2015. Bankruptcy prediction using terminal failure processes. *European Journal of Operational Research* 242(1): 286–303.

Duda, M., Schmidt, H., Asgharian, H., 2010. Bankruptcy Prediction: Static Logit Model versus Discrete Hazard Models Incorporating Macroeconomic Dependencies.

El Kalak, I., Hudson, R., 2016. The effect of size on the failure probabilities of SMEs: An empirical study on the US market using discrete hazard model. *International Review of Financial Analysis* 43, 135–145.

Elkamhi, R., Ericsson, J., Parsons, C.A., 2012. The cost and timing of financial distress. *Journal of Financial Economics* 105, 62–81.

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

Emrouznejad, A., Anouze, A. L., & 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(1), 297-304.

Emrouznejad, A., & Yang, G. L. (2017). A survey and analysis of the first 40 years of scholarly literature in DEA: 1978–2016. *Socio-Economic Planning Sciences*.

Etemadi, H., Anvary Rostamy, A.A., Dehkordi, H.F., 2009. A genetic programming model for bankruptcy prediction: Empirical evidence from Iran. *Expert Systems with Applications* 36, 3199–3207.

Färe, R., Grosskopf, S., Lindgren, B., Roos, P., 1992. Productivity Changes in Swedish Pharamacies 1980–1989: A Non-Parametric Malmquist Approach, in: Jr, T.R.G., Lovell, C.A.K. (Eds.), *International Applications of Productivity and Efficiency Analysis. Springer Netherlands*, pp. 81–97.

Färe, R., Grosskopf, S., Norris, M., Zhang, Z., 1994. Productivity Growth, Technical Progress, and Efficiency Change in Industrialized Countries. *The American Economic Review* 84, 66–83.

Fethi, M.D., Pasiouras, F., 2010. Assessing bank efficiency and performance with operational research and artificial intelligence techniques: A survey. *European Journal of Operational Research* 204, 189–198.

Frydman, H., Altman, E.I., Kao, D., 1985. Introducing recursive partitioning for financial classification: the case of financial distress. *The Journal of Finance* 40, 269–291.

Gestel, T.V., Baesens, B., Suykens, J.A.K., Van den Poel, D., Baestaens, D.-E., Willekens, M., 2006. Bayesian kernel based classification for financial distress detection. *European Journal of Operational Research 172*, 979–1003.

Hillegeist, S.A., Keating, E.K., Cram, D.P., Lundstedt, K.G., 2004. Assessing the probability of bankruptcy. *Review of Accounting Studies* 9, 5–34.

Huang, Z., Chen, H., Hsu, C.-J., Chen, W.-H., Wu, S., 2004. Credit rating analysis with support vector machines and neural networks: a market comparative study. *Decision Support Systems*, Data mining for financial decision making 37, 543–558.

Kim, M.-J., Kang, D.-K., 2010. Ensemble with neural networks for bankruptcy prediction. *Expert Systems with Applications* 37, 3373–3379.

Kumar, P. R., & 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.

Laitinen, E.K., Suvas, A., 2016. The effect of national culture on financial distress prediction modelling: evidence from European countries. *International Journal of Accounting and Finance* 6, 299–318.

Lane, W.R., Looney, S.W., Wansley, J.W., 1986. An application of the Cox proportional hazards model to bank failure. *Journal of Banking & Finance* 10, 511–531.

Li, H., Sun, J., 2011. Predicting business failure using forward ranking-order case-based reasoning. *Expert Systems with Applications* 38, 3075–3084.

Li, H., Sun, J., 2009. Predicting business failure using multiple case-based reasoning combined with support vector machine. *Expert Systems with Applications* 36, 10085–10096.

Li, Z., Crook, J., Andreeva, G., 2017. Dynamic prediction of financial distress using Malmquist DEA. *Expert Systems with Applications* 80, 94–106.

Li, Z., Crook, J., Andreeva, G., 2014. Chinese companies distress prediction: an application of data envelopment analysis. *Journal of Operational Research Society* 65, 466–479.

Liang, D., Lu, C.-C., Tsai, C.-F., Shih, G.-A., 2016. Financial ratios and corporate governance indicators in bankruptcy prediction: A comprehensive study. *European Journal of Operational Research* 252, 561–572.

Luoma, M., Laitinen, E.K., 1991. Survival analysis as a tool for company failure prediction. *Omega* 19, 673–678.

Lyandres, E., & Zhdanov, A., 2013. Investment opportunities and bankruptcy prediction. *Journal of Financial Markets* 16(3), 439-476.

Maddala, G.S., 1986. *Limited-dependent and qualitative variables in econometrics*. Cambridge university press.

Malmquist, S., 1953. Index numbers and indifference surfaces. *Trabajos de Estadistica* 4, 209–242.

Martin, D., 1977. Early warning of bank failure: A logit regression approach. *Journal of Banking & Finance* 1, 249–276.

McKee, T.E., Lensberg, T., 2002. Genetic programming and rough sets: A hybrid approach to bankruptcy classification. *European Journal of Operational Research* 138, 436–451.

Meyer, P.A., Pifer, H.W., 1970. Prediction of bank failures. *The Journal of Finance* 25, 853–868.

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

Mukhopadhyay, A., Tiwari, S., Karmaker, A.N.B.R., 2012. A New Approach to Predicting Bankruptcy: Combining DEA and Multi-Layer Perceptron Layer Perceptron Layer Perceptron 9, 71–78.

Nam, C.W., Kim, T.S., Park, N.J., Lee, H.K., 2008. Bankruptcy prediction using a discrete-time duration model incorporating temporal and macroeconomic dependencies. *Journal of Forecasting* 27, 493–506.

Neves, J.C., Vieira, A., 2006. Improving bankruptcy prediction with Hidden Layer Learning Vector Quantization. *European Accounting Review* 15, 253–271.

Ohlson, J.A., 1980. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of accounting research* 18, 109–131.

Oral, M., & Yolalan, R., 1990. An empirical study on measuring operating efficiency and profitability of bank branches. *European Journal of Operational Research*, 46(3), 282-294.

Ouenniche, J., Tone, K., 2017. An out-of-sample evaluation framework for DEA with application in bankruptcy prediction. *Annals of Operations Research* 1–16.

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

Paradi, J.C., Yang, X., others, 2014. Data Envelopment Analysis of Corporate Failure for Non-Manufacturing Firms Using a Slacks-Based Measure. *Journal of Service Science and Management* 7, 277.

Paradi, J.C., Zhu, H., 2013. A survey on bank branch efficiency and performance research with data envelopment analysis. *Omega* 41, 61–79.

Pastor, J. T., & Lovell, C. K., 2005. A global Malmquist productivity index. *Economics Letters*, 88(2), 266-271.

Piesse, J., Lee, C.-F., Kuo, H.-C., Lin, L., 2006. Corporate failure: Definitions, methods, and failure prediction models, in: *Encyclopedia of Finance*. Springer, pp. 477–490.

Pille, P., & Paradi, J. C., 2002. Financial performance analysis of Ontario (Canada) Credit Unions: An application of DEA in the regulatory environment. *European Journal of Operational Research* 139(2), 339-350.

Pindado, J., Rodrigues, L., & de la Torre, C., 2008. Estimating financial distress likelihood. *Journal of Business Research* 61(9), 995-1003.

Portela, M. S., Thanassoulis, E., & Simpson, G., 2004. Negative data in DEA: A directional distance approach applied to bank branches. *Journal of the Operational Research Society* 55(10), 1111-1121.

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, 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, 620–626.

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

Seballos, L.D., Thomson, J.B., others, 1990. Underlying causes of commercial bank failures in the 1980s. *Economic Commentary* (Sep).

Sharp, J. A., Meng, W., & 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(12), 1672-1677.

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, 1988–1997.

Shin, K.-S., Lee, T.S., Kim, H., 2005. An application of support vector machines in bankruptcy prediction model. *Expert Systems with Applications* 28, 127–135.

Shin, K.-S., Lee, Y.-J., 2002. A genetic algorithm application in bankruptcy prediction modelling. *Expert Systems with Applications* 23, 321–328.

Shumway, T., 2001. Forecasting bankruptcy more accurately: A simple hazard model. *The Journal of Business* 74, 101–124.

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(2), 324-351.

Sueyoshi, T., 2004. Mixed integer programming approach of extended DEA–discriminant analysis. *European Journal of Operational Research* 152(1), 45-55.

Sueyoshi, T., 2006. DEA-discriminant analysis: methodological comparison among eight discriminant analysis approaches. *European Journal of Operational Research* 169, 247–272.

Sueyoshi, T., Goto, M., 2009. DEA–DA for bankruptcy-based performance assessment: Misclassification analysis of Japanese construction industry. *European Journal of Operational Research* 199, 576–594.

Sueyoshi, T., Goto, M., 2009. Can R&D expenditure avoid corporate bankruptcy? Comparison between Japanese machinery and electric equipment industries using DEA–discriminant analysis. *European Journal of Operational Research* 196(1), 289-311.

Sueyoshi, T., Goto, M., Omi, Y., 2010. Corporate governance and firm performance: Evidence from Japanese manufacturing industries after the lost decade. *European Journal of Operational Research* 203, 724–736.

Tinoco, M.H., Wilson, N., 2013. Financial distress and bankruptcy prediction among listed companies using accounting, market and macroeconomic variables. *International Review of Financial Analysis* 30, 394–419.

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

Tone, K., 2001. A slacks-based measure of efficiency in data envelopment analysis. *European journal of operational research* 130, 498–509.

Trujillo-Ponce, A., Samaniego-Medina, R., Cardone-Riportella, C., 2014. Examining what best explains corporate credit risk: accounting-based versus market-based models. *Journal of Business Economics and Management* 15, 253–276.

Tsai, M.-C., Lin, S.-P., Cheng, C.-C., Lin, Y.-P., 2009. The consumer loan default predicting model – An application of DEA–DA and neural network. *Expert Systems with Applications* 36, 11682–11690.

Wu, Y., Gaunt, C., Gray, S., 2010. A comparison of alternative bankruptcy prediction models. *Journal of Contemporary Accounting & Economics* 6, 34–45.

Xu, X., Wang, Y., 2009. Financial failure prediction using efficiency as a predictor. *Expert Systems with Applications* 36, 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, 1535–1541.

Zhou, L., Lai, K. K., & Yen, J., 2012. Empirical models based on features ranking techniques for corporate financial distress prediction. *Computers & Mathematics with Applications* 64(8), 2484-2496.

Zhou, L., 2013. Performance of corporate bankruptcy prediction models on imbalanced dataset: The effect of sampling methods. *Knowledge-Based Systems* 41, 16–25.

Zhou, L., 2014, A comparison of dynamic hazard models and static models for predicting the special treatment of stocks in China with comprehensive variables. *Journal of the Operational Research Society* 66(7), 1077-1090.

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

Zopounidis, C., Doumpos, M., 2002. Multi-criteria decision aid in financial decision making: methodologies and literature review. *Journal of Multi-Criteria Decision Analysis* 11, 167–186.

Chapter Six

Conclusion

6.1 Summary of Findings and Conclusion

The performance evaluation of competing failure prediction models is the common concern of both academics and practitioners in the field of corporate credit risk. The conventional performance evaluation exercise of failure prediction models has been an exercise that is unidimensional in nature, in one hand, and static, on the other hand. Consequently, conflicting rankings from one performance criteria to another are frequently stated in other comparative studies. Also, a single or a very restricted number of criteria are only applied, and therefore the "big picture" is not considered.

The first project (chapter two) contributes to the methodology by proposing an orientation-free super-efficiency DEA model to overcome this methodological issue. Also, the study performed a comprehensive comparative analysis of the most popular six bankruptcy modelling frameworks organised into four categories; namely, original models, original models refitted, reworking models in a logit framework with the same original explanatory variables, and new models. The research used four performance criteria, which are used in the literature; namely, the discriminatory power, the calibration accuracy, the information content, and the correctness of categorical prediction. The study has taken to account several measures for each criterion to find out about the robustness of multidimensional rankings on different combinations of measures. The empirical findings suggested that first, the multidimensional framework provides a valuable tool to delivers a single ranking based on multiple performance criteria. Second, in contrary to the unidimensional rankings, the multidimensional rankings of the best and the worst models are not too sensitive to the changes in most combinations of performance measures. Third, empirical results suggest that the survival analysis model tends to 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. It is worth mentioning that also in theory the dynamic models outperform the static ones because they are able to take into account the time-varying features of firms. 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. These empirical results support the theory of efficient market hypothesis, which suggest that in an efficient market a firm's stock price carries all available information about a firm, i.e. past, current and insider information.

Though, within the super-efficiency DEA framework, the reference benchmark changes from one prediction model evaluation to another one, which in some contexts might be viewed as "unfair" benchmarking. Therefore, the second project (chapter three) overcomes this issue by using a slacks-based context-dependent DEA framework to assess the performance of competing distress prediction models. The numerical results suggest that first, the rankings of DPMs under orientation-free SBM-super efficiency and orientation-free SBM-CDEA are very similar, the latter one, however, does not suffer from the changes of reference benchmark from one prediction models to another. Second, the results reveal that amongst the dynamic models, which are always superior in performance, DD_VEX and DD_1/ln(age) that use the volatility of exchange rate (VEX) and 1/ln(age) as a time-varying baseline, respectively, followed by DIWOB tend to be superior. These results suggest that incorporating macroeconomic indicators and firm's characteristics as the proxy of baseline rate in developing dynamic prediction models improved the performance of models. Third, empirical results suggest that amongst the static models, LPA and PA models are superior to others. In theory, under Logit analysis framework the restricted underlying assumptions of Discriminant analysis are relaxed, which has led to better results in many empirical studies. Finally, developing new models using the most recent accounting, market and macroeconomic information improves the performance of DPMs. In fact, economic cycle lead to changes in the trend of financial ratios overtime. Therefore, developing new models using the most recent information would improve the performance of models.

The third project (chapter four) contributed to the literature in several ways. First this chapter proposed a multi-period performance evaluation framework based on an orientation-free super-efficiency Malmquist DEA index that provides a single ranking based on multiple performance criteria over time. Second, this study performed an exhaustive comparative analysis of the most cited static and dynamic distress prediction models. Same as last two projects, this study used several measures under four commonly applied performance criteria (i.e., the discriminatory power, the information content, the calibration accuracy, and the correctness of categorical prediction) in the literature. Third, this project considered the effect of information, sample type and sample period length on the performance of static and dynamic DPMs during the years with higher distress rate (HDR). The findings suggest that the proposed multi-criteria dynamic framework is a useful tool in evaluating the relative performance of DPMs over time and provides more consistent results. Also, the empirical findings support the last two projects' results that most static and dynamic models perform better when fed with market information. Also, dynamic models, specifically DDWTDB_1/ln(age) and DDWTDB_ln(age) are always amongst the best DPMs under different combinations of measures.

The last project (chapter five) focused on the recent trend in the application of DEA in credit scoring and distress prediction. This project contributed to the literature by providing a comparative analysis between static and dynamic two-stage DPMs, and analysing the discriminatory power of different DEA efficiency scores as a predictor in DPMs. For this, it estimated the efficiency of companies on market (using market information as input and output) and managerial (using accounting information as input and output) perspective. It used CCR, BCC and SBM DEA models to estimate cross-sectional efficiency measures and Malmquist-DEA models to estimate dynamic efficiency measures. Also, it proposed to decompose overall static, and dynamic SBM scores into PTE, SE and ME scores, and overall static and dynamic TE score into PTE and SE scores and integrate them with accounting, market, and macroeconomic ratios to develop distress prediction models. The numerical results suggest that managerial efficiency measures have more discriminatory power in predicting distress. However, market efficiency measures provide valuable information in predicting distress, especially using ME score as a predictor in models. These findings verify the significant relationship between the firm's managerial efficiency and the probability of distress. Further, the results suggest that market efficiency of firms, which is estimated using market information of firms as inputs and outputs of DEA models, is a valuable feature in developing distress prediction models.

Also, the empirical results indicate that incorporating dynamic scores in a dynamic framework is the best approach to improve the accuracy of DPMs. This is because the dynamic models are capable to incorporate time-varying features of a firm. Further, the findings indicate that decomposition of TE (respectively, SBM efficiency scores) into PTE and SE (respectively, PTE, SE and ME) enhances the prediction accuracy of models. This is because through decomposition of scores, more features of firm's efficiency scores are taken into account.

6.2 Research Limitations

Same as other research this study has faced several limitations. Time has been one of the main limitations of this study. Reviewing the literature, collecting and organising dataset, understanding different types of models, programming in different statistical packages for estimating models, performance efficiency scores and efficiency measures, analysing and writing up results have been time-consuming processes. Therefore, I restricted this study to statistical failure prediction models. Further, this research is limited to the listed UK companies in LSE since I had no access to the information of non-listed UK companies.

6.3 Future Research

The focus of this study has been on statistical failure prediction models. Future studies could apply the proposed multi-criteria performance evaluation frameworks to assess the performance of non-parametric bankruptcy and distress prediction models such as artificial intelligence, neural network, and operations research. Also, this study used the sample of listed companies in London Stock Exchange (LSE). The future studies could apply the proposed failure prediction models and evaluation frameworks to other countries. The last but not the least, the two-stage distress prediction models could be extended to other DEA models in the first stage and other classifiers in the second stage. Also, other combinations of inputs and outputs could be used in DEA models.

The future studies could compare the performance of models with different combinations of inputs and outputs for DEA models.