LUND UNIVERSITY
School of Economics and Management

Comparative study of bankruptcy prediction model

Master Thesis in Finance

May 2018

Authors：

Siqi Chen

Shengxiang Xu

Supervisor：

Jens Forssbaeck

Master's Programme in Finance

Abstract

Corporate bankruptcy prediction is usually conducted through empirical research. By using econometric methods to derive a standard model, the established model is used to predict the probability of bankruptcies. The probability of falling into a financial distress can be estimated, so that the company can take effective measures to improve the operation of the company and prevent the occurrence of the crisis. The bankruptcy forecast can also help investors to identify company's condition and avoid loss of property.

Many bankruptcy prediction models have been proposed in history. According to information based in the model, they can be divided into three types: accounting-based model, market-based models and mixed model. Research shows that the mixed model is the best among three models. This paper evaluates the predictive ability of two recently proposed mixed bankruptcy prediction models, Chava & Jarrow (2004) and Campbell, Hilscher, and Szilagyi (2008) respectively. Through using the data of bankrupt companies and non-bankrupt companies in North America listed on NYSE, AMEX and NASDAQ from 1993 to 2013, this study implements logit regression to estimate the parameters, and evaluate the performance of the two models by ROC curve and CAP curve.

The result shows, the model from Campbell et al is better. The reason behind might be the enhanced explanatory power of extended three variables in Campbell et al's model MB, CASHMITA and PRICE and more market information are included in Campbell et al's model.

Keywords: Bankruptcy prediction model, Chava & Jarrow (2004) model, Campbell et al (2008) model, Logit regression, ROC curve and CAP curve.

Acknowledgments

We would like to thank our fantastic supervisor Jens Forssbaeck for great guidance and advice during the process of writing this thesis. Thank you, Jens, your help was invaluable.

Table of Content

List of Abbreviations

Abbreviations	Meaning
UDA	Univariate Discriminant Approach
MDA	Multivariate Discriminant Approach
BSM	Black-Scholes-Merton option-pricing theory
PD	default probability
ROC	Receiver operating characteristic
AUC	Area under the Curve
CAP	Cumulative accuracy profile
AR	Accuracy ratio
NYSE	New York Stock Exchange
AMEX	American Stock Exchange
NASDAQ	National Association of Securities Dealers Automated Quotation
ETF	Exchange Trade Funds
SIC	Standard Industry Classification
S&P 500	Standard & Poor's 500
LR	Likelihood Ratio
GFC	Global financial crisis

List of Regression

List of Equations

List of Charts

List of Tables

1 Introduction

1.1 Background

The field of corporate bankruptcy prediction is always topical. It has been drawing a great attention ever since the collapse of many large firms around the world because of the effects of bankruptcy. From a practical perspective, decision-makers such as creditors and investors especially shareholders who are not in the prior liquidation sequence are largely influenced by bankruptcy. In addition, bankruptcy also results in serious social problems. Mass unemployment and financial recession are likely to be triggered by too many corporate bankruptcies. By most accounts, in 2008, Lehman Brothers' filing a bankruptcy made financial recession of the United States reach a fever pitch, ripples of which profoundly influenced the world economy. Specifically, the number of unemployed people reached to 590000 in the United States in January 2009, setting a record high in the last 17 years. In this case, investigating the probabilities of firms' bankruptcies is crucial for academics, market participants and regulators.

A bankruptcy prediction model is preliminary thought on evaluating a firm's performance. It not only helps managers to find out the inner or outer reasons why the firm faces a high probability of bankruptcy so as to make some relative adjustments, but also alarms investors firms' financial positions to avoid property loss.

In all simplicity, models used to forecast bankruptcy are roughly classified into three categories: accounting-based models, market-based models and mixed models. Scoring models illustrated below extract their data from the market. Dating back to the middle term of last century, financial ratios were in the widespread used as predictors of failure. According to Beaver (1966), who found that some specific ratios such as Cash flow/Total asset, Net income/ Total asset and Total debt /total assets in a single variable discriminant model were good predictors of bankruptcy. Altman (1964) had put a “Z-score” model. Based on the financial report, it uses profitability, leverage, liquidity, solvency and activity to predict the likelihood of bankruptcy. Ohlson (1980) used a large sample of balance sheet ratios as bankruptcy indicators to implement logit regression. O-score was generated. The market-based models of bankruptcy prediction were originally developed by Merton (1974). The underlying concept of Merton model is that the firm would default if it is unable to service its financial liabilities. The equity of a levered firm could be a call option. The firm would be easy to be insolvent if its liabilities is more than asset value. In this manner, the Merton model incorporates market information rather than accounting information. A more advanced model considering both accounting information and market information was raised up firstly by Shumway (2001). Shumway (2001) proposed a time discrete hazard model that explicitly account for time. Structural models mentioned above ignore that the explanatory variables change over time, which would generate biased approximation. It is a breakthrough in the field of bankruptcy prediction model. Based on Shumway’s model, Chava & Jarrow (2004) further validated the superiority of Shumway’s model over scoring models. Moreover, industry effects have been taken into consideration. Campbell et al (2008) expanded explanatory variables in the basis of Chava & Jarrow’s model, aiming to comprehensively explore the determinants of corporate failure. A mixed model combining market information with accounting information is regarded to have a better performance in forecasting the probability of bankruptcy.

1.2 Purpose

Bankruptcy prediction models are gradually bringing forth the new through the old and a better performed model has higher capacity to predict the probability of bankruptcy. This paper mainly conducts comparative analysis between different prediction models.

Two models have been selected from the type of mixed models, Chava & Jarrow's model, Campbell et al's model respectively. Several concerns involved in implementing these two models to do comparative analysis. The quality of accounting-based prediction models are questioned for excluding the relatively important source of information-market information (Hillegeist, Keating, Gram and Lundstedt, 2002). Accounting data are historically acquired from company's financial statement, somehow, they are inconsistent with newly updated information of company. In this manner, accounting based models may be deficient in predicting bankruptcy. Although market-based models are tested to be more accurate in predicting bankruptcy than accounting-based models, they still have some drawbacks such as the limitations of model's assumptions and the need to back out asset value and volatility (Hillegeist et al, 2004). Chava & Jarrow's model and Campbell et al's model, denoted as mixed models, take both accounting information and market information into consideration. These two models are frequently used and moderate on predicting bankruptcy compared with scoring models. Moreover, the basic logistic regression used in two models are analogical, except Campbell et al's model having three more expanding variables.

Overall, model of Chava & Jarrow and Campbell et al are relatively accurate models in terms of research actuality. By proceeding logit regression and accuracy test, a better performed model can be evaluated.

1.3 Delimitations

This paper is subject to limitations such as the impact of financial crisis. The time period of our data is from 1993-2013. The data collected after 2007 went through a time span where the world's economy was affected by financial recession. Unusual change in systematic environment could generate deviation from the expected empirical results. In addition, with a lack of bankrupt data, it is capable but ineffective to test the industry effects on bankruptcy prediction, which is contained in Chava & Jarrow's model.

1.4 Outline of the thesis

This paper is outlined as follows: Section 2 mainly concentrates on literature review where the definition of bankruptcy and the bankruptcy prediction model classified by three categories are stated. In section 3, it is research method used in this paper including logistic regression and assessing techniques (ROC and CAP). Section 4 mainly concentrates on implementing a logit regression, such as sample collection and variable selection. Section 5 elaborates the result of empirical study and section 6 conducts comparative analysis for two models to evaluate their performance. Section 7 is the conclusion.

2 Literature review

2.1 Definition of bankruptcy

Enterprise bankruptcy refers to the situation where its liabilities exceed assets owned by company in the process of production and operation, and the due debt cannot be paid. There are mainly three types of bankruptcy according to federal laws, chapter 7, chapter 11 and chapter 13 respectively.

According to U.S. law, bankruptcy has the following definitions (Zingarelli Law Office, 2018): In normal case, chapter 7 involves a condition of liquidation. a bankruptcy trustee collects the debtor's non-exempt property and converts it into cash for the benefit of unsecured creditors. Chapter 11 and chapter 13 concern with a case of rehabilitation or reorganization. Chapter 11 usually applies to individual debtors with excessive or complex debts, or to large commercial entities like corporations. Chapter 13 usually applies to individual consumers and to some small businesses with debts that fit within the boundaries of Chapter 13.

2.2 Literature on prediction models

2.2.1 Accounting based models

Bankruptcy prediction models based on accounting information basically compare one or more ratios derived from financial statements, such as liquidity, and repayment ratios with benchmarks. Three relevant approaches would be discussed as following.

2.2.1.1 Univariate Discriminant Approach (UDA)

William Beaver (1966) proposed the Univariate Discriminant Approach. According to William Beaver, a company's failure is defined as bankruptcy, bond default, an overdrawn bank account, and nonpayment of a preferred stock dividend. Besides, the financial ratio has potential predictive ability of financial statement, so it was an effective approach to predict the failure of the company. To testify his argument, 79 failure companies and healthy companies with same amount are randomly selected, inside which a set of financial ratios are chosen. It turns out that some ratios of bankrupt companies become worse in the last five financial years before it filed a bankruptcy and are pretty bad right in the fifth year. Thus, he thinks those worsening financial ratios could be treated as early bankrupt warning signals. Besides, ratios such as cash flow to total asset and net income to total asset perform better than the other. Although William Beaver’s test proves that financial ratios might be an important component in the prediction of failure, not all financial ratios are testified to be capable of predicting the failure and results might vary from different ratios. William Beaver still thinks more financial ratios can be added to predict failure.

2.2.1.2 Multivariate Discriminant Approach （MDA）

The Univariate Discriminant Approach obviously cannot proceed an accurate prediction, American financial expert Edward Altman (1968) presents the Multivariate Discriminant Approach, the Z-score model equally. He chooses 66 manufacturing enterprises as research sample, which consists of 33 bankrupt companies and 33 non-bankrupt companies. 22 ratios are chosen as variables in terms of potential relevance. Those ratios are diversified into five types: liquidity, profitability, lever, solvency and activity. On the basis of ratios ’contribution and relation, five variables are determined in the last listed below. Altman model was a kind of multivariate linear decision equation. Differing from UDA, this model revealed that when a company is nearly collapsed, multiple ratios will change at the same time due to the bankruptcy is caused by multiple reasons. Moreover, it solves the problem that a univariate ratio can introduce different results in different companies. Edward Altman further established the following equations through the MDA.

$$\mathrm{Z} = 1.2\mathrm{X}_1 + 1.4\mathrm{X}_2 + 3.3\mathrm{X}_3 + 0.6\mathrm{X}_4 + 1.0\mathrm{X}_5$$
$$\mathrm{X}_1 = \text{working capital/total asset}$$
$$\mathrm{X}_2 = \text{retained earing/total assets}$$
$$\mathrm{X}_3 = \text{earning before interest and taxes/total assest}$$
$$\mathrm{X}_4 = \text{maket value equity/book value of total debt}$$
$$\mathrm{X}_5 = \text{sales/total assets}$$

Z score is the criterion.

Z > 2.99 - "Safe" Zone, the company in this area can be clearly defined as a non-bankrupt company.

1.81 < Z < 2.99 - "Gray Zone", the company in this area cannot be clearly defined as bankruptcy or not bankruptcy, when the value of Z is equal to 2.675, the bankrupt and non-bankrupt probability is 50%, 50% respectively. Thus, Altman Z = 2675 is the best dividing line to distinguish whether a company is going to file a bankruptcy or not.

Z < 1.81 - "Distress" Zone, the company in this area can be clearly defined as a bankrupt company.

Altman then tested the model with data spanning from 1969 to 1999 several times, and the accuracy rate was between 80% and 90%. In addition, the Z-score model was implemented for non-production enterprise, for which the variable Sale could not be included. Thus, the equation kept other four variables as Z-score model for manufacturing enterprises.

$$\mathrm{Z} = 6.65X_1 + 3.26X_2 + 6.72X_3 + 1.05X_4$$

Edward Altman et al (1997) put up with ZETA model, an extension of Z-score model. ZETA model had seven variables, which reflected different characteristics of enterprises. These characteristics were: Return on asset, Stability of earing, Debt service, Cumulative profitability, Liquidity, Capitalization, and Size. ZETA model is more widely applied in practice and is better at predicting bankruptcy than the original Z-score model. The accuracy rate of ZETA model could be over 90% (Edward Altman, 2000).

2.2.1.3 Ohlson's O-score

Ohlson (1980) raised a series of questions on Z-score model, of which some underlying assumptions might be required, such as: variables in Z-score model should be normal distribution and both the bankrupt and the non-bankrupt group should have the same variance – covariance. Nevertheless, those assumptions are difficult to be satisfied in practice. He thinks Z score does not have a specific meaning and cannot predict the probability of bankruptcy. Ohlson then put up with a new model: O-score, which had a similar form with z-score and is based on the key proportion of accounting information. Differing from Z-score, O-score firstly puts logit regression to use. 105 bankrupt companies and 2058 non-bankrupt companies from 1970 to 1976 are chosen. 9 independent variables are included. The model was formulated as follows:

$$\begin{aligned} \mathrm{O-score} = {} & -1.32 - 0.407\mathrm{SIZE} + 6.03\mathrm{TLTA} - 1.43\mathrm{WCTA} + 0.0757\mathrm{CLCA} \\ & - 1.72\mathrm{OENEG} - 2.37\mathrm{NITA} - 1.83\mathrm{FUTL} + 0.285\mathrm{INTWO} \\ & - 0.521\mathrm{CHIN} \end{aligned}$$

Probability of bankruptcy is sited as $\mathrm{P(B)}$:

$$\mathrm{P(B)} = \frac{1}{1 + \exp(-(\mathrm{O - score}))}$$

1. ***SIZE*** = log (total assets/GNP price-level index).
2. ***TLTA*** = Total liabilities divided by total assets.
3. ***WCTA*** = Working capital divided by total assets.
4. ***CLCA*** = Current liabilities divided by current assets.
5. ***OENEG*** = One if total liabilities exceed total assets, zero otherwise.

6. ***NITA*** = Net income divided by total assets.
7. ***FUTL*** = Funds provided by operations divided by total liabilities.
8. ***INTWO*** = One if net income was negative for the last two years, zero otherwise.
9. ***CHIN*** = (NIt – NIt-1) / (|NIt| + |NIt-1|), where NI is net income for the most recent period. The denominator acts as a level indicator. The variable is thus intended to measure change in net income.

O-score is more powerful than Z-score at predicting bankruptcy. The accuracy rate of using the critical value of P(B)=0.5 to predict whether a company will file a bankruptcy is up to 96.12%. Ohlson still thinks that applying information of bankrupt companies will give a rise to the accuracy rate.

2.2.2 Market based models

Although Z-score and O-score model are still widely used in empirical study, many researchers have found market-based models surpass these accounting-based models in the accuracy of bankruptcy forecasting. Market information such as capital market data and asset price volatility are crucial for assessing companies ’operating conditions, while those are excluded in accounting-based models. The Black-Scholes-Merton framework would mainly be discussed below, as the representative of market-based models.

2.2.2.1 Black-Scholes-Merton option-pricing theory（BSM）

Based on the Black Scholes option pricing model, Merton (1974) proposed the first market-based models for bankruptcy prediction. The purpose behind this model is to measure the relative bankrupt risks of a leveraged firm. Merton’s model is predominantly applied to stock option pricing, whereas some clues of bankruptcy prediction could still be tracked. Merton demonstrates that a company’s asset could be regarded as European call option, whose exercise value is equal to the face value of debt. Consequently, shareholders are not obligated to pay off the company's liabilities. That is to say, shareholders are insulated from unlimited debt. If the market value of asset of asset is greater than the book value of debt at maturity, shareholders would exercise the right and vice versa. Merton used the Black Scholes model to calculate stock values

$$E = AN(d_1) - Ke^{-rT}N(d_2)$$

$$d_1 = \frac{Ln(A/K) + (r + \sigma_A^2/2) \times T}{\sigma_A \times \sqrt{T}}$$

$$d_2 = d_1 - \sigma_A \times \sqrt{T}$$

E: equity value, *A*: assets value, *K*: face value of debt, *r*: risk-free rate, *T*: time horizon and σ_A: assets volatility.
The formula of default probability (PD) is:

$$\mathrm{DD} = \frac{Ln(A/K) + (r - \sigma_A^2/2) \times T}{\sigma_A \times \sqrt{T}}$$

$$\mathrm{PD} = \mathrm{N}(-\mathrm{DD})$$

However, Merton's model can only be used under a series of restrictive assumptions, one of which is that the bond should be considered as a zero-coupon bond whereas bond has interest in practice. Moreover, volatility of assets and risk-free return are supposed to be constant under assumptions. Consequently, Merton's model tends to be more like a theoretical model, but in the basis of Merton's model, there is lot of empirical applications. Bharath & Shumway (2008) compared the Merton model and its naive model. Result showed both the original model and the naive model had well-predicting ability, but naive model was better than original model. Afik et al (2016) did close testing that compared Merton model, Down-and–Out Option pricing model and a simple naive model, and their conclusion showed original model had well-predicting ability, though Down-and–Out Option pricing model and a simple naive model were better.

2.2.3 Mixed models

With the in-depth study of bankruptcy prediction models, more researchers have found that under certain conditions, neither accounting based models nor market-based models are capable of forecasting bankruptcy accurately. Hypothesis of accounting information added into market-based models as a supplement was raised, therefore, a model mixed with both accounting information and market information was conceived to have better performance than the other in predicting bankruptcy.

2.2.3.1 Discrete time hazard model

Shumway (2001) called Multivariate Discriminant Approach a static model which he believed was not suitable for bankruptcy prediction. The characteristics of healthy company often changes at different time points, while previous studies tended to choose observing data in a particular time point such as one year before filing a bankruptcy. Thus, an unnecessary selection bias could be created by using a single time point data. In his empirical study, he chose 3182 companies (excluding financial institution) from 1962 to 1992, 300 bankrupt companies included. He then implemented a logit estimation program which could be used to calculate maximum likelihood to establish the discrete time hazard model. Compared with static model, the discrete time hazard model could better reflect the characteristics of the company's dynamic change in different time period rather than a year before bankruptcy.

Due to the time factor was considered, more data could be collected within the same amount of company samples, so the estimation would be more accurate. Moreover, Shumway hold the view that accounting variables used in previous study were limited to explain the bankruptcy, and part of variables were low correlated with bankruptcy probability. He then added three other market variables into the model which were market size, past stock returns, and the idiosyncratic standard deviation of stock returns. Those market variables were more sensitive to the change of the enterprise, the model could be better explained.

2.2.3.2 Chava & Jarrow's model

Chava & Jarrow (2004) studied discrete time hazard model. They substituted three newly identified market variables: excess return (the difference between firm's stock return and index return), relative size (the ratio of firm's market value to index market value), and sigma (the stock's volatility).

$$P_{t-1}(Y_{it} = 1)$$
$$= 1/1 + \exp(-\alpha - \beta_1 \text{TLTA} - \beta_2 \text{NITA} - \beta_3 \text{SIGMA} - \beta_4 \text{EXRET} - \beta_5 \text{RSIZE})$$

$$\text{NITA} = \text{Net income}/\text{Total asset}$$
$$\text{TLTA} = \text{Total lability}/\text{Total asset}$$
$$\text{SIGMA} = \text{stock's volatility}$$
$$\text{EXRET} = \text{different between firm's return and index'return}$$
$$\text{RISZE} = \log(\text{total firm equity value}/\text{total equity value in market})$$

The result of their study indicated that Shumway's discrete time hazard model was superior to Altman's Z-score model. Differing from what Shumway had done, Chava & Jarrow chose both yearly and monthly data, which was tested to improve the performance of model. Besides, an achievement of their research was that they firstly studied the effect of enterprises type to bankruptcy prediction. By grouping companies into four sections, which were financial institution (finance, insurance and real estate), transportation, communications and utilities, manufacturing and minerals, other industries respectively, they found out enterprises type had significant influence on the intercept and coefficient of predicting models. Therefore, an approach to optimize model could be the adjustment according to different kinds of companies.

2.2.3.3 Campbell et al's model

Based on the previous study of Shumway (2001), Chava & Jarrow (2004), and others, Campbell et al (2008) proposed a new predicting model. Two major breakthroughs had been taken by them. First one is the expanding of explanatory variable. Studies shown that the use of the asset's market value was better than the book value which was frequently used in previous studies. Because the market value contained the latest market information which could better reflect the company's prosperity, the company's intangible assets evaluation would be more accurate. Motivated by this result, they substituted the previous (Net income)/(total asset) & (total lability)/(total asset) of (Net income)/(market value of total asset) & (total lability)/(market value of total asset). More importantly, in Campbell et al's model, not only the company's cash and short - term assets to the market value of its assets was added as an expanding variable to reflect the company's liquidity situation, market - to - book a well.

$$P_{t-1}(Y_{it} = 1) = \frac{1}{1 + \exp(-\alpha - \beta_1 \text{TLMTA} - \beta_2 \text{NIMTA} - \beta_3 \text{CASHMTA} - \beta_4 \text{SIGMA} - \beta_5 \text{EXRET} - \beta_6 \text{RSIZE} - \beta_7 \text{MB} - \beta_8 \text{PRICE})}$$

$$\text{NIMTA} = \text{Net income/Market value of total asset}$$
$$\text{TLMTA} = \text{Total lability/Market value of total asset}$$
$$\text{SIGMA} = \text{stock's volatility}$$
$$\text{EXRET} = \text{different between firm's return and index'return}$$
$$\text{RISZE} = \log(\text{total firm equity value/total equity value in market})$$
$$\text{CASHMTA} = \text{Cash and shout asset/Market value of total asset}$$
$$\text{MB} = \text{Market value equity/Book value equity}$$
$$\text{PRICE} = \text{Log price per share}$$

Another breakthrough was that they found some generality of bankrupt companies: low profitability, high leverage, low liquidity, high market-to-book ratio, high volatility and low price per share. Campbell et al and others believed that their model fitted the general time pattern quite well, and it had a stronger explanatory power than the models of Shumway (2001), Chava & Jarrow (2004).

2.3 Assessment on three types of models.

It is reasonable to conclude that model based on accounting information has a weaker performance than it based on market information. Agarwal and Taffler (2008) argued that accounting information could be more effective in evaluating past conditions of a firm, but the data derived from financial statements have limited ability to predict bankruptcy. Hellegeist et al (2004) indicated that in lack of some fundamental market information such as market value of stock price and volatility of assets, accounting based models were not as accurate as market-based model in bankruptcy forecasting.

However, the market-based model is still imperfect. Saunders and Allen (2002) put up with the main deficiency of this model could be the limitation of assumptions. For instance, stock returns are required to be normal distributed in market-based models, which is virtually impossible to be satisfied. Overall, according to research reality, mixed model figures out the main problem of using accounting information or market information separately and is testified to be a better kind of bankruptcy predicting model.

3 Methodology

The paper aims to do comparative analysis about two models, Chava & Jarow (2004), Campbell et al (2008) which mentioned above. When conducting the study, two approaches would be adapted to evaluate performance of two models. The probability of bankruptcy will be transformed by the logit regression function. Another one is verification method. Two validation techniques (ROC and CAP curve) would be implemented to rate different models.

3.1 Logit regression

A normal formulation of a logit regression is as follows:

$$Y_{it} = \alpha + \beta_j X_{ijt-1} + \varepsilon_{it-1}$$

$$P_{t-1}(Y_{it} = 1) = 1/1 + \exp(-\alpha - \beta_j X_{ijt-1})$$

where Y_{it} is a virtual variable representing the bankruptcy of the firm i. $Y_{it} = 1$ means the firm i went bankrupt at time t. $Y_{it} = 0$ means the firm i not went bankrupt at time t. X_{ijt-1} representing the j independent variables of firm i at time t-1. α is a constant term.; β_j is regression coefficient; ε_{it-1} is an error term. $P_{t-1}(Y_{it} = 1)$ is the predicting probability of a bankruptcy while a company truly filed a bankruptcy.

To be more specific, ten independent variables are included in two target models of this paper separately, which are listed above.

$$P_{t-1}(Y_{it} = 1) = \frac{1}{1 + \exp(-\alpha - \beta_1 \text{TLTA} - \beta_2 \text{NITA} - \beta_3 \text{SIGMA} - \beta_4 \text{EXRET} - \beta_5 \text{RSIZE})}$$

Chava & Jarrow's model

$$P_{t-1}(Y_{it} = 1) = \frac{1}{1 + \exp(-\alpha - \beta_1 \text{TLMTA} - \beta_2 \text{NIMTA} - \beta_3 \text{CASHMTA} - \beta_4 \text{SIGMA} - \beta_5 \text{EXRET} - \beta_6 \text{RSIZE} - \beta_7 \text{MB} - \beta_8 \text{PRICE})}$$

Campbell et al's model

A logistic regression does not have any particular distribution requirements for variables, thus it can be widely used and get a robust parameter estimation. Besides, the probability of bankruptcy is easy and direct to get in this case.

3.2 Rating methodologies (ROC and CAP curve)

3.2.1 ROC curve

ROC (Receiver Operating Characteristic) curve is a graphic plot that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied. In this paper, it is an approach to assess model's predictive ability. Y axis represents the true positive rate, which is defined as the rate when a company is forecasted to be bankrupt and it has been bankrupt in reality. The X-axis represents the false positive rate referring to a situation where a company is forecasted to be bankrupt but not in actual. The same model under different threshold value of bankruptcy probability such as 90% and 50% will perform different. If the curve is on the upside of random guess, the corresponding model is indicated to have a relatively accurate prediction for bankruptcy. More close to the upper left corner the curve is, more accurate the prediction.

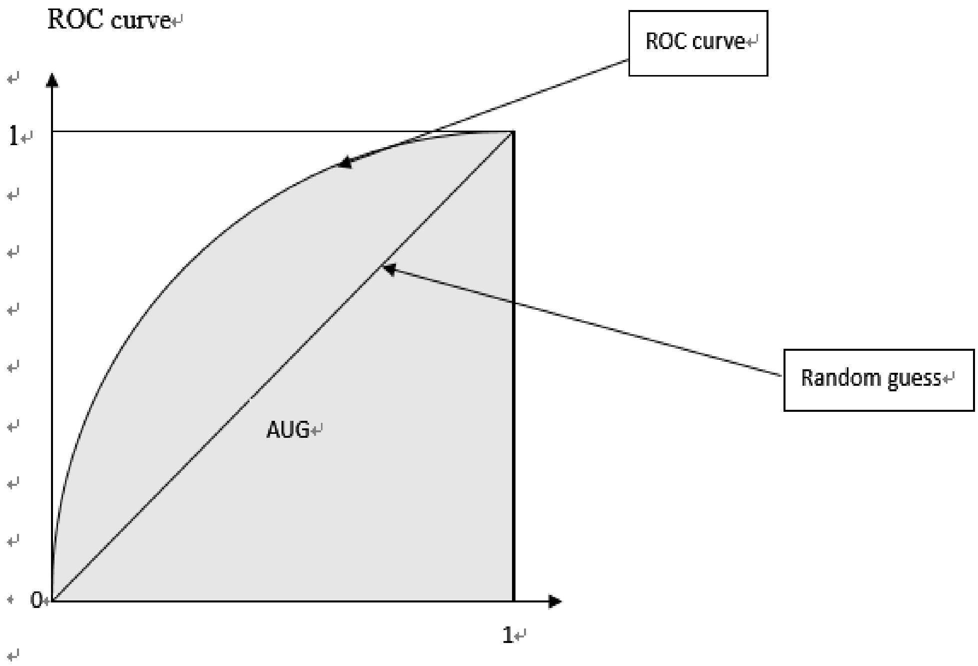


Chart 1: Explanation of ROC curve

AUC stands for the Area Under the Curve and usually refers to the area under the ROC curve. AUC can be regarded as, under all the possible cutoffs, the model's capability to distinguish bankrupt companies and normal companies. The value of AUC is supposed to vary from 0 to 1. Hosmer and Lemeshow (2000) gave a detailed account of AUC. If the value of AUC is below 0.5, the predicting model fail to identify two kinds of companies. When the value of AUC is between 0.5 and 0.7, model's predicting ability is average. A value from 0.7 to 0.8, a well-predicting model can be defined. If the value of AUC exceeds 0.8, in normal case, the model is almost perfect with strong predicting ability.

3.2.2 CAP curve

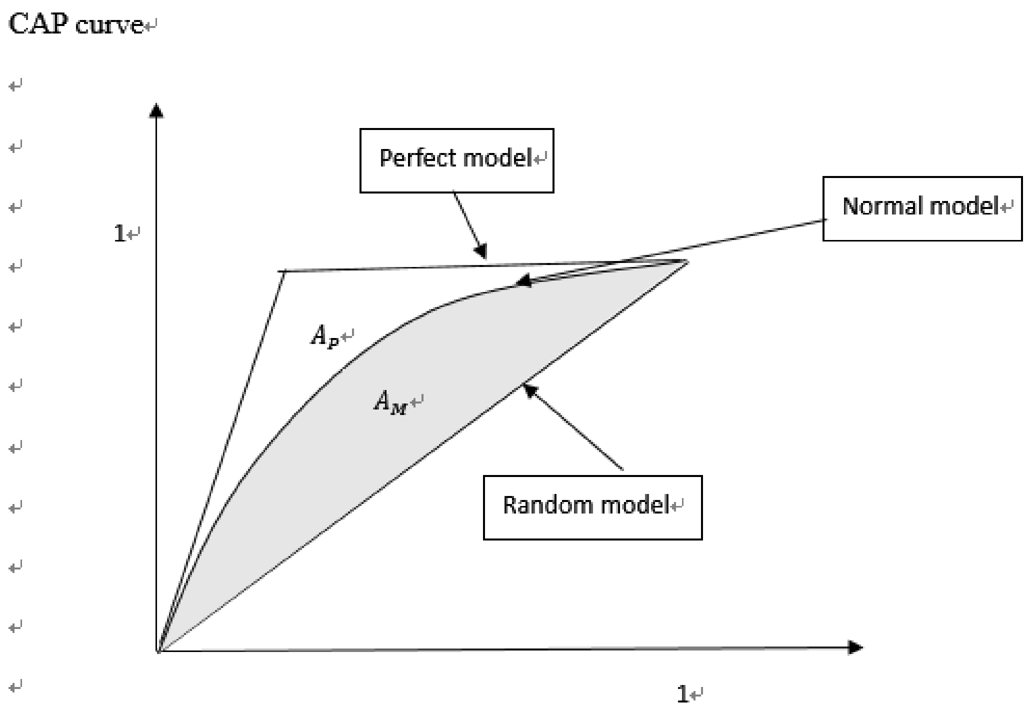


Chart 2: Explanation of CAP curve

The cumulative accuracy profile, denoted as CAP curve, is a statistic tool to evaluate the performance of predicting models. It graphically represents the accuracy of bankruptcy prediction models. To demonstrate the CAP curve, firstly, bankrupt risk of each company estimated by logit regression should be arranged in a pecking order from high to low. The X-axis illustrates alarm rate of bankrupt risk which is prior to X%. The Y-axis shows cumulative hit rate of bankrupt companies, which is literally the percentage of bankrupt companies correctly distinguished by model to all bankrupt companies. All points are connected by straight lines, and a CAP curve is depicted like chart 2.

A perfect model is capable of generating higher bankruptcy probability for all bankrupt companies. Its CAP curve has a very high slope in initial, and Y value will soon increase to 1. While a random model denoted as a straight line which does not have the ability to forecast, it would randomly give any companies a probability of bankruptcy no matter what the actual situation is. In practice, most models are not in the extreme situation to be perfect or random. The performances of those models are normally between a perfect model and a random model as the chart shows. Additionally, a single indicator named Accuracy Ratio (AR) is used to measure the efficacy of CAP curve on the model. AR can be calculated by the ratio of $\mathrm{A_m}$ to $\mathrm{A_p}$, where $\mathrm{A_m}$ is the area between normal model and random model, $\mathrm{A_p}$ represents the area between perfect model and random model.

CAP curve is similar but not identical to ROC curve. Englemann, Hayden and Tasche (2003) proved that CAP's accuracy ratio AR and the area under ROC curve AUC are related by the equation AR=2ROC-1. Moreover, CAP curve emphasizes on the relation of hit rate to all alarms, while ROC curve concentrates more on false alarms.

4 Empirical study

Shinong Wu and Xianyi Lu (2001) tested the performance of Fisher's linear discriminant analysis, multiple linear regression and logistic linear regression on predicting bankruptcy. It turns out that all three models could forecast accurately but logistic regression is superior to the other. In this section, a logit regression is proceeded to estimate the probability of bankruptcy for each company.

4.1 Sample selection

In acquiring the sample of this research, some principles are formulated briefly as follows. Data source is COMPUSTAT. For defining bankruptcy, chapter 11 and chapter 7 are adapted, which are represented by delisting code 2 and 3 respectively. The subjects are bankrupt companies and non-bankrupt companies in North America listed on New York Stock Exchange (NYSE), American Stock Exchange (AMEX) and National Association of Securities Dealers Automated Quotation (NASDAQ). The corresponding exchange code for NYSE, AMEX and NASDAQ is 11, 12 and 14. Sample's geographic scope is limited to study the North America market, where has a sufficient number of observations on bankruptcies of listed firms.

The time period of data spans from 1993 to 2013. The latest bankrupt information is not included due to inaccessibility in data. In the data collection of bankrupt companies, the last 5 years bankrupt data ever since 2014 are excluded because some data are partly missing, and data of financial institutions such as ETF, non-operating companies with Standard Industry Classification Code 9995 are pointless. Thus, only 45 companies filing a bankruptcy between 1993 and 2013 are remained to be effective samples at last. As for non-bankrupt companies, 180 non-bankrupt companies are selected by random sampling in terms of 4 times larger than the amount of bankrupt companies in the same year. That is to say, if 10 bankrupt companies are available in 1993, correspondingly, 40 non-bankrupt companies are stochastically selected in 1993.

As it is mentioned before, 5 variables and 8 variables are included in two models respectively. For each company in each model, normally only one set of yearly data is necessary. To maximally validate these two models, four sets of quarterly data containing all these variables are needed. The four quarterly data ordered by time for bankrupt companies should better be the nearest to the year when the company goes bankrupt, but not more than two and a half years distance from the year filing a bankruptcy. It is basically the same principle for randomly chosen non-bankrupt companies. If a non-bankrupt company is selected as a sample in 1993, the four quarterly data should be ranged from January 1992 to December 1992. This paper contains 888 sets of quarterly data in total, consisting of 168 sets of bankrupt quarterly data and 720 sets of non-bankrupt quarterly data.

Data between 1993 and 2003 would be used to carry out in-sample estimation, and data spanning from 2004 to 2013 would be used to proceed out-sample prediction. Accordingly, every in-sample quarterly data is applied to estimate the coefficients of each model in logit regression.

4.2 Sample description

The samples of bankrupt companies are counted up in terms of the exchange markets. The result shows as following that most of them listed on NASDAQ. The number of companies listed on AMEX and NYSE are the same.

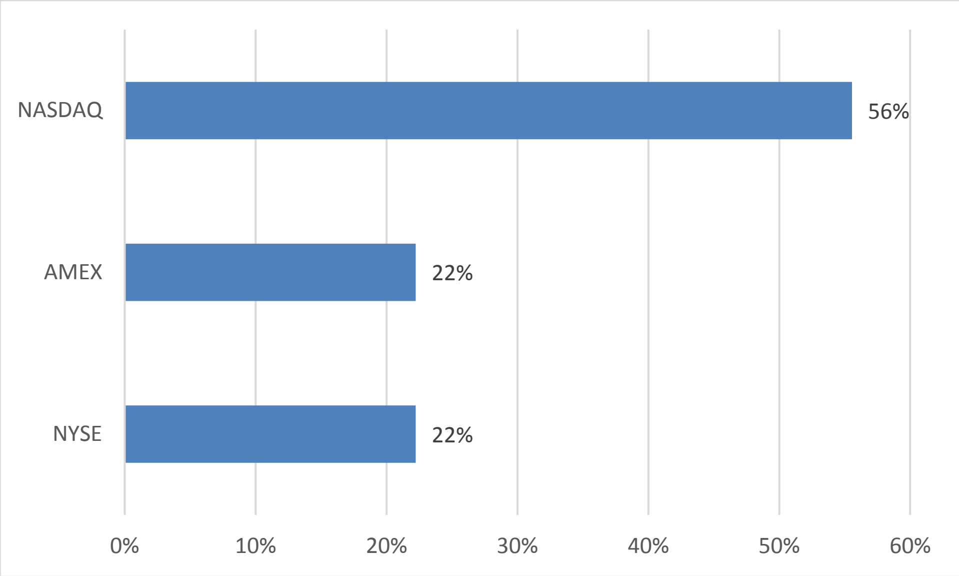


Chart 3: Bankruptcy by exchange market

Chava & Jarrow (2004) studied industry effect on bankruptcy prediction models. They found that enterprises type had significant influence on the intercept and coefficient of predicting models. Accordingly, samples classified by industry would have different regression equation and accuracy of models would be improved. In this paper, the distribution in various industries of 45 bankrupt companies are counted in conformity with Standard Industry Classification Code (SIC). Although financial institution has the largest share, accounting for almost 36% of total bankrupt companies, it is still insufficient to conduct logit regression by industries. Less than 350 sets of data could be implemented a regression even in a financial industry and in service industry 80 sets of data are avaliable. Inadequete data would result in bias estimation. Therefore, the logit regression in this study would be performed considering no industry effect.

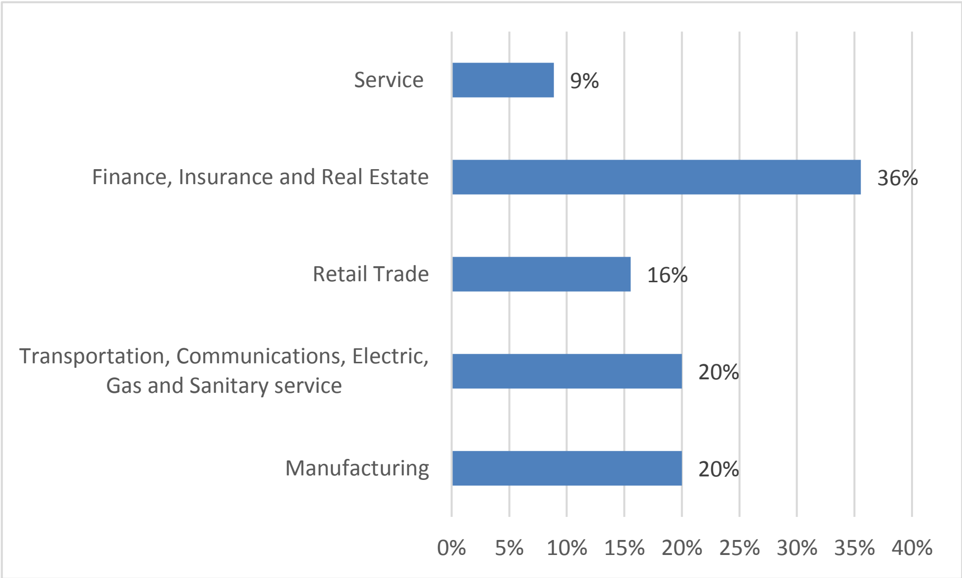


Chart 4: Bankruptcy by industry

4.3 Variable selection

In Chava & Jarrow's model, the logit regression is formulated as:

$$P\,(Y_{it} = 1) = \frac{1}{1 + \exp(-\alpha - \beta_1 \mathrm{TLTA} - \beta_2 \mathrm{NITA} - \beta_3 \mathrm{SIGMA} - \beta_4 \mathrm{EXRET} - \beta_5 \mathrm{RSIZE})}$$

And in Campbell et al's model, it is:

$$P\,(Y_{it} = 1) = \frac{1}{1 + \exp(-\alpha - \beta_1 \mathrm{TLMTA} - \beta_2 \mathrm{NIMTA} - \beta_3 \mathrm{CASHMTA} - \beta_4 \mathrm{SIGMA} - \beta_5 \mathrm{EXRET} - \beta_6 \mathrm{RSIZE} - \beta_7 \mathrm{MB} - \beta_8 \mathrm{PRICE})}$$

Totally, 10 explanatory variables are involved. Those are listed in detail below.

$$\mathrm{NITA} = \text{Net income/Total asset}$$
$$\mathrm{TLTA} = \text{Total lability/Total asset}$$
$$\mathrm{NIMTA} = \text{Net income/Market value of total asset}$$
$$\mathrm{TLMTA} = \text{Total lability/Market value of total asset}$$
$$\mathrm{SIGMA} = \text{stock's volatility}$$
$$\mathrm{EXRET} = \text{different between firm's return and index'return}$$
$$\mathrm{RISZE} = \log(\text{total firm equity value/total equity value in market})$$
$$\mathrm{CASHMTA} = \text{Cash and shout asset/Market value of total asset}$$
$$\mathrm{MB} = \text{Market value equity/Book value equity}$$
$$\mathrm{PRICE} = \text{Log price per share}$$

In the database of COMPUSTAT, data like market value of total asset are missing. Therefore, in this paper, a practical approach raised by Campbell et al (2008) is applied. The market value of total asset would be equivalent to the sum of book value of liability and market value of equity. S&P 500 would be proxy for index.

Variables TLTA, NITA, EXRET, and RISZE are used in both models, while CASHMTA, MB, PRICE is only used in Campbell et al's model.

4.4 Statistic description of variables

The statistical properties of ten given variables are displayed below.

	TLTA	NITA	TLMTA	NIMTA	MB	CASHMTA	PRICE	EXRET	SIGMA	RESIZE
Group A: Non-Bankrupt Firm										
Mean	0.404103	-0.00036	0.340086	-0.00127	2.121957	0.271645	1.114767	-0.01751	0.034582	-10.6439
Median	0.353579	0.007726	0.274649	0.006864	1.539039	0.208662	1.187513	-0.00998	0.028197	-10.6811
Std.Div	0.266973	0.043754	0.260079	0.043998	2.025315	0.243511	0.499501	0.146339	0.024615	1.029329
Min	0.011698	-0.32337	0.007022	-0.61408	-6.37271	0.000112	-0.60206	-1.34121	0.006465	-13.1632
Max	1.047084	0.177454	0.957717	0.211001	25.24725	1.487944	2.531479	0.712081	0.178168	-7.79706
Group B: Bankrupt Firm										
Mean	0.602363	-0.01134	0.444748	-0.01712	7.039887	0.193818	0.533477	-0.0531	0.062167	-11.4954
Median	0.569207	0.002204	0.399884	0.001575	1.048715	0.038658	0.536171	-0.0288	0.037799	-11.4466
Std.Div	0.661135	0.282476	0.340657	0.145427	59.47873	0.387152	0.540615	0.226926	0.067833	0.768624
Min	0.00805	-2.21646	0.001276	-1.09641	0.015926	0	-1.39794	-1.00266	0.00382	-13.6529
Max	6.966523	2.095547	0.998896	0.710248	772.125	2.889776	1.451018	0.8617	0.48576	-9.86544

Table 1: Statistics summary of all variables

Group A refers to 180 non-bankrupt companies. From table, it is easy to find out the variable NIMTA. The average of NIMTA with value -0.00127 is almost zero and lower than its median. Thus, it can be inferred that the probability distribution of NIMTA is negatively skewed. A negatively-skewed distribution often called left-skewed distribution, which has a long-left tail. Normally, in this situation, most data are located in the right side of average. It is the same case for variable NITA. As it can be seen, the average of NITA and NIMTA are negative, which means a large portion of selected companies have a negative net margin. In other words, those companies are in loss. For the variable TLTA, average of TLTA is 0.4041013. Assets acquired by leverage is about 40% on average. On the other side, average of TLMTA is 0.340086, which is lower than TLTA by 6%. It can attribute to a high MB ratio. When MB ratio is larger than 1, market value of stock surpasses its book value. The average of variable EXRET is -0.01751, indicating that non-bankrupt companies have poor performance in late sample with a weighted average near zero. Besides, average of variable SIGMA is 0.034582. the stock volatility is very low. As for the variable CASHMTA with average 0.271645, market value of total assets is slightly higher than cash and short-term invest.

Group B refers to 45 bankrupt companies. Overall, ten variables individually in group B have the same skewness with those in group A. Thus, a difference comparison would emphasize on the average of two groups. For the variables TLTA and TLMTA, which reflect leverage ratio, group B's average is 60% and 44%, higher than the value in group A. Bankrupt companies are more likely to have a high leverage. For variable MB, its average 7.039887 is much larger than the value in group A. One explanation is that stocks of bankrupt companies are overvalued. Most companies in group B are in loss, the same case but worse than companies in group A. To some extent, it indicates that the performance of bankrupt companies do not meet investors 'expectations'. From the variable CASHMTA, cash liquidity is weak for bankrupt companies compared with the value in group A, and a low cash flow liquid means a low turnover of current capital once facing with a financial distress. Stock return of companies in group B is lower than companies in group A on average, referred from variable EXRET. Moreover, it is obvious that stock price of bankrupt companies is more volatile than non-bankrupt companies. Thus, stocks of bankrupt companies seem to have high volatility and low return.

4.5 Implementation of logit regression

In conducting the logit regression, three exercises are generated. Firstly, for time period 1993 to 2013, all sets of data formulated by 10 variables are employed in two models to proceed significance test. Secondly, data within 1993 to 2003 are applied to the so called in-sample estimation, which give the estimated coefficients of variables for two models. Finally, using the estimated coefficients to implement an out-sample prediction with data ranging from 2004 to 2013. The forecasting probability of bankruptcy could be computed.

Specifically, when conducting logit regression, for every set of explanatory variables in bankrupt database, the corresponding dummy variable is 1, bankruptcy normally is a situation that the company suffers from a financial distress for a bit long time. It is acceptable to value all dummy variable is in four sets of each model of every bankrupt company. This is illustrated in a logistic regression as following:

$$Y_{it} = \alpha + \beta_j X_{ijt-1} + \varepsilon_{it-1}$$

Y=1, when a set of independent variables comes from bankrupt company.

Y=0, when a set of independent variables comes from non-bankrupt company.

This paper contains 888 sets of quarterly data in total, consisting of 168 sets of bankrupt quarterly data with dependent variables Y equal 1, and 720 sets of non-bankrupt quarterly data with dependent variables Y equal 0.

5 Result analysis of logit regression

5.1 Analysis of Chava & Jarrow's model

Period		1993 to 2013
Observations		888
Variable	**Coefficient**	**Significant (P-value)**
TLTA	2.094343	0.00%
NITA	1.127231	6.22%
EXRET	-0.058428	91.75%
SIGMA	8.995582	0.19%
RESIZE	-1.004873	0.00%
Constant	-13.96122	0.00%
McFadden R-squared		0.199
LR statistic		171.71

Table2: Result of logit regression of Chava & Jarrow's model with simple from 1993 to 2013

Look at the P-value of each variable, coefficient of EXRET is not significant. Coefficient of NITA is significant at 10% confidence level, and coefficients of other variables are significant at 1% confidence level. 5% confidence level is set as the principle in this paper, therefore, NITA is regarded as insignificant.

The positive coefficients of variables TLTA and SIGMA indicate that these two variables are positively correlated with predicting probability of bankruptcy. If these two variables are increased or decreased by one unit, the logarithm of bankruptcy prediction probability will increase of decrease accordingly. Furthermore, SIGMA has a greater impact on probability of bankruptcy than TLTA due to a higher value of coefficient. Due to the largest coefficient of SIGMA, stock volatility is predominantly influential on bankruptcy prediction. RESIZE is considered to be negatively correlated with bust-up risks, which means, a small size company is easier to go broke than big companies.

Model's goodness of fitness can be assessed through McFadden R-squared. When

McFadden R-squared is between 0.2 and 0.4, independent variables in this model are believed to be capable of explaining the dependent variable. From the regression result, McFadden R-squared of Chava & Jarrow's model is 0.199, implying that model interpretation is roughly good.

5.2 Analysis of Campbell et al's model

As it is mentioned before, the main difference between Chava & Jarrow's model and Campbell et al's model is the replacement of TLTA and NITA by TLMTA and NIMTA in Campbell et al's model. Campbell et al (2008) thought compared with the book value of total asset, market value of total asset could better reflect a company's real properties.

Period		1993 to 2013
Observations		888
Variable	**Coefficient**	**Significant (P-value)**
TLMTA	0.857795	3.52%
NIMTA	0.434343	73.21%
CASHMTA	-2.409244	0.00%
MB	0.136404	0.00%
PRICE	-2.049720	0.00%
EXRET	0.394755	49.70%
SIGMA	-0.657924	83.36%
RESIZE	-0.650490	0.00%
Constant	-6.964596	0.00%
McFadden R-squared		0.277
LR statistic		238.77

Table 3: Result of logit regression of Campbell et al's model with simple from 1993 to 2013

According to the table above, variable SIGMA, NIMTA and EXRET are not significant, which quite deviates from the result of Campbell et al's study, where all the variables were significant. TLMTA and MB is positively correlated with bankruptcy probability, while CASHMTA is negatively correlated with it. From Campbell et al's point of view, a company with a high CASHMTA can pay interest by current assets. Therefore, if the company gets better during this period, the probability of bankruptcy could be reduced. The probability of going bankrupt is sensitive to Price. The higher the price per share, the lower the probability of bankruptcy. It can be well explained by the investors 'expectations. For Campbell et al's model, the McFadden R-squared is 0.277, which implies that it is better fitted in comparison with Chava & Jarrow's model.

5.3 Conclusion for empirical result

Coefficients of NITA and EXRET are unusually neither significant in Chava & Jarrow's model nor in Campbell et al's model. Considering of variable NITA, one explanation could be that selected data are too short to reflect a company's profitability. In this paper, only four quarterly data are chosen for each company, while it is not sufficient for defining a company's profitability. For instance, an investment would somehow lead to a loss when it is beginning. After the invested program growing to be mature, return would then be relatively high. Thus, a quite long process is necessarily needed to see how the net income would affect the probability of bankruptcy. And normally four quarterly data are not enough. This may be one underlying reasons why variable NITA is not significant in both two models. As for variable EXRET, one guess could be that, high stock return not only implies a steady-developed company, is correlated with highly risky companies as well. Basically, the higher the yield, the higher the risk, which is the property of bankrupt companies. The trend in both directions may lead to variable EXRET insignificant in two models. However, these explanations still need to be testified. As for stock volatility, referred as SIGMA, the estimation in Campbell et al's model is very different from it for the Chava & Jarrow's model (negative and insignificant in one, positive and highly significant in the other). The reason behind might be the substitution of market value in Campbell et al's model, which constitutes of more market information than Chava & Jarrow's model. In this manner, stock volatility could reasonably be insignificant in Campbell et al's model.

It is worth to mention that data used in this regression experience financial crisis. The two models validated in this paper are mixed models, which consist of market information and accounting information. During the financial crisis, bankrupt risk is systematically high. Not surprisingly, robustness of models is inevitable to be affected by the financial crisis. However, the same sample is used in two models. To some extent, it would not insert that much influence on the comparative analysis, but whether the financial crisis produce more negative effect on any models is still unknown. Ignoring the influence of financial crisis could be a delimitation of this paper.

In terms of the regression results, Campbell et al's model apparently fits better than Chava & Jarrow's. Although both of them passes likelihood ration test, model of Campbell et al has a larger LR statistic value than another, 238 and 171 respectively. Besides, McFadden R-squared could be regarded as an explanation for the reason why Campbell et al's model is better-performed in empirical part. In addition, with two more variables CASHMTA and Mb which are verified to be effective, Campbell et al's model has stronger explaining power and would be more perfectly implemented than Chava & Jarrow's model.

6 Comparative analysis of two models

This paper uses ROC curve and CAP curve to perform out-of-sample forecasting, so as to measure the predictive power of two models. The results of estimated parameters for two models are separately presented below, which would then be used to predict the probability of bankruptcy for companies spanning from 2004 to 2013.

Period		1993 to 2003
Observations		642
Variable	**Coefficient**	**Significant (P-value)**
TLTA	2.431188	0.00%
NITA	0.692481	29.91%
EXRET	0.616832	30.50%
SIGMA	5.489656	9.55%
RESIZE	-1.262071	0.00%
Constant	-16.44148	0.00%
McFadden R-squared		0.241
LR statistic		156.26

Table 4: Result of logit regression of Chava & Jarrow's model with simple from 1993 to 2003

Period		1993 to 2003
Observations		642
Variable	**Coefficient**	**Significant (P-value)**
TLMTA	0.754229	9.76%
NIMTA	-0.664149	68.80%
CASHMTA	-3.376003	0.00%
MB	0.208199	0.00%
PRICE	-1.496617	0.00%
EXRET	0.371803	54.90%
SIGMA	0.560908	68.80%
RESIZE	-1.037925	0.00%
Constant	-11.445063	0.00%
McFadden R-squared		0.277
LR statistic		193.2

Table 5: Result of logit regression of Campbell et al's model with simple from 1993 to 2003

Accordingly, two predicting models can be formulated as following:

$$P(Y_{it} = 1) = \frac{1}{\begin{array}{c}1 + \exp(16.44147616 - 2.431188 * \text{TLTA} - 0.692481 * \text{NITA} \\ -0.616832 * \text{EXRET} - 5.489655735 * \text{SIGMA} + 1.262071406 * \text{RSIZE})\end{array}} \quad (1)$$

Chava & Jarrow's model

$$P(Y_{it} = 1) = \frac{1}{\begin{array}{c}1 + \exp(11.445063 - 0.754229 * \text{TLMTA} + 0.664149 * \text{NIMTA} \\ +3.376003 * \text{CASHMTA} - 0.560908 * \text{SIGMA} - 0.371803 * \text{EXRET} \\ +1.037925 * \text{RSIZE} + 0.208199 * \text{MB} + 1.496617 * \text{PRICE})\end{array}} \quad (2)$$

Campbell et al's model

6.1 Analysis of ROC curve

In the ROC curve, the Y-axis indicates the true position rate that the forecasting result of a company is bankrupt, and it has actually filed a bankruptcy. For instance, ordinate value is 0.5, implying 50 bankrupt companies are forecasted to be bankrupt out of 100 bankrupt companies. The X-axis refers to the false position rate that a company is predicted to be bankrupt while it does not go bankrupt. The 45-degree line represents a random model without predictive power.

The graph below is the ROC curve for Chava & Jarrow's model, which is above the 45-degree straight line. AUC of it is 0.7871. In this case, the model is obviously better than stochastic model. Besides, on the basis of Hosmer and Lemeshow (2004), a model with AUC between 0.7 and 0.8 can be considered as a well-predicting model.

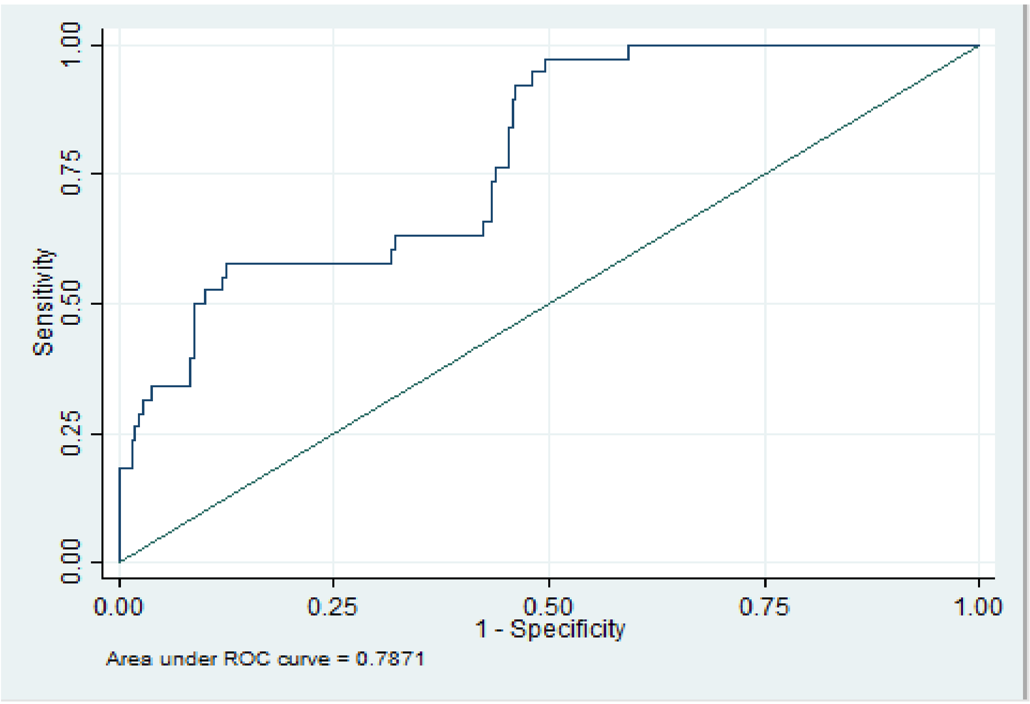


Chart 5: ROC curve of Chava & Jarrow's model

It is basically the same case for Campbell et al's model, except it has a higher AUC with value exceeding 0.8, which can be regarded as a perfect model.

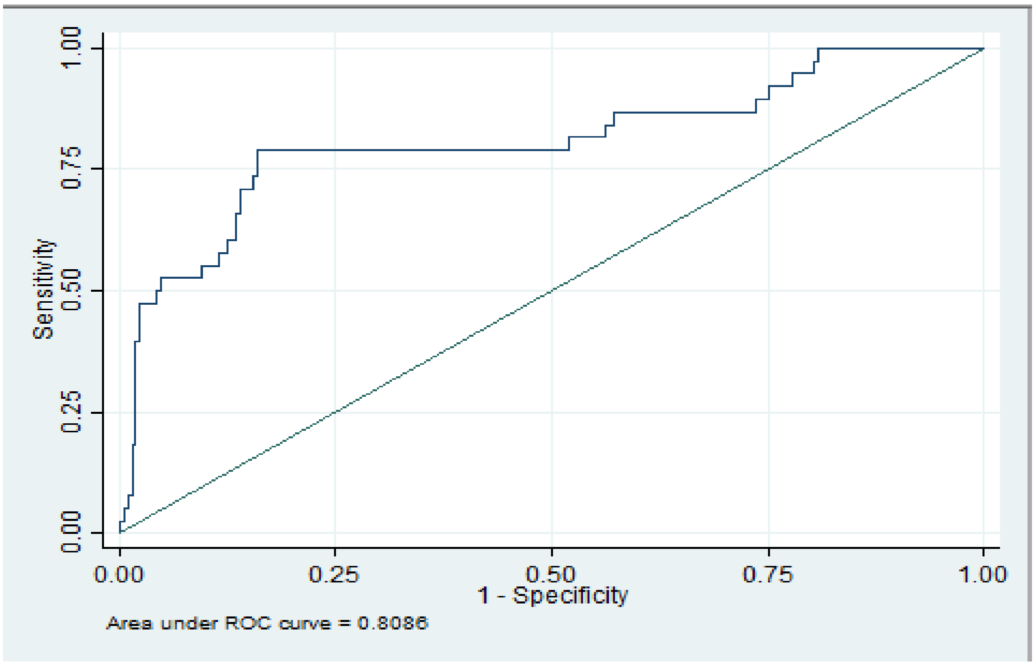
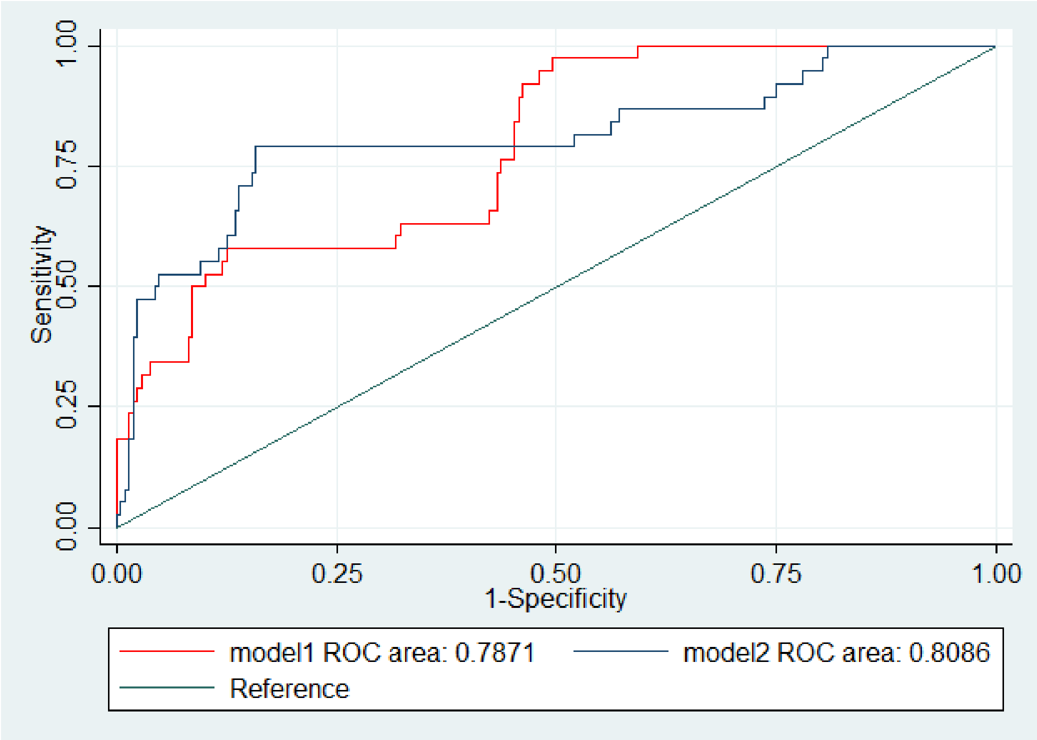


Chart 6: ROC curve of Campbell et al's model

A diagram of combining the ROC curve of two models together is displayed below.



(**Model 1**: Chava & Jarrow's model **Model 2**: Campbell et al's model)

Chart 7: ROC curve by comparing two models

In the case of high threshold, specifically referring to the front part of the curve. Campbell et al's model performs better than Chava & Jarrow's. With the decrease of threshold, two curves tend to plateau. In the low-threshold phase, two curves start to climb but curve of Chava & Jarrow's model climbs faster than it of Campbell et al's model. Chava & Jarrow's model has a better performance in the second half of the curve. Normally the situation in area with a false position rate under 0.5 could be more discussed. Combining with a higher AUC value, Campbell et al's model is considered to perform better, which is consistent with the result of logit regression.

6.2 Analysis of CAP curve

The summary statistic of CAP curve and the accuracy ratio, AR, can be calculated directly from AUC (Engelmann et al 2003). The CAP curve is also denoted as power curves, which is used for visually evaluating model performance. For a given percentage of the total number of companies, the CAP curve is constructed as the y(x) percentage of the bankrupt companies whose non-bankrupt probability is lower than or equal to x. The X-axis is sorting the companies by increasing the non-bankrupt probability.

Chart 8 represents the CAP curves for both two models. As it can be seen, when the forecasting probability of non-bankruptcy is 20%, x=0.2, equivalently, the portion of the bankruptcies predicted by Campbell et al's model is 40%, while it is only 20% in Chava & Jarrow's model. Campbell et al's model can identify more bankrupt companies than another until the probability of non-bankruptcy exceeding 50%, or X>0.5 Accuracy ratio should be considered. The model with a higher accuracy demonstrates good ability of bankruptcy prediction. Table 8 shows that AR of Campbell's model is slightly higher than it of Chava & Jarrow's. Campbell et al's model performs better than Chava & Jarrow's model in this part.

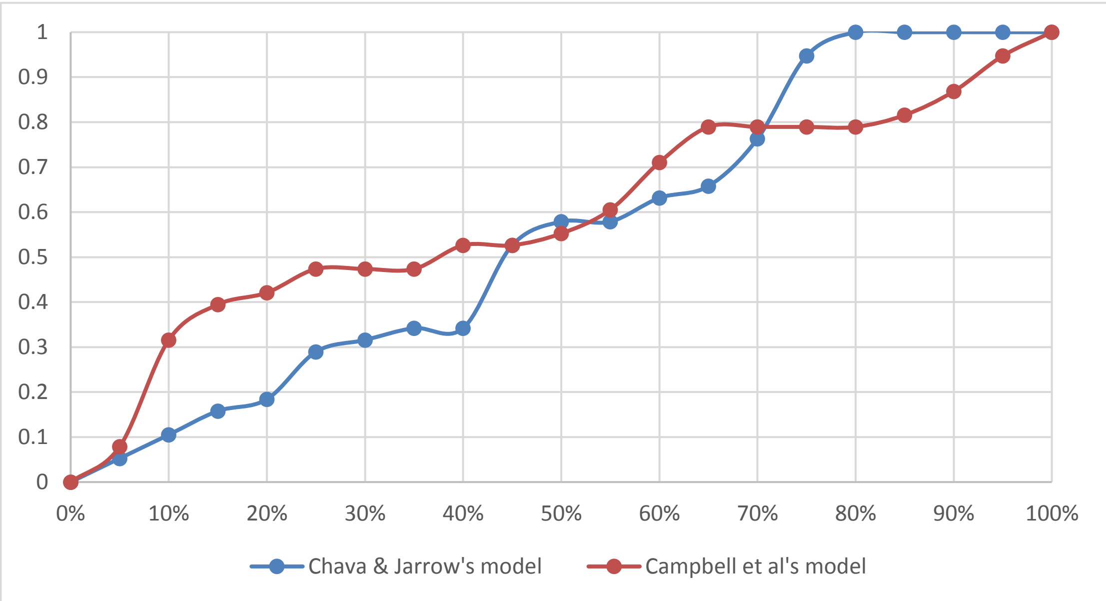


Chart 8: CAP curve by comparing two models

Model	AR (Accuracy Ratio)
Chava & Jarrow	0.5742
Campbell et al	0.6172

Table 6: Result of accuracy ratio of two models

6.3 Conclusion for comparative analysis

ROC curve and CAP curve are approaches to assess the discriminative power of two models. Both results indicate that Campbell et al's model is more powerful at predicting bankruptcy than Chava & Jarrow's model. It complies with the conclusion draw in logit regression. Overall, Campbell et al's model is better performed compared with Chava & Jarrow's model.

7 Conclusion

Throughout the paper, data derived from NYSE, NASDAQ and AMEX between 1993 and 2013 are selected to validate the predicting accuracy of two models, Chava & Jarrow's model (2004) and Campbell et al's model (2008). Two techniques are implemented to evaluate these models. In the logit regression, coefficients are estimated with in-sample data. Total liability to total assets, denoted as TLTA and TLMTA is tested to have significant influence on bankruptcy predicting. RSIZE is negatively related with the probability of bankruptcy, which means a relatively larger size company has a lower probability to go bankrupt than a small company under the same other circumstances. The stock volatility has the strongest explanatory power in Chava & Jarrow's model, but the opposite case in Campbell et al's model. The difference could be partly explained by the substitution of market value in Campbell et al 's model, but it still needs to be testified.

In terms of the regression results, Campbell et al's model apparently fits better than Chava & Jarrow's model for a higher LR statistic value and McFadden R-squared. In ROC curve and CAP curve. Although Campbell et al's model is not superior to Chava & Jarrow's all the time, especially when alarm rate exceeding 0.8 in CAP curve, it still has a slightly higher AUC and AR, which indicate a better performed model. Consequently, a conclusion is rational to be conducted as: With the same sample selected in this paper, Campbell et al's model is better than Chava & Jarrow's model.

The reason behind might be considered from two parts. One is the extended three variables in Campbell et al's model MB, CASHMITA and PRICE are significant. Furthermore, variable CAHSMITA and PRICE have relatively high explanatory power for the model, with coefficients equal to -2.409244 and -2.049720. In consequence, the performance of Campbell et al's model is improved. The other is that more variables reflecting market information are included in Campbell et al's model. Agarwal and Taffler (2008) indicate accounting information could be effective in evaluating past conditions of a firm, but market information is more sensitive to instant operational status of a company, which is predominant when forecasting the probability of bankruptcy. In this manner, a further inspiration could be draw as a bankruptcy prediction model with larger range of market information is likely to be more accurate in forecasting bankruptcy.

References

Afik, Z., Arad, O., & Galil, K., (2016). Using Merton model for default prediction: An empirical assessment of selected alternatives, Journal of Empirical Finance, Vol 35. Pp.43-67.

Agarwal, V., & Taffler, R., (2008). Comparing the performance of market-based and accounting-based bankruptcy prediction models, Journal of banking and Finance, Vol 32 (8). Pp. 1541-1551.

Altman, E. I. (1968). Financial ratio, Discriminant Analysis and the Prediction of Corporate, Journal of Finance, Vol 23 (4). Pp. 589-609.

Altman, E. I. (2000). Predicting Financial distress of companies: Revisiting the Z-score and ZETA Models, Journal of Finance. Pp. 428-456.

Altman, E. I., Haldeman, R. G., & Narryanan, P. (1997). ZETA Analysis: A new model to identify the bankruptcy risk of corporations, Journal of banking and Finance, Vol 1 (1). Pp. 29-54.

Bharath, S. T., & Shumway, T., (2008). Forecasting default with the Merton model, Review of Financial Studies, Vol 23 (3). Pp. 1339-1369.

Campbell, J. Y., Hilscher, J., & Szilagyi, J. (2008). In Search of Distress Risk, Journal of Finance, Vol 63 (6). Pp. 2899-2939.

Chava, S., & Jarrow, S. (2004). Bankruptcy Prediction with Industry Effects, Review of Finance, Vol 8. Pp. 589-609.

Engelamann, B., Hyaden, E., & Tasche, D. (2003). Testing rating accuracy, RISK, Vol 16. Pp. 82- 86.

Evangelos, C. C. (2013). On the prediction of corporate financial distress in the light of the financial crisis: Empirical evidence from Greek listed Firms. Bank of Greece, Athens.

Hillegeist, S. A., keating, E. K., Cram, D. P., & Lundstedt, K. G. (2004). Assessing the Probability of Bankruptcy, Vol 9 (1). Pp. 5-34.

Hosmer and Lemeshow (2000) Applied logistic regression, Second edition. John Wiley & Sons, Inc., New York.

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

Ohlson, J. A. (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy, Journal of Accounting Research, Vol 19 (1). Pp.109-131.

Sannders, A., & Allen, L. (2002). Credit risk management: New approached to value at risk and other paradigms, Second edition. John Wiley & Sons, Inc., New York.

Shinong Wu and Xianyi Lu (2001) A Study of Models for Predicting Financial Distress in China' s Listed Companies, Vol6. Pp. 46-55.

Shumway, T. (2001). Forecasting Bankruptcy More Accurately: A Simple Hazard Model, Journal of Business, Vol 74 (1). Pp. 101-124.

William, H. B. (1996). Financial Ratios as Predictors of Failure. Journal of accounting research, Vol 4. Pp. 71-111.

Zingarelli Law Office, L. (2018). *Types of Bankruptcy: Chapter 7, Chapter 11 & Chapter 13*. [online] Zingarellilaw.com. Available at: https://www.zingarellilaw.com/resources/types-bankruptcy/ [Accessed 13 May 2018].

Appendix

Appendix 1: Results from analysis software

Logit regression from EViews 10 of Chava & Jarrow's model (2004)with data from 1993 to 2013

Dependent Variable: Z
Method: ML - Binary Logit (Newton-Raphson / Marquardt steps)
Date: 05/11/18 Time: 14:03
Sample: 1 888
Included observations: 888
Convergence achieved after 6 iterations
Coefficient covariance computed using observed Hessian

Variable	Coefficient	Std. Error	z-Statistic	Prob.
TLTA	2.094343	0.320604	6.532498	0.0000
SIGMA	8.995582	2.889894	3.112772	0.0019
NITA	1.127231	0.604521	1.864667	0.0622
EXRET	-0.058428	0.552132	-0.105822	0.9157
RESIZE	-1.004873	0.118988	-8.445126	0.0000
C	-13.96122	1.378548	-10.12748	0.0000

McFadden R-squared	0.199329	Mean dependent var	0.189189
S.D. dependent var	0.391880	S.E. of regression	0.351332
Akaike info criterion	0.790236	Sum squared resid	108.8691
Schwarz criterion	0.822594	Log likelihood	-344.8650
Hannan-Quinn criter.	0.802605	Deviance	689.7299
Restr. deviance	861.4402	Restr. log likelihood	-430.7201
LR statistic	171.7103	Avg. log likelihood	-0.388361
Prob(LR statistic)	0.000000		

Obs with Dep=0	720	Total obs	888
Obs with Dep=1	168		

Logit regression from EViews 10 of Campbell et al's model (2008) with data from 1993 to 2013

Dependent Variable: Z
Method: ML - Binary Logit (Newton-Raphson / Marquardt steps)
Date: 05/11/18 Time: 14:09
Sample: 1 888
Included observations: 888
Convergence achieved after 6 iterations
Coefficient covariance computed using observed Hessian

Variable	Coefficient	Std. Error	z-Statistic	Prob.
TLMTA	0.857995	0.407466	2.105684	0.0352
SIGMA	-0.657924	3.132291	-0.210046	0.8336
RESIZE	-0.650490	0.151060	-4.306158	0.0000
PRICE	-2.049720	0.316001	-6.486435	0.0000
NIMTA	0.434343	1.268623	0.342374	0.7321
MB	0.136404	0.041381	3.296291	0.0010
EXRET	0.394755	0.581215	0.679189	0.4970
CASHMTA	-2.409244	0.476738	-5.053599	0.0000
C	-6.964596	1.864350	-3.735670	0.0002

McFadden R-squared	0.277175	Mean dependent var	0.189189
S.D. dependent var	0.391880	S.E. of regression	0.335180
Akaike info criterion	0.721476	Sum squared resid	98.75202
Schwarz criterion	0.770013	Log likelihood	-311.3354
Hannan-Quinn criter.	0.740029	Deviance	622.6708
Restr. deviance	861.4402	Restr. log likelihood	-430.7201
LR statistic	238.7694	Avg. log likelihood	-0.350603
Prob(LR statistic)	0.000000		
Obs with Dep=0	720	Total obs	888
Obs with Dep=1	168		

Logit regression from EViews 10 of Chava & Jarrow's model (2004) with data from 1993 to 2003

Method: ML - Binary Logit (Newton-Raphson / Marquardt steps)
Date: 05/12/18 Time: 13:23
Sample: 1 642
Included observations: 642
Convergence achieved after 7 iterations
Coefficient covariance computed using observed Hessian

Variable	Coefficient	Std. Error	z-Statistic	Prob.
TLTA	2.431188	0.373966	6.501087	0.0000
SIGMA	5.489656	3.293305	1.666914	0.0955
RESIZE	-1.262071	0.163603	-7.714252	0.0000
NITA	0.692481	0.666836	1.038459	0.2991
EXRET	0.616832	0.601352	1.025742	0.3050
C	-16.44148	1.817081	-9.048291	0.0000

McFadden R-squared	0.241552	Mean dependent var	0.202492
S.D. dependent var	0.402170	S.E. of regression	0.351777
Akaike info criterion	0.782961	Sum squared resid	78.70300
Schwarz criterion	0.824686	Log likelihood	-245.3305
Hannan-Quinn criter.	0.799154	Deviance	490.6610
Restr. deviance	646.9280	Restr. log likelihood	-323.4640
LR statistic	156.2670	Avg. log likelihood	-0.382135
Prob(LR statistic)	0.000000		
Obs with Dep=0	512	Total obs	642
Obs with Dep=1	130		

Logit regression from EViews 10 of Campbell et al's model (2008) with data from 1993 to 2003

Dependent Variable: Z
Method: ML - Binary Logit (Newton-Raphson / Marquardt steps)
Date: 05/12/18 Time: 14:13
Sample: 1 642
Included observations: 642
Convergence achieved after 6 iterations
Coefficient covariance computed using observed Hessian

Variable	Coefficient	Std. Error	z-Statistic	Prob.
TLMTA	0.754229	0.455229	1.656811	0.0976
SIGMA	0.560908	3.546746	0.158147	0.8743
RESIZE	-1.037925	0.225739	-4.597893	0.0000
PRICE	-1.496617	0.409906	-3.651124	0.0003
NIMTA	-0.664149	1.653957	-0.401551	0.6880
MB	0.208199	0.056116	3.710174	0.0002
EXRET	0.371803	0.620453	0.599245	0.5490
CASHMTA	-3.376003	0.704085	-4.794883	0.0000
C	-11.44506	2.716029	-4.213896	0.0000

McFadden R-squared	0.298654	Mean dependent var	0.202492
S.D. dependent var	0.402170	S.E. of regression	0.335001
Akaike info criterion	0.734767	Sum squared resid	71.03877
Schwarz criterion	0.797355	Log likelihood	-226.8603
Hannan-Quinn criter.	0.759057	Deviance	453.7207
Restr. deviance	646.9280	Restr. log likelihood	-323.4640
LR statistic	193.2073	Avg. log likelihood	-0.353365
Prob(LR statistic)	0.000000		

Obs with Dep=0	512	Total obs	642
Obs with Dep=1	130		

Appendix 2: Additional statistics and instructions

Bankruptcy Statistics by SIC code

SIC Code	Industry name	Number of bankruptcy firm	Percentage of total bankruptcy firm
0100-0999	Agriculture, Forestry and Fishing	0	0%
1000-1499	Mining	0	0%
1500-1799	Construction	0	0%
1800-1999	Not used	0	0%
2000-3999	Manufacturing	9	20%
4000-4999	Transportation, Communications, Electric, Gas and Sanitary service	9	20%
5000-5199	Wholesale Trade	0	0%
5200-5999	Retail Trade	7	16%
6000-6799	Finance, Insurance and Real Estate	16	36%
7000-8999	Services	4	9%
9100-9729	Public Administration	0	0%
9900-9999	Nonclassifiable	0	0%

Calculation of true positive rate and false positive rate in ROC curve

$$\text{TPR (true positive rate)} = TP/(TP + FN)$$

$$\text{FPR (false positive rate)} = FP/(FP + TN)$$

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="2">Fact</td></tr>
<tr><td>Bankruptcy</td><td>Non-Bankruptcy</td></tr>
<tr><td rowspan="2">Forecast</td><td>Bankruptcy</td><td>TP</td><td>FP</td></tr>
<tr><td>Non-Bankruptcy</td><td>FN</td><td>TN</td></tr>
</table>

TP (Ture positive) = company is forecasted to be bankrupt and has been bankrupt in reality

FP (False positive) = company is forecasted to be bankrupt but not in actual

FN (False negative) = company is forecasted to be non-bankrupt but not in actual

TN (Ture negative) =company is forecasted to be non-bankrupt and it is not bankruptcy in actual