The performance of insolvency prediction and credit risk models in the UK: A comparative study, development and wider application.

Submitted by Anthony Paul Wood to the University of Exeter

as a thesis for the degree of

Doctor of Philosophy in Accountancy

In September 2012

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signature: …………………………………………………………..

ABSTRACT

Contingent claims models have recently been applied to the field of corporate insolvency prediction in an attempt to provide the art with a theoretical methodology that has been lacking in the past. Limited studies have been carried out in order to empirically compare the performance of these "market" models with that of their accounting number-based counterparts. This thesis contributes to the literature in several ways: The thesis traces the evolution of the art of corporate insolvency prediction from its inception through to the present day, combining key developments and methodologies into a single document of reference. I use receiver operating characteristic curves and tests of economic value to assess the efficacy of sixteen models, carefully selected to represent key moments in the evolution of the art, and tested upon, for the first time, post-IFRS UK data. The variability of model efficacy is also measured for the first time, using Monte Carlo simulation upon 10,000 randomly generated training and validation samples from a dataset consisting of over 12,000 firm-year observations. The results provide insights into the distribution of model accuracy as a result of sample selection, which is something which has not appeared in the literature prior to this study. I find overall that the efficacy of the models is generally less than that reported in the prior literature; but that the theoretically driven, market-based models outperform models which use accounting numbers; the latter showing a relatively larger efficacy distribution. Furthermore, I obtain the counter-intuitive finding that predictions based on a single ratio can be as efficient as those which are based on models which are far more complicated – in terms of variable variety and mathematical construction. Finally, I develop and test a naïve version of the down-and-out-call barrier option model for insolvency prediction and find that, despite its simple formulation, it performs favourably compared alongside other market-based models.

The author would like to express his love and appreciation to his beloved family; Beccy, Tara, and Matthew, for their understanding and boundless love, throughout the duration of his study. The author also wishes to express his gratitude to his supervisor, Prof. Richard Jackson who has offered invaluable assistance, support and guidance throughout this process.

Table of Contents

List of Tables

List of Figures

Abbreviations

AR. Administrative Receivership

AZ. Altman Z-score

AZN. Altman Z-score variables run through a neural network

AZU. Altman Z-score with updated coefficients

B/M. Book to Market

BM. Brownian Motion

BS. Bharath and Shumway's Naïve European Call option

BV. Beaver's best performing ratio – cash-flow / total debt

CA. Current Assets

CGC. UK Corporate Governance Code

CF. Cash-Flow

CDF. Cumulative Density Function

CFTL. Cash-Flow divided by Total Liabilities

CHIN. Change in Net Income

CL. Current Liabilities

CLCA. Current Liabilities divided by Current Assets

CPM. Conditional Probability Model

CR. Current Ratio (current assets / current liabilities)

CVA. Company Voluntary Arrangement

CVL. Creditors Voluntary Liquidation

DOC. Down-and-Out Call option

EBIT. Earnings Before Interest and Tax

EBITTA. Earnings Before Interest and Tax divided by Total Assets

EU. European Union

FRC. Financial Reporting Council

FUTL. Funds provided by operations divided by Total Liabilities

GCQ. Going Concern Qualification

GDP. Gross Domestic Product

GNP. Gross National Product

HKCL. Hillegeist, Keating, Cram, and Lundstedt European Call option model

IMF. International Monetary Fund

IR. Income Ratio (net income / working capital)

INTWO. One if NI was negative for the last two years, zero otherwise

JW_DOC. Jackson and Wood Naïve DOC model

LF. Likelihood Function

LGD. Loss Given Default

LLF. Log Likelihood Function

LMDA. Linear Multivariate (or Multiple) Discriminant Analysis

LDF. Logistic Distribution Function

LPM. Linear Probability Model

LSE. London Stock Exchange

LSPD. London Share Price Database

MDA. Multivariate (or Multiple) Discriminant Analysis

ML. Maximum Likelihood

MVL. Members Voluntary Liquidation

NCI. No-Credit Interval

NI. Net Income

NITA. Net Income divided by Total Assets

NLV. Net Liquidation Value

NN. Neural Network

OENEG. Dummy variable. One if total liabilities exceeds total assets, zero otherwise

OL. Ohlson's logit model

OLN. Ohlson's logit model variables run through a neural network

OLS. Ordinary Least Squares

OLU. Ohlson's logit model with updated coefficients

QMDA. Quadratic Multivariate (or Multiple) Discriminant Analysis

QR. Quick Ratio (cash or cash equivalents / current liabilities)

RETA. Retained Earnings divided by Total Assets

RMA. Robert Morris Associates

ROC. Receiver Operating Characteristic

ROI. Return on Investment

SBA. Small Business Associates

SEC. The Securities and Exchange Commission

SIZE. Deflated Market Value of Equity

STA. Sales divided by Total assets

TA. Total Assets

TL. Total Liabilities

TLTA. Total Liabilities divided by Total Assets

TZ. Taffler Z-score

TZN. Taffler Z-score variables run through a neural network

TZU. Taffler Z-score with updated coefficients

UK. United Kingdom (of Great Britain and Northern Ireland)

US. United States (of America)

VE. Market value of Equity

VETL. Market value of Equity divided by Total Liabilities

WCTA. Working Capital divided by Total Assets

WRDS. Wharton Research Data Service

1 Introduction

The prediction of financial distress and corporate insolvency and the assessment of credit risk have been the subject of much academic and professional research over the last half century (e.g. Beaver, 1966; Altman, 1968; Ohlson, 1980; Taffler, 1984; Coats and Fant, 1993; Charitou et al, 2004; Reisz and Purlich, 2007; Bharath and Shumway, 2008; and Agarwal and Taffler, 2008). As when a pebble is thrown into a lake, the ripples or shockwaves reach far beyond that of the initial impact. If a company becomes financially distressed or insolvent, there are adverse consequences for its diverse stakeholder groups – investors, managers, employees, customer, suppliers, etc. – which impact onward into other firms, the wider economy and society. According to the president of R3 (the UK association of business recovery professionals)[1] Steven Law (2010):

“Any future increases in corporate insolvencies are likely to affect others as Insolvency Practitioners estimate that around 27% of corporate insolvencies are triggered by another company's insolvency - the ‘domino effect’.”

In a similar vein, a member of the Turnaround Management Association UK asserts in Cooper (2010) that:

“As business struggle to survive into 2010, they are likely to put increasing pressure on their suppliers. Payments will be withheld for as long as possible. If and when a company fails, it is likely that the other businesses it owes money to will get little or nothing in return. Unfortunately, the knock on effect will be that other firms will also be starved of cash and more will find themselves under financial pressure.”

The area of distress/insolvency prediction is of high economic significance, in terms of the number of firms and individuals affected; and the implications for investment and lending decision making. Studies such as Blochlinger and Leippold (2006) estimate that the loss given default (LGD) approximates to 40% of the investment, suggesting that on average only 60% of the value of loans are recovered upon default. Agarwal and Taffler (2008) use an LGD of 45% under the assumption that lending institutions follow the

[1] Termed “R3” due to their slogan; “Rescue, Recovery, Renewal”.

Basel II Foundation Internal Ratings-based approach, and that all loans are unsecured senior debt. This latter approach makes more intuitive sense as its successor, Basel III, is set to be introduced from 2013 with wider international adoption and inheriting much of the Basel II capital adequacy and LGD estimates. Under Basel II we are able to estimate, therefore, that the amount of investment which is able to be recovered given default or insolvency is approximately 55%.

Up-to-date research is mandated by the present era of global financial difficulty which follows recent periods of financial sector institutional failure and credit crunch. Figure 1.1 shows the number of UK insolvencies annually since 1960. The glut of insolvencies during the economic downturn of the early 1990s is clearly visible; as is a sharp rise in the incidence of insolvencies since 2007.[2]

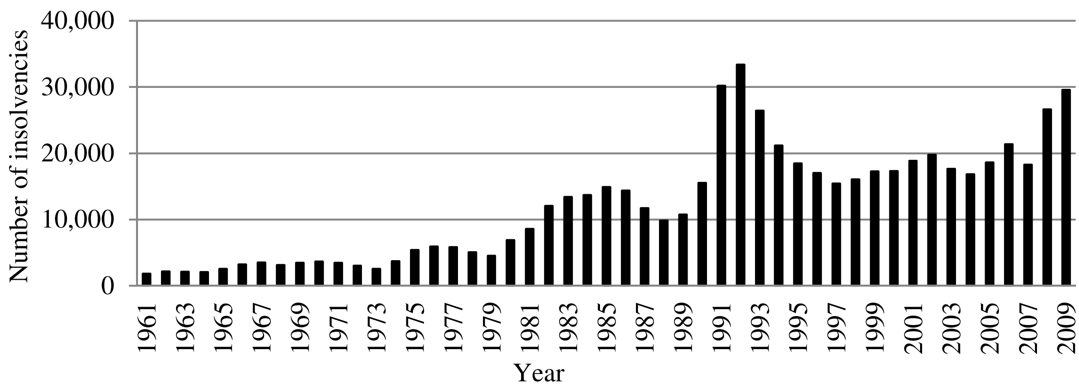


Figure 1.1: Total number of company insolvencies in the UK per year

Source: www.insolvency.gov.uk

As regards the current prognosis for the incidence of insolvencies over the coming period, R3 (2010) has the following sober message:

"We would still expect a spike in the number of insolvencies in the five or six quarters following a recession because many businesses that suffered during the recession find it hard to borrow as lending requirements tighten. Many of them will see 'green shoots' but will not be able to fund expansion, especially if interest rates increase. And when a recession ends and assets rise in value, creditors are encouraged to move ahead with more aggressive debt collection."

[2] A more detailed evaluation of the number and rate of insolvencies can be found within Section 4 of this thesis.

This thesis evaluates a number of different methods which have been popularly employed in the prior literature to assess firm health and some more recent extant approaches attributable to credit rating agencies. In the current economic climate of global financial turmoil, I seek to assess, using data from recent insolvency cases and post-IFRS implementation, how these various methods and approaches perform for the UK. As a result, this thesis will be able to determine the most accurate methods available to predict insolvency and assess credit risk, aid financial institutions in their lending decisions by highlighting weaknesses in their existing methods and, as a result, forgoing losses attributable to default.

The comparison of corporate insolvency prediction models is by no means a novel idea, and is something which is widespread throughout the literature. Zavgren (1983) provides a discussion regarding the "state of the art", describing and contrasting several different modelling approaches. Similarly Scott (1981) discussed a variety of models attempting to find a theoretical rational for the use of some of the more popular variables found within multivariate insolvency prediction models. Jones (1987) and Balcaen and Ooghe (2004) also discuss a variety of different modelling methodologies, comparing and contrasting their benefits and weaknesses. Although these papers compare and contrast many different models between them, what I consider lacking from the literature is an independent empirical analysis of modelling techniques, not simply a discussion of their associated strengths and weaknesses.

The empirical analysis of insolvency prediction and credit risk models is also nothing new in itself. The general trend within the literature is geared towards producing a new model, which is generally purported to provide a higher degree of accuracy than any model before it (a poorly performing model would of course not warrant inclusion in any academic journal therefore, as a general rule, only models with high accuracy claims are normally found in the literature). To demonstrate these claims, the new model is often pitted against a popular predecessor, with the new model usually coming out on top with regards to the associated accuracy tests. For example, Altman et al (1977) compared the ZETA model with that of the original Altman (1968) Z-score, the ZETA model showing improved accuracy; Reisz and Purlich (2007) compared contingent claims models again with the Altman Z-score and show that in all but one instance, the contingent claims models are superior in accuracy; and Agarwal and

Taffler (2008) compared the Taffler Z-score to two contingent claims models and suggest that it is the Taffler Z-score which is superior.

Other papers empirically compare modelling technologies. Salchenberger et al. (1992) for example, compared neural networks to logistic regression and found that neural networks out-performed logistic regression considerably with regards to their classification accuracy. A comparison of neural networks and discriminant analysis was conducted by Altman et al (1994). Their study did not definitively demonstrate a clear winner when it came to classification accuracies; but it was surprisingly discriminant analysis which yielded marginally better results – contrary to the findings of prior studies such as Odom and Sharda (1990).

These studies and many others are described in detail within the literature review of chapter 5 which traces and describes the evolution of the literature to date. What is abundantly clear from the literature review is that there is no general consensus as to what constitutes the best approaches/methods/and indeed variables which should be used to predict insolvencies or determine credit risk.

The thesis contributes to the insolvency prediction and credit risk literature in several ways: Firstly by independently considering a wider variety of models for empirical evaluation than has been seen prior to this study, I am able to evaluate the accuracy claims of a large number of models and techniques; Secondly, developing several new methods for testing model accuracy (a Monte Carlo approach to determine the distribution of model accuracy over multiple sample splits, and an elimination test based upon economic value), this thesis provides an in-depth descriptive and empirical analysis of the evolution of insolvency and credit risk modelling from its inception through to the present day. The thesis also presents two new models; the first of these models forms a natural extension to the existing credit risk literature, based upon sound theoretical arguments and well defined formulae; the second model is derived in an arbitrary manner in order to demonstrate some of the inherent problems associated with the more traditional statistically-based modelling techniques. The latter model shows that given an arbitrary choice of variables and modelling technology, a model can easily be constructed to fit the data and out-perform extant models from the literature. The inherent problem with this arbitrary approach, as will be discussed in length during this thesis, cannot (and will not) guarantee similar success in alternate or subsequent datasets. Therefore as a generalisation, I show that it is the models which rely upon a sound theoretical underpinning which ultimately, over time, outperform those which do not.

The timing of this thesis encompasses the current financial crisis which provides an opportunity to empirically evaluate extant models under the conditions of economic and financial turmoil. If these models are adequately able to distinguish between healthy and non-health firms in a period of financial instability, then this can act as reassurance to their various user groups, particularly to lenders within a loan decision process, as to the robustness of these models over time. The timeliness of this study therefore provides an additional robustness factor with regards to model evaluation which would otherwise be lacking in times of economic stability.

The current global financial crisis is the most significant financial meltdown the world has seen in a generation and is arguably only eclipsed by the "great depression" born out of the 1929 Wall Street crash. The current crisis stems from the realisation of problems within the sub-prime mortgage lending markets of the United States in April 2007. Sub-prime mortgages were securitised by banking institutions and traded globally. When customers began defaulting on these mortgages, banks were forced to increase interest rates on mortgages for those customers who had not defaulted in a bid to cover their losses. Consequently more mortgage customers began defaulting, snowballing unsustainably, which ultimately led to these sub-prime securities losing the majority of their market value. With many financial institutions holding securities of this type, the result was a reluctance to lend within the financial sector. The severity of the problem was bought to public attention some months after the sub-prime lending revelations. On the 9th of August 2007, investment bank BNP Paribas announced a "complete evaporation of liquidity" in the financial market [3]. The ensuing shortage of inter-bank borrowing saw a number of high profile UK banking institutions experience financial difficulties eventually resulting in government intervention at Northern Rock, The Royal Bank of Scotland and Bradford and Bingley, for example. Despite the Bank of England cutting the base rate gradually to just half of one percent, the lowest level in its 315 year history, the UK economy officially went into recession on 23rd January

[3] Source: http://www.bloomberg.com/apps/news?pid=20601087&refer=worldwide&sid=aNIJ.UO9Pzxw.

2009 when the government reported a second consecutive quarter of negative economic growth[4] and subsequently entered a double-dip recession after the first quarter of 2012.

It is not for this thesis to provide a detailed analysis of the causes of the financial crisis but a brief discussion of some key areas is required to place the overall importance of this study into context. It is suggested that many institutional failures, particularly in the financial sector, have the true causes of their problems rooted deep in their governance policies. The failure of many such companies, it is posited, has come about because of poor risk management, a failure to detect and mitigate such risks and excessive remuneration for directors[5]. The UK has in place the UK Corporate Governance Code (CGC) which through its various guises has attempted to provide sufficient guidance to manage and mitigate the risks of large companies for the last seventeen years and it has been suggested that the governing bodies and companies alike have been sufficiently happy with the content of the code. [6]

"When the FRC reviewed the implementation of the 2003 Code in the second half of 2005, we found that it was generally felt to be bedding down well. There was no demand for major changes; rather, there was a widespread desire for a period of stability". Sir Christopher Hogg, Chairman of the FRC (June 2006).[7]

It may be said that although these guidelines are in place, albeit on a "comply or explain" approach, many companies may simply treat their governance requirements a "box-ticking" exercise rather than a serious attempt to properly understand the real risks that they might face in the future, as suggested by Cassidy (2003) for example. This can be seen in the case of Northern Rock, whose annual report was fully compliant with the CGC and full of facts and figures indicating the company was experiencing increased growth and profitability. The corporate governance section of the annual report gave details of the work undertaken by the audit committee, risk committee and internal control committee. The risk committee, which met four times, reviewed "the key risks inherent in the business" and concluded;

[4] Source: http://news.bbc.co.uk/1/hi/business/7846266.stm.
[5] See for example the "no free lunch" column of the Investors Chronicle by Alistair Blair http://www.investorschronicle.co.uk/Columnists/NoFreeLunch.
[6] The current 2012 version of the code has evolved from earlier work by the Cadbury committee (1992) and Greenbury committee (1995) for example and is updated and reviewed regularly (e.g. Higgs, 2003; Smith, 2003; Turnbull, 2005). The latest version can be found here: http://www.frc.org.uk/Our-Work/Codes-Standards/Corporate-governance/UK-Corporate-Governance-Code.aspx.
[7] http://www.frc.org.uk/press/pub1130.html.

"Northern Rock is a well-controlled, risk-averse business that continues to adopt a prudent stance in the management of risk". Northern Rock Annual Report (2006).[8]

Whether the CGC requirements were treated simply as a box-ticking exercise or the management really failed to identify a major flaw in their risk management operations, the impression that the board reports is not very dissimilar to the thousands of reports published throughout the UK every year.

"...it is an example of how a company can have its corporate governance gleaming like a Rolls-Royce but have the satnav set on a course for disaster". (Jamieson 2008).

It is not just the management who should accept blame for this failure. Why did nobody else identify this major risk? The investors, auditors, Financial Services Authority (FSA), government and the Bank of England all failed to identify the problems with Northern Rock and indeed, with many other companies which have suffered recent financial distress. A subsequent report by the FSA (2008) suggests that the overall supervision of Northern Rock by the tripartite[9] was not in line with its risk profile, it was initially looked after by the insurance group department before being transferred several times, which led to the admittance by the tripartite supervisory regime that there was an overall "lack of knowledge of Northern Rock". (Financial Services Authority - Internal Audit Division, 2008)

In light of this financial crisis through cases like Northern Rock and others, whether down to poor governance, risk management or simply the insolvency ripples resonating through the economy, many UK companies are facing financial difficulties today.[10]

UK lending institutions have for some time, since the onset of the current financial crisis, and like their counterparts around the globe, been urged to re-expand their

[8] http://companyinfo.northernrock.co.uk/downloads/results/res2006PR_AnnualReportAndAccounts.pdf p.16.
[9] Bank of England, The Treasury and the FSA make up the tripartite that regulates and supervises the banking sector in the UK. At the time of writing this thesis, the government intends to abolish the tripartite system. This will result in the FSA ceasing to exist in its current form, and the establishment of three new regulatory bodies in early 2013: the Financial Policy Committee (FPC), the Prudential Regulation Authority (PRA) and the Financial Conduct Authority (FCA). The PRA and the FCA will be responsible for the majority of the functions currently performed by the FSA.
[10] For example the retail sector has shown a 91% increase in firms facing financial difficulties. http://www.retail-week.com/retail-sectors/number-of-retailers-facing-financial-difficulties-soars/5004495.article

willingness to lend in order to alleviate the current credit crunch and promote economic growth;[11] and under Project Merlin, as announced on 9th February 2011 by George Osbourne, UK Chancellor of the Exchequer, the four largest UK lending banks agreed, inter alia, to lend £190 billion during 2011 – with the Bank of England to monitor achievement of lending targets.

In this regard it is apparent that extreme care should be taken in relying, wholly or in part, on information obtained from a firm health-determining model for the purposes of financial or economic decision making particularly with regards to the perceived accuracy of insolvency prediction and credit risk models. My study provides a stark warning to lending institutions seeking to re-expand their lending portfolios as to the technologies they might employ to assess insolvency and credit default risk, a key aspect of lending. The literature shows that many insolvency prediction and credit risk models have shown good *ex post* efficacy in past studies; my findings suggest that many of these models (which are largely statistically-based) prove to have a more modest ability to predict insolvency or financial distress in the current economic downturn, although alternate models based upon sound theoretical underpinnings perform relatively well. It is clear therefore that it is the latter type of modelling technique which should be adopted or revised by the lending institutions in order to provide a sound basis for future lending and, ultimately, economic recovery.

This thesis continues as follows: section 2 discusses the history of credit, interest and debt, and describes the evolution of the credit rating institution in order to contextualise this thesis; section 3 defines what constitutes failure, bankruptcy and insolvency and section 4 discusses in detail the formal avenues available to individuals, private and public companies along with any potential avenues for research; section 5 provides an in-depth literature review, tracing and providing discussion upon the evolution of the art of insolvency prediction and credit risk modelling; section 6 details the objectives and research questions underlying this thesis; section 7 describes selection and development of a sample of firms and firm-level data against which to test the different models; section 8 provides an overview of the methods and models tested within my study along with the development of the proposed new methods by which to compare the models,

[11] UK Chancellor of the Exchequer, 20th May 2010: 'There is an urgent priority that is getting lending going to small- and medium-sized businesses. That is an absolute urgent priority". http://uk.reuters.com/article/marketsNewsUS/idUKGOVT20MAY20100520.

and section 9 present the results of the comparison along with an evaluation of the distributional properties of the model accuracies; section 10 provides an extension to the research, namely an attempt to construct a “best of breed” model based upon the results obtained in section. Section 11 integrates the results and provides concluding remarks along with recommendations for future research.

2 History and context of this study

2.1 The origins and evolution of credit, interest and debt

Since the first concepts of trade, commerce and monetary values became common place in society, individuals have been forced, through necessity, to come to terms with credit, interest and debt along with their inevitable associated problems. It would be somewhat foolish to believe that the various financial crises which have transpired over the last one hundred years are a new occurrence, a direct result of the industrial revolution and the rise of the capitalist economy. In fact, history tells us that the struggle with credit and debt can be traced back several thousands of years with the earliest recorded documentation suggesting that the human race has faced much more difficult periods of financial woe than that which we see today.

Historians suggest that systems of credit and debt were in place as much as 5,000 years ago, long before the earliest evidence for the written word. At this time, a system of counting tokens, possibly an early form of money, was used to keep track of various trades and obligations. Not only tokens were used as currency, but cattle, gold and grain were given values and used for trade. It is believed that the art of writing came about as a necessity to record and understand these financial dealings. [12]

The earliest known recordings of financial transactions were written in clay tablets around 3,000 BC in the city of Uruk, now Warka, which is situated south of Baghdad, Iraq. The tablets recorded information regarding who owed what to whom, contracts, and fiscal records.

The concept of interest, the idea that a lender should be returned an amount greater than the initial loan value, is thought to have originated with the Sumerians around 2000-3000 BC. The concept arose with the lending of cattle which were the primary source of currency at that time. If a herd of cattle was lent to an individual for a period of say, one year, then the lender would expect to receive more cattle back at the end of

[12] Several insights and examples are taken from the Code of Hammurabi (1792-1750BC) of which reprints are readily available through library and online resources. Other insights can be gained from Davies (2002) and Van De Mieroop (2004).

the period as the herd multiplies. This idea of "giving birth" to your assets transferred into other commodities where equal recompense would be expected.

The first known example of regulatory systems put in place to control interest, credit, debt and forgiveness were developed during the reign of Hammurabi in ancient Babylonia (1792-1750 BC). Interest was seldom charged on any form of cash advance, providing that the loan was paid back as agreed. Failure to pay back the debt as per the terms of the contract resulted in high interest rates or worse:

"If anyone owes a debt for a loan, and a storm prostrates the grain, or the harvest fail, or the grain does not grow for lack of water; in that year he need not give his creditor any grain, he washes his debt-tablet in water and pays no rent for this year.

If any one fails to meet a claim for debt, and sell himself, his wife, his son, and daughter for money or gives them away to forced labor: they shall work for three years in the house of the man who bought them, or the proprietor, and in the fourth year they shall be set free.

If any one fails to meet a claim for debt, and he sells the maid servant who has borne him children, for money, the money which the merchant has paid shall be repaid to him by the owner of the slave and she shall be freed.

If a inn-keeper gives 60 liters of beer on credit she will receive 50 liters of barley at harvest-time". (The code of Hammurabi)

Some 1,500 years after Hammurabi we have the first known example of an interest agreement written in 172 BC:

"Theokles, son of Euboulides, Macedonian of the epigone, has lent to Aristokles, son of Herakleides,Corinthian of the epigone, 3 talents 780 drachmai of bronze money produced to view, for 13 months from the above date, with interest at 2 drachmai the mina per month. Aristokles shall repay to Theokles the aforestated loan, namely 3 talents 780 drachmai in bronze money, plus the interest accruing thereon, in the month of Apellaios, or in the Egyptian calendar, Pauni, in the tenth year.

If he does not repay it, he shall pay as penalty the loan plus fifty per cent and simple interest, and Aristokles shall have the right of execution upon Aristokles or his surety, *whichever of them he may choose, or upon both, as if by court decision"*. (Cited in Davies, 2002)

These early credit lending examples paved the way for banking institutions such as the Greek Bank of Delos (323-30 BC) to evolve as a direct result of the desire for credit. Delos's influence was to replace physical cash transactions with credit receipts and payments made on simple instructions with accounts kept for each client. The Model employed by the Bank of Delos was adopted by the Roman Empire throughout their territorial reign over much of Europe, Africa and Asia between 27 BC and 476 AD. The importance of the credit granting banking system, however, was limited during this period due to the Roman preference for cash transactions with coins. The fall of the Roman Empire meant that banking was all but forgotten and had to be re-invented much later.

The re-invention of banking occurred during the time of the crusades, approximately 1100-1300 AD.

"The Crusades gave a great stimulus to banking because payments for supplies, equipment, allies, ransoms etc. required safe and speedy means of transferring vast resources of cash. Consequently the Knights of the Temple and the Hospitallers began to provide some banking services such as those already being developed in some of the Italian city states... In Italian city states such as Rome, Venice and Genoa, and in the fairs of medieval France, the need to transfer sums of money for trading purposes led to the development of financial services including bills of exchange. Although it is possible that such bills had been used by the Arabs in the eighth century and the Jews in the tenth, the first for which definite evidence exists was a contract issued in Genoa in 1156 to enable two brothers who had borrowed 115 Genoese pounds to reimburse the bank's agents in Constantinople by paying them 460 bezants one month after their arrival". (Davies 2002).

The development of banks and financial growth of these northern Italian cities throughout the 14th, 15th and 16th centuries evolved into what we might term the "modern" banking system. The most famous and largest of the early banks is the Medici bank, established in 1397. The Medici banking system re-invented and overhauled the profession of accounting making notable contributions towards the improvement of ledger control by developing and improving the double-entry book keeping system in order to keep track of credit and debits more effectively.

The 17th century saw an explosion of banking, finance and commerce, centred on the lucrative East Indian trade companies of Amsterdam, London and Hamburg. Individuals could purchase bills of credit, stock, which meant that for the first time both banks and companies, were being owned by the public as opposed to a few wealthy individuals. By the 18th Century the publicly owned bank was commonplace, indeed, many of the banks we see today along with their credit policies were founded around this time.

The first example of a “cash-credit” system was introduced by the Royal Bank of Scotland upon its formation in 1727. Certain applicants could apply for loans to withdraw cash as they required, with interest being charged only on the amount which was withdrawn. This concept was adopted by other banks in Scotland, then England and throughout Europe and is the origin of the bank overdraft.

2.2 Development of the financial ratio and credit rating institutions

The history of insolvency prediction and credit assessment has its roots embedded within the concept of financial statement analysis. It could be proposed that financial statement/ratio analysis to a large extent *is* insolvency prediction, with the primary use of financial ratios being that of evaluating the performance of companies as a going concern, driven by the needs of both the lending and investing stakeholder groups. As far back as 1849, credit rating institutions such as Dun & Bradstreet have been providing an independent opinion of firm credit worthiness[13], although the ratings and rating agencies akin to the likes of which we see today did not begin to evolve until the early 20th century when financial statements became commonplace.

The establishment of the UK Companies Act in 1900 and the introduction of the compulsory balance sheet audit by non-insiders provided an increase in publicly available financial statements. The introduction of the Act was particularly necessary as the economic and industrial evolution seen in Western Europe was being sought by

[13] The first publication of commercial credit ratings was produced by the Bradstreet organisation in 1849.

America as it strived for industrial maturity which leads to a shift in economic power to financial institutions as more and more firms were now seeking financial backing in the form of credit. Commercial banks began to request financial statements for lending purposes as early as the 1870s; but this did not become a widespread practice until the 1890s. During the 1890s, the volume and flow of financial information increased greatly (Horrigan, 1968), driving the need for greater comparability and reliability of financial statements. The result of the new Companies Act requirements ensured that this was achieved, providing a greater volume of comparable and reliable information which could now be readily obtained and analysed by institutions, investors, academics and other interested parties.

The influence of this Act on the evolution of the credit rating sector can be demonstrated if I refer back to the conception and beginnings of several of the leading and best known credit rating agencies that are in existence today, each becoming prominent around the turn of the century.[14]

2.2.1 Fitch ratings

John Knowles Fitch founded the Fitch Publishing Company in 1913. Fitch published financial statistics for use in the investment industry via "The Fitch Stock and Bond Manual" and "The Fitch Bond Book." In 1924, Fitch introduced the AAA through D rating system that has become the basis for ratings throughout the industry. With plans to become a full-service global rating agency, in the late 1990s Fitch merged with IBCA of London, subsidiary of Fimalac, S.A., a French holding company. Fitch also acquired market competitors Thomson BankWatch and Duff & Phelps Credit Ratings Co.

2.2.2 Moody's Investors Service

John Moody and Company first published "Moody's Manual" in 1900. The manual published basic statistics and general information about stocks and bonds of various industries. From 1903 until the stock market crash of 1907, "Moody's Manual" was a national publication. In 1909 Moody began publishing "Moody's Analyses of Railroad Investments", which added analytical information about the value of securities.

[14] The proceeding short histories are taken from Finney (2009).

Expanding this idea led to the 1914 creation of Moody's Investors Service, this, in the following 10 years, would provide ratings for nearly all of the government bond markets at the time. By the 1970s Moody's was rating commercial paper and bank deposits, becoming the full-scale rating agency that it is today.

2.2.3 Standard & Poor's

Henry Varnum Poor first published the *History of Railroads and Canals in the United States* in 1860, the forerunner of securities analysis and reporting to be developed over the next century. Standard Statistics formed in 1906, which published corporate bond, sovereign debt and municipal bond ratings. Standard Statistics merged with Poor's Publishing in 1941 to form Standard and Poor's Corporation, which was acquired by The McGraw-Hill Companies, Inc. in 1966. Standard and Poor's has become best known by indexes such as the S&P 500, a stock market index that is both a tool for investor analysis and decision making, and a U.S. economic indicator.

2.3 The use of ratio analysis for credit assessment

The financial ratio and financial statement analysis have been integral to the evolution of credit assessment. It may even be stated that financial ratio analysis to a large extent *is* insolvency prediction, with the primary use of financial ratios being that of evaluating the performance of companies as a going concern, driven by the needs of both the lending and investing stakeholder groups. This proposition can be demonstrated by the conception of the financial ratio itself, noticing that the evolution of the ratio is closely tied in history to the evolution of the credit rating institution.

Although the primary cause of the evolution of ratio analysis in general was Euclid's rigorous analysis of the properties of ratios in *Book V* of his *Elements* in about 300 B.C., the adoption of ratios as a tool of financial statement analysis is a relatively recent development (Horrigan, 1968). During the latter decade of the 19th century, financial institutions would compare a firm’s current assets to that of its current liabilities, being the predominant factor in assessing a firm’s credit worthiness. The current ratio, as it was termed, became the first example of the modern financial ratio and a firm was considered in good health if its current assets outweighed its current liabilities 2:1. Although other ratios were developed around that time, the current ratio bore a long-lasting and significant impact upon financial statement analysis and credit assessment, indeed it is the current ratio above all other which was first cited in academic publication by Rosendale (1908) in his article “Credit Department Methods”.

The early 20th century saw several studies which investigated the use of different financial ratios for the determination of a firm’s financial position. Most noticeably a “Study of Credit Barometrics” by Wall (1919), examined a large sample of financial statements, now freely available due to regulatory requirements, and scrutinised seven different ratios. Wall determined that there were large variations in ratios between different geographical areas, and although his results would come under much criticism by modern standards of analysis; his study was historically significant because it was a widely-read, overt departure from the customary usage of a single ratio with an absolute criterion (Horrigan, 1968). Wall’s study sparked great interest in the use of the financial statement and ratios for firm analysis throughout the 1920s, notably Lincoln (1925) who discussed and illustrated forty different financial ratios, the largest number of ratios discussed during this period. It was, however, Wall himself who provides possibly the first example of combining financial ratios to create on overall picture of a firm’s health. Wall and Duning (1928) attempted to quench the explosion of ratio usage by developing a ratio index. The index was effectively a weighted combination of several different ratios with the weights being arbitrarily chosen by the authors resulting in a naive model approximating the first real linear multivariate discriminant function.

From this moment onward, the empirical evaluation of financial statements and their contained ratios became widespread, driven by the need of financial institutions and investors to better understand and assess firm health and credit worthiness, embedded in the foundations of the capitalist economies that we see in the present day.

This section has provided an introduction to the evolution, conception and foundations of insolvency prediction and credit assessment methodologies which we are familiar with today. The literature review of section 5 examines in detail the evolution of these models, methods, and technologies through the latter half of the 20th century through to the present, with respect to the particular research avenues that are available for this thesis. The field of insolvency/bankruptcy prediction, credit assessment (company and individual), detecting financial distress, academic/institutional failure research, call it what you will, is vast. One could spend a lifetime empirically researching all of the avenues potentially available for study. Section 4 investigates some of these avenues of possible research in order to determine which particular areas provide the possibility and/or warrant further empirical work. Before this discussion however, an introduction must be made of some of the terms and generalisations that will be used throughout the remainder of this thesis.

3 Bankruptcy, insolvency and financial distress

3.1 Bankruptcy and insolvency

In this thesis I will be using legislative terminology predominately from a UK perspective and the data for my empirical analysis will be from UK companies. Many academic papers to which I refer are from US authors who use the term "bankruptcy" whereas I will refer to the process as "insolvency". The difference in these terms is born through legal differences. In the United States, where the majority of insolvency prediction literature originates, the term "bankruptcy" refers to the legal insolvency process used for both companies and individuals. In the UK, "bankruptcy" is the process for individuals only; companies in the UK will enter one of several legal insolvency processes which are discussed in section 4.1.[15]

3.2 Financial distress

Financial distress is a term which is used to describe a firm which is suffering from short-term financial difficulties. If the "symptom" of financial distress is not "treated" effectively and efficiently, then this can often lead to the onset of insolvency proceedings; although financial distress can often be detected many months before any formal prescription or legal action need be sought, and therefore one might say that with appropriate intervention insolvency can often be avoided altogether. In accounting and financial terms, financial distress could be thought of as being one of the following scenarios.

- Company operating returns are lower than its cost of capital
- Costs are greater than revenue
- Actual returns are less than expected returns
- Outflows are greater than inflows

[15] Although I have stated that my analysis, explanations and thesis will be focusing on UK, the legal distinctions need to be classified and compounded even further as Scotland has its own form of legalities when dealing with insolvency. In fact I will be referring to the insolvency practices of England and Wales only throughout the remainder of this thesis but refer to them as UK practices for convenience.

These scenarios are unlikely individually to cause the company to fail and through adequate management skills these problems more often than not can be addressed before they become out of control and irreversibly lead to legal insolvency process. Traditionally, insolvency prediction models aim to pick up signs of these financial distress signals contained within past accounting data and assess the likelihood of the distressed company becoming formally insolvent.

A theoretical explanation of distress perhaps originates with Walter (1957) who perceived a company to be akin to a reservoir of liquid assets (or working capital) with this reservoir being supplied by monetary inflows and drained by its outflows. He describes financial distress as the exhaustion of this liquid asset reservoir which acts as a cushion or bolster, dampening the impact of variations in monetary flows. Taffler (1983) suggests that with all else being equal, the probability of failure is greater when:

- The size of the reservoir is smaller
- The inflow of funds from operations is smaller
- The claims on resources by creditors is larger
- The funds flow required for business operations is larger and
- The variation of inflows, outflows and claims on the company is higher

These ideas form the basis of much academic literature on insolvency prediction by Beaver (1966), Altman (1968) and Taffler (1983), for example, who, by using accounting data and financial ratios, have been able to detect stark differences between the financials of failed and non-failed companies.

3.3 Causes of financial distress

The R3 conducts regular surveys that consistently show that many insolvency proceedings could have been avoided if professional advice has been obtained earlier, and the longer a company delays seeking professional help the higher the chance that the business cannot be rescued.[16]

[16] For example the R3's Ostrich's guide to business survival (2002).

Economic failure has no real concrete definition but in economic terms, research by the R3 suggests the following common reasons why a company may become insolvent:[17]

- Loss of market: where companies have not recognised the need to change in a shrinking or changing marketplace, because their margins have been eroded or because their service has been overtaken technically (29%)
- Management failure to acquire adequate skills, either through training or buying them in, over-optimism in planning, imprudent accounting, lack of management information (22%)
- Bad debts (10%)
- Lack of working capital (20%)
- Other reasons including excessive overheads, new venture/expansion/acquisition are also attributed.

The decision of the management team is critical if any signs of financial distress are detected, but all too often there is a reluctance to seek professional help which may ultimately save the business. Foreman (2003) provides evidence that floored business plans and poor execution are the root causes of corporate failure given that traditional financial ratios concerning relative profitability, capital structure, and the power to finance growth internally almost completely explain which firms would fail within 2 years.

"*Early intervention can be a tough decision to make… Often the thought of taking advice from insolvency professionals is unpalatable to the management team as they fear that insolvency professionals are like vultures – circling prior to the terminal insolvency event. If insolvency is to be avoided, it is about taking appropriate advice as the earliest opportunity*". Cawkwell (2008).[18]

"*It appears that financial distress happens to companies whose operating decisions perform below expectations. Bankruptcy, in contrast, seems to be a result of a decision*

[17] Figures taken from Ostrich's guide p.17. (Percentage of total insolvency cases in parenthesis).
[18] Andrew Cawkwell, head of corporate recovery and insolvency with Watson Burton *(www.**watsonburton**.co.uk).*

that companies make to relieve themselves of problems such as excess debt levels". Platt and Platt (1994, p.155).

3.4 Technical/legal insolvency

The Companies Act (2006)[19] is the legislation that governs all companies within the UK and refers to the definition of insolvency in accordance with that of the Insolvency Act (1986) Section 123.[20] Insolvency is described as the inability of a company to pay its debts and is defined in two ways, in terms of cash-flow and in terms of the balance sheet:

"123. Definition of inability to pay debts
(1) A company is deemed unable to pay its debts - [...]

(a) if a creditor (by assignment or otherwise) to whom the company is indebted in a sum exceeding £750 then due has served on the company, by leaving it at the company's registered office, a written demand (in the prescribed form) requiring the company to pay the sum so due and the company has for 3 weeks thereafter neglected to pay the sum or to secure or compound for it to the reasonable satisfaction of the creditor. [Cash-flow insolvency].

(2) ...if the value of the company's assets is less than the amount of its liabilities, taking into account its contingent and prospective liabilities". [Balance sheet insolvency].

Legal insolvency thus occurs when a company cannot meet its immediate liabilities, or when the total liabilities are in excess of its total assets. If insolvency is detected then it is the responsibility of the company directors to ensure that appropriate formal steps are pursued. If the company continues to trade after this point the directors must take every step thereafter to minimise the potential loss to the company's creditors, otherwise wrongful trading may be established and the director may be ordered to contribute to the

[19] http://www.opsi.gov.uk/acts/acts2006/pdf/ukpga_20060046_en.pdf.
[20] http://www.insolvency.gov.uk/insolvencyprofessionandlegislation/legislation/uk/insolvencyact.pdf.

assets of the company for the benefit of the creditors and/or face disqualification. The court will use both objective and subjective tests to determine whether wrongful trading has occurred. A director, therefore, has to match the standards of performance "reasonably expected" of directors having similar functions, else a judgement of fraudulent trading may be established which is a criminal offense. Unlike fraudulent trading which includes an element of dishonesty or recklessness with intent to defraud creditors, wrongful trading is not a criminal offence.

Insolvency in England and Wales is governed by the Joint Insolvency Committee (JIC) which is assembled from representatives from each of the insolvency Recognised Professional Bodies (RPBs) and from The Insolvency Service.[21,22]

The role of the RPB is to regulate Insolvency Practitioners (IPs) to which they hold licence. An IP is authorised by one of the RPBs to carry out and act in insolvency matters, often working in a specialist insolvency department within a leading accountancy firm. The largest regulator of IPs in England and Wales is the ICEAW. The guidelines to which all IPs must act are the Statements of Insolvency Practice (SIPs) which are commissioned by the JIC and produced by the Association of Business Recovery Professionals, the R3.

Once insolvency has been detected there are several courses of action that are available which will be discussed in the next section. To ensure that there is no further depletion of assets or increase in liabilities during these proceedings, the affairs of the company are usually taken in control by an IP who also has the responsibility of reporting to the Department of Trade and Industry (DTI) regarding the conduct of the company's directors.

Section 4 now discusses insolvency processes in detail in order to understand and investigate the avenues of possible research, by way of determining which particular areas provide the possibility of and/or warrant further empirical work.

---

[21] The Insolvency Service is part of the Department of Business, Enterprise and Regulatory Reform (BERR) which licenses some Insolvency Practitioners but mainly focuses on the monitoring of the RPBs.
[22] The seven member bodies of JIC are listed in Appendix A.

4 Formal processes, research avenues and limitations

4.1 Individuals

Individual insolvencies within the UK (or bankruptcy as it should be correctly termed), have reached an all-time high in recent months. Mounting personal debts, tighter restrictions on credit granting, falling house prices and rising unemployment have been the main contributors to the overall number of bankruptcies. There are several different legal proceedings and avenues which may be pursued by either the individual or their creditor(s, should that individual find themselves in a financially distressed situation and are unable to meet their debts as they fall due.

4.1.1 Bankruptcy Orders

Bankruptcy is an option that often has to be considered when you cannot pay your debts as they fall due. The bankruptcy proceedings free you from overwhelming debts so you can make a fresh start and make sure that your assets are shared out fairly among your creditors. A court makes a bankruptcy order only after a bankruptcy petition has been presented. It is usually presented either yourself (debtor's petition), or by one or more creditors who are owed at least £750 by you and that amount is unsecured (creditor's petition). Once a bankruptcy order is issued by the court, an official receiver is appointed to administer the bankruptcy, protect the individual's assets and act as a trustee to the bankruptcy. (Source: The Insolvency Service. www.insolvency.gov.uk)

4.1.2 Debt Relief Orders (From 6th April 2009)

DROs provide debt relief for individuals but are subject to some restrictions. They are suitable for people who do not own their own home, have little surplus income and assets and have less than £15,000 of debt. An order lasts for 12 months. In that time creditors named on the order cannot take any action to recover their money without permission from the court. At the end of the period, if your circumstances have not changed you will be freed from the debts that were included in your order. DROs do not involve the courts. They are run by The Insolvency Service in partnership with skilled debt advisers, called approved intermediaries, who will help you apply to The Insolvency Service for a DRO. (Source: The Insolvency Service. www.insolvency.gov.uk)

4.1.3 Individual Voluntary Arrangements

This is a formal version of the Debt Relief Order. An individual voluntary arrangement (IVA) begins with a formal proposal to creditors to pay part or all of the individual's debts. An application is made to the court and an insolvency practitioner is assigned to the individual. Any agreement reached between the individual and his creditors will be binding on them. The IVA process is as follows:

- First, find an authorised insolvency practitioner prepared to act for you. (Your local court can give the names of local practitioners.) A list is also available for you to look at in your local Official Receiver's office.
- Then you apply to the court for an "interim order". This prevents your creditors from presenting or proceeding with, a bankruptcy petition against you while the interim order is in force. It also prevents them from taking other action against you during the same period without the permission of the court.
- The insolvency practitioner tells the court the details of your proposal and whether in his or her opinion a meeting of creditors should be called to consider it.
- If a meeting is to be held, the date of the meeting and details of the proposals are sent to your creditors. Only those creditors who had notice of the meeting are bound by the arrangement, so it is important that you have accurate records of all your creditors' names and addresses. Otherwise, the arrangement might fail because the practitioner cannot contact all the creditors and, therefore, bind them to it.
- At the meeting, the creditors vote on whether to accept your proposals. If enough creditors (over 75% in value of the creditors present in person or by proxy, and voting on the resolution) vote in favour, the proposals are accepted. They are then binding on all creditors who had notice of, and were entitled to vote at, the meeting.
- The insolvency practitioner supervises the arrangement and pays the creditors in accordance with the accepted proposal. (Source: The Insolvency Service. www.insolvency.gov.uk).

4.1.4 Deeds of Arrangements

Deeds of Arrangement are very simple to draft. The Debtor decides on the offer he wishes to make to his creditors and these are forwarded in the form of a Deed to them. Only a majority in value is required for the arrangement to be approved, however, unlike IVA's only those creditors who voted in favour of the arrangement are bound by its terms. The dissentients remain entitled to continue or commence enforcement action against the debtor irrespective of the arrangement which can result in its failure. A Deed of Arrangement is a relatively little used procedure today, but can be highly appropriate when dealing with only a handful of creditors. In these circumstances, it is an under-utilised procedure in favour of IVA's or the instruction of debt management companies. (Source: www.businessrecoveryadvice.co.uk)

After the Insolvency Act 1986 was brought into force, the act of entering into a deed of arrangement could no longer automatically form the basis of a bankruptcy petition and therefore reduced the threat that any dissenting creditor might petition for bankruptcy.

Although Deeds are not used as a means of proposing an arrangement with creditors as frequently as Individual Voluntary Arrangements under the Insolvency Act 1986, they do have the advantage of only requiring the approval of a simple majority of creditors in number and value; and they may in some circumstances be cheaper and simpler as there is no nominee, no report to court and it is not necessary to call a meeting of creditors. On the other hand, the level of protection is much lower; a dissenting creditor is not prevented by the relevant legislation from taking action against the debtor. (Source: The Insolvency Service. www.insolvency.gov.uk)

4.1.5 Individual bankruptcy rates and economic significance

Figure 4.1 shows the total number of individual bankruptcy cases from 1960-2009 and Figure 4.2 shows the number of bankruptcies cases split into its component parts.

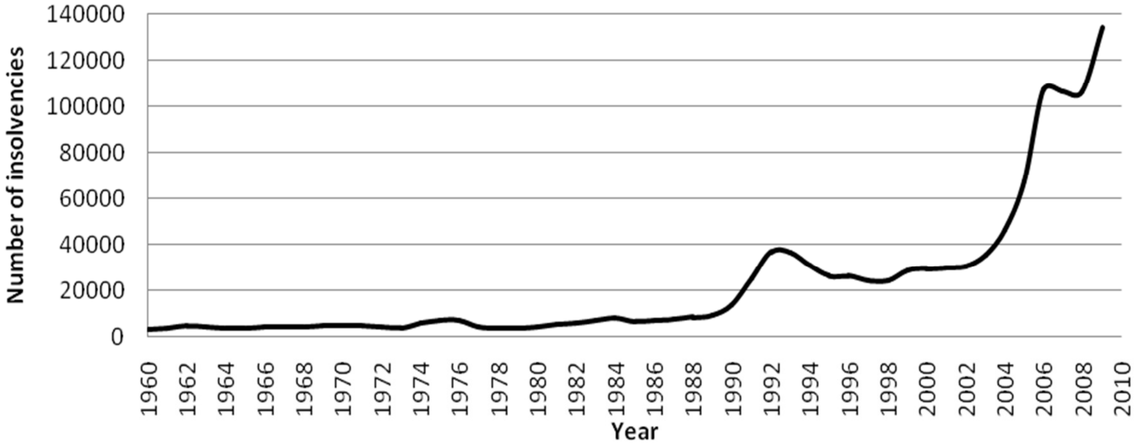


Figure 4.1: Total number of UK individual bankruptcies 1960-2009

Source: www.insolvency.gov.uk

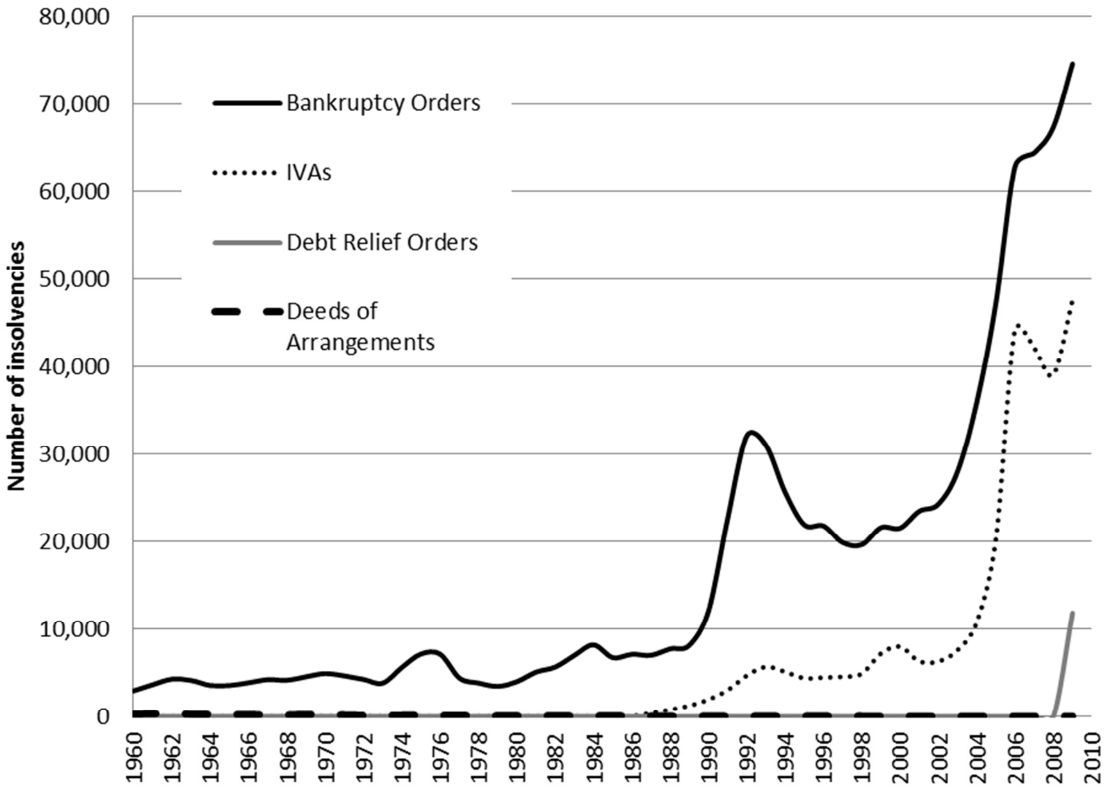


Figure 4.2: Component individual bankruptcy procedures 1960-2009.

Source: www.insolvency.gov.uk

Figures 4.1 and 4.2 show that personal bankruptcy levels have rocketed to an all-time record high total of 134,142 bankruptcies as at the end of 2009. Individual bankruptcy orders are the predominant contributor totalling 74,670 (an increase of 10.7% on 2008), and IVAs contribute 47,641 cases (increasing by 21.8% on the previous year). The remaining contributing cases consist of 11,831 debt relief orders (a new alternative to bankruptcy coming into force on the 6th April 2009). Only one Deed of Arrangement case has been recorded since 1999 and only a handful since the introduction of the Insolvency Act in 1986 showing that Deeds of Arrangement are indeed a rarely used procedure.

From the above figures the individual bankruptcy boom is evident. Up until 1986 and the introduction of the Insolvency Act (allowing for a cheaper and more attainable variety of processes), bankruptcies numbered around 5,000 per year, a small fraction of the levels which we see today. A rise in bankruptcies can clearly be seen, not only as the Insolvency Act takes effect in the late 1980s but leading to a sharper rise as we hit the recession of the early 1990s. This rise however is nothing compared to that which can be seen in recent years with the number of bankruptcies being almost five times larger today than in 2002. This may, in part, be attributed to the reform of the Insolvency act in 2002, now superseded by the Enterprise Act 2002[23]. The most considerable change to the individual bankruptcy process to come out of the introduction of the new Act is the bankruptcy period itself. Before its introduction, bankruptcy would typically last for a period of three years, but it is now usual for a bankruptcy to be discharged after just 12 months. Perhaps it is for this reason that the number of bankruptcies has risen quite dramatically since the Act's introduction, with the bankruptcy "punishment" seemingly a more attractive proposition to the individual than it was previously.

The total bankruptcy levels that we see today are now a cause for economic concern with an average of 368 individuals becoming bankrupt each day.

Mark Allen at Grant Thornton Accountancy suggests that

"*We are seeing people with higher salaries coming to us as many who could previously cope with their monthly commitments and considered themselves relatively well off in the past are hit by a reduction in their monthly disposable income and falling house prices, which leaves them with few alternatives when facing up to their financial problems."* (www.insolvencynews.com)

Mark Sands, director of personal insolvency at RSM Tenon, suggests that:

[23] http://www.opsi.gov.uk/acts/acts2002/ukpga_20020040_en_1.

“Despite the recent news that the recession has come to an end, the impact of unemployment and falling incomes mean that we will see levels of personal insolvencies continue to increase into 2010 as the effects of the recession continue to linger. We should therefore expect to see a further rise of around 12% in annual insolvency figures this year to 150,000 and these levels are likely to remain until at least 2012". (www.insolvencynews.com).

Alec Pillmoor, head of personal insolvency at Baker Tilly, concurs:

"Whilst the recent unemployment statistics have been more favourable, many families are finding it difficult to manage their finances with reduced hours of work. Our major concern is for 2010 as the predicted increase in interest rates and reductions in public sector spending may result in even more individuals entering insolvency." (www.insolvencynews.com).

4.1.6 Research limitations

Although I have demonstrated that individual bankruptcies have a large economic significance, one overriding factor inhibits the conduction of any thorough academic research in this area. This limiting factor is the availability of appropriate data to conduct such a study. Financial data on individuals is not publicly available and is guarded by the Data Protection Act 1998 which governs and protects the use of personal data information by businesses and other organisations. A study of individual bankruptcy prediction or individual credit assessment by academia, therefore, is not possible due to the proprietary nature of the data required which is held predominantly by the banking sector. The banks themselves use and analyse this data in order to produce their own proprietary credit rating systems on which loan decisions are made. From the outside it is difficult (perhaps impossible) to gauge the methodology, extent and accuracy of such systems, and whether there are major differences between each bank’s proprietary system. This avenue of potential research is therefore closed to the researcher seeking trans-sector samples of publically available information and we must turn our attention to company insolvencies.

4.2 Private companies limited by shares

4.2.1 UK corporate insolvency processes

If insolvency is deemed inevitable then there are a number of remedies available to both the company and its creditors, these range from the more formal and restrictive actions such as administration, compulsory liquidation and administrative receivership, to the informal, less restrictive voluntary arrangements. The details of each of the different processes are described below and the exact requirements and procedures according to the Enterprise Act 2002 are provided in Appendix B[24].

4.2.1.1 Administration

The purpose of administration has been changed by the introduction of the Enterprise Act 2002 and is now an out-of-court procedure although the old form of administration is still available in certain circumstances (see Appendix B.II). The new purposes of administration follow a strict hierarchy as follows:

- Rescue the company as a going concern and if not reasonably practicable then;
- Achieve a better result for the company's creditors as a whole than would be likely if the company were wound up and if not reasonably practicable then;
- Realise the company's property in order to make a distribution to one or more secured or preferential creditors.

The appointment of an administrator can be made by the company, its directors or a qualifying floating charge-holder, and is it usual for the period of administration to last 12 months.

[24] Also applicable to Public Limited Companies.

4.2.1.2 Administrative receivership

Administrative Receivership (AR) is usually a consequence of a company failing to meet its long term borrowing obligations where an administrative receiver is appointed by the credit granting institution (e.g. a bank)[25]. The administrative receiver takes control of the company and its affairs usually with a view to sell the company as a going concern rather than selling off assets individually. The number of ARs has declined substantially and will eventually be phased out altogether when there are no longer any pre-Enterprise Act (15 September 2003) floating charges in existence.

4.2.1.3 Company voluntary arrangements

The company may propose a Company Voluntary Arrangement (CVA) which is a formal agreement between itself and its creditors which satisfies both parties regarding the payment of debt. This usually will involve a delayed or reduced payment, capital restructuring or an ordered disposal of assets. These arrangements must satisfy at least 75% of total creditors to be legally binding, warranting very little involvement by the courts and is solely governed by a proposal which sets out the details of the scheme. The Proposal can be made:

- Where the company is not in liquidation or subject to administration by its directors;
- Where an administration is in force, by the administrator; or
- Where the company is being wound-up, by the Liquidator

The company remains under the control of the directors but the scheme itself is under the control of an IP who acts as a supervisor. The CVA procedure is designed primarily as a mechanism for business rescue, but is often used instead of liquidation as a means of distributing funds upon the conclusion of an administration.

[25] Note that it is generally no longer an option to appoint an administrative receiver for charges made after 15th September 2003.

4.2.1.4 Creditors' voluntary liquidation

A Creditors' Voluntary Liquidation (CVL) is the process whereby the directors of a company would meet with an insolvency practitioner to discuss the details of their financial problems and seek his/her advice. Should that advice be that the company be liquidated, then the following procedure should occur. The winding-up of the company is initiated by a resolution of the members of the company, in agreeance that the company is insolvent and is to be placed into liquidation. This decision is then ratified by its creditors (rather than by court order), during a formal meeting with attendance by the way of a "Form of Proxy" receipt and acceptance by each creditor concerned. During the meeting, the directors must detail all of the company's assets and liabilities in a "Statement of Affairs", the insolvency practitioner is formally appointed, and the terms of the liquidation are detailed.

4.2.1.5 Compulsory liquidation

Compulsory liquidation (CL) (or compulsory winding up) is a liquidation ordered by the court, made as a result of a creditor petition, incomplete registration or failure to commence business. A winding up order is also, more often, made as a result of the company becoming insolvent in accordance to the definitions within the Insolvency Act 1986 and is usually preceded by a period of attempted rescue such as the administration process. The Secretary of State for Trade and Industry may also present a winding up order to the court on the grounds of public interest.

4.2.1.6 Members' voluntary liquidation

Although members' voluntary liquidation (MVL) is a formal liquidation process for the purposes of this thesis it cannot be categorised as being part of the failure/insolvency process, although it must be mentioned for completeness. For legitimate reasons, a director(s) might wish to wind down the affairs of a company whilst it is still solvent, for example if the director wishes to retire, or if the company is deemed non-profitable. An MVL is only applicable if the company is able to pay all of its debts within one year and the director(s) make a declaration of solvency. The MVL process therefore will not be included, or enter further into any discussion regarding failure and insolvency processes.

4.2.2 Private limited firm insolvency rates

The total number of private limited firm insolvencies from 1960-199 are shown in Figure 4.3, whereas Figure 4.4 splits the total number of limited firm insolvencies into the five component parts as described in section 4.2.1.

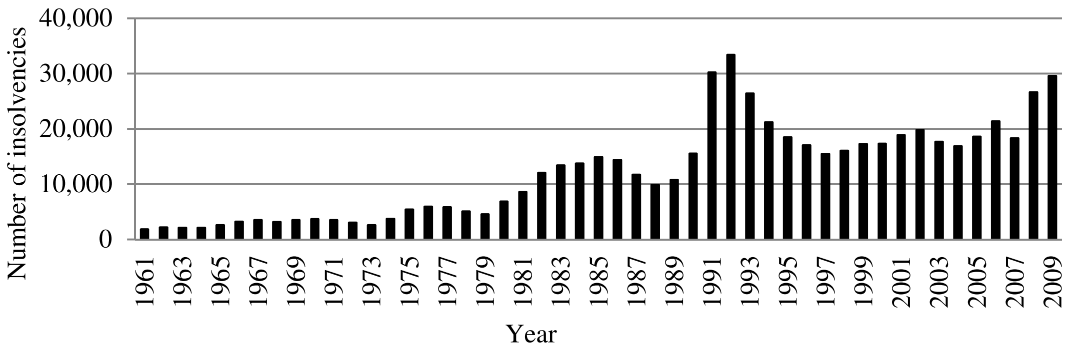


Figure 4.3: Total number of private firm insolvencies in the UK per year 1960-2009

Source: www.insolvency.gov.uk

As with the bankruptcy of individuals, we observe a large increase in these insolvencies circumventing the recession of the early 1990s. The introduction of the Enterprise Act in 2002 does not seem to cause a large increase in insolvencies like that of the individual, but what we can see is the rise in administration processes due to the increased availability and entry routes that the Act conceived. This rise in administrations appears to be associated with a fall in the number of creditor liquidations from 2002 to 2008. Administrations rise from 643 in 2002, to 2,512 by 2007 (an increase of 390%), whereas the number of creditors' liquidations falls 27.1% from 10,075 to 7,342 over the same period. 2007-2009 shows an increase in the total number of insolvencies and Figure 4.4 shows that the main driver for this increase is the sharp rise in creditors' liquidations although the number of administrations is also continuing to climb.

The Enterprise Act 2002, although offering a simplified and relaxed route for the restructuring of a failing company as compared to its Insolvency Act 1986 predecessor, offers no legal barrier for the directors of a failed firm to continue directorship with an existing or new start-up company. Only in extreme cases will a director be disqualified and this decision is made on an individual basis by the courts rather than the insolvency practitioner. The Company Directors Disqualification Act 1986 seeks to ban "unfit" directors from being directors of a limited liability company. This action is taken only when it is proven the director has acted wrongfully, fraudulently, or just very badly. A government agency will first have to prove wrongful trading or "unfit behaviour" before any action can be taken. The period of disqualification can range from 2-15 years suggesting that the punishment for unfit behaviour can to some extent prevent further insolvencies from occurring within the immediate future. The overall impact that this has on the number of insolvencies, although difficult to measure, is likely to be negligible as disqualification is a rarity rather than the norm.

.

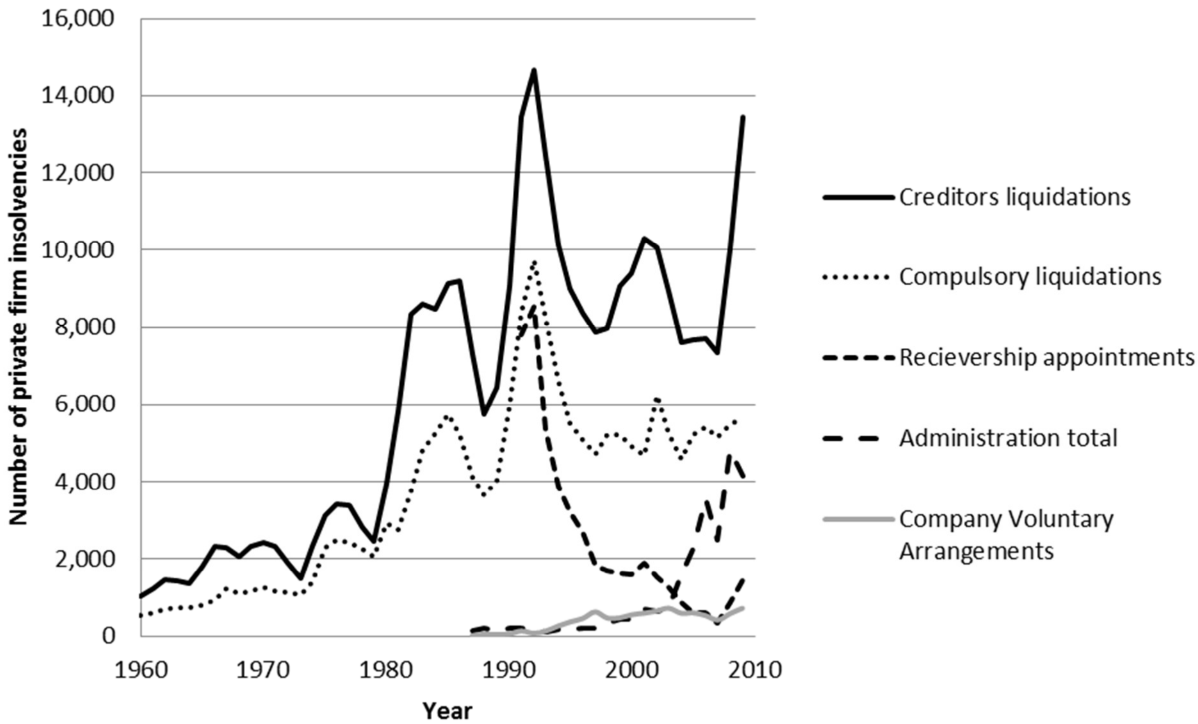


Figure 4.4: Component private firm insolvency procedures 1960-2009.

Source: www.insolvency.gov.uk[26]

[26] The data obtained from the insolvency website required some adjustments:

(i) The numbers obtained from the downloaded excel spreadsheet contained statistics for ALL UK companies. Deductions were therefore required to remove those companies which were not privately owned (i.e. Public Limited Companies). The number of PLC insolvencies were obtained from the London Share Price Database.

(ii) Receivership appointments comprise administrative receivers appointed under the 1986 Insolvency Act (and the 1989 Order for Northern Ireland) and certain other receiver appointments, for example

If we calculate the rate of insolvencies as a percentage of the total number of firms, the resulting chart again shows an increasing trend over the previous decade 2000-2009, with the rate showing the greatest increase in the final two years. This result is shown in figure 4.5. The rate of insolvency shows a 0.22 percentage point increase from 2007 to 2009, represented by an increase in total insolvencies of 11,307, offset by an increase in firm population of 154,965. The rise in the rate of insolvencies is what we would expect given a period of recession.

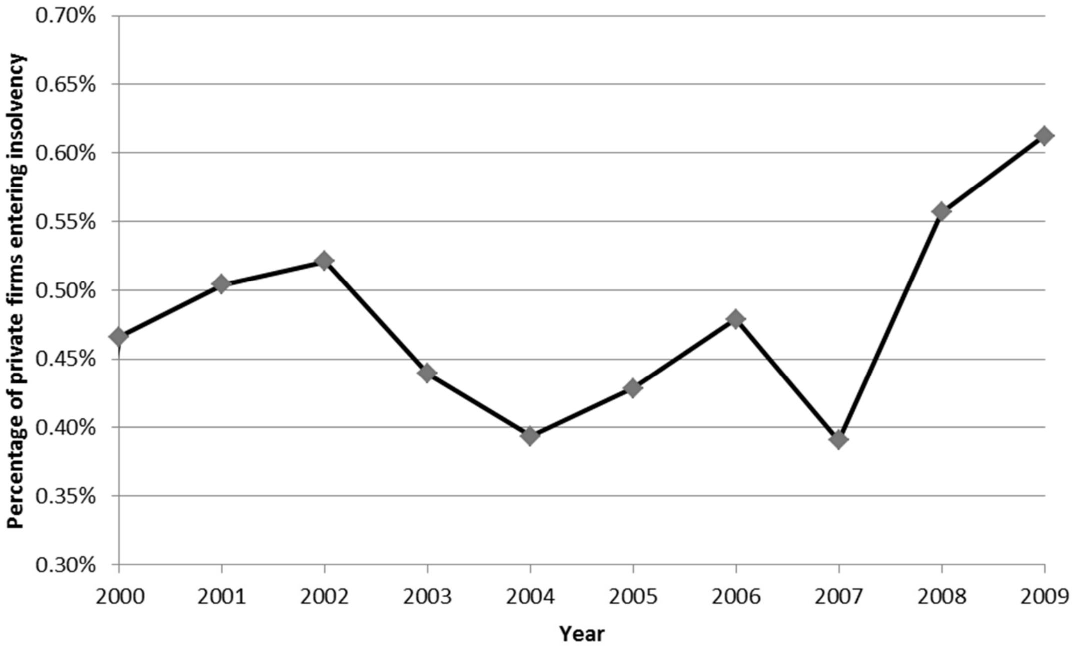


Figure 4.5: Rate of private firm insolvencies 2000-2009.

Source: www.insolvency.gov.uk

4.2.3 Economic significance

With 29,545 private limited firms entering insolvency in 2009 it is clear that these numbers have economic significance. If we are to believe the R3 in that 27% of all insolvencies are a direct result of another firm failing, then in 2010 almost 8,000 of these private sector firms will be directly sent dominoing into failure by the 2009

under the Law of Property Act 1925 - due to the use of the same statutory documentation for different types of receivership, it is not possible to give a breakdown between them. (www.insolvency.gov.uk).

(iii) Administrations include both the old administration procedure under Part II of the Insolvency Act 1986, and the more streamlined inclusion of additional entry routes under Schedule B1 of the Enterprise Act 2002.

insolvency glut. As the ripples of the recession continue to flow through the economy we would expect that the total number of private firm insolvencies in 2010 to continue to rise.

Are we able however, to place a value on these 29,545 private firm insolvencies? Exact statistics for this type of investigation are not apparent but we are able to make an approximate estimation using data obtained from the Office for National Statistics[27]. A statistical press release (October 2009) by the Department for Business Innovation and Skills provides a number of statistics on Small and Medium-sized Enterprises (SME) which we can use to estimate the economic significance of private firm (approximated by SME) insolvencies. 4.8 million[28] private sector enterprises are estimated to exist by the publication (99.9% of all UK firms) which suggests that during 2009 the percentage of these firms entering insolvency is just 0.6%. We can therefore apply this percentage to get a rough estimate of the costs to the economy in terms of turnover and employment. Total employment within the SME sector is an estimated 13.7 million people, and had an estimated combined annual turnover of £1,500 billion. These numbers give us a rough approximation of the economic costs of private sector SME insolvency of 82,200 job losses and £9 billion of lost economic revenue.

4.2.4 Limitations for academic research

The vast majority of insolvency prediction literature focuses on large companies, i.e. those which are listed on, and, priced by the market. There have been a relatively small number of prior academic studies which examine the usage and effectiveness of insolvency prediction and credit scoring models with reference to the smaller private firms. The most notable of these studies are that of Edminster (1972), Peel & Peel (1987), and Falkenstien (2000).

The reasons for the apparent deficiency of such papers throughout the literature are most probably due to both the difficulty in obtaining sufficient data for empirical analysis and, the reliability of any such data should it be acquired. These difficulties stem from the way the data is prepared and reported compared to that of the larger

[27] http://www.statistics.gov.uk, rounded to the nearest 100,000.
[28] Consisting of 3 million sole proprietors (306,000 with employees), 462,000 partnerships and 1.3 million companies.

corporation. For most (if not all) of the global markets and economies, far more emphasis is placed on the regulation and governance of the larger publicly traded firm than that of the smaller private company. Although regulation exists for limited company accounting they are not subject to the more stringent rules faced by the listed firm, the differences born through the ever evolving needs and requirements of markets, the economy, and shareholder protection.

Similarly the application of models derived through the analysis of public firm data should not be extended to assess the health of a private limited company.

"*The relationship between financial variables and default risk varies substantially between public and private firms. An important consequence of this is that default models based on public firm data and applied to private firms will likely misrepresent actual default risk*". Falkenstein (2000).

I now examine and discuss the main causes of the lack of private data for empirical academic analysis and the reasons why it may not be as reliable when used in the construction of insolvency prediction models compared to the data obtained from the public firm sector.

4.2.4.1 Accounts preparation and audit exemption

All UK private limited liability companies are required to keep records of the company's financial transactions. The records must contain sufficient detail to enable the financial position and performance of the company to be determined, and the directors must provide profit and loss account and balance sheet accounting statements in compliance with the Companies Act 2006. The records should contain details of any income and expenditure and a record of the company's assets and liabilities. It is a statutory requirement for all companies to submit financial statements to Companies House within 9 months after the accounting year end. There is no legal obligation for these accounts to be prepared by a certified accountant but ultimately it is the director's responsibility to ensure that the annual accounts are true and fair. The accounts prepared by the limited company must include:

- A profit and loss account, or an income and expenditure account if it is a company not trading for profit
- A balance sheet, signed by a director
- Notes to the accounts
- Group accounts - if appropriate

The accounts are normally accompanied by an auditor's report, unless the business is exempt from audit. For limited companies, the accounts must also be accompanied by a director's report, with a business review if the company does not qualify as small or medium. To qualify as a small or medium company and be eligible to file abbreviated accounts and negate the auditor's report, the company must satisfy at least two out of the following three conditions (for financial years on or after the 6th April 2008):[29]

To qualify as a small company;

- Annual turnover must be no more than £6.5 million
- Balance sheet total must be no more than £3.26 million
- Average number of employees must be no more than 50

To qualify as a medium-sized company;

- Annual turnover must be no more than £25.9 million
- Balance sheet total must be no more than £12.9 million
- Average number of employees must be no more than 250

Under the Companies Act 2006, total audit exemption exists if the qualifies as a small company as described above. For the large majority of companies, externally audited accounts are not filed and therefore may be viewed with a degree of uncertainty with regards to their reliability. Currently in the UK (2010), 86.3% of non-dormant companies do not have their individual accounts audited.[30]

---

[29] Note that for financial years starting before the 6th April 2008 the requirements are different but in most cases are now irrelevant.

[30] Companies House, Statistical Tables on Companies Registration Activities 2009-10, Table F2.

It is perceived that a lack of audit inhibits the inherent usefulness of the accounting data to make survival and credit-granting decisions. Melumad and Thomon (1990) demonstrate that in some cases, although audit is non-mandatory, private firms may appoint an audit team because lending institutions conclude that it is only borrows with a high credit risk that would choose not be audited. This voluntary appointment of auditors therefore acts as a signal to the lender that the company is of a low-risk nature. Lennox and Pittman (2011) suggest that under a voluntary audit regime, companies which are audited receive credit ratings which are on average 14 points higher than those which are not.

4.2.4.2 Availability of accounting data

If we negate the suggestion, for the moment at least, that the accounting numbers contained within private company accounts, due to their regulatory lack of external scrutiny, may not be as reliable as the financial statements produced by the larger companies who are subject to compulsory audit; and also believe in the ideology that the an efficient market would dictate these accounts to indeed be true and fair; then we may wish to investigate where the relevant data items may be obtained. Unlike public companies, accounting data is not widely available for empirical analysis through the use of electronic financial databases and is more difficult to obtain. Some online database resources such as Hemscott Company Guru Academic do provide information and data from UK private non-listed companies. From experience, however, the data supplied is often missing, incomplete, or subject to large rounding errors, rendering the use of such resources extremely problematic. The data items provided by these resources are also geared towards those larger firms of which audit is either compulsory or volunteered (13.7% of all private firms) and therefore not entirely representative of the true population.

A further problem associated with private firm data is that all the numbers are accounting-based, that is to say, by their very nature private firms shares are not traded by the public and therefore no market-based data will be available. Many of the recent insolvency prediction and credit risk models to be published (e.g. Sobehart and Stein, 2000; Hillegeist et al., 2004; Bharath and Shumway, 2004), use market data as a fundamental part of their analyses. Without such data, the more modern approaches to insolvency prediction modelling will not be applicable to the private limited company. Any analysis that is undertaken will therefore be restricted to the use of accounting-number based data.

To obtain the most accurate and complete accounting data for these firms, the most direct approach would be through Companies House where the records of all UK companies are filed. From the Companies House website (www.companieshouse.gov.uk), it is possible to obtain copies of the accounts for any registered company, albeit for a small fee of £1 per document. This method however would prove to be quite expensive and very time consuming given the number of data items required for a thorough empirical analysis. In a potential analysis, therefore, there needs to be a trade-off between cost, time, and accuracy when collecting private limited company data, perhaps this is why there are only a small number of studies of this nature within the literature.

4.2.5 Summary

Although the lack of data and its reliability issues inhibit a study of privately owned firms by the academic, some proprietary systems do exist to assess the credit worthiness of private limited companies and are generally found within the banking sector. The banking sector, where small business loans are granted for example, may not treat or assess the loan applicants by generalised number-based system similar to that of the individual, but are often treated on a case-by-case basis. This can be demonstrated following the recent criticism faced by banks for their unwillingness to lend to small businesses as shown by the current failure to meet lending requirements as set out within project Merlin[31]. These feelings are rebuffed by the Chief Executive of the British Bankers Association (BBA) who stated that "Banks have made repeated commitments to support business. There are funds available to lend to firms with a viable business plan".[32]

[31] http://www.bbc.co.uk/news/business-17009985.

[32] "BBA statement regarding banks profitability" available at http://www.bba.org.uk/bba/jsp/polopoly.jsp?d=145&a=18035

Credit rating agencies also have proprietary systems for evaluating the smaller firm such as Moody's "RiskCalc" for private companies as detailed in Falkenstein (2000). Moody's are able to utilise their proprietary database of firms in order to model default risk for private companies.

"*The model's key advantage derives from Moody's unique and proprietary middle market private firm financial statement and default database (Credit Research Database), which comprises 28,104 companies and 1,604 defaults*". Falkenstein (2000).

For the reasons stated in this section, the ability to obtain and rely upon the necessary data for an empirical study of private firms is problematic. The focus of this thesis therefore will now move onto the public limited company where the majority of research in the field of insolvency prediction and credit risk assessment has occurred to date.

4.3 Public limited companies

4.3.1 PLC insolvency rates

The number of public companies which become insolvent each year is shown in Figure 4.5. We can see that the absolute number of firms which are flagged as insolvent is miniscule compared to the private company figures presented in the previous section. This is not particularly surprising as public companies tend to be the largest firms with extensive systems of governance, operating under greater scrutiny from its shareholders and regulators. Figure 4.5 shows that on average less than 10 PLCs become insolvent each year which increases as we enter times of economic downturn such as the early 1990s and the latest recession. Although the absolute number of insolvencies seems quite small, we must remember that they are drawn from a relatively small population of approximately 2300 firms each year which are listed on the London Stock Exchange's Main Market. The rate of insolvency occurrence as a percentage of the total population of firms is provided in figure 4.7. In 2000 the insolvency rate amongst public companies was approximately 0.5%, which is not dissimilar to the average rate shown by private firms. This rate rose briefly in 2001 and 2002 to 1.8% in accordance with the high number of administrations which occurred in those years. The rate then fell from

2003-2007 to a minimum of 0.46% in 2005. Post-recession and by 2009, the rate of PLC insolvencies grew to over 2%, mush higher than the 0.6% attributable to private firms in the same year. This suggests that although these larger firms are subject to greater accounting, governance and legal regulations which are primarily aimed at protecting the shareholder and the company's very existence, the rate of insolvency is far higher than that of the smaller, private firms, which seems to be somewhat counterintuitive.

If we examine the different types of insolvency proceedings for public firms we can note several points of interest. Firstly, since the introduction of the Insolvency Act 1986 we can see that liquidations have been all but phased out, being replaced by the more popular administration process. The number of receiverships has also declined since the introduction of the Enterprise Act 2002 where it is generally no longer an option to appoint an administrative receiver for charges made after 15th September 2003. Currently therefore it is the administration process which is the most popular insolvency route for PLCs seemingly because of its rescue mandate and lowered entry barriers post-Enterprise Act 2002.

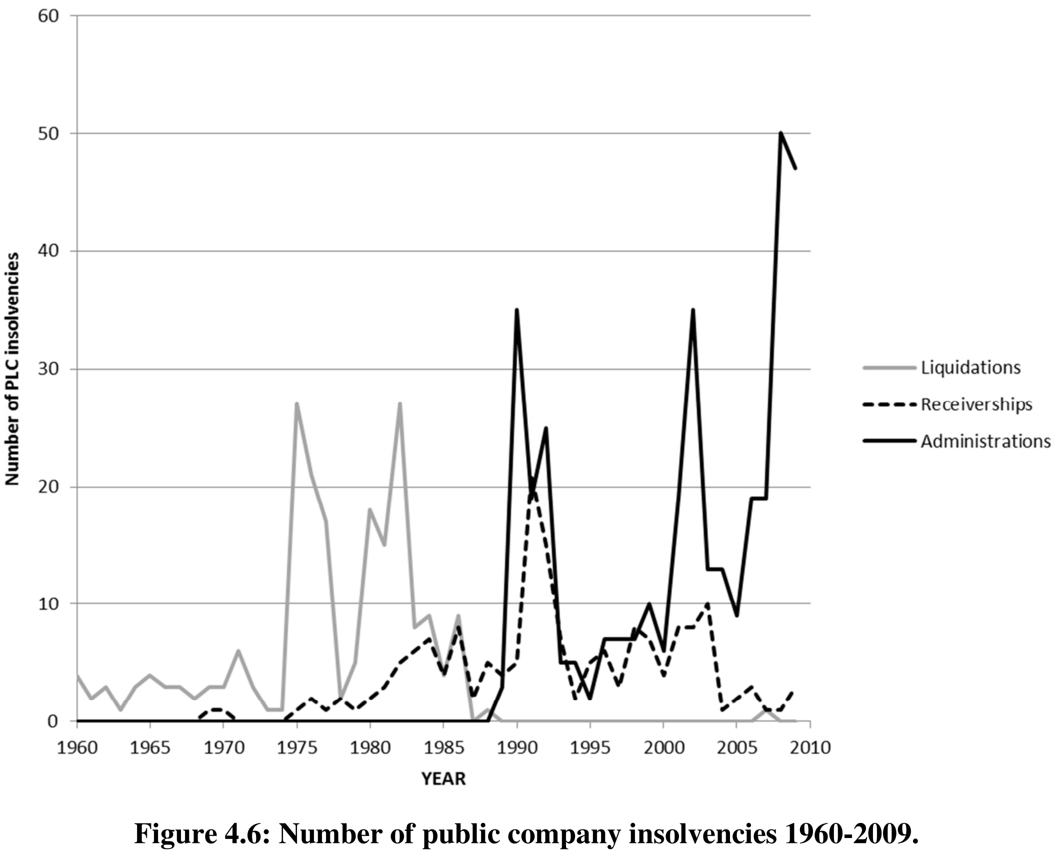


Figure 4.6: Number of public company insolvencies 1960-2009.

Source: London Share Price Database

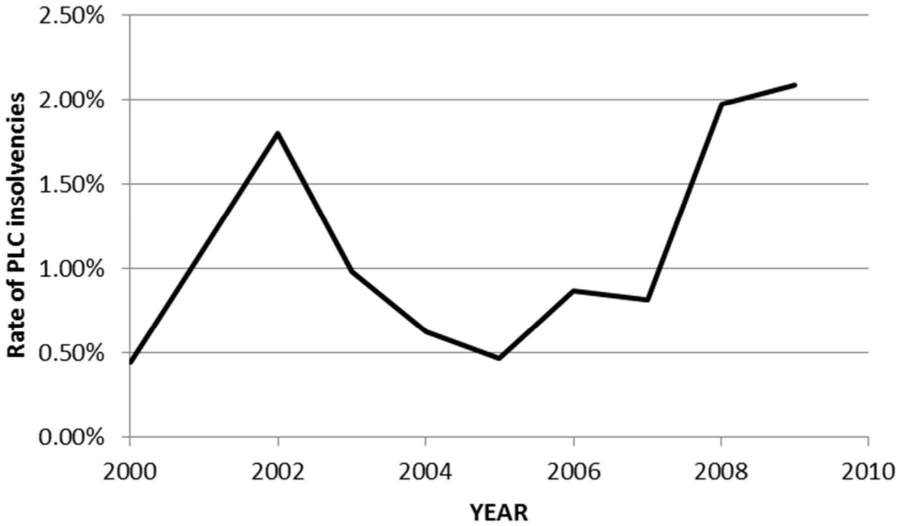


Figure 4.7: Rate of public company insolvencies 2000-2009.

Source: London Share Price Database

4.3.2 Costs to society and economic significance

Data collected from the Thomson One Banker database provides information regarding the 47 UK PLC insolvencies which occurred during 2009. The loss of these companies meant that approximately 40,000 jobs and £4.5 billion of turnover were lost during this period. Comparing these figures to that of 82,000 jobs and the £9 billion of lost turnover associated with the private limited company sector we can see that in terms of economic significance, PLC insolvencies contribute roughly half of the value attributed to private firms. This is unsurprising given the absolute number of firms which are observed for each sector class (47 vs 29,545). Interestingly, we can calculate the economic value (based on employee and turnover losses) per firm for both sectors which gives us an approximate per-firm estimate. The per-firm value estimates provide significant differences between the two groups[33]. For smaller private firms the number of employees lost per insolvency is 2.7 compared to 851 that are lost for every public-firm insolvency. In terms of turnover, the private limited firms contribute approximately

[33] Based on 2008 figures.

£300,000 of lost turnover per insolvency which compares to £9.5 million attributable to each listed firm.

We can see that although private companies provide around double the economic losses in real terms for the economy compared to the listed firms, this is not unsurprising considering the sheer number of insolvencies which occur within the private limited company sector. When I scale these figures down to the value attributed to each individual firm, again it is unsurprising that a public company has more economic significance should it enter into insolvency than its private limited counterpart. This is one possible reason, perhaps, why there has been a great deal more academic attention focused on the public company, although other factors are also prevalent and will be discussed shortly. Before this discussion however, I feel it necessary to provide a brief overview of the insolvency prediction and credit risk research/literature that has been directed at the public listed company, in order to both contextualise this thesis and to provide some understanding of the history and magnitude of public firm insolvency prediction and credit risk analysis.

4.3.3 A wealth of academic research

The work of Beaver (1966) seeded the modern literature on insolvency prediction with a univariate approach, treated further in Tamari (1966) and Beaver (1968). Methods in the 1970s centred on a multivariate framework, with the widespread use of multivariate discriminant analysis (MDA) models and the production of the 'Z-score' and similar technologies (Altman, 1968; Deakin, 1972; Edminster, 1972; Blum, 1974; Diamond, 1976; Taffler and Tisshaw, 1977; Taffler, 1983). Criticisms relating to violations of the statistical assumptions underlying the MDA approach,[34] however, led researchers of the 1980s to concentrate their efforts on the development of conditional probability models, the most popular being the logit[35] (Ohlson, 1980; Hamer 1983; Zmijewski, 1984; Zavgren, 1985; Keasey and Watson, 1987).

[34] See, for example, Joy and Tollefson (1975, 1978), Eisenbeis (1977), Moyer (1977), and Altman and Eisenbeis (1978).

[35] Refer to section 8.1 for an evaluation of methodological popularity.

Technological developments and an increase in the availability and power of computer processing allowed the insolvency prediction researchers in the 1990s to adopt a wider range of methods. Much of the work in the 1990s concentrated on artificially intelligent systems such as neural networks, genetic algorithms, case-based reasoning and recursive partitioning (Odom and Sharda, 1990; Coats and Fant, 1993; Boritz et al., 1995; Charitou et al., 2004). Despite the voluminous research, however, insolvency prediction suffered much criticism because of a lack of theoretical underpinning – giving rise to problems associated with the classical paradigm (choice of definition of firm failure, non-stationarity and instability of data, and sample selection) and the arbitrary selection of variables and modelling method.[36]

The latest modelling developments are the contingent claims models which are based on option pricing theory as set out in Black and Scholes (1973) and Merton (1974). Such models are known by various names: 'BSM' (after Black, Scholes Merton), 'KMV-Merton' (after KMV Corp., a company specialising in tools for managing corporate credit risk, and Merton), and 'Merton model' for example. Prior to the acquisition of KMV by Moody's Corp. in April 2002 very little was known publicly about KMV's proprietary credit risk appraisal methodologies – which included contingent claims models. After the acquisition, a large selection of papers written by various practitioners became publicly available to download from the Moody's KMV website,[37] and contingent claims models have since attracted a deal of interest from academics. Comparisons between these models and traditional accounting number-based models have been performed by Bharath and Shumway, 2004, Hillegeist et al., 2004, Vassalou and Xing, 2004, Campbell et al., 2006, Reisz and Purlich, 2007, and Agarwal and Taffler, 2008.

[36] See Balcaen and Ooghe (2006) for an excellent discussion of the problems.

[37] http://www.moodyskmv.com/research/New%20Research_sectionl.html.

4.3.4 Reasons for this wealth of research

4.3.4.1 Accounts preparation and audit process

The annual reports of public companies are a more attractive proposition for analysis than that of the private firm for two key reasons. Firstly, the reports themselves are, for the norm, readily and easily obtainable, not only by the academic but also the general public. Any stakeholder of a particular company will, quite quickly, be able to find detailed financial information on that firm by either visiting their corporate website or, by visiting one of many websites dedicated to providing for free, databases of UK and international company financial content[38]. These free resources may fulfil the needs of the private investor, but often lack the detailed information and mass-download-ability that an academic would require for a thorough empirical analysis. Larger and more concise databases are available, but often accompanied by a hefty price tag. These include Datastream, Thomson-Reuters, and Bloomberg. The price of access to these databases varies; for instance a Bloomberg terminal would set you back somewhere in the region of £20,000 per annum, not something the average member of the public could afford, but a valuable asset to a top university. The advantages of holding a license for one of these databases is the number of companies covered, the amount of information which is available for each company, and the ability to create custom reports for required data items and download the report into a single file.

Secondly, we have the requirement of the compulsory audit. All listed company accounts must be prepared in conjunction with a qualified audit team. This differs to that of a private firm where a compulsory audit is not required should it meet the conditions as discussed in section 4.2. I therefore suggest that the use of public firm data would be more reliable and complete than that of the private firm because the financials have been validated by an external source and not simply left to the director's discretion. The reasons for the external audit are born through regulatory and governance evolution as mandated by the Financial Reporting Council (FRC) with key factors influencing reporting evolution including size, economic significance, public interest, and investor protection.

[38] For example Yahoo Finance (uk.finance.yahoo.com) , Hemmington-Scott (www.hemscott.net), and your standard Iphone stocks application.

4.3.4.2 Linking going concern and insolvency prediction

A key feature of the compulsory external audit, and particularly relevant to this thesis, is the going concern qualification (GCQ). Part of an auditor's duty to the company is the requirement of the audit team to assess the firm as a going concern. Financial statements are prepared on the assumption that the firm is a going concern and will continue to operate throughout the next reporting period. Therefore, annual reports issued on a going concern basis have already been judged to survive the following year and remain solvent. In reality however, studies show that it is extremely unlikely for a failing firm to be issued with a GCQ (e.g. Citron and Taffler, 1992). The lack of GCQs, is perhaps the reason why accounting number-based insolvency prediction models have apparently been able to detect signs of impending failure from accounting records, even if by definition the accounts represent a solvent company perpetual in existence.

The International Auditing and Assurance Standards Board (IAASB) provide guidance to auditors on how to evaluate a client's viability status within ISA 570. The auditor must assess the client's ability to meet its obligations as they become due without having to liquidate its assets, restructure debt, be forced by outsiders to revise its operations, or other similar actions.

Conditions or events that raise doubts about the client's ability to continue in existence include:

- Negative trends e.g., recurring operating losses, working capital deficiencies, negative cash-flow from operating activities, adverse key financial ratios.
- Indicators of possible financial difficulties e.g., default on loan or similar agreements, arrearages in dividends, denial of usual trade credit from suppliers, restructuring of debt, non-compliance with statutory capital requirements, need to seek new sources or methods of financing or to dispose of substantial assets.
- Internal matters e.g., work stoppages or other labour difficulties, substantial dependence on the success of particular projects, uneconomic long-term commitments, or need to significantly revise operations.
- External matters e.g., legal proceedings, legislation, or similar matters that might jeopardize an entity's ability to operate; loss of key franchise, license, or patent; loss of

a principal customer or supplier; or uninsured or underinsured catastrophe such as a drought, earthquake or flood.

If the aggregate effect of these conditions and events suggest that the client may have continuity problems, the auditor must consider and evaluate the feasibility of management's plans for dealing with these adverse effects. Management plans include:

- Plans to dispose of assets and the effects of such disposals;
- Plans to borrow money or restructure debt and the effects of such plans on existing covenants;
- Plans to reduce or delay expenditure and the effects of such delayed expenditure on operations; and
- Plans to increase ownership equity.

If substantial doubt about the entity's ability to continue as a going concern remains, the auditor's report must include an explanatory paragraph with the audit opinion that comments on the going concern uncertainty. (Accounting, Finance & Tax)[39]

Citron and Taffler (1992) suggest that unless the chances of a company failing are very high, the probability of a GCQ is very low. The reason for low GCQ rates is likely to be based on the independence of the auditor and the possible reliance on future economic benefits gained should the firm continue to exist untarnished by a detrimental audit report. Literature in this area by Kida (1980), Altman (1982) and Moizer (1985) suggests that there are a number of cost factors that may lead to a reduction in auditor independence which can be summarised as:

- The value of the auditor's economic interest in the client
- The likelihood that this economic interest will be lost either due to the client switching auditors or failure brought about by the GCQ itself (self-fulfilling prophecy hypothesis)
- The likelihood that the client will sue if the report is qualified and the client does not fail.

[39] http://accounting-financial-tax.com.

Conversely they show that there are certain cost factors which lead to an *increase* in independence, namely:

- The loss of future revenues due to loss of reputation should the auditor not qualify and the client fails.
- The likelihood of lawsuits by third parties if there is no GCQ and the client fails.

In a sample of UK quoted companies observed over the decade 1977-86, Citron and Taffler (1992) find some evidence in support of an association between the presence of a GCQ and auditor switching. No evidence was found to support the self-fulfilling prophecy hypothesis, and a further finding was that GCQ rates are not affected by firm size.

Kida (1980) conducted a survey of audit firms, requiring them to rank companies on a six point scale of failure given past data from forty anonymous manufacturing companies. His results showed that on average the auditors (using a five variable discriminant model) correctly predicted the outcome of 33.2 out of the 40 firms. In reality, however, studies often paint a different picture. Further evidence by Citron and Taffler (1992) suggests that only 26% of failed firms are qualified on a going concern basis in the period prior to insolvency and were not subject to a GCQ unless failure was imminent.

Several authors have examined the link between insolvency prediction models and the issue of the GCQ. Altman and McGough (1974) suggested that a model to predict bankruptcy may help auditors judge a company's ability to continue its operations by alerting the auditor to certain problems that may be difficult to detect using traditional auditing procedures. Likewise Koh (1991) posits that insolvency prediction models can be useful to auditors in making going concern assessments. Paquette (1996) suggests that bankruptcy prediction models can assist auditors in assessing the validity of the going concern assumption by directing attention to indicator ratios as well as by providing a method for combining the information contained in the ratios.

The notion that insolvency prediction models may be used as a tool by the auditor to detect impending failure is unsurprising. The determination of a company's health, whether it be for audit GCQ's, investment appraisal, mergers and acquisitions, credit ratings, or credit risk; are all intrinsically linked by the need to evaluate all of the available data and information to make an informed judgement. Some will be privileged to more information than others on which to base their decisions, but many of the processes and technologies will remain the same. Gaganis (2007) provides an excellent review of the technologies employed within the field of audit GCQ's and the similarities to the insolvency prediction literature are clearly evident.

Perhaps surprising on initial consideration, there are examples within the literature which suggest that a going concern qualification acts in a positive manner when appraising survival chances. Taffler and Tseung (1984) found that only 20% of firms which received a GCQ subsequently failed. "*A going concern qualification paradoxically is associated with the continuance of an enterprise not its failure*" (Taffler and Tseung, 1984, p.25). This evidence was further supported by Peel and Peel (1987) who find that when used as a dummy variable in a predictive context, the GCQ had a sign which was wrong *a priori* in predicting failure

4.3.5 Summary

This section has discussed in some depth the process of insolvency along with potential research avenues and limitations for the three broad groups; individuals, private companies, and public companies. In summation it is evident that the relative availability (and perceived accuracy) of information makes the study of PLCs a more empirically sound option than its counterparts. The difficulties discussed earlier in this section inhibit the efficacy and accuracy of such a study for both private firms and individuals. I have also provided evidence as to the economic importance of a failing PLC, in that although the frequency of PLC insolvencies is far less than for private firms, the value of these firms is considerably more on an individual basis in terms of both monetary and social impact. One further point to note is that credit risk theory relies heavily on market data thus a study on private firms (or individuals) is not possible if one wishes to incorporate credit risk models into the research. As one of the primary aims of this study is to empirically evaluate a comparison between accounting number-based models and contingent claims market-based models, it is evident that public company data must be used.

5 The prediction of corporate insolvency: a literature review

5.1 1928-1968: the univariate approach to insolvency prediction

Section 2.2 traced the evolution of the accounting ratio and concluded with a naive model by Wall and Duning (1928) approximating the first example of real linear multivariate discriminant function. Attempting to thwart an explosion of ratios and ratio usage which was prevalent at this time, Wall and Duning created a ratio index, a weighted combination of several different ratios with the weights being arbitrarily chosen by the authors and providing a numerical snapshot of any given firm. Although the ratio index is possibly the first example of a multivariate use of ratios, because of the inherent lack of any sound statistical methodology during the weight selection process, it cannot be truly classed as a multivariate model. Although Wall and Duning's work was specifically directed towards the area of insolvency it still remains, however, a forerunner to the ratio analysis and insolvency prediction models that we are familiar with today.

The first example of an academic study of financial ratios directed solely towards the field of bankruptcy came four years after Wall and Duning's pioneering ratio index. Fitzpatrick (1932) investigated the differences between ratios of "successful industrial enterprises with those of failed firms". In his study he analysed twenty matched pairs of firms from the period 1920-1929 and discussed the usage of thirteen of their accounting ratios as predictors of insolvency. The evidence that Fitzpatrick presented was that there were persistent differences in the ratios of the two groups of firms for at least three years prior to failure. He suggested that all of the ratios predicted failure to some degree but the best indicators were net profit to net worth, net worth to debt, and net worth to fixed assets. Although this analysis has a clear shortcoming in the size of the sample (twenty firms), this paper was significant within the field of insolvency prediction as it provides the first example of insolvency prediction methodology and the ability of the accounting ratio to predict failure.

In a similar vein, Smith and Winakor (1935) analysed a sample of twenty-nine companies which had experienced financial difficulties and distress during the period 1923-1931. They investigated the trends of twenty-one accounting ratios, analysing the mean of each ratio up to ten years prior to the occurrence of the financial difficulty, (although what is deemed to constitute a "financial difficulty" is somewhat unclear from the text). They concluded that the ratio of net working capital to total assets was the most accurate predictor of failure, with its declination evident a full ten years before the event. Their data also demonstrated (but not emphasised) that the ratio of net worth to fixed assets was also one of their best indicators, agreeing with the findings of Fitzpatrick (1932). The study of Smith and Winakor (1935) has, however, a serious drawback. The timeliness of this study and the data that it incorporates encompasses the Wall Street crash of 1929, and the lack of a comparative selection of healthy firms leaves the findings of this study somewhat ambiguous.

Merwin (1942) was described in Horrigan (1968) as being "...the first really sophisticated analysis of ratio predictive power". Merwin overcame some of the shortcomings of the previous studies by Fitzpatrick (1932) and Smith and Winakor (1935) by analysing a far larger sample of firms. His study investigated the ratio properties of 936 small manufacturing corporations from the period 1926-1936 with the sample separated into "continuing" and "discontinuing" firms. Firms were considered insolvent if the corporation discontinued making reports of income tax at some time before the end of 1936. This study suffers the same problem in timeliness as the Smith and Winakor study in that the sampling period envelopes the Wall Street crash of 1929. It is quite surprising however when Merwin reports that from the 939 firms drawn as a sample in 1926, 558 (approximately three-fifths) of the companies become insolvent by the end of the sample period. 237 of these firm discontinuations occurred from 1927-1929 and a further 189 between the years 1930-1932. Withstanding the economic conditions surrounding the study, the percentage of firms apparently becoming insolvent seems surprisingly large.

The glut of discontinuing firms and the timeliness of the study may have been fortuitous, providing the author with a larger sample of failing firms than had previously been examined. Merwin found that from an undisclosed "large" number of ratios which were examined, three ratios were "especially symptomatic of impending discontinuance"; current assets to current liabilities (quick ratio), net worth to total debt, and net working capital to total assets. These three ratios "showed a decidedly unfavourable trend for all five industries when measured by the empirical standards of the remaining solvent corporations". Merwin was able to detect differences in the ratio means for as much as six years before discontinuance, with net working capital to total assets being the most sensitive indicator of ultimate discontinuance.

It is not for another twenty-four years that any further direct and dedicated insolvency prediction studies become apparent in the literature. A number of studies do however arise within this period which continues to evaluate the financial ratio as a useful decision tool. These studies largely focus on the need for identifying good credit and the pricing of bonds in the corporate market. Hickman (1958) analysed the default experience of corporate bond issues during 1900-1943 and found that the net profit to sales and the times-interest-earned ratios were the best predictors of default. Other studies such as Saulnier et al. (1958) found evidence that borrowing firms with poorer current ratios and net worth to debt ratios were more prone to loan defaults. These ratios are the same best indicators as suggested by Merwin (1942) which demonstrates the intrinsic linkage between insolvency prediction and credit appraisal.

In 1966 William H. Beaver published his seminal paper "Financial Ratios as Predictors of Failure" which is widely regarded as being the forefather of modern insolvency prediction literature. Encouraged by the findings of the earlier studies of Fitzpatrick (1932), Smith and Winakor (1935), and Merwin (1942), Beaver hoped to incorporate the potential of the ratio as a predictive tool, with the liquid reservoir theoretical explanation of insolvency provided by Walter (1957) discussed in section 3.2 of this thesis. The primary objective of Beaver's work was to determine the usefulness of financial ratios in the general sense, but declared that ratios could only be regarded as being "useful" if successfully tested in relation to some particular purpose, in this instance, he chose the prediction of company failure. Beaver (1966) warrants particular discussion as I feel it provides a sound platform for the understanding of the proceeding evolution of the literature.

Having gone through the process of data collection for my own empirical research, I begin with expressing my highest regard to Beaver and, indeed, all authors of similar studies around this time. They were able to gather, analyse and process voluminous amounts of financial data without the use of modern tools such as computers and online databases, which has proved very time consuming and laborious for myself even with these recent technological advancements. Using Moody's Industrial Manual[40], Beaver gathered a sample of financial ratios from 79 failed firms over an eleven year period (1954 to 1964 inclusive) and a comparator selection of non-failed firms. The selection of non-failed firms was made according to a matching pair analysis based on industry and firm size, that is to say, that for every failed company in the sample, a non-failed firm was selected from 12,000 leading US corporations matched on the basis of both industry and firm size. This methodology of sample matching has been adopted by many studies to date .

Littleton (1926) was the first to provide empirical evidence that ratio distributions are different between industries and the theoretical underpinnings for the matching pair design may be traced back as early as Bliss (1923) who suggested that for any complete ratio analysis, consideration of industry factors must be included.

The matching of firm size is also important because, as Beaver (1966) posited, "there are certain statistical reasons for believing that asset size alters the relationship between ratios and failure". He suggests that even if two companies have identical ratios, if one company is larger than the other then that company will have less of a chance of failure than the smaller one. Statistical evidence of the time supports this idea and showed that the return on assets is subject to less variability, the larger the size of the company.[41] A more recent study by Falkenstein et al. (2000) shows that the distributional properties of listed and private firm ratios are significantly different: "A model fit using public data will deviate systematically and adversely from a model fit using private data, as applied to private firms." (p.46).

Initially, Beaver intended to simply compare the ratios of each matched pair. By using this paired approach however, Beaver suggests that there was one "serious drawback". He noted that this particular method is not able to draw inferences regarding a single observation but can only show the relationship between pairs of observations as found in a previous study of his, where he could only make restrictive statements using matched pairs such as "A is more solvent than B" which leads to the question: but how

[40] Moody's Industrial Manual contained financial statement data for industrial, publicly owned US corporations.

[41] For example, Alexander (1949).

solvent *is* B? Such ambiguity lead to Beaver dropping the idea of matched pair analysis in favour of analysing the samples as a whole, where averages (or profiles) of each ratio were compared between the two groups of firms[42].

From a pool of thirty financial ratios selected on the basis their popularity in nineteen financial analysis texts[43], the thirty ratios were classified into six common element groups and one ratio from each group was selected for further analysis[44]. The six ratios chosen were selected so that each ratio provided; (a) as much additional information as possible, (b) reflected the liquid reservoir theory described by Walter (1957), and (c) provided the best results in mean or "profile" tests.

The final ratios selected for analysis (representing group in parenthesis) were; cash-flow to total debt (cash-flow ratios), net income to total assets (net income ratios), total debt to total assets (debt to total asset ratios), working capital to total assets (liquid asset to total asset ratios), the current ratio (liquid asset to current debt ratios), and the no credit interval (turnover ratios). [45]

Using the liquid asset reservoir theory as a backdrop for the insolvency process, it is possible to make inferences regarding the differences between the mean ratio values of the failed firm group and the values obtained from the non-failed firms. For example, if I refer to the cash-flow to total debt ratio, I would expect that the non-failed group of firms would have a higher mean profile than that of the failed firm group. The reason for this is that the liquid reservoir theory dictates (with respect to this ratio and all else being held constant) that a particular firm is more solvent than the next if either the reservoir inflows are greater, or that the reservoir outflows are smaller. For the cash-flow to total debt ratio, cash-flow acts as the inflow and total debt (more precisely the debt repayments) act as the outflow, stipulating therefore that the mean ratio should be

[42] Very large companies were omitted from the analysis in accordance with the idea that very large companies have little or no variability in their ratios.

[43] Beaver suggests that the popularity of those ratios might be detrimental to the study as their popularity may attract the most amount of management manipulation.

[44] Some were mere transformations of others and many shared common numerators or denominators so only one from each group was selected.

[45] The No-Credit interval is the time a company can continue its operations from its immediate assets if all other forms of short term finance are cut off. Measured in days it is more directly it is defined as (immediate assets (current assets – stock) – current liabilities) divided by daily operating costs excluding depreciation.

higher for the non-failed group of firms. The inferences for each of the six ratios are presented in Table 5.1. By analysing and comparing the ratio means for each group for each year up to five years before failure, Beaver was able to confirm these predictions which are shown graphically in Figure 5.1.

Table 5.1: Beaver (1966) expectations of ratio performance

Ratio	Prediction of mean values
Cash-flow to total debt (CFTL)	Non-Failed > Failed
Net income to total assets (NITA)	Non-Failed > Failed
Total debt to total assets (TLTA)	Failed > Non-Failed
Working capital to total assets (WCTA)	Non-Failed > Failed
Current ratio (CR)	Non-Failed > Failed
No credit interval (NCI)	Non-Failed > Failed

Figure 5.1 demonstrates that indeed there are noticeable differences between the samples of failed and non-failed firms suggesting that financial ratios may be possible to predict impending firm failure. These illustrations, however, provide no information on the distribution or variation of the data. For a particular ratio it would be a perfect system if all failed companies have a value lower than a particular number (cut-off point) and non-failed firms are higher in value (or vice versa), the overlapping distributions of the data however almost always ensure that this is not the case. Given this principle it can be seen that wherever the cut-off point is placed, some firm's profiles will appear the wrong side of the "line" thus causing a misclassification error. The underlying necessity of any prediction model is to find an appropriate position to place the cut-off point which minimises these misclassifications.

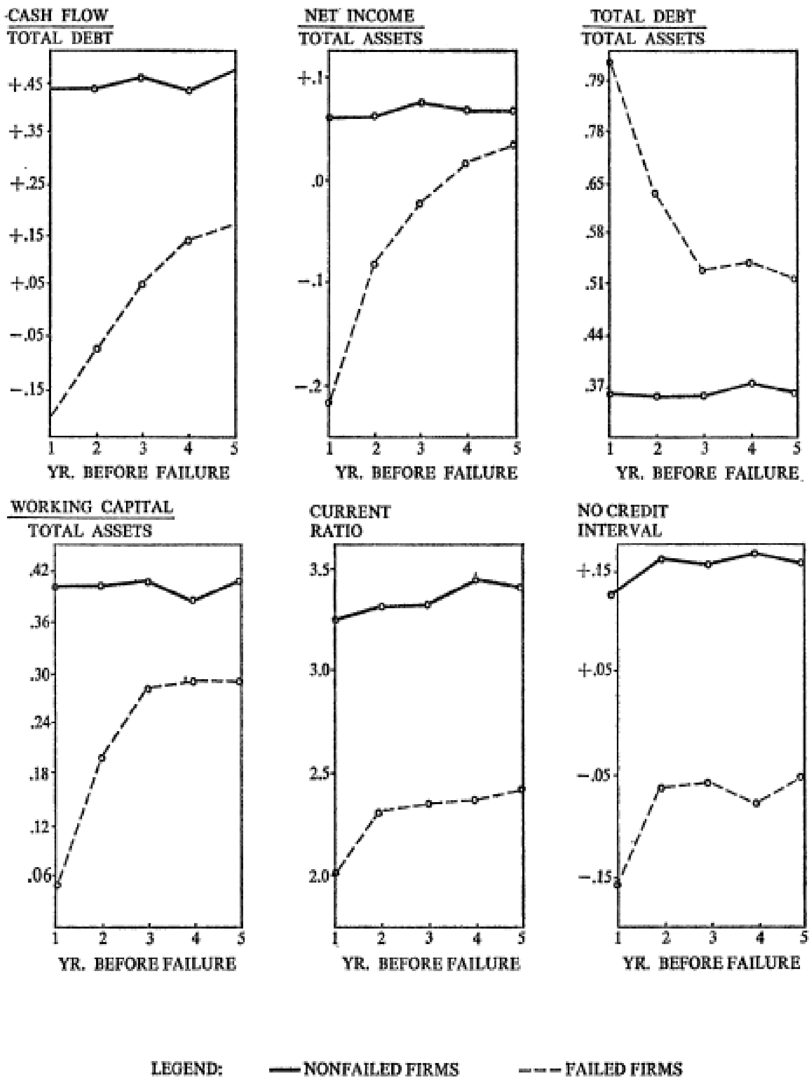


Figure 5.1: Graphical representation of Beaver's six ratio means.

Source: facsimile from Beaver (1966, p.82).

Misclassifications in any statistical process can be of two types. A type I error is rejecting a null hypothesis which is true and a type II error when you fail to reject a null hypothesis which is false. In the case of insolvency prediction the null hypothesis is usually that the company has failed[46]. Therefore predicting a firm to be non-failed when it actually failed would be a type I error and predicting a firm as failed when it has not failed would be a type II error.

[46] This is not a requirement but is used, one suspects, to be consistent with the seminal work of Altman (1968).

The importance of determining a suitable cut-off point which minimises these errors can be explained by use of the small example of a credit lending institution. The institution wishes to assess their client's abilities to meet their repayment obligations should a loan be granted. A prediction model is being developed to determine which clients (if any) will become insolvent in the near future. The priority of the institution is to not lend any money to those firms which it suspects will become insolvent and default on its obligations i.e., reduce the number of type I errors. It would be possible to completely remove any chance of such an error by raising the cut-off point of the model so high the institution in fact will not lend to any firm thus failing to provide any loans to worthy applicants (type II errors) and not doing any business at all. Thus, although reducing type I errors is of high importance it must not be done at the expense of creating an unnecessary number of type II. A careful balance must be sought where the institution must be prudent with its lending without being too restrictive on its overall operations, placing a high importance on the determination of the appropriate cut-off point of the prediction model.

In his analysis, Beaver (1966) found that the distributions of his best predicting ratio cash-flow to total debt showed an increasing overlap between the sets of firms the further back in time the comparison was made. Once the appropriate cut-off point was established, the ratio's one year prior to failure provided an 87% accuracy in predictive ability but deteriorated to just 78% using data five years prior to failure, indicating that the further back in time you go, the harder it is to find any distinction between the groups of failed and non-failed companies. This is somewhat expected as many factors influence the life-cycle of a company. Two companies which had identical profiles twenty years ago will, over the years, see their profiles change reflecting their environment and management team strategies.

A note must be made here of a critical assumption made by Beaver and other insolvency prediction model authors: that all accounting data is normally distributed and which is a necessity for some of the statistical techniques to hold. Evidence suggests that this property is extremely unlikely (and impossible in some instances); but it may be possible, however, to transform the data to normal[47]. This matter will be discussed in more depth later in this thesis.

Although the evidence provided by Beaver indicates that individual financial ratios can be used to measure differences between sets of failed and non-failed companies, he suggests that the predictive ability of these ratios may be overstated. Two good reasons for this overstatement stem from the fact that financial ratios were, and still are, widely used as measures of insolvency. If these ratios are used to determine the financial solvency or "illness" of a company, there will be many cases where this illness is detected sufficiently earlier enough that the company can be saved. The companies which appear in the sample of failures will largely include those whose problems were not detectable by ratio analysis. Alternately, credit lending institutions may fail to grant credit to companies who do not meet the required profile standard thus influencing the future solvency of the company in question. Thus Beaver suggests that the use of ratio analysis for predictive purposes should be used with some caution. He suggested that insolvency prediction using financial ratios would perhaps be better served using a multivariate approach instead of his single ratio methodology, using a combination of several ratios and their changes over time. Attempts by Beaver to develop such a model proved fruitless and seemed to provide no greater predictive ability than his single best predictive ratio, cash-flow to total debt.

Beaver (1968) extended this work to include market-based comparators to the accounting ratio predictor and investigated two key areas. Firstly, Beaver determined that the extent of the change in market pricing of stocks could also be used to predict failure. Secondly, investors have reliance upon the financial ratio when assessing the solvency position of firms.

To determine whether market data could be used in a similar fashion to the financial ratio in order to detect impending firm failure, Beaver (1968) analysed five years of return data for the same sample of firms as Beaver (1966). He showed that in each of the five years before failure, the median of the failed returns is below that of the non-failed, and that this difference increases as the failure date draws nearer. Table 5.2

[47] See Deakin (1976) and Watson (1990) for examples. Unless otherwise stated, this paper will assume that the normal distribution of these ratios applies.

demonstrates this finding along with the observation that "although the non-failed firms exhibit no upward or downward trend, the median of the failed firm's drops over time with the largest price decline occurring in the final year".

Table 5.2: Comparison of median values of stock returns

Year before failure	Medians		
	$\bar{F}$	F	Difference
5	.02	.02	+.00
4	.02	−.03	+.05
3	.11	.00	+.11
2	.12	−.08	+.20
1	.03	−.26	+.29

$\bar{F}$ denotes non-failed firms and F denotes failed firms.

Source: facsimile from Beaver (1968, p.182).

Beaver noted that the performance of the market returns coincided with the results obtained from his earlier ratio analysis, and that investors appear to adjust to the new solvency positions of the firms continuously over time. The largest unexpected deterioration still occurs in the final year before failure, indicating that investors are still surprised at the occurrence of actual failure. This similarity with the ratio performance is evident suggesting that there is a relationship between returns and ratios, which is unsurprising as investors are likely to incorporate financial ratio analysis into their investment appraisal methodologies.

To what extent investors use, or rely upon, financial ratios as part of their investment appraisal, Beaver undertook a time series analysis to determine the comparative failure prediction lead times between market returns data and four financial ratios; cash-flow to total liabilities, net income to total assets, total debt to total assets, and working capital to total assets. Table 5.3 presents Beaver's results in which the number of firms correctly predicted as failing for each of the five years prior to actual failure is reported.

Failure was predicted at a time when the ratio or return showed a substantial deterioration in solvency.

"This point in time was operationally defined as the year when the first large drop in the market price occurred. A large drop was defined as a negative rate of return so large that very few non-failed firms ever incurred such a return during the time period studied. Examination of the data indicated that if the return is equal to -0.50, only six non-failed firms fell below either of these values and all but six failed firms fell below one of these values at least once. The analysis was based upon 42 non-failed and 34 failed firms for which return data were available for all five years before failure. The point in time at which the financial ratios predicted failure was similarly defined. For the cash flow to total debt ratio, the year failure was forecast was defined as the year in which the ratio fell below 0.02 and never rose above 0.08 subsequently. The critical values for the net income to total assets ratio were -0.01 and 0.05, respectively; for the total debt to total assets ratio, 0.65 and 0.45; and for the working capital to total assets ratio, 0.10 and 0.30" Beaver (1968, p.187).

Table 5.3: Year in which failure was predicted

Year before failure	Number of firms for which failure prediction was made				
	Market	*CF/TL*	*NI/TA*	*DBT/TA*	*WC/TA*
5	6	4	5	5	3
4	4	2	2	1	1
3	3	5	4	1	0
2	6	6	8	3	2
1	4	8	6	4	5
0[a]	6	4	4	15	18
Total forecasts	29	29	29	29	29
Mean (years)	2.45	2.17	2.31	1.45	.97

[a] This row indicates number of failed firms for which the variable never predicted failure.

Source: facsimile from Beaver (1968, p.189).

Beaver's results show that investors forecast firm failure earlier than that of the ratio which is consistent with the notion that analysts rely upon ratios in their decision making process. This evidence therefore suggests that investors rely upon sources of information other than the financial ratio in order to aid their decision process, or that the ratio may not be considered in isolation but as a combined effect/deterioration of multiple ratios. Beaver himself suggests that "...research in progress tentatively suggests that a multi-ratio model, consisting of the most recent value of the cash-flow ratio and the first differences of the previous values, possesses greater predictive power than any single ratio.... more knowledge is needed about the non-ratio information investors rely upon and how investors use the ratio information in a multivariate context."

5.2 1968-1980: the step from univariate to multivariate discriminant analysis

5.2.1 Introductory comment and overview

The move away from univariate analyses towards combinations of variables into a multivariate function is perhaps, given hindsight, an obvious progression. In 1968 Altman revolutionised corporate insolvency prediction when his seminal paper "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy" which for the first time combined several ratios into one single linear function, the Z-score, popularising the art and pushing it into mainstream financial literature.

Z-score models consist of a linear combination of variables estimated through Multivariate Discriminant Analysis (MDA)[48]. MDA was initially developed in the 1930s for use in the field of biological and behavioural sciences and its methodology was first described in Fisher (1938). MDA is in essence a reversed multiple analysis of variance (MANOVA). MDA classifies a particular observation into a particular category or *a priori* group; in this case there are two groups, failed firms and non-failed firms[49]. The outcome is a linear combination of the different data characteristics which best discriminates between the groupings; the characteristics in this particular case are usually financial ratios or other qualitative or categorical variables.

The scores obtained from the resulting discriminant function allow for an ordinal ranking of firms, and the higher the score usually denotes the better the predicted solvency[50]. The application of a cut-off point score then divides firms between those predicted to fail, and those predicted to survive (if a firm scores higher (lower) than the

---

[48] The functional form does not have to be linear (e.g., Altman et al., 1977 use quadratic discriminant analysis). Linear discriminant analysis, however, is by far the most popular method in prior literature and, therefore, we use 'MDA' to refer to the linear methodology.

[49] A full derivation of the discriminant process, function, and maximum likelihood estimation technique can be found in Appendix C.

[50] Studies such as Ooghe et al. (1994) define their model alternatively, higher scores denoting higher probability of failure. Later in this thesis, I change the signs on my tested Z-score models not only to match such ranking systems but also directly to compare these to other methodologies which, instead of rankable scores, provide probability of failure estimates. A later section of this thesis explains this transformation in more detail..

cut-off point, then it is classified as being predicted non-failed (failed)). The choice of the cut-off point is somewhat arbitrary and leads to several restrictive assumptions and requirements associated with MDA models.

First, many researchers have chosen cut-off points in order to attempt to minimise the overall error rate of the model which may result in the misleading or incorrect estimates of model accuracy – although the differing economic impacts and values attached to the individual type I errors (incorrect classification a failing firm as a non-failing firm) and type II errors (incorrect classification of a non-failing firm as failing) has often been acknowledged (Edmister, 1972; Eisenbeis, 1977; Deakin, 1977; Zavgren, 1983). Some studies such as Altman (1977) investigate differing error costs and others such as Taffler (2008) look at economic value when misclassification costs are different. Second, the classic insolvency prediction paradigm is that the dependent variable is dichotomous, treating failure as being discrete, non-overlapping and identifiable; and thus not reflecting the true nature of financial distress and the various insolvency procedures which may or may not ensue. Third, two statistical requirements of the MDA model are of multivariate normality and that the two *a priori* groups to be separated have identical variance-covariance matrices. These statistical requirements are rarely satisfied by the data, thus concerns as to bias pertain.

A consequence of ignoring the foregoing is that the MDA modelling technique is often applied in an inappropriate way, with the resulting models being unsuitable for generalization (Joy and Tollefson, 1975; Eisenbeis, 1977, Balcaen and Ooghe, 2006). These issues are addressed further in section 5.5 of this thesis.

5.2.2 Development of the Z-score models: Altman (1968)

Edward Altman, Professor of Finance at New York University, encouraged by the findings of Beaver (1966) remarked that there was a "definite potential of ratios as a predictor of bankruptcy" (Altman, 1968), when developing what is probably the most widely recognised insolvency prediction model, the multivariate "Altman's Z-score".

For the purpose of his analysis Altman purported that the most appropriate way to develop a multivariate model would be to use multivariate discriminant analysis (MDA). MDA is a statistical technique which classifies a particular observation into a particular *a priori* group; in this case there are two groups, failed companies and non-failed companies. By collecting data for all observations within mutually exclusive groups, MDA is used to establish a linear combination of the different characteristics of the data which best discriminates between the groupings; the characteristics in this case are financial ratios.

By considering an entire profile of ratios it is possible to compute which are the predominant determinants of the group differences, not only on an individual ratio basis but by constructing a variance-covariance matrix one can observe each ratio's interaction with each other and ratios which are have high correlation with each other can be reduced to a single variable. By using this technique Altman (1968) suggested it would remove some possible ambiguities and misclassifications that where observed in earlier multivariate attempts which used integer and linear programming. The MDA methodology results a single linear function which best classifies the data into its two groups with the form:

$$Z = V_1X_1 + V_2X_2 \dots \dots + V_nX_n \tag{5.1}$$

where V_n are the discriminant coefficients and X_n are the independent variables (ratios).

A full derivation of the discriminant function along with any associated statistical assumptions is provided in Appendix C of this thesis. This derivation has particular relevance to the critical assessment of MDA which pertains within the following sections of this study.

To develop his model Altman (1968) gathered data from 33 failed manufacturing companies during the twenty year period 1946-1965[51] (that filed a bankruptcy petition under Chapter X of the National Bankruptcy Act during the period), and following the methodology of Beaver (1966), selected a matched sample of 33 non-failed companies

[51] As ratios tend to vary over time, Altman was somewhat reluctant of his restricted choice.

(based on firm size and industry) whose financial data was collected from Moody's Industrial Manual. Large companies were omitted due to their high improbability of failure at that time.

Regarding the selection of appropriate variables, Altman (1968) derived a list of 22 potential variables that were found to be significant indicators of financial distress based on previous studies (although the single best ratio as determined by Beaver (1966), cash-flow to total debt, was not selected as an appropriate variable because Altman considered there to be a lack of consistency and imprecise data on depreciation measures used in its calculation). The 22 ratios were then categorised into five distinct categories, liquidity, profitability, leverage, solvency and activity. By repeatedly running the MDA analysis exploring different combinations of these ratios, their individual significance, inter-correlations and the predictive accuracy of each combination, a final function was derived as being the best discriminator between the two groups of companies which is now known as the Altman Z-score:

$$Z = 0.012X_1 + 0.014X_2 + 0.033X_3 + 0.006X_4 + 0.999X_5 \tag{5.2}$$

where

X_1 = working capital / total assets (WCTA) (%),
X_2 = retained earnings / total assets (RETA) (%),
X_3 = earnings before interest and tax / total assets (EBITTA) (%),
X_4 = market value of equity / book value of total liabilities (VETL) (%),
X_5 = sales / total assets (STA) (number of times), and
Z = the overall index or "Z-score".

The more common iteration of the formula forgoes the percentage terms and is found expressed in the more convenient form:

$$Z = 1.2X_1 + 1.4X_2 + 3.3X_3 + 0.6X_4 + X_5 \tag{5.3}$$

The expectation for each of the ratios considered is that they will be higher in value for the non-failed firm thus the lower the score, the greater the chance is that the company will fail. A problem thus arises as to how determine an appropriate cut-off point for classification; that is to say, where should the line be drawn (score X) on this scoring system that categorically designates a firm into either the predicted non-failed class of firms should its associated score be above the cut-off point X, or into the predicted failed class of firms should its score be below X.

5.2.2.1 Determining the optimal cut-off point X

Table 5.4 is a contingency table (which for the remainder of this thesis shall be termed a confusion matrix in accordance with the majority of literature in the field of predictive analytics). The confusion matrix (sometimes termed the matching matrix) is used as a visualisation tool, making it easy for the researcher to determine whether the model is confusing two or more classes.

Table 5.4: The confusion matrix

	ACTUAL		
PREDICTION	Failed	Non-failed	TOTAL
Failed	TP	FP	TP+FP
Non-failed	FN	TN	FN+FP
TOTAL	TP+FN	TN+FP	TP+FP+TN+FN

The predicted class of an observation (as determined by the model) is compared to its actual class membership (observed). The observations are each classified as to whether or not the model successfully predicted their correct class membership. TP (true positive) states the number of observations which were correctly predicted as being a failed company. FP (false positive) states the number of observations which were incorrectly predicted as failed but actually did not, (also known as a type II error). TN (true negative) states the number of observations which were correctly predicted as being a non-failed company. FN (false positive) states the number of observations which were incorrectly predicted as non-failed but actually failed, (also known as a type I error).

To determine an optimum cut-off point for a particular model, a cut-off must be selected in order that both the type I and type II error rates are kept to a minimum. The error rates are calculated as the proportion of incorrect classifications for each of the classes, thus the type I error rate is defined as FN / (FN+FP), and the type II error rate is defined as FP / (TP+FP). Traditionally a balance is sought which minimises the overall error rate, i.e. FP+FN / (TP+FP+TN+FN) treating the cost of each type of error as being the same. In practice however an equal cost of errors is unlikely to be satisfied which can be explained by example of a credit lending institution.

The institution wishes to assess their client's abilities to meet their repayment obligations should a loan be granted. A prediction model is being developed to determine which clients (if any) will become insolvent in the near future. The institutional priority is to not lend any money to those firms which it suspects will become insolvent and default on its obligations i.e., reduce the number of type I errors. It would be possible to completely remove any chance of such an error by raising the cut-off point of the model so high the institution in fact will not lend to any firm thus by failing to provide any loans to worthy applicants (type II errors) and not doing any business at all. Thus, although reducing type I errors is of high importance, it must not be done at the expense of creating an unnecessary number of type II errors. A careful balance must be sought where the institution must be prudent with its lending without being too restrictive on its overall operations, placing a high importance on the determination of the appropriate cut-off point of the prediction model.

Section 8.3 of this thesis discusses some alternate methods for model evaluation in the situation where these misclassification costs do differ in their economic value, and whether the difference in economic error costs has a significant impact when comparing a selection of insolvency prediction models. For now though the discussion follows with the assumption that these costs remain equal as we continue to follow the evolution of the literature.

Through use of the confusion matrix, Altman (1968) was able to determine that companies having a Z-score above 2.99 clearly fall into the non-bankrupt group whilst those with a score below 1.81 are all bankrupt. However, there is a grey area or "zone of ignorance" which lies between the two values, and that in this area, classification must be treated with caution; although it is probable that the company is suffering some form of financial distress, even if it has not yet lead to formal bankruptcy procedures.

By calculating the score for each of the 66 companies from the accounts one year prior to failure, each company could be classified into one of the groups. The classifications (or predictions) were compared to the actual failed or non-failed status of the companies and the total predictive accuracy of the model was established. Altman (1968) presents the results in a style similar to the now conventional confusion matrix, incorporating hits (H) for correct prediction and misses (M) for incorrect prediction. M1 corresponds to type I errors and M2 for type II. (Table 5.5). By using a cut-off point of 2.99 the initial sample provided an overall combined accuracy of 95% correct classification. (Table 5.6)

Table 5.5: Altman's confusion matrix

	Predicted Membership	
Actual Membership	Bankrupt	NonBankrupt
Bankrupt	H	M1
NonBankrupt	M2	H

Table 5.6: Prediction results for Altman's Z-score

	Predicted Membership	
Actual Membership	Bankrupt	NonBankrupt
Bankrupt	31	2
NonBankrupt	1	32

The model was tested again with accounting data from the same sample in the second, third, fourth and fifth years prior to failure. Beaver (1966) and others[52] had suggested that a firm exhibits signs of failure as much as five years prior to the actual failure date although Altman (1968) suggests that there is little mention of the true significance of such statements. Altman's results show a mark drop of in the accuracy of his model over five years prior to failure, from years one to five the percentage of companies that were predicted correctly were 95%, 72%, 48%, 29% and 36% respectively. He concluded that the Z-score has little or no predictive ability when the

[52] FitzPatrick (1932) and Merwin (1942) for example.

lead time is larger than two years and that the change in results from year to year has little or no meaning past this point.[53]

To test the predictive ability of his model further, Altman (1968) applied it to a hold-out sample of companies. This follows an earlier paper by Frank et al. (1965) who described two validation procedures for determining an unbiased estimate of the predictive power of a multivariate discriminant function. They suggested that the split or "holdout" sample approach was the preferred method when a large sample is available[54]. Altman selected 66 firms, within the same range of firm size as his training sample, but this time the sample was chosen so that a large majority of the companies has suffered negative profits in the previous 2-3 years (symptoms of financial distress). For the 25 failed companies whose firm sizes were in the same range as the first sample, 24 (96%) were identified correctly by the model (a type I error rate of 4%). For the sample of non-failed companies, 14 were unsuccessfully classified by the model (31% type II errors), although Altman noted that ten of them lay within the grey area with scores of between 1.81 and 2.99, suggesting that they were suffering from some form of financial distress even though they had not met the legal definition of insolvency.

Altman had successfully demonstrated that multiple financial ratios could be combined into a single linear function whilst improving on the accuracy achieved by Beaver's best performing ratio, cash-flow to total debt, for a one-year-ahead forecast. Correct predictions made by Altman's model however, deteriorated considerably when evaluating firm health in a timeframe any further than two years prior to failure. Both Beaver (1966), and Altman (1968), sparked widespread interest in the field of corporate insolvency which remains embedded in the art to this very day. The next section underlines the contribution made by these two authors by discussing the literature from the following decade, the nineteen seventies, and how, despite advances in methodological technique, computer processing power, and credit risk theory, the MDA methodology (particularly the Altman Z-score) has remained a regular contestant for academic study and benchmarking to the present day.

[53] A later model by Altman, the Zeta score, is discussed later in this dissertation which proposed a higher accuracy over a longer period of time.

[54] The second method (the simulated sample approach) is treated later in Edminster (1972).

5.2.3 Further application and adaptation of the Z-score model

The first notable paper to assess the work of both Beaver (1966; 1968) and Altman (1968), was Deakin (1972). Deakin's study sought to evaluate the methodologies used in the aforementioned papers and propose an alternative model which predicts failure as early as possible. He noted that "*Although Beaver's empirical results suggest that his method has greater predictive ability, the method used by Altman has more intuitive appeal*". In his study he analysed the same fourteen ratios as Beaver (1968), but conducting the analysis a smaller sample of failed firms than Beaver (32 failed compared to Beaver's 78). Although the sample period adopted by Deakin was just six years (1964-1970) compared to Beaver's ten (1954-1964), the criteria required to meet the definition of failure as proposed by Deakin was fundamentally different and more in tune with Altman's definition (filing a bankruptcy petition under Chapter X of the National Bankruptcy Act).

"…the term failure was defined here to include only those firms which experienced bankruptcy, insolvency, or were otherwise liquidated for the benefit of creditors. Beaver included firms which defaulted on loan obligations or missed preferred dividend payments. But unless the nonbankrupt firms were matched by debt structure as well as by size and industry, there could be a potential bias in certain of the ratios". Deakin (1972).

Table 5.7 shows the results of Deakin's analysis and the error rate for each of the ratios from one to five years prior to failure with Beaver's results in parentheses. The Spearman's rank correlation coefficient denoted by r_s describes the relationship between the relative predictive ability of each variable across the two studies.

Table 5.7: Fourteen variables: Deakin vs Beaver

Percentage Error for 14 Ratios on Dichotomous Classification Test

Ratio	Year before failure				
	5	4	3	2	1
Non-liquid asset group:					
cash flow/total debt	27 (22)	24 (24)	28 (23)	16 (21)	20 (13)
net income/total assets	38 (28)	43 (29)	30 (23)	20 (21)	27 (13)
total debt/total assets	33 (28)	33 (27)	22 (34)	22 (25)	20 (19)
Liquid asset to total asset group:					
current assets/total assets	47 (49)	45 (47)	41 (48)	47 (48)	36 (38)
quick assets/total assets	52 (40)	52 (48)	47 (36)	53 (42)	34 (38)
working capital/total assets	34 (41)	39 (45)	34 (33)	28 (34)	30 (24)
cash/total assets	47 (38)	44 (36)	47 (30)	42 (29)	33 (28)
Liquid asset to current debt group:					
current assets/current liabilities	41 (45)	36 (38)	25 (36)	27 (32)	28 (20)
quick assets/current liabilities	44 (37)	39 (34)	41 (40)	28 (32)	30 (24)
cash/current liabilities	41 (38)	45 (38)	36 (36)	28 (28)	36 (22)
Liquid asset turnover group:					
current assets/sales	50 (51)	56 (49)	52 (48)	61 (51)	48 (44)
quick assets/sales	53 (44)	59 (52)	55 (45)	59 (47)	52 (46)
working capital/sales	47 (40)	52 (46)	45 (42)	31 (33)	28 (26)
cash/sales	48 (45)	53 (43)	52 (36)	42 (24)	47 (34)
r_s	.76	.80	.56	.82	.88

Data in parentheses are from Beaver (1968, p. 118).

Source: facsimile from Deakin (1972, p.169).

From the results it is clear that there is a definite correlation between the two studies with the highest rank correlation coefficient of 0.88 coming from a one-year-ahead forecast, although the correlation is much lower at three years than at any other point. Deakin commented on the apparent drop in correlation at three years. Noting a considerable rise in the total assets of failed firms in the third and fourth years before failure, he suggested that a rapid expansion of the fixed asset base, financed by increased debt rather than common stock or retained earnings, led to the failure of these firms if they were unable later to generate sufficient sales and net income. Another notable difference between the two sets of results is the cash to sales ratio.

"While most of the other ratios tended to be consistent with those observed by Beaver, this ratio was consistently and significantly different. One possible explanation is that corporations tended to invest more of their cash reserves during the late 196O's when interest rates were high. Thus a low cash/sales ratio may have been due to good money management rather than to general company mismanagement". Deakin (1972).

Following the MDA methodology employed Altman (1968), Deakin inputted the fourteen variables into his "discriminant analysis program", combining the variables into single linear functions (in this case, one distinct function for each of the five years before failure). Having a separate model for each of the five years is far less appealing than a function, such as Altmans, which is applicable to, and encompasses all years of data. We need to ask ourselves how are these models to be applied in practice? Does each company need to be tested on each model individually? How are the results to be interpreted if a company scores well in some years and not so well in others?

Practical applications aside, Deakin's results nevertheless indicated that each of the variables added significant value to each model beyond the 0.001 critical level for the first three years before failure, and were still significant in years four and five at the 0.01 and 0.05 critical levels respectively. This finding indicates that each of the fourteen ratios considered by both Beaver and now Deakin, (considered at inception both because of their popularity in literature and a previous pilot study by Beaver), have significant discriminatory power when separating failed and non-failed firms.

In a deviation from these past methodologies, Deakin forgoes the selection of an error-minimising arbitrary cut-off point to classify his companies. Deakin noticed that the largest number of misclassifications occurred close to the cut-off point of which an example is provided in Figure 5.2.

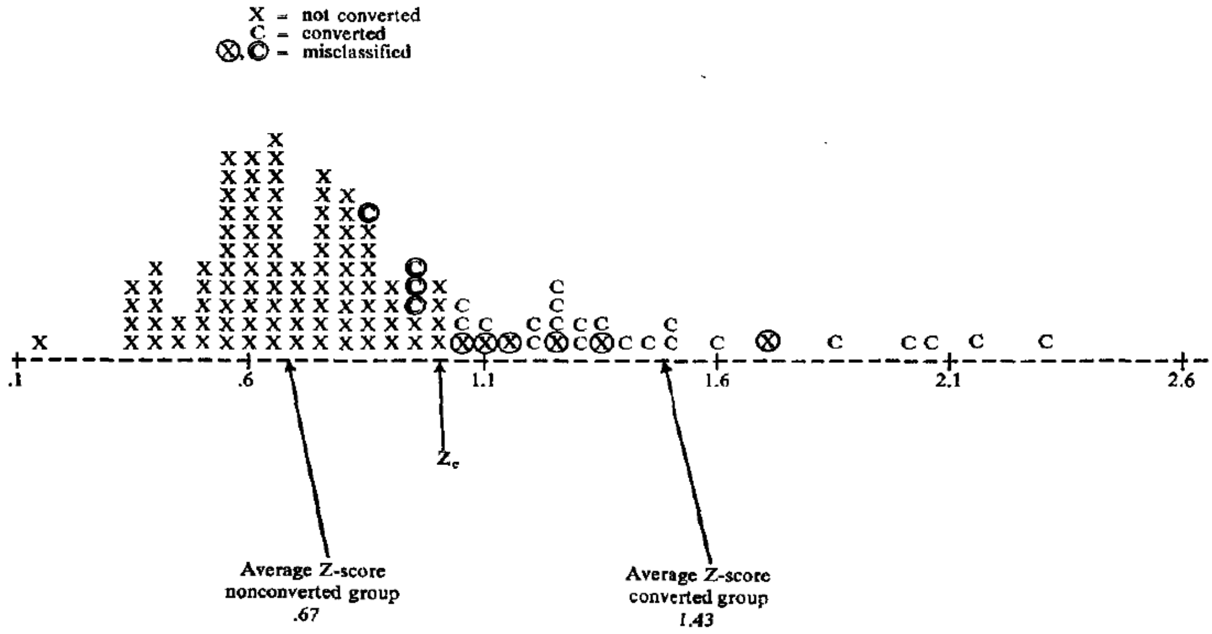


Figure 5.2: Distribution of Z-scores within sample.

Source: facsimile from Frank and Weygandt (1971, p.123).

Instead of the cut-off point, Deakin assigns each company to a particular class by assigning it a probability based on a multivariate extension of the univariate *Z* test. To correctly apply this method, two assumptions must be made. Firstly the vectors of discriminant scores follow a p-variate normal distribution and secondly, the variance-covariance matrix of the two classes matches the variance-covariance matric for the population. If these conditions are satisfied then the following statement is valid and companies are assigned according to the probability distribution.

$$d'\Sigma^{-1}\, d \sim \mathcal{X}_n^2 \tag{5.4}$$

where d' = the row vector of deviations from the mean score

d = the column vector of deviations from the mean score

Σ = the population variance-covariance matrix

n = number of observations

Using this classification method, Deakin's results showed that by using a combination of all fourteen variables he was able to successfully classify 97%, 95%, and 95% of the companies using one, two, and three year-ahead forecasts, with accuracy dropping to 79% and 83% in years four and five, much higher than any ratio evaluated in isolation. These results however, are misleading as they only report firms which were used to construct the model, i.e. the original training sample. In a second test Deakin used a validation sample consisting of 11 failed and 23 non-failed companies to assess his model. The resulting accuracies of 78%, 94%, 88%, 77%, and 85% for the five years before failure respectively, hold up reasonably well despite an obvious deterioration in the first year[55]. I would not expect the accuracies to be quite as high those achieved with the training sample by convention, but the provision of such a small validation sample means that we should not make too many inferences regarding these results.

What we can gain from Deakin (1972) is that by combining variables into a single multivariate function we can expect to improve on the discriminatory power that we would otherwise have had if evaluating each variable in isolation. Although Altman's (1968) multivariate results demonstrated a lower misclassification rate than that of Beaver's (1966; 1968) univariate approach, only one of Altman's variables (debt to total assets) appears in the Beaver paper(s). That is to say that until Deakin (1972) there was no direct comparison between a set of single ratios and there combined functional form, Deakin provided empirical evidence for the first time that the latter yields a greater predictive ability.

The next paper I would like to discuss is that of Edminster (1972), who conducted an interesting study on small business failure prediction. Besides being the first of a limited number of small firm insolvency prediction analyses, the evolutionary merit of this paper comes in three forms, firstly in the use of "trend-level" ratio analysis, and second in model validation techniques. Similar to Altman (1968) and Deakin (1972), Edminster also provides a version of the Z-score model created through MDA but unlike the previous authors Edminster uses a binary coding system for the discriminant variables

Edminster's methodologies focused upon testing four hypotheses: (1) a ratio's level and/or relative level (compared to an industry standard) as a predictor of small business failure; (2) the three-year trend of a ratio as a predictor of small business failure; (3) the three year average of a ratio as a predictor of small business failure; (4) a combination

[55] Deakin himself commented that "…the deterioration in the first year is rather severe and cannot be explained by the presence of any unusual events peculiar to the sample used".

of the industry relative trend and the industry relative level for each ratio as a predictor of small business failure.

The combination of the industry relative trend and industry relative level for each variable (trend-level) is an interesting concept originally suggested by Weston and Brigham (1966) who noted: "*Although Composite Company's current ratio is below the industry average, a credit analyst would not mark the firm down badly on this count, in view of the favourable trend*". A trend-level analysis based on industry may be able to detect patterns and distinctions across data classes which would not be apparent when performing a market-wide company comparison.

Two samples were selected from a population of Small Business Administration (SBA) borrowers over the period 1954-1969. The first (small) sample consisted of 42 borrowers of whom three consecutive years of adequate company data were available. The second (large) sample consisted of 562 borrowers that provided just one year of adequate data. The large sample would be subject to the split (holdout) sample approach following Altman, and the small sample would be subject to the Frank et al. (1965) alternate validation technique, the simulated sample approach, as such a small sample of firms inhibits the meaningful use of holdout sample analysis.

Using 19 common ratios advocated by theorists or found to be significant predictors of business failure in prior studies, Edminster sought to determine their discriminatory power both in single and combined form, reporting that the ratios predictive power is cumulative in that no single ratio predicts nearly as well as a small group. Edminster does not report the finding of his univariate analysis, only those of his discriminant function. Also he was not able to determine with any confidence a discriminant function derived from the large sample. He concluded that:

"…the small business function fails to discriminate when only one statement is available, whereas Altman, and Beaver show that one financial statement is sufficient for a highly discriminant function for large businesses. This leads the author to qualify his conclusion above with the provision that at least three consecutive financial statements be available for analysis of a small business". Edminster, (1972).

For his three-year-statement discriminant function Edminster's MDA program employed a step-wise procedure in which a variable was not allowed to enter the final function if it had a correlation higher than 0.31. Altman (1983) remarked on this approach stating that:

"*A major shortcoming with this approach is that important explanatory power may be excluded from the regression equation because of the arbitrary correlation coefficient cut-off point*".

Another interesting deviation from Altman's methodology was the use of a binary coding system where rather than the variables entering the equation in their raw form; each ratio is transformed into a qualitative zero-one variable based upon an arbitrary cut-off point. Seven variables entered the final equation; the details of each and their cut-off criterion are defined in Equation 5.5.[56]

$$Z = 0.951 - \underset{(4.24)}{0.423X_1} - \underset{(2.82)}{0.293X_2} - \underset{(4.51)}{0.482X_3} + \underset{(2.61)}{0.277X_4}$$

$$- \underset{(2.60)}{0.452X_5} - \underset{(1.68)}{0.352X_6} - \underset{(7.11)}{0.924X_7} \qquad (5.5)$$

$R^2 = 0.74$ F = 14.02   (t-statistics in parentheses)

Z = 1 for a successful (non-loss) business and Z = 0 for a failed (loss) business.

X_1 = Annual funds flow (net profit before tax plus depreciation) to current liabilities. It equals one if the ratio is less than 0.05 and zero otherwise.

X_2 = Equity to sales ratio. It equals one if the ratio is less than 0.07, zero otherwise.

X_3 = Net working capital to sales ratio divided by its respective RMA average. It equals one if the ratio is less than -0.02 and zero otherwise.

X_4 = Current liabilities to equity divided by its respective SMA average. It equals one if the ratio is less than 0.48 and zero otherwise.

[56] SBA refers to the SIC industry average as per the Small Business Administration. RMA refers to the industry average as per Robert Morris Associates' Annual Statement Studies.

X_5 = Inventory to sales ratio divided by its respective RMA average. It equals one if the ratio shows an upward trend but remains less than 0.04. Zero otherwise.

X_6 = Quick ratio divided by its respective RMA average quick ratio. It equals one if the ratio shows a downward trend and its level is less than 0.34 just prior to the loan.

X_7 = Quick ratio divided by its respective RMA average quick ratio. It equals one if the ratio shows an upwards trend and zero otherwise.

Edminster's function was able to correctly classify 39 out of 42 private companies (93%) within his sample, and provided an interesting discussion as to the determination of the required cut-off point.

He noted that the cut-off should be selected so as to equate the probability of type I and type II errors with the ratio of the explicit cost of accepting a failure to the opportunity cost of rejecting a success. Edminster represented this by using the following equality:

$$\frac{P_{II}}{P_I} = \frac{MC_I}{MC_{II}} \tag{5.6}$$

where:

P_{II} = the probability of rejecting a loan to a successful firm (type II error),

P_I = the probability of accepting a loan to a failing firm (type I error),

MC_{II} = the marginal opportunity cost of rejecting a loan to a successful firm, and

MC_I = the marginal actual cost of accepting a loan to a failing firm.

Anderson (1962) shows us that the optimal cut-off point under certain assumptions (equal variance-covariance matrices and multinormal populations) is therefore:

$$CP = \ln \frac{q_{\mathrm{I}} C_{\mathrm{I}}}{q_{\mathrm{II}} C_{\mathrm{II}}} \tag{5.7}$$

where:

CP = the optimal cut-off point,

q_I, q_{II} = the prior probability that a firm will be either classified as failed (q_I) or non-failed (q_{II}),

C_I = the cost of accepting a loan to a failing firm

C_2 = cost of rejecting a loan to a successful firm

One problem arising when trying to determine the optimal cut-off point is, therefore, what costs are attributable to each type of error? The associated costs are dependent on who the user is, what costs are relevant to the user, and how these costs should be determined. Section 8.3 covers this issue in some detail when I evaluate a selection of models based on their economic value when misclassification costs differ.

Edminster noted the impracticality of this cost determination and proposed a practical alternative which he termed the "black-gray-white" method, reservedly treating misclassification costs as being equal.

Figure 5.3: The black-gray-white method.

Source: facsimile from Edminster (1972).

A sample of Z-score frequencies may appear as in Figure 5.3. Scores less than 0.47 (the black area) are attributable only to failed firms, and therefore rejecting all loan applicants with scores less than this can only benefit the lender. Scores above 0.53 (firms within the white area) should be accepted. Firms falling within gray area (similar to Altman's zone of ignorance) should be subject to additional analyses over and above the Z-score. Edminster suggested:

"*Nonfinancial statement information regarding management, markets, plant and equipment, and other factors would be scrutinized in order to classify the applicant as a loss or nonloss borrower*".

Edminster's optimal cut-off score of 0.53 represents a need to minimise loan losses, with the lender unable to grant a loan to any firm with a score below point.

To conclude the contribution made by Edminster, I refer the reader back to section 4.2.2 of this thesis which described the number of private-firm insolvencies over the past fifty years. What is apparent from the insolvency data pertaining to these smaller firms is that they account for approximately 99.998% of all company insolvencies in the UK and 66% of the economic value lost to the process. Although the perceived difficulty in obtaining private-firm data somewhat inhibits a thorough analysis upon this population, there remains a need to identify the characteristics of financially distressed private companies. Edminster was able to show (albeit on a very small number of firms) that discrimination between the two classes of private firms can be achieved with accuracy comparable to that of previous studies. This may however be due his data transformations. Altman (1983) commented on this inherent problem attached to small business failure prediction:

"*…since the financial ratios for small businesses usually are dispersed widely, it is difficult to obtain a meaningful data set without some sort of adjustment*". Altman (1983, p.159).

Blum's (1974) Failing Company Model (FCM) was developed in an attempt to predict failure as defined in the failing company doctrine[57]. One of only a few defences to a US merger prosecution at the time, the doctrine could be invoked when one of two merging companies is failing and the failing company receives no offer to merge from a company with which a merger would have been legal. Blum noted two uncertainties in applying the doctrine: (1) the point at which a company is considered failing, and (2) the manner of ascertaining absence of good-faith purchasers for the failing company. Legally the definition of failure was ambiguously defined as a "grave probability of failure" and Blum's FCM sought to address this uncertainty by developing an MDA model which would quantify this probability.

Blum proposed that such a model should be based on the liquid reservoir of cash-flows as described by Walter (1957) and detailed in section 3 of this thesis. Akin to Beaver (1966), Blum proposed a general framework for variable selection, dividing the reservoir theory into three component parts: liquidity, profitability, and cash-flow

[57] *International Shoe* v. F.T.C, 280 U.S. 291 (1930).

variability. He proposed the use of 12 variables for his model which are presented in Table 5.8 below.

Table 5.8: Blum's (1974) failing company model

I. Liquidity:
- *A*. Short-Run Liquidity
  - Flow: 1. The "quick flow" ratio[a]
  - Position: 2. Net quick assets/inventory
- *B*. Long-Run Liquidity
  - Flow: 3. Cash flow/total liabilities
  - Position: 4. Net worth at fair market value/total liabilities[b]
  - 5. Net worth at book value/total liabilities

II. Profitability: 6. Rate of return to common stockholders who invest for a minimum of three years[c]

III. Variability: 7. Standard deviation of net income over a period
- 8. Trend breaks for net income[d]
- 9. Slope for net income[e]
- 10–12. Standard deviation, trend breaks, and slope of the ratio, net quick assets to inventory; variables 10, 11, and 12 are only used at the first and second year before failure.

---

[a] Cash + Notes Receivable + Market Securities + (Annual Sales ÷ 12) ÷ (Cost of Goods Sold − Depreciation Expense + Selling and Administrative Expense + Interest) ÷ 12.

[b] Fair market value was measured by the harmonic mean of the bounds to the range of stock prices during a year. The implicit weighting system of the harmonic mean is inverse to the size of the observation (the reciprocal of the arithmetic average of reciprocals)—thus speculative upsurges in market value will not be as influential as in the case of the arithmetic mean.

[c] The market rate of return accrues to a common stockholder who bought his shares at an average price at the beginning of a given time span (e.g., from the fifth to the first year before failure) and sold them at an average price during the last year of the span. The rate of return is based on the stockholder's gain or loss and cash dividends received, all adjusted for temporal location by present-value analysis. The internal rate of return to an investor and all variables using net worth at fair market value were adjusted for capital changes. To compare entities more nearly similar, shares issued in mergers or offered to the public were added to prior totals of shares outstanding, adjusted for stock splits and dividends.

[d] A trend break is defined as any performance by a variable less favorable in one year than in the preceding year, such as a decline in income from $10,000 to $1,000 from year three to year four before failure.

[e] Slope of a "trend" line fitted to the group of observations by the method of least squares.

Source: facsimile from Blum (1974, p.16).

The FCM was derived using data from the period 1954-1968 and consisted of 115 failed firms and a matched sample of non-failed firms based on industry, size and fiscal year. Validation was achieved by using the holdout sample approach where half of the sample was used to train the model and derive the discriminant coefficients, and the remaining half used to apply and validate the model.

Because of the "variability" variables contained within the FCM, Blum's model and the pertaining empirical results are difficult to digest. The results are presented as separate trend ranges (twenty one sets of coefficients in total), with trends calculated from data covering between three to eight consecutive years. The predictive power of the FCM was also separated into one through six years prior to failure. These results are shown in Table 5.9

Table 5.9: Blum's FCM validation and (training) sample accuracy.

Interval of years composing the range	More recent year before failure					
	1	2	3	4	5	6
3	.87 (.92)	.79 (.84)	.72 (.76)	.74 (.79)	.67 (.80)	.57 (.81)
4	.95 (.93)	.80 (.88)	.70 (.80)	.80 (.86)	.69 (.83)	
5	.93 (.95)	.80 (.90)	.70 (.82)	.70 (.83)		
6	.95 (.92)	.78 (.95)	.70 (.87)			
7	.64 (.96)	.74 (1.00)				
8	.69 (1.00)					

Source: facsimile from Blum (1974, p.17).

The accuracies for the four, five, and six year trend ranges are very similar, being on average 94%, 80%, and 70% for one, two, and three years prior to failure respectively. These accuracies are consistent with the finding of the previous studies discussed within this Section, despite the FCM's greater complexity regarding the use of ratio trends. He also noted that the cash-flow to total debt ratio, found to be the best predictor of failure by Beaver (1966) received "generally high rankings" and played an important role within the model.

Blum included ratios which he felt were consistent with the liquid reservoir theory and criticised previous versions of the Z-score model for "searching" for the best variables from an increasingly large set of possibilities. However, in his concluding remarks Blum suggests that alternative ratios should be considered in future models. This arbitrary selection of variables is an inherent problem with MDA and other non-theoretically based modelling technologies. The liquid reservoir "theory" actually has no concrete theoretical underpinning as to what the choice of variables should be, and is therefore more of a concept than a theory. A further discussion of the issues and problems associated with these traditional modelling approaches can be found in section 5.6 of this thesis.

Although Libby's study does not pertain to be a member of the MDA assemblage, he does provide an interesting study on the practical use of the fourteen ratios previously analysed by Beaver (1966) and Deakin (1972). The study was constructed to jointly ascertain the predictive power of ratio information and the ability of loan officers to evaluate that information in the business failure prediction context. Libby postulated that previous empirical studies had provided sufficient support for the notion that certain ratios are highly related to business failure. Feltham and Demski (1970, p.637) proposed that once the signal's ability to reflect a priori relevant events has been established, the next logical extension is to move from analysis of the signal to analysis of action induced by the signal. A prior study to that of Libby was Abdel-Khalik (1973) who applied this extension in order to determine whether analysts could successfully use accounting ratios. Abdel-Khalik concluded that his results "cast doubt upon the usefulness of financial ratios as predictors of failure in an ex-ante prediction situation". Libby suggested that given the limited number of firms analysed within in this study, added to the contrary nature of the finding to other research, "further examination of the issue is in order".

In addition to the analysis of Beaver and Deakin's variables, Libby's sample consisted of 60 out of the 64 firms analysed by Deakin; selected at random as 30 failed and 30 non-failed companies. The ratios were then computed (at random) for one of the three years before failure, to satisfy that the sample consisted of 10 failed and 10 non-failed firms for each of the three years.

Using principal components analysis followed by a varimax rotation, Libby sought to reduce the number of ratios, remove redundancy, and identify the salient financial dimensions represented by the ratios. The analysis identified five sources (or dimensions) of variation within the fourteen variable set, namely; (1) profitability, (2) activity, (3) liquidity, (4) asset balance, and (5) cash position. To represent these five dimensions, net income to total assets, current assets to sales, current assets to current liabilities, current assets to total assets, and cash/total assets were selected respectively through use of the rotation factor matrix.

It is interesting to note here that Beaver's (1966) best performing ratio, cash-flow to total debt, does not appear in the reduced set of ratios. Libby suggests that as this ratio is highly correlated with others contained within the reduction and would have little or no effect on predictive ability should it be included. To test the performance of the reduced set of ratios compared to that of the full set using double cross-validation[58].

The results showed that the original fourteen ratio set was able to correctly predict 54 of the 60 firms (using the derivation sample) and 41 of 60 firms using double cross validation. The reduced set of five variables correctly predicted 51 and 43 firms out of 60 using the derivation sample and double cross validation respectively. Libby deemed the reduced set of variables to be satisfactory for his loan officer experiment given only the slight reduction in overall predictive ability (the smaller model interestingly performs better in holdout tests despite the removal of nine of the ratios). Given this result and the greater manageability of the reduced variable set, these were the ratios given to the loan officers.

Libby's experiment was to provide forty-three commercial loan officers with seventy ratio data sets, one per firm, with each data set containing five ratios. These seventy firms consisted of the sixty sample firms placed in random order with ten repeat firms placed as every forth firm following the first thirty cases. The addition of the repeat firms allowed Libby a measurement of test-retest reliability over the two intervening time intervals while effectively hiding the identity of the repeat firms.

The loan officers were told that half of the firms experienced failure within three years of the statement date in order to set the *a priori* probabilities, and were instructed to work independently and complete the cases within one week. To assess the accuracy of the loan officer's prediction, they were told that a correct prediction was worth one point and an incorrect prediction cost one point.

Libby found that the accuracy of the loan officers was superior to a random selection model, and determined that the ratios had been utilised correctly. Correct predictions ranged from 27 to 50 correct out of 60 predictions and the average predictive ability of

[58] Double cross-validation is simply a cross validation (splitting the sample into a training and a holdout (validation) sample) performed twice. The average of the two validation tests is taken to be the pertaining accuracy of the model.

the loan officers was 44.4 out of sixty (74%). The experimental results allowed Libby to pinpoint four useful outcomes of his research and concluded that (1) there was no significant difference between the mean predictive accuracy of loan officers from small banks to those officers from large banks; (2) there were no significant correlations between predictive accuracy and loan officer characteristics, such as age and experience; (3) there were no differences in short-term, test-retest reliability between user subgroups; and (4) there was a relatively uniform interpretation of the ratios across all officers.

It is encouraging that results of this study suggest that loan officers are reasonably accurate in making predictions upon the likelihood of failure of a potential loan candidate. Libby's evidence suggests, however, that only three-out-of-four loan decisions are correct and therefore lending would carry high risk in a real-world setting. It is evident then that loan officer's decisions are unlikely to be determined based on a small set of accounting ratios, and are more likely to be based on wide range of accounting, economic, and market variables.

Although a useful experiment, one factor of its construction leads to concern, and that is regarding the *a priori* probabilities. The loan officers were told before conducting the experiment that half of the firms had failed, yet information of this nature is unavailable to analysts. A further study by Casey (1980) found that loan officers who were not informed about failure frequencies could only accurately predict 27% of bankrupt firms.

Almost a decade after Altman (1968) published his popular Z-score model, a new ZETA model was developed by Altman, Haldeman, and Narayanan (1977) in conjunction with a private financial firm. This model was marketed by *ZETA Services, Inc*. (Mountainside, New Jersey). The empirical findings of the ZETA analysis were made available to practitioners as a statistical service on a subscription only basis. By the mid nineteen-eighties over three dozen institutions (mainly financial institutions) had subscribed to the service, providing:

"*tangible evidence that the model has had a degree of acceptance in practical circles as well as a scholarly journal*", (Altman, 1983).

Following the literature evolution of the preceding decade (of which the main research has been discussed within this section: primarily aimed towards the improvement and extension of the insolvency prediction and classification problem), Altman proposed that a new model was required due to the changes which had occurred within the financial and business environment since the formulation of his original Z-score model. He suggested that there were at least five reasons for his expectation that a new model could improve on the efficacy of prior studies:

1. The size of firms along with their financial profiles had seen a dramatic change over the preceding decade. Many of the prior studies had utilised relatively small firms to develop prediction models, and any new model should be as relevant as possible to the population to which it was intended to be applied[59]. The average failed firm size utilised by the ZETA analysis was approximately $100 million, with no firm having less than $20 million in assets. The *largest* firm analysed by Altman's (1968) study had assets of less than $25 million.
2. A new model should be based on data which is as current as possible. The ZETA model was derived from firms failing in the prior seven years (1969-1975) with just three exceptions.
3. The majority of previous models had been based upon the wide-ranging classification of the manufacturers or even specific industries. Altman wanted to broaden this selection to include retail companies which he felt were particularly vulnerable to failure and could be analysed on an equivalent basis with manufacturers.
4. The study was developed to include the most recent changes in financial reporting standards and accepted accounting practices. An example of this was to include changes to the treatment of lease capitalisation which was brought into effect several years after the formulation of the initial version of the model. By taking into account the latest reporting practices it was intended that that the model was not only relevant to past failures, but also data that would appear in the future. The predictive power and classification accuracy was implicit to their efforts.
5. A new model would enable the authors to test and assess several of the aspects, pitfalls and validity of the MDA process with light to recent literature of the time. A discussion of this literature is provided in section 5.2.6 of this thesis.

[59] Exceptions to this include Altman's (1973) railroad study, and Sinkey's (1975) commercial bank study.

The ZETA analysis was based upon 53 failed firms and a sample of 58 non-failed firms matched by industry and the year of data, of which half of the firms were from the manufacturing sector and half from the retail sector. 27 variables were analysed which included a number of financial ratios and "other measures" which had been found to be significant predictors of failure in previous empirical studies. Using an iterative process the number of variables was reduced to just seven which was deemed not only to classify the test sample well, but also proved reliable in various validation procedures. Adding additional variables to the model did not improve on its predictive ability and likewise, no model with fewer variables performed as well as the seven variable model. The seven variables included in the model are as follows:

- X_1, **return on assets**, which is measured by EBIT divided by total assets.
- X_2, **stability of earnings**, a normalised measure of the 10-year trend in X_1.
- X_3, **debt service**, measured by EBIT divided by total interest payable.
- X_4, **cumulative profitability**, measured as retained earnings divided by total assets.
- X_5, **liquidity**, defined as the current ratio.
- X_6, **capitalisation**, measured by the five-year average of the market value of common equity divided by total capital.
- X_7, **size**, depicted by the firms total tangible assets.

Although the variables for the model have been provided, because of its nature (a marketed subscription only service), the coefficients of the model have never been made available for public viewing or analysis. This, of course, makes the ZETA model in its original form somewhat difficult to test or analyse by academics, although an analysis of proprietary coefficients based upon the researchers sample, utilising the seven variables, would be entirely possible.

With respect to model assumption (5) as listed above, one requirement of the linear form of the discriminant function is that the variance-covariance matrices of the failed and non-failed samples of firms are equal. This proposition is discussed further in a later section of this thesis (5.2.6). For the ZETA analysis Altman sought to investigate this further by comparing two version of the model; (1) the linear form of the model where the dispersion matrices are assumed equal, and (2) the quadratic form of the model where the matrices are deemed to be different.

The findings of the ZETA analysis showed that the model was able to successfully predict failure with almost 93% accuracy for one year prior to failure using both the linear and quadratic versions of the model. Validation tests of the model using holdout samples reduced the accuracy of the model from 91% for a one-year-ahead forecast, to approximately 77% for forecasts up to five years. These results compare favourably to those gained from the original Z-score where accuracy declined dramatically after two years prior to failure. Comparing both the linear and quadratic versions of the model, the linear model was superior in its validation results. The quadratic version misclassified over 50% of the failed firms five years prior to failure with the linear model out performing its counterpart by an average of approximately 5%.

The ZETA study continues to present its results based solely on the linear version of the model. Following Edminster (1972) a suitable cut-off point was sought using the formulation as per Equation (5.7). Altman (1977, pp. 43-44) implicitly assumes equal prior probabilities and equal costs of errors, which results in a zero cut-off score, compared to the arbitrarily chosen cut-off point of 2.99 associated with the original (1968) Z-score. A visualisation of the ZETA's score distribution is shown in Figure 5.4.

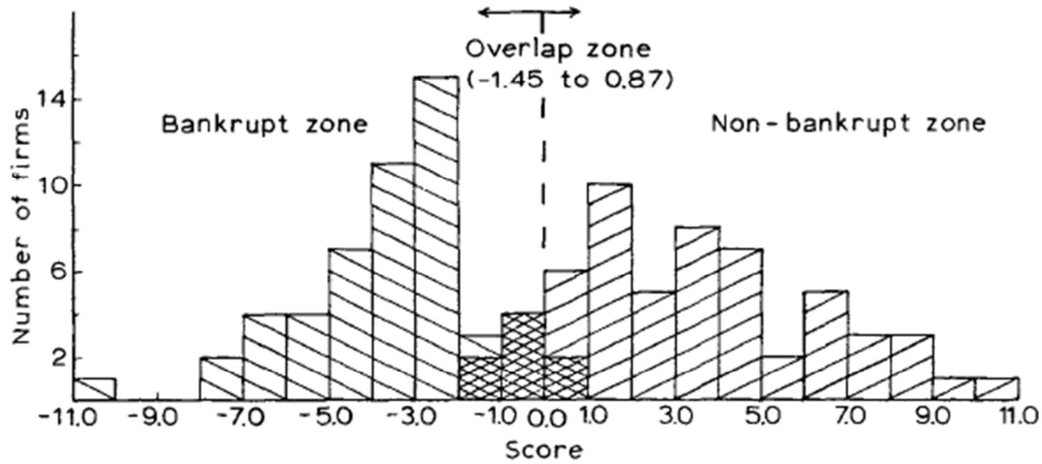


Figure 5.4: Distribution of Zeta scores one year prior to failure.

Source: facsimile from Altman et al. (1977, p.49).

The linear ZETA model was compared and benchmarked against the Altman (1968) Z-score both in its original form and a version with updated coefficients based upon the ZETA data sample of firms. The results of the comparison are provided below in Table 5.10.

Table 5.10: Comparison of ZETA and Z-score models: percentage of correct predictions.

Model	Years before failure				
	1	2	3	4	5
ZETA	91	89	83.5	79.8	76.8
Z-Score (1968)	95.45	82.9	48.3*	28.6 *	36 *
Z-score – ZETA sample [a]	84.6	86.15	81	73.85	71
Z-score- Updated [b]	88.5	84.6	81.2	70.25	63.15

[a] Represents Altman's 1968 model and parameters applied to the ZETA sample of firms.
[b] The coefficients of the 1968 model are updated using the ZETA sample of firms.
* Accuracy based upon failed firms only. Non-failed firm accuracy not provided in original study.

We can observe from the table that the newer ZETA model is much more accurate than the original Z-score from two to five years prior to failure with the original model the slightly better performer in the initial year (because the 1968 Z-score results from years three through five are based upon failed firms only, I am reluctant to draw too many inferences for these years). What is more interesting, and perhaps more appropriate, is the application of the original model to the later ZETA sample of firms and the coefficient-updated model comparison to that of the newer ZETA model. In the case of the original model's application to the new sample, one can observe that the ZETA model shows a higher predictive ability for each year prior to failure, being especially superior in years one and five. Similarly, this is the case for the updated version of the Z-score, with the ZETA model outperforming in each year. An interesting observation is that the latter model with the updated coefficients performs less well than the original Z-score (derived from much older data, and constructed ten years prior) in years two, four and five prior to failure and almost identical in year 3. The only year in which the updated coefficients out-perform the original is in the first year. These results are somewhat counter-intuitive as one would expect the latter model to be superior in all years as it is derived from firms within the same sample population as the test data.[60]

In an interesting direction taken by the ZETA analysis, the study provides the first example of an accuracy estimation based upon differing misclassification costs, where all previous studies had assumed the costs to be equal. As I have mentioned earlier in this section, it is far more realistic to assume that the cost of misclassifying a future

[60] This anomaly also occurs in my own study where one of the original models which I test, outperforms that of the updated version. A discussion of possible causes is contained within the results of section 9.3.

failed firm as non-failed would be of greater economic significance than if I misclassify a non-failed firm as being failed.

To be able to estimate the misclassification cost component of the discriminant analysis and determine a better estimate of a model's predictive ability, requires the specification of the decision maker's role. The ZETA analysis utilises the commercial bank loan function as its framework where errors occur upon the unsuccessful judgement by the lender when granting (or not granting) loans. A type I error (C_{I}) occurs when an accepted loan defaults, and a type II error (C_{II}) occurs when a loan is rejected that would have otherwise resulted in a successful payoff. I can therefore specify the error costs of the function as being:

$$C_1 = 1 - \frac{LLR}{GLL}, \qquad C_{II} = r - i \tag{5.8}$$

where:
LLR = amount of loan losses recovered,
GLL = gross loan losses (charged off),
r = effective interest rate on the loan,
i = effective opportunity cost for the bank.

The cost of a type I error is therefore dependent upon how successful the bank is in recovering the loan amount should a default occur. Altman et al. note that additional costs such as legal fees, practitioner fees, lost interest, and transaction charges are relevant to the process and suggest that these would obviously increase the error cost although they are omitted from the ZETA analysis. LLR and GLL were determined from both financial statements from 26 large commercial banks, and 33 usable responses obtained from questionnaires which were sent to smaller regional banks. The resulted estimation of a type I error was determined to be approximately 70%.

To determine the value of C_{II}, despite its simplified formulation, there are several considerations to take into account. If the decision is made not to lend to a firm, then the loss may be mitigated by lending the funds to an alternative company at the same risk level which is deemed to be a successful loan. In this instance the opportunity cost for the bank is equal to the interest rate of the loan and therefore the cost of a type II error would be extremely small or indeed zero. Realistically however, the rejected firm would probably carry a higher risk than the alternative secure firm and $r - i$ will therefore be positive but still fairly low. At the very least, the bank would be able to invest in a form of Treasury bond or other government "riskless" security, in which case Altman suggests that $r - i$ would be somewhat higher at around 2-4%. For the ZETA analysis, an arbitrary approximation of 2% was assumed for C_{II}. The revised cut-off point for the ZETA model could therefore be defined as:

$$CP = \ln \frac{q_{\text{I}} C_{\text{I}}}{q_{\text{II}} C_{\text{II}}} = \ln \frac{0.02 \cdot 0.70}{0.98 \cdot 0.02} = -0.337 \tag{5.9}$$

where the prior probability of failure was estimated to be 2%.

By changing the cut-off point from zero to – 0.337, the predictive accuracy of the model also changes. Applying this new cut-off and taking into account the different error costs, the accuracy of the ZETA model is affected, increasing the rate of type I errors from 3.8% to7.6%, whilst at the same time reducing the rate of type II errors from 10.3% to 7%. Although the number of misclassifications by the model (8 in total) remains the same, it is evident to me that a decrease in type II errors to the detriment of type I errors is unsatisfactory for a models construction, albeit a more realistic approach. Nevertheless, Altman suggests that the cut-off point should be determined on a proprietary basis by individual lenders, based on the prior probabilities and error costs attributable to that particular institution and changing economic conditions.

5.2.4 A comparison of the different models: Scott (1981)

The aim of Scott (1981) was to review and evaluate and several insolvency prediction models that the previous decade had witnessed. As well as considering univariate and multivariate empirical models, Scott also investigated several insolvency prediction theories which had been developed in recent years, most notably the theoretical gambler's ruin model[61], and option pricing formulae as derived and discussed in both Black and Scholes (1973) and Merton (1974)[62].

The concern of Scott was that even though the development of empirical models which were able to discriminate between solvent and insolvent firms, was an "important accomplishment in modern finance", said models did not fully command professional acceptance, in part because they "lack the underpinnings of an explicit and well-developed theory". That is to say that the majority of the models derived in prior studies had been done so by a statistical search through a number of possible financial variables.

An exception to this was the gambler's ruin model which applied an established theoretical framework to the task of insolvency prediction.

"*Although the theory specified a functional form for the probability of ultimate ruin, Wilcox found that this probability was not meaningful empirically. It could not be calculated for over half of his sample because the data violated the assumptions of the theory: firms that were supposed to be bankrupt were actually solvent. Faced with this problem, Wilcox (1976) discarded the functional form suggested by the theory and used the variables it suggested to construct a prediction model. It is hard to assess the resulting model since he did not test it on a holdout sample. Nevertheless, Wilcox's work is notable as the first attempt to use explicit theory to predict bankruptcy*". Scott (1981, pp. 323).

Scott sought to investigate the similarities between the variables contained within the theoretical strands of both Wilcox's research and option pricing formulae (Black and Scholes, 1973; Merton, 1974) with that of the empirical strands of insolvency prediction to provide a well-defined theoretical framework to assist the direction of future research. The empirical models discussed and evaluated in Scott's work were Beaver (1966), Altman (1968), Deakin (1972), Sinkey (1975), and Altman et al. (1977). He

[61] The gambler's ruin problem was originally developed by Feller (1968) and its application to the problem of insolvency prediction was championed later by Wilcox (1976) and Santomero and Vinso (1977). A brief guide to the gambler's ruin theory is provided in Appendix D.

[62] A more thorough review of the option pricing literature and its relationship with insolvency prediction can be found in section 5.6 of this thesis.

noted that the low misclassification rates of these studies suggest that bankruptcy prediction is possible but is hard to determine which of the models discriminates the best, as different data and differing procedures underlie the reported test results. The best multivariate models seemed to outperform the best univariate models, but the best univariate models outperformed some of the multivariate models which could (probably) be attributed to a degree of over-fitting. Scott also commented that earnings or cash-flow variables appear in all of the models, and debt appears in several. All of the models rely on accounting data which suggest that there is useful information to be gained from company accounts, but stock price data was also an important factor.

"Of the multidimensional models, the ZETA model is perhaps most convincing. It has high discriminatory power, is reasonably parsimonious, and includes accounting and stock market data as well as earnings and debt variables. Further it is being used in practice by over thirty financial institutions. As a result, although it is unlikely to represent the perfect prediction model, it will be used as a benchmark for judging the plausibility of the theories discussed in the following sections." (Scott (1981, pp. 324-325).

To compare and contrast the findings of the prior studies into a single theoretically-driven framework, Scott first comments on the similarities between the "simplest" insolvency prediction model (the single period model) and the Black-Scholes option pricing formulae, and how they can (or cannot as the case may be) explain the success of empirically driven models.

The single period model states that a firm will become insolvent at the end of a period of trading if its liquidation value (V_1) is less than the amount which it owes to its creditors (X_1).

$$V_1 < X_1 \tag{5.10}$$

If V_1 has a probability distribution with mean μ_v and standard deviation σ_v, standardising both sides implies that insolvency occurs if:

$$\frac{V_1 - \mu_v}{\sigma_v} < \frac{X_1 - \mu_v}{\sigma_v}$$

(5.11)

This inequality is depicted graphically in Figure 5.5, where insolvency occurs should the standardised liquidation of the firm value falls into the shaded region.

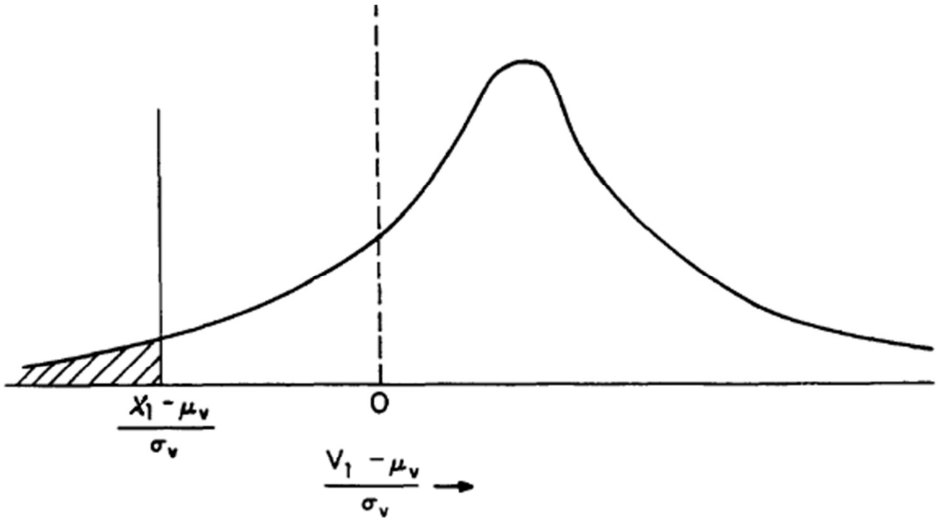


Figure 5.5: Graphical representation of failure probability

Source: facsimile from Scott (1981, p.327).

Based on this simple theory, Scott noted that the empirically derived models did not contain variables which related to this single period model. Each of the empirical models relied on accounting values and earnings variables which do not appear in equation (5.11), which only contains stock values. Hence it is clear that the single period model is unable to explain the success of the empirical insolvency prediction models.

The Black-Scholes (1973) option pricing model is similar to this single period model. The model views equity as a European call option on a firm's assets with a strike price equal to the face value of its liabilities. The option expires at time T when the debt matures. At this point the equity holders either; exercise their option and pay off the debt (if the firm's assets (V_1) are of greater worth than the face value of its liabilities (X_1)) or, let the option expire (if the assets are not sufficient to cover the cost of the maturing debt). If the option is left to expire then the debt holders will file for liquidation and the assets will be transferred to them without cost. Merton (1973) assumes that the firm's debt consists of a single zero-coupon bond and therefore equates to that of equation (5.11) in its criteria. As with the single period model, the options pricing model does not contain earnings or accounting variables and therefore cannot explain the success of empirically-based models. Scott comments that:

"*... in principle, the option pricing methodology could be used to determine bankruptcy probabilities. However, it has not been used to date*".

In a later section of this thesis I discover that the option pricing methodology is later adopted by the art, driven by research within the credit rating fraternity.

The theory of the gambler's ruin model assumes that a firm has a given amount of capital or net liquidation value of stakeholder's equity (NLV), and those changes in NLV, *S*, occur at random. Positive operating cash-flows lead to positive changes in NLV, and losses require the firm to liquidate assets[63]. Insolvency occurs when NLV becomes negative (NLV+S < 0). Scott suggests that if we are willing to assume that accounting variables can serve as surrogates for liquidation values, then the gambler's ruin model goes some way in explaining the predictive accuracy of the empirically-based ZETA model.

Under this interpretation, the NLV is represented by the book value of stakeholder's equity, and *S* is represented by the change in retained earnings. The normalised equation therefor implies that a firm will become insolvent if:

$$\frac{S - \mu_s}{\sigma_s} < \frac{-(\mu_s + NLV)}{\sigma_s}. \tag{5.12}$$

If we then divide the numerator and denominator by total assets (*TA*) and multiply by minus one, then the probability of failure will increase with decreases in

$$\frac{\left(\frac{\mu_s}{TA}\right) + \left(\frac{NLV}{TA}\right)}{\sigma_s / TA} \tag{5.13}$$

[63] The model assumes that the firm is detached from any security market and that its losses must be funded through the selling of assets. The firm therefore can sell neither debt nor equity.

The variables contained within this equation are similar to those found within the ZETA model and affect the probability of failure in the same direction. ZETA variable X_4, (cumulative profitability, measured as retained earnings divided by total assets) is close to NLV/TA. X_1, (return on assets, measured by EBIT divided by total assets) is close to μ_s/TA, where μ_s is the expected change in retained earnings. A third variable, X_2, (stability of earnings, a normalised measure of the 10-year trend in X_1) is also very similar to σ_s/TA. Of the similarities in the variables, Scott comments that:

"*… although the gambler's ruin model helps explain the empirical success of ZETA, the explanation is imperfect, for ZETA contains additional variables, and even where the variables are similar, they are not identical. Some of the additional variables in ZETA may be due to statistical over-fitting; however they may also represent factors important in bankruptcy prediction, which do not appear in the relatively simple gambler's ruin model*". Scott, (1981, pp.329).

Further in his paper, Scott (1981) describes numerous expansions of the gambler's ruin model and alternative theoretical propositions which converge with many of the variables which had been empirically derived as being good predictors of failure. Among the expansions, Scott evaluates several generalisations and model variants based upon various considerations such as fixed floatation costs, imperfect secondary markets for real assets, differences between accounting and economic values, risk aversion, personal taxes, inter-temporal dependence in earnings, and debt functions. These expansions make interesting reading, although too numerous to detail in this thesis. What can be gained from Scott (1981), however, is that there is a connection between empirically based models and those which are theoretically driven, although the overlap of the models is somewhat imperfect. Variables which are relied upon in theory, often find themselves in the results of economic practice, determined through the means of statistical searching. Scott suggests that by using the variables and functional forms contained within insolvency theory, it may be possible to improve the predictive accuracy of future empirical work. There does however need to be a certain amount of trade-off between the complex nature surrounding a real company experiencing real financial distress or pending insolvency, with that of a parsimonious model able to capture all of the important information within a small number of variables. A model which is able to effectively capture, in detail, a company's state of solvency would need to include far more financial, economic, and political variables than would be feasible to analyse in a practical sense and such a model, in any case, would be subject to a large degree of over-fitting. Scott refers back to Beaver (1966) and concludes that:

"...in this respect it is useful to remember that Beaver's work suggests simple models may work very well". (Scott, 1981, p. 341).

5.2.5 UK application of Z-score models

The discussion of the literature thus far has centred upon US studies which dominated the field throughout the 1960s and 1970s. The success of these studies inevitably led to their methodological application to alternative markets and different populations of firms. Many countries were subject to studies in line with the methodologies devised by Altman (1968) and Beaver (1966). Developed and developing countries, non-capitalist countries, smaller economies and large industrial nations alike, have a vested interest in the avoidance of corporate failure in both the public and private sectors. Altman participated directly in the formulation of insolvency prediction models in many countries most notably Brazil, Canada, and Australia[64], and other applications can be seen in numerous international studies of the time[65]. An overview of this literature can be found in Altman (1984).

The focus of this thesis is not, however, an international critique of insolvency prediction models and methodologies, but is primarily concerned with insolvency prediction and credit risk models, their application the UK financial market, and the evolution of the associated literature. Although US and UK markets are probably the most similar in terms of the quality and availability of financial information, in theory the UK literature should provide a more appropriate comparison and be more relevant to my own empirical research. I therefor turn the focus away from the US literature to notable studies which centre on the development of Z-score prediction models within the UK.

---

[64] See Altman et al. (1979, Brazil); Altman and Lavallee (1981, Canada); Altman and Izan (1982, Australia).

[65] Examples include Cahill (1981, Ireland); Bilderbeek (1977, Netherlands); Van Frederikslust (1978, Netherlands); Collongues (1977, France); Mader (1975, 1979, France); von Stein (1968, Germany); Beermann (1976, Germany); Castaga and Matolesy (1981, Australia); Takahashi et al. (1979, Japan); Ko (1982, Japan).

It is also appropriate to mention here that throughout the 1970s and early eighties, the rate of business failures within the UK were possibly the highest in the world. The high number of corporate failures was the result of the 1973-75, and 1980-82 recessive periods, and with the relatively tight monetary policy (aimed at reducing inflation), inevitably leading a declination of company earnings by an average of 35%[66]. The rate of failures within the UK for this period (0.76%) was, on average, twice the failure rate of firms within the US (0.35%).[67]

The most prominent figure in the field of corporate distress prediction within the UK is Richard Taffler who was one of the first pioneers of the application of MDA to discriminate between failed and non-failed companies here in the UK. Although Taffler is widely recognised as being the first to successfully apply MDA to the discrimination of failed and non-failed companies within the UK, the very first UK study of this kind (and possibly the first outside of the US) was that of Lis (1972) which is described in Bolitho (1973). Lis's model correctly predicted 24 out of 30 companies, but further evaluations of the models performance are non-existent. Curiously, the four variables within this model are identical to those found within Altman's (1983) Z' score for private companies, which in itself has had limited exposure. Taffler (1984) provides details and discussions of several other early UK studies upon the prediction of failure of UK companies (Earl and Marias, 1979; Mason and Harris, 1978; Datastream, 1980; Betts and Belhoul, 1982, 1983; El Hennaway and Morris, 1983).

The number of early UK studies, particularly circa 1980, demonstrates the popularity of the art which is still evident today. Many authors can be seen to be competing for dominance in the field, with each trying to demonstrate that his/her discriminant model has a better predictive ability than the next. For whatever reasons however, it is the Taffler version of the UK Z-score which has stood the test of time, being the most documented and recognisable insolvency prediction model within the UK over the past thirty years.

---

[66] Source: *The Financial Times* 20th November 2008.

[67] Source: Altman (1983, pp.339).

It is evident from the early literature that there were in fact two competing versions of Taffler's UK Z-score (which it has since become known), one version authored by Taffler (1976, 1982), and the second version first appearing in Taffler and Tisshaw (1977), and again in Taffler (1983) onwards. The first model is one containing five variables and first appears in Taffler (1976) (although it is noted in Taffler (1983) as being completed in the summer of 1974). The model is finally, and concisely, described in Taffler (1982). There are several peculiar observations surrounding Taffler (1982) which I ought to mention at this point. Firstly, Taffler states that:

"The only UK-based work fully documented in the journals is that of Mason and Harris (1979), relating to the construction industry" (Taffler, 1982, p.342).

It is interesting that there is no mention of his own earlier study (1976), nor does he acknowledge the earlier work by Taffler and Tisshaw (1977) which discusses the two competing UK Z-score models (favouring a four-variable version). Curiously however, the Mason and Harris (1979) paper *does* cite Taffler (1976) in its references, but oddly there is no mention of the study in the body of the text. The failure to acknowledge two of his previous papers in a review of past literature seems to me rather peculiar. I can only assume given the similarity between the early Taffler papers and the indecision between the two (five and four variable) models of which I shall discuss within this section; that the author was on a constant quest to improve upon and update the UK Z-score as new data became available, and to reflect changes in economic conditions.

The second oddity of Taffler (1982) is that it is later described in Taffler (1983) as being used for many years by a firm of London stockbrokers to "provide a service to their clients". Initially I thought that this may explain why the five-variable model does not make another appearance in the literature after this publication. The model may have been commercially distributed at this point, but this would make no sense as the disclosure of the model variables and coefficients was made available in Taffler (1982). Perhaps this disclosure is the reason for the discontinuation (or at least the perceived discontinuation) of the stockbroker's use of this version of the UK Z-score in practice. Officially a footnote in Taffler (1984) suggests that:

"…it appears that the service finally ceased not primarily through operational problems but because 'the style of stockbrokers' research does not lend itself particularly well to Z-score analysis". (Financial Weekly, 24th June, 1983, p.4).

Interestingly it is the coefficients of the alternative four-variable model (when is generally perceived as *the* UK Z-score model), which were not disclosed in publication until recently in Agarwal and Taffler (2007).[68]

Upon an analysis of 23 failed firms (1968-1973) and 45 non-failed firms (1972-1973), Taffler (1982) considers an enormous number of variables, perhaps more than any other paper within the field. 152 variables were selected on the basis of their effectiveness in previous and related studies, popularity in the literature, theoretical arguments (such as Walter's liquid reservoir theory, 1952), and suggestions by financial analysts. Fifty-eight ratios were calculated from profit and loss accounts and balance sheet items, forty-four were four-year trend measures, and forty-eight ratios computed from the funds statement.[69]

Reducing the variables using step-wise discriminant analysis (where variables are eliminated from the Z function if they are deemed not to add sufficient information to the model or are highly correlated with another), the resulting model consisted of just five variables; X_1, earnings before interest and tax divided by the previous year's (or opening) total assets; X_2, total liabilities divided by net capital employed; X_3, quick assets over total assets; X_4, working capital over net worth; and X_5, stock turnover. This five-variable version of the UK Z-score has the following functional form:

$$Z = 0.71X_1 - 0.93X_2 + 0.32X_3 + 0.49X_4 + 0.53X_5 \tag{5.14}$$

[68] It is possible that the four-variable UK Z-score was used in practice by the credit scoring company Syspas Limited for a number of years before it was dissolved in March 2001. It is also possible that Taffler was a director of this company, however, only access to the last accounts (1997) would be likely to confirm these statements with certainty, which I have so far been unable to obtain.

[69] Statement of sources and application of funds is a predecessor of the cash-flow statement in the UK. The forty-eight cash-flow ratios were later dropped due to instability.

Applying this discriminant function to the initial 68 firms (used for the training of the model), resulted in an almost perfect number of correct predictions. Only one firm (Rolls Royce) was misclassified as being solvent when it had actually entered into receivership. Taffler notes however that

"…whether Rolls Royce was strictly speaking insolvent at the date of appointment of the receiver, February 4th, 1971, is open to question. All creditors' claims were subsequently met in full, debenture holders were paid together will all accrued interest and equity holders eventually received four times the figure the shares stood at when dealings were finally suspended. These facts add weight to the argument that Rolls Royce entered into receivership as the only way to break its ruinous RB211 jet engine contract with Lockheed allowing renegotiation of terms. Consequently, it might be reasonably maintained that the solvent classification was a fair reflection at least of the financial state of Rolls Royce as at December 31st, 1969, the end of the last financial year for which audited accounts were available." Taffler (1982, p.347).

Given this information, despite the training and validation samples being one and the same, and that the sample sizes are relatively small compared to contemporary empirical studies, this is the first and, perhaps only time that an insolvency prediction model has been purported as being 100% accurate. Further tests of the model on an ex-post holdout sample of 33 insolvent firms (1974-1976) and 10 non-failed firms (1972-1973) were also presented to assess the model's predictive ability. The results of this holdout examination showed that from the 43 firms in question, 4 were incorrectly classified as being solvent when they actually failed, a type I error rate of 12.1%. None of the solvent firms were misclassified as being failed, giving an overall accuracy of approximately 91%.

In a further validation test, Taffler sought to determine the number of companies which were actually at risk of failure in order to provide "essential" information for any user of the Z-score model. The model was applied to 1257 quoted industrial companies with accounts held on the *Jordan Dataquest* database in the year 1975. The results indicated that 10.7% (135) of the companies in question were at risk of failure. It is unfortunate that Taffler did not provide any further information regarding these "at risk" companies, such as what percentage actually failed within the following period(s). A similar case is that of the Altman et al.'s (1977) ZETA model, which was reported in

Business Week (March 24th, 1980, p.106) as classifying 520 (20%) of 2600 US companies listed on *Compustat* as being possible failures. What percentage of these companies actually failed within the proceeding years is not reported.

One study that did provide a follow-up population study was Deakin (1977) who examined the five variables which were applied in Libby's (1975) loan officer experiment in an extension to Deakin (1972). The validation of the derived model was carried out upon 1980 companies on the *Compustat 1800* file for the fiscal year 1971, and results suggested that 290 companies (16.3%) had profiles which were similar to those of the failed group of firms. Unlike Taffler, Deakin provides a follow-up analysis of these firms for a subsequent three-and-a-half year period 1972-1975. He found that of the 260 companies in question, only 18 (6.2%) actually failed within the period. Altman (1983) however, identifies some interesting arguments regarding the definition of failure which is used by Deakin. The 18 failures which Deakin identifies were based upon failure being defined as bankruptcy, liquidation, or reorganisation. Altman noted that if the seven mergers which occurred within the group and within the period, happened under distress conditions, then the total number of correctly predicted failures would rise to 25 (8.6%). Further, defining failure to include dividend cuts or omissions causes the number of correct failure predictions to rise rapidly to 224 (77.2%).

Notwithstanding the inclusion of the particularly weak definition failure suggested by Altman (1983), it can be seen from Deakin (1977) that the application of insolvency prediction models in practice, when they are derived using a small number of (hand-picked) companies, and applied upon the large population of firms, does not yield particularly strong results. Later studies utilise a far greater number of companies for their analyses, in fact it is often possible to monitor and evaluate most of the population of firms in any western financial market. That the practical application of these early models is rarely followed-up and evaluated in the early literature, is not too problematic as modern computing power and databases make it possible to provide true ex-post empirical studies of these models.

Taffler and Tisshaw (1977) again used the empirical methodology of MDA to formulate a predictive function derived from, and to be used upon, UK company data. The main focus of Taffler and Tisshaw (1977) was centred upon the apparently low number of audit GCQs associated with the sample of firms under their analysis. They suggested that only 22% of the failed companies were actually issued with a GCQ which is very similar to the results of Altman and McGough (1974) in the US. From their experience they also point out that that it is only around 15-20% of companies that display a financially distressed profile that will actually fail in the UK, due to the relatively large amount of government aid which is potentially available to keep a firm alive, compared to the US where this practice is very rare.

Using a sample of 46 failed[70] and 46 non-failed manufacturing companies from 1968 to 1976, matching by firm size and industry, eighty different ratios were investigated. Using a step-wise discriminant package, the eighty variables were reduced to a model containing just four variables; X_1, Profit Before Tax/Current Liabilities; X_2, Current Assets/Total Liabilities, X_3, Current Liabilities/Total Assets, and X_4, No-credit Interval[71]. The model (with undisclosed coefficient estimates) was reported to have a one-year-ahead forecast accuracy of 97%.

Taffler (1983) reverts back to the Taffler and Tisshaw (1977) four-variable model in a concise discussion of the models formulation, parameters and practical application, pausing only briefly to mention his earlier five-variable version.

The Coefficients of the model were not published at the time but later became available in Agarwal and Taffler (2007). The functional form of the UK Z-score is as follows:

$$Z = 3.2 + 12.18X_1 + 2.5X_2 - 10.63X_3 + 0.029X_4 \tag{5.15}$$

The cut-off point for this model is zero with companies providing a score of less than zero being those deemed failed as per the definitions described in Taffler and Tisshaw (1977) above. Companies with a score above zero are deemed to be healthy and thus survive. The predictive ability of this model was evident in a near perfect class

[70] Failure was defined as those companies entering receivership, CVL, compulsory winding up, and interestingly government bailouts.

[71] The No-Credit interval is the time a company can continue its operations from its immediate assets if all other forms of short term finance are cut off. Measured in days it is more directly it is defined as (immediate assets (current assets – stock) – current liabilities) divided by daily operating costs excluding depreciation.

identification of almost 98% accuracy in a Lachenbruch hold-out test[72]. Although the Lachenbruch (1967) jack-knife test is not a holdout test in the strictest sense of the word, it is often used to produce almost unbiased estimates of the true model misclassification rates when actual holdout data is not available. Two samples of failed and non-failed firms with sizes n_1 and n_2 are used for the jack-knife procedure. For the common delete-one jack-knife, a single observation m_i is omitted from the training procedure and is then classified according to the resulting function. The procedure is repeated for each m for i = 1 to $(n_1 + n_2)$. The misclassification rates can then be calculated.

The high accuracy is perhaps why this model had been used widely throughout credit grating institutions within the UK, in particular over the last 30 years[73]. Originally designed for use with industrial companies (with separate models developed for retail and service enterprises), Taffler (1995) acknowledges that the model can be applied with equal confidence to all non-financial companies and hence makes a good candidate for further study in this thesis[74].

The main aim of Taffler (1983) was to:

"*...test the model's true ex ante predictive ability and the results from its use in practice in the several years subsequent to its development*".

The UK Z-score (as it has become known) certainly boasts an impressive track record in its predictive ability over time. Taffler (1995) provided evidence for this over the eighteen year period from when the model was first developed. Within this period Taffler identified 152 outright failures amongst UK public industrial companies[75] with his model being able to identify 150 of these correctly[76]. Agarwal and Taffler (2007) offers further verification of the UK Z-score's robustness over time providing much

[72] One of the two companies which were deemed to have survived when they actually failed was Rolls Royce as previously discussed in Taffler (1982).

[73] See Agarwal and Taffler 2007 for details.

[74] Altman's Z-score model was also originally developed from samples of industrial manufacturing companies but has likewise conventionally been applied across the whole spectrum of non-financial companies.

[75] Defined as receivership, administration, creditors' voluntary liquidation or wind-down.

[76] One of the two companies not detected was Polly Peck, now notorious for its scandalous accounting.

discussion of theory and strong statistical evidence to suggest that this method of detecting financial distress does indeed have true ex-ante predictive ability over long periods. Tests over a 25 year sampling period showed that the UK Z-score continued to have an impressive 96% type I accuracy, a higher rate than the longevity results shown by Altman (2000) in which the original Altman Z-score showed a type I accuracy ranging from 82-94% over a 31 year period (1969-1999).

Both the five-variable and four-variable models of Taffler, utilise similar methodology and purport similar predictive ability but it is the four-variable version which has received the most exposure and testing within the academic literature to date due in part to the popularisation and promotion by the author himself. Taffler has showed that the prediction of corporate failure within the UK can be achieved with highly accurate results. Indeed these results seem to persist over time as shown in Taffler (1995), Agarwal and Taffler (2007), and the latest examination of the UK Z-score provided in Agarwal and Taffler (2008).

The empirical section of this thesis critically evaluates several popularised insolvency prediction models which include both the Altman and Taffler Z-scores. Although (as I shall discuss) MDA has its critics and is perhaps considered an out-of-date methodology which is seldom used in modern finance application, I am interested to discover whether these models still provided the predictive ability that they once claimed. By providing a critical and impartial evaluation of these models using (for the first time) post-IFRS implemented company data, I will be able to determine whether these models are sufficiently relevant to today’s financial climate and whether they still hold any predictive ability at all.

5.2.6 Discussions on the validity and pitfalls of the MDA methodology

Despite the perceived popularity of the MDA approach and the resulting “forest” of Z-score-type models, each trying to leave its competitors in the shade, several papers are evident from the 1970s which suggest that MDA should be used with caution. The MDA methodology gained criticism from various detractors who sought to provide evidence that the method was floored, be it statistically, theoretically or otherwise. This section outlines some of the key papers and discussions regarding the validity of MDA as an insolvency prediction tool.

I have briefly mentioned earlier in this section that although many authors (including myself) use the abbreviation MDA as representing Multivariate (or Multiple) Discriminant Analysis, what I am in-fact referring to is Linear Multiple Discriminant Analysis (LMDA). For any classification or prediction exercise, be it separating failing and non-failing firms, or otherwise, assumptions pertinent to the MDA methodology dictate whether or not LMDA actually provides an optimal solution to the classification problem.

If one takes a population of firms separable into two categories (failed and non-failed), each firm being characterised by a vector of attributes or variables, then these attributes must (a) arise from multivariate normal populations and, (b) the two categories must have identical variance-covariance matrices. If these two conditions are met and the variables exhibit differing mean values between the two categories of firms, then LMDA will provide an optimal solution to the classification problem. If on the other hand the variables arise from multivariate normal populations, but the variance-covariance matrices are not identical then it is Quadratic Multivariate Discriminant Analysis (QMDA) which will yield the optimal solution. This quadratic form of discriminant analysis is gained through the derivation of the MDA methodology as provided in Appendix C and is shown that the quadratic (true) form of the model will collapse into the less complex linear form of the model only when the two variance-covariance matrices are equal (equations C.15 and C.16). Theoretically speaking, this suggests that QMDA should therefore yield the better classification results as it is unlikely that the equal matrices assumption will hold in reality; in practice however Altman et al. (1977) for example, demonstrate that (in their analysis) it is the linear form of the model which is attributed the greatest classification accuracy.

If the variance-covariance matrices dictate optimal results by the correct application of either quadratic or linear discriminant analysis, the problem persists that both of these discriminant model forms rely on the assumption that the variables used to characterise each firm are multivariate normally distributed. In reality however, financial variables attributing multivariate normality are likely to be considered a rarity rather than the norm. Indeed, dichotomous variables which are frequently employed along-side continuous variables within the discriminant function are by their very definition (being either zero or one) will not exhibit multivariate normal distributions. Chang and Afifi

(1974) derive and examine classification methods for models containing mixtures of continuous, discrete, and dichotomous variables. They propose that the most appropriate cause of action would be to split the sample based on either the values of the discrete variables, or where possible, the discrete variable itself. This proposition (to me at least) seems somewhat impractical. If I was to follow Chang and Afifi (1974) then I would effectively be constructing many different discriminant functions for application to any given sample of firms. The number of different functions would depend on the number of dichotomous and discrete variables that the author wishes to analyse rising dramatically as the number of dichotomous and discrete variables increases. It makes more practical sense to evaluate the effects of non-normality statistically to determine what effects this assumption violation would have on the resulting model.

Eisenbeis (1977) notes that violations of normality assumptions are likely to bias any tests of significance or estimated error rates pertinent to the MDA methodology, and further asks, to what extent does the relaxation of this assumption have on the classification. He suggests that one reason for the lack of attention to this problem was that most of the available normality tests were for univariate and not multivariate normality, although several tests have been documented and discussed.[77]

The application of various transformations to data subsequent to the discriminant analysis has been widely adopted with the most popular transformations being that of the natural and standard logs. The thought process behind these transformations is that many of a firm's variables, such as the market value of equity, book value of debt, or employee numbers, tend to exhibit positive skewness with a small number of firms producing large out-lying variables whilst the majority tend to be smaller. The consequence of the natural log transformation is that the variable's marginal distribution is more symmetrical and is bounded below by zero. Therefore, depending upon the variable(s) in question, the log transformation certainly has intuitive appeal because the resulting distribution will be more symmetric and, in all probability, more normal. An obvious detraction however, is that negative numbers cannot be transformed this way.

Watson (1990) provides a concise examination of multivariate distributional properties of a set of financial ratios and concludes that financial ratios are likely to

[77] See for example Malkovich and Afifi (1973), Sharipo and Francia (1972), and Koziol (1982).

include multivariate outliers and to differ substantially from multivariate normality. Watson was able to determine two procedures which could be used to successfully induce an approximation for multivariate normality.

1. Multivariate outliers (an average of four percent) were identified and deleted using a generalised distance measure that accounted for the multivariate structure of the data.
2. Modified power transformations were applied to the data after the multivariate outliers were deleted.

With regards to these power transformations Watson follows Box and Cox (1964) in changing a variable y to $y[\lambda]$ by the family of transformations:

$$y[\lambda] = \frac{[y^{\lambda}-1]}{\lambda}, \quad \text{for } y \neq 0,\ y > 0$$
$$y[\lambda] = \ln[y], \quad \text{for } y = 0,\ y > 0 \tag{5.16}$$

λ is determined such that the logarithm of a normal likelihood function is maximised as per Box and Cox (1964) which results in a value of λ which is appropriate to induce approximate normality. Watson deems this to be a better approximation for normality that otherwise would be induced by using logarithmic or square root transformations for financial ratios that may be skewed (e.g. Deakin, 1976; Bates, 1973).

A further consideration is that if the data does not exhibit multivariate normality then the researcher much derive alternative joint probability density functions and fit this within the existing methodology, albeit a seemingly impossible task. This is perhaps why most researchers are content to accept the notion that the results, estimations and significance tests of the discriminant analysis methodology provide a reasonable approximation to their true values should multivariate normality exist. Lachenbruch et al. (1973) evaluate the robustness of both the linear and quadratic discriminant forms for three non-multivariate normal distributions, log normal, logit normal and inverse hyperbolic sine normal. They found that linear procedures were quite sensitive to non-normality but the problems were not so great when the distributions were bounded as found for example in the logit normal. Overall Lachenbruch et al. (1973) found that the estimated overall classification accuracy rates were not, however, affected to any significant degree. Similarly to the later Altman et al. (1977) paper, the authors found somewhat surprisingly that the results of quadratic analysis were not significantly greater than their linear counterparts, and in many cases they were inferior.

Finally in this section I would like to discuss a series of papers which focus and discuss some further points on the application of discriminant analysis. These papers (Joy and Tollefson, 1975; Eisenbeis, 1977; Altman and Eisenbeis, 1978; Joy and Tollefson, 1978, and Scott, 1978) largely focus upon the construction of samples, the design, application and definition of classification and prediction methods, and the importance and interpretation of the resulting explanatory variables within the discriminant function.

Joy and Tollefson (1975) argue that the predictive abilities of many insolvency prediction models tend to be overstated because the authors confuse ex-post classification results with ex-ante predictive abilities. They note that many failure prediction studies use ex-post discrimination in a validation sample (cross validation) where prediction accuracy is based upon classifying members of a time-coincident holdout (validation) sample.

"Prediction means to foretell the future. Ex post discrimination may provide a useful foundation for explanation of the past, but it does not provide sufficient evidence for concluding that the future can be predicted. However, this fact is virtually unrecognized in financial LDF (LMDA) *studies".* (Joy and Tollefson, 1975).

However, as Joy and Tollefson also note, under the assumption of stationarity within the population over time, ex-post discrimination may be seen as being tantamount to prediction. Altman and Eisenbeis (1978) emphasise that this concept of prediction is not strictly appropriate in relation to an MDA model because MDA does not explicitly incorporate a time dimension. Joy and Tollefson (1978) retort that if indeed this is the case then MDA should not be used in studies that require, or propose, the concept that a prediction can be made about the future regardless of application focus. Altman and Eisenbeis (1978) note that Lachenbruch (1975) criticises the use of a holdout sample all-together, in favour of the Lachenbruch jack-knife validation approach previously discussed within this section. This is also supported by Scott (1978) who notes that studies in finance such as these, where data and sample sizes are often small, the use of the split sample techniques can lead to poor estimates and erroneous conclusions regarding the error rates with respect to the discriminant function. Regardless of the validation technique employed by each author, it is the erroneous and misleading use of the term "prediction" that Joy and Tollefson have issue with, and not any particular validation procedure, although I would concur with Scott (1978) in saying that reducing the estimation sample in order to provide a unique validation sample is somewhat counter-productive when a larger data population is likely to achieve a more realistic model.

More recently, studies have been able to forego these small quibbles on phrasal technicalities by constructing studies which can, as more data has become available through time, evaluate some of these earlier models on a true ex-ante basis. Heine (2000) and Agarwal and Taffler (2007) are prime examples of such studies where the performance of their original Z-score models are evaluated using new data from up to 30 and 25 years later respectively. These two studies provide results that indicate only a marginal drop in "predictive" ability compared to their original counterparts suggesting that discriminant analysis can be fairly robust over time, despite the associated problems and discrepancies.

With regards to the importance and interpretation of the variables within the discriminant function, it is important to note that the coefficients of the discriminant model do not indicate the hierarchic relationship between the constituting variables because they cannot be construed or interpreted in the same way as the coefficients of a regression. It may be the case that in a univariate context, one or more variables may prove insignificant in their discriminatory power, but add significant value when combined into a multivariate framework such as an MDA model (Altman, 1968). It is also possible that some variables may display counter-intuitive signs (Ooghe and Verbaere, 1985). It is therefore unwise (and incorrect) to attach too much importance to the magnitude (and sign) of the coefficients within the discriminant function.

Joy and Tollefson (1975); Eisenbeis, (1977); Scott (1978), for example suggest to the contrary that we can indeed attach coefficient importance to the individual variables, if the coefficients are standardised. Nevertheless, we need to recognise that the attempt to assess the role of the individual coefficients is inappropriate in view of the purpose of MDA (Zavgren, 1985).

Two further points of interest which are worth mentioning at this time are that the dependent variable is treated as being dichotomous, treating failure as being discrete, non-overlapping and identifiable thus not reflecting the true nature of financial distress and the various insolvency procedures.

"In reality, corporate failure is not a well-defined dichotomy. Therefore, dichotomizing failure seems inappropriate and the classic statistical modelling techniques, which are based on a dichotomy assumption, are applied inappropriately to the topic of corporate failure prediction. In fact, the arbitrary definition of failure may have serious consequences for the resulting failure prediction model". Balcaen and Ooghe (2006).

I have also previously mentioned that having an arbitrary choice of variables for inclusion in these discriminant functions, fails to allow any such to exactly follow any theoretical based arguments for corporate failure. These latter points will be referred to in greater depth as I progress through this section which culminates in a discussion regarding the main issues arising from the traditional approaches of corporate insolvency prediction and the "classical paradigm".

5.2.7 Summarising remarks

Despite its proven predictive accuracy, the Z-score often comes under fire from detractors who through various means seek to provide evidence that it is floored, be it statistically, theoretically or otherwise. Taffler (1995) is keen to point out that most of the concern regarding the Z-score is through a lack of understanding of what the score is designed to do. The Z-score model simply evaluates a company's financial profile and suggests whether it is more in line with that of a failed company or a non-failed one. Taffler (1995) argues that criticisms by Morris (1997) and others who suggest that the Z-score has no relevance because it is not based on any sound theoretical underpinnings are largely unjustified.

These models are not explanatory theories of failure but pattern recognition devices; by sifting through voluminous accounting data they are able to provide a readily interpretable communication tool portraying the likely distinctions between failed and non-failed company groups. Its “power” comes from the principle that the whole is worth more than the sum of the parts, and by considering the many different facets of economic information that is contained within a set of financial accounts simultaneously; it provides a great deal more information than conventional univariate ratio analysis. The liquid asset reservoir model (Walter, 1957) provides an adequate underpinning for the Z-score and if a more advanced theoretical framework for the Z-score could be developed, it would prove unnecessary move away from its actual purpose of being a pattern detecting device.

MDA procedures also use the assumption that the groups being investigated are identifiable and discrete. In the case of insolvency prediction this is questionable as the analysis usually classifies on the grounds of complete failure in the legal insolvency sense but as I have already discussed, failure comes in many forms and can often be rectified prior to any insolvency process.

Literature has shown us that sometimes it can be difficult to completely satisfy the classification problem (e.g. Altman’s grey area) and could therefore theoretically limit any practical application of the model, although Eidleman (1995) suggests that the use of these models is widespread.

A further critique of Z analysis is that judgements are made on past accounting data and do not necessarily reflect the actual current financial position of the company. In times of economic downturn such as today, rapid changes in the financial security of a company may be prevalent thus rendering the use of such techniques inappropriate. The empirical study contained within this thesis hopes to illustrate this fact and provide further evidence as to the validity of basing prediction models solely on past accounting data. These models also negate to reflect any changes in accounting standards or reporting practice which may occur, thus making year-on-year comparisons of the models again inappropriate.

The requirement of any Z approach to insolvency prediction modelling is therefore to continually redevelop and asses the models using ratios that measure more appropriately the different dimensions of a company’s current financial profile, performance measures, accounting regimes and current economic challenges.

Questions have also been asked as to the validity of some of the statistical techniques used in the development of the Z-score model and in particular the use of MDA as an appropriate technique[78]. One of the main arguments is that the use of MDA relies heavily on the assumption that the data is normally distributed; financial ratios however are generally not normally distributed[79] and tend to be highly skewed thus affecting the validity and significance of any estimated error rates used within the MDA technique. Certain transformations of data have been used in the past to try and approximate ratio distribution to normal, the most frequent being that of the standard or natural logs (e.g. Bates 1973). This transformation of the accounting data however may detract and change the information regarding the interrelationships between the variables being analysed and may affect the relative positions of observations within the group (Eisenbeis 1977). Gilbert (1968) compared three alternate models with that of the MDA function and concluded that the assumption of normality makes little or no difference to the predictive ability.

Efron (1975) interestingly found that classification models that use MDA under the condition of multivariate normality are more efficient than those employing logistic regression which are less restrictive. Contrary to this finding, the next section discusses the move away from MDA by researchers, possibly to overcome some of the statistic problems, towards conditional probability models such as logit and probit.

[78] For instance Joy and Tollefson (1975) and Eisenbeis (1977).

[79] And would actually be an impossibility for some.

5.3 The 1980s: conditional probability, a new contender

5.3.1 Introduction

The conditional probability model (CPM) comes in multiple forms depending on the Cumulative Density Function (CDF) assigned within the estimation process. Appendix (E) to this thesis provides a derivation of two of these forms, the logit and the probit models. For the purposes of corporate failure or insolvency prediction, the most popular choice of CPM is the logit model.[80]

The logit model is a conditional probability model which uses the non-linear maximum log-likelihood technique to estimate the probability of firm failure under the assumption of a logistic distribution. The parameter estimates are obtained using the logit model's maximum likelihood method as derived in Gujarati (2003). The resulting model is of the following form:

$$P_i = E(Y = 1|X) = \frac{1}{1 + e^{-(\beta_1 + \beta_2 X_i)}} = \frac{1}{1 + e^{-z_i}} \tag{5.17}$$

where P_i is the probability that firm i will fail ($Y = 1$) given a vector of attribute variables X (ratios, categorical or qualitative variables; for ease of exposition I simply denote just one variable X_i) for firm i, and the β_j are parameter estimates for each X_i. The function $(\beta_1 + \beta_2 X_i)$ for such models is often simply referred to as Z_i.

The logistical function ensures that the probability estimates are bounded between 0 and 1. To adhere to the majority of prior literature I refer to the model as being constructed using a dummy variable taking the value of 1 for a failed company, and 0 for a non-failed company. A high logit score indicates a high failure probability and poor financial health; and low logit score indicates good financial health and a low probability of failure. The underlying logistic function of the model implies that an

[80] Section 8 of this thesis provides evidence for the frequencies of the logit and probit modelling technology choice apparent in a selection of key academic literature. Academic studies employing logit estimations account for approximately 21% of the literature compared to only around 1% for probit modelling.

extremely healthy (weak) company, as compared to a firm of average financial health, must experience a proportionally larger deterioration (amelioration) of its attributes in order to deteriorate (ameliorate) its logit score (Laitinen and Kankaanpää, 1999).

The previous section of this thesis discussed some of the key pitfalls and assumption violations associated with the popular tool MDA (Edmister, 1972; Eisenbeis, 1977; Altman and Eisenbeis, 1978; Joy and Tollefson, 1975; Joy and Tollefson, 1978; Ooghe et al., 1994). As the literature enters the 1980s we observe a marked shift away from MDA in favour of the conditional probability model, possibly in an attempt to avoid some of these pitfalls and violations. In particular, the logit model neither requires multivariate normally distributed variables and secondly, nor does it rely on equal variance-covariance matrices of the two classification groups (Ohlson, 1980; Zavgren, 1983).

However, the logit model (as with all CPM's) and similarly to MDA, still relies on statistical features inherent in the classical prediction paradigm, namely: the dependent variable is dichotomous, treating failure as being discrete, non-overlapping and identifiable thus not reflecting the true nature of financial distress and the various insolvency procedures; the input variables are chosen arbitrarily, thus not conforming to any theoretical arguments; and the use of the term "prediction" could be construed as being incorrectly applied.

5.3.2 Ohlson: a first application of a logit model to insolvency prediction

The first published research within the field of corporate insolvency prediction to use the econometric methodology of conditional logit analysis was that of Ohlson (1980), although Ohlson himself alludes to several unpublished papers incorporating failure probabilities which pre-date his own study.[81]

Ohlson was under no delusion that the choice of a logit analysis would avoid some of the problems associated with the popular MDA methodology of which I have previously discussed. Furthermore, Ohlson suggests other reasons why conditional probability

[81] The papers which Ohlson refers to are White and Turnbull (1975a;1975b), and Santomero and Vinso (1977). The latter utilising the gambler's ruin approach (Appendix D).

models should be preferred: (1) the output (Z-score) of the MDA model has little intuitive interpretation, since it is simply an ordinal ranking device. If the researcher is able to ascribe a probability of failure to a particular firm then this would provide a greater intuitive appeal; (2) there are notable problems associated with the matching paired sample approach which is evident in all MDA studies. Pairs of failed and non-failed firms are usually matched by arbitrary criteria such as size and industry and Ohlson notes that:

"*It is by no means obvious what is really gained or lost by different matching procedures, including no matching at all. At the very least, it would seem to be more fruitful actually to include variables as predictors rather than to use them as matching purposes*". (Ohlson, 1980, p.112).

This latter point is demonstrated by Ohlson's empirical sample sizes. Populations of 105 failed firms and 2,058 non-failed firm years (from the period 1970-1976) are evidently larger (and more recent) than any previous study within the field of corporate failure prediction. Although he would have like to included more firm year observations for his empirical study, time restraints limited the number of firm years to just one per each company that was listed on the Compustat tape (database). Data for failed firms was collected up to four years prior to bankruptcy, in order that a one or two year-ahead forecast was potentially possible.

One further important consideration by Ohlson was regarding the timing of the annual report data relative to the bankruptcy date of the failed firms. The structure and make-up of the data sample is certainly a critical one, and the issue of report timing had not been explicitly considered prior to this study. Previous studies had simply taken annual report data as documented, and presumed available, at the company's fiscal year-end. In reality this is not the case. The Securities and Exchange Commission (SEC), responsible for implementing financial regulations in the US, allow for an account preparation time of four months after the fiscal year-end[82]. In the UK the statutory

[82] Four months is the preparation time allowed by the SEC at the time of writing this thesis. I am unsure as to what statutory requirements were in 1980 but can only assume something similar. Ohlson (1980) notes that many cases within his data have at least four and one-half months delay between the fiscal year-end and the date of the audit report. He suggests that it is possible for firms to apply to the SEC for an extension, although in many cases of bankruptcy, firms do not subsequently file any reports at all.

preparation time is a longer period of seven months as dictated by the Financial Reporting Council (FRC). Therefore in a real-world situation, a company's accounts may not be made available until several months after the official fiscal year-end. An issue therefore arises, when one is trying to accurately detect future insolvencies, if the derived model is based upon corporate failures which occur somewhere between the company year-end and the point at which the accounts are actually published. It is not unfeasible given the preparation lag time that such an occurrence would take place, and this occurrence may possibly lead to data items within the accounts being over, or understated, in order to appease investors or other stakeholders with a vested interest in the solvency of the company.[83]

One finding of the Ohlson (1980) study, with regards to the concept of this timing issue, is that previous empirical research seems to have overstated the predictive accuracy or discriminatory power of the resulting models. Ohlson suggests that evidence indicates that it will be easier to "predict" insolvency if models employ predictors derived from data within financial statements released after the event of bankruptcy.

To counter this issue of timing I need simply to ensure that accounting data is ascribed to the failed population of firms from annual reports which would have been available prior to the insolvency announcement. In most cases this simply means that the previous year's accounts are used as a substitute. A direct result of this circumvention of the timeliness issue means that the lead time in Ohlson's study (the time between the fiscal year-end date of the appropriate accounting data and the date of the insolvency), was fairly lengthy in comparison to previous studies, being an average of thirteen months, but nevertheless reflects more accurately the real world conditions that a decision maker would encounter.

The logit model itself follows the maximum likelihood estimation methodology as described in the introduction to this section and as defined in Appendix (E). Nine independent predictor variables were included in three models, representing one, two and a one-or-two-year-ahead forecasts. The nine variables were selected, not through

[83] Ohlson uses the term "back-casting" – a method in which the future desired conditions (insolvency) are envisioned and steps are then taken to achieve those conditions.

any theoretical arguments (observing that prior literature had also precluded such an approach), but largely due to their popularity and frequency in said literature. Ohlson also notes that no attempt was made to develop any "new or exotic" ratios, but they were simply chosen for simplicity. The nine ratios are as follows:

1. SIZE = Log (total assets/GNP).[84]
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 exceeds 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 = Change in net income.[85]

The coefficients of the resulting logit model (Ohlson's "model 1" for a one-year-ahead forecast) adhered to prior expectations regarding their respective signs and are shown in Table 5.11 along with the associated t-statistics regarding significance to the model. With a higher "score" or probability indicating a higher chance of failure, the three variables which I would expect to have positive signs (TLTA, CLCA, and INTWO), do indeed do so. The remaining variables (with the exception of OENEG)[86] all display negative signs, matching prior expectations that having a ratio of a higher value, holding all else equal, would lower the probability that a particular firm would become insolvent.[87]

---

[84] The index assumes a base value of 100 for 1968. Total assets are reported in dollars. The index year is as of the year prior to the year of the balance sheet date. The procedure assures a real-time implementation of the model.

[85] CHIN is calculated as being (NIt – NIt-1) / (|NIt|+|NIt-1|), where NIt is net income for the most recent period. The denominator acts as a level indicator and therefor the variable is intended to measure the change in net income.

[86] Ohlson notes that OENEG "serves as a discontinuity correction for TLTA. A corporation which has a negative book value is a special case".

[87] I must note a small mistake in the Ohlson paper. The author notes that three variables (WCTA, CLCA, and INTWO) have t-statistics that are less than two and are therefore not significant at a respectable level. Whilst this observation is of course correct, there is a failure to mention that NITA is also insignificant.

With regards to the accuracy of this particular model, using an initial arbitrary cut-off point of 0.5 where failure occurs above this value, the percentage of firms correctly "predicted" (within the validation sample itself) was equal to 96.12 percent[88]. This result is comparable to the high accuracies claimed by previous studies, but Ohlson is keen to point out several problems with this simplified evaluation of the model.

Table 5.11: Coefficients and t-statistics of Ohlson's variables

Variable	Estimate	t-statistic
SIZE	-0.407	-3.78*
TLTA	6.03	6.61*
WCTA	-1.43	-1.89
CLCA	0.0757	0.761
OENEG	-1.72	-2.45*
NITA	-2.37	-1.85
FUTL	-1.83	-2.36*
INTWO	0.285	0.812
CHIN	-0.521	-2.21*

* Denotes significance at the 5% level.

Firstly it must be noted that given the unequal populations of the failed (105 firms) and non-failed (2058 firms) groups, if I were to simply classify all of the firms as being non-failed then the overall accuracy of the model would equate to 91.15 percent (2058/2163). Therefore, unlike previous studies, due to the disproportionate samples sizes I have a clear portrayal that the marginal unconditional prior probability of failure is an important factor when interpreting the predictive accuracy of a failure prediction model.

[88] Ohlson notes that it would have been preferable to an "error analysis" on a fresh data set (use a validation sample), but deemed it impossible given the lack of data beyond the estimation period. An obvious option would have been to divide the sample into two and use one half for training and one half for validation, but he deemed that the primary purpose of the paper was to minimise the errors in the coefficients rather than getting a precise evaluation of a predictive model.

Secondly there is no reason to suggest that the cut-off point of 0.5 is adequate or indeed justified. Edminster (1972) showed that the optimal cut-off point should take into account both misclassification costs and prior probabilities of failure and therefore not rely on any arbitrary choice of cut-off, although I am of course now dealing with probabilities rather than continuous ordinal ranking and therefor 0.5 has some intuitive appeal. As I have also discussed in the previous section(s) with reference to earlier univariate and MDA methodological studies, the assessment of model accuracy has predominantly been based on the classification (or confusion) matrix as previously described and as shown in Table 5.5. The general assumption is that the two types of misclassification errors are treated as being equal due to the difficulty of cost assignment, and that these two error percentages have an additive property. The "best" model has therefore been determined as minimising the overall summation of these two errors.

Ohlson's model appraisal sought to challenge this somewhat traditional view of how model accuracy should be determined by evaluating his "model 1" over the whole range of possible cut-off points[89].

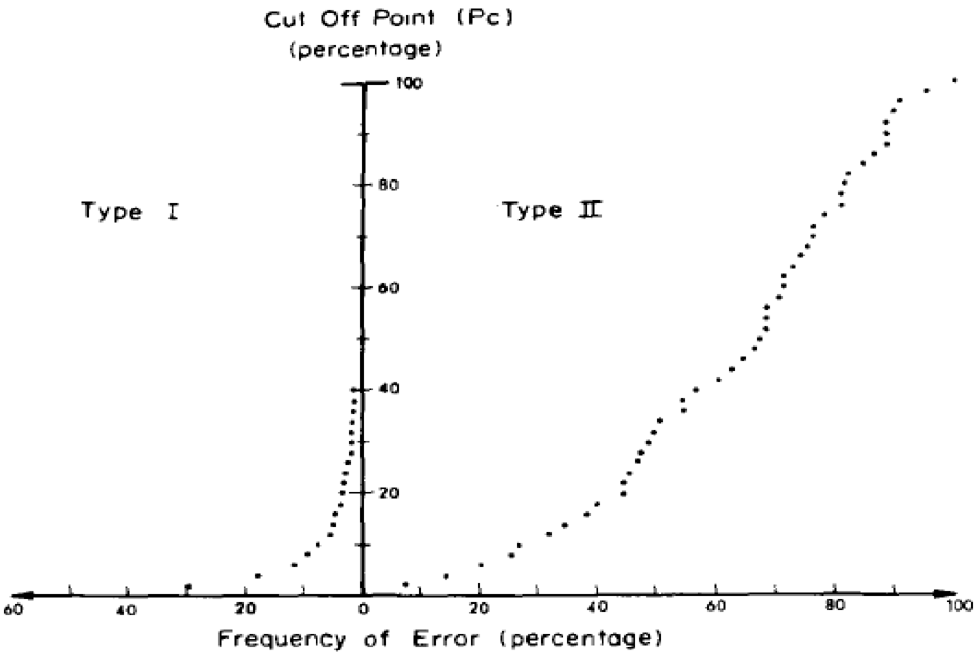


Figure 5.6: Ohlon's examination of error rates verses cut-off point

Source: facsimile from Ohlson (1980, p.127).

[89] This type of analysis is not too dissimilar to the receiver operating characteristic (ROC) curves which frequent modern insolvency prediction literature. ROC curves are discussed in section 8.3 of this thesis.

Figure 5.6 plots the error rates of Ohlson's model over various cut-off points between zero and 100. Contrary to my previous discussions on type I and type II errors and their occurrence in the prior literature (a type I error being the misclassification of a failed firm as being non-failed, and type II errors being the misclassification of a non-failed firm predicted as failed), Ohlson views his errors as the reversal. For this particular study (and I am not particularly sure why prior literature convention is not adhered to), a type I error is said to occur if the probability of failure of a non-failed firm is greater than the cut-off point, and a type II error occurs when a failed firm has a failure probability less than the cut-off point.

The cut-off point at which the sum of the errors is minimised is 0.038 which is somewhat difficult to determine by a simple glance at Figure 5.6. What we can observe is the expected shifts of the error rates as I reduce the cut-off point from 100 to 0. Type II errors drift from 100% misclassification down to zero because I would like all of the failed firms to lay above the cut-off point, and type I errors are the reversal. Notice however the difference in the curvatures of the two line plots. The plot of type II errors on the right hand side is almost linear in nature, curving slightly as it approaches the lower cut-off range. The line plot of the type I errors is much more concaved, curving in towards zero.

Remember that in absolute terms, that if we simply assigned all firms as being non-failed then the resulting accuracy would be 91.15 percent (2058/2163). This fact is one which is not reflected in Figure 5.6 as the percentage of failed firms is viewed equally to that of the non-failed group, although in outright values the two groups differ by a factor of twenty. We can therefore see that assigning all firms as being non-failed, the average misclassification rate would be fifty percent (and fifty percent accurate) when we use this appraisal technique, which reflects more appropriately the need to minimise both of the errors.

The error minimising cut-off point (3.8 percent) provides error rates as follows: type I errors rates show that 17.4 percent of non-failed firms would be misclassified, and type II error rates show that 12.4 percent of failed firms would be misclassified. Therefore given an infinite populations of firms of which half of the firms are failed and the other half are non-failed, then the expected combined error rate for the model would be 14.9 percent, or a model with an accuracy of 85.1 percent.

Ohlson notes that this result is one which cannot easily be compared to the prior literature, given the number of innovative methods which have been employed along with differences in the data sampling periods. Although Ohlson's logit methodology appears to be more appealing than discriminant analysis, and despite using a larger number of variables and firm populations than any previous study, the one-year-ahead model failed to deliver in-sample accuracy close to the high-ninety percentages that we have previously seen in the literature.

Three notable studies can compared and contrasted to the results gained by Ohlson as they utilise data from a similar sample period. These studies are Altman and McGough (1974), Moyer (1977) and Altman et al. (1977). Altman and McGough (perhaps the most similar study to Ohlson's research in terms of sample period and failure timeframe) applied the Altman (1968) Z-score to 28 failed firms within the period 1970-73 which resulted in an error rate of 18 percent, far from the misclassification rate of five percent within the original study; errors for the non-failed group of firms are not discussed but the results gained are reasonably similar to that of Ohlson. Altman et al. (1977) also uses data from the 1970s and as I have discussed in a previous section, the resulting accuracies were as high as 96 percent (93 percent holdout) for a one-year-ahead forecast. Ohlson suggests that accounting data lead times could explain the difference in performance, although in general he was unable to account for this difference noting that "In sum, differences in results are most difficult to reconcile" (Ohlson 1980, p.128), and further noted than another study of the Altman (1968) Z-score by Moyer (1977) resulted in a misclassification of 25 percent, indicating that there are many factors which are likely to affect the accuracy of an insolvency prediction model.

"There is always the possibility that an alternative estimating technique, other than the logit model used, could yield a more powerful discriminatory device. Unfortunately, a priori reasoning appears to be of no use in finding such an "optimal" estimating technique. All one can do is to try some alternatives. One approach I tried, MDA, produced results which were somewhat "worse" than those previously reported, in that the minimum average error rate was 16 percent. More generally, I would hypothesize

that many "reasonable" procedures will lead to results which will not differ too much". Ohlson (1980, p.129).

5.3.3 Comparing and contrasting MDA and logit methodologies

Following the work of Ohlson (1980), the most notable research to be published in the subsequent half-decade was perhaps that of Zavgren. I would like to continue by primarily discussing two of her papers, Zavgren (1983) and Zavgren (1985).

Zavgren (1983) is a small discursive paper describing several preceding insolvency prediction models whilst highlighting various statistical and methodological weaknesses attached to each of them. The "state of the art" paper covers four of the papers previously discussed within this thesis: Beaver (1966), Altman (1968), Deakin (1972), and Edminster (1972) and also two models that I have previously not mentioned, that of Diamond (1976) and Zavgren (1982)[90]. Zavgren (1983) has been frequently cited in the academic literature as providing a good overview (state of the art) of the literature at that point in time, and providing a comparative study with respect to their classification error rates. With promotion of this nature, it was with bated breath that I waited for a copy of this journal article, due to it being a "special" non-university-subscribed source. Upon delivery I must admit that I was a little disappointed. What I expected was a thorough, independent, empirical examination upon several key corporate insolvency prediction models, based on a new population of data. Instead, the paper is discursive, providing opinion and critique upon the models previously stated. Despite my disappointed expectations, I do agree that the paper is an excellent summary of previous work, particularly as it provides the first known comparison of 1960s / 1970s discriminant analysis modelling approaches with that of embryonic conditional probability application to corporate insolvency prediction.

Jones (1987) also offers a comparative paper, but again there is no empirical evaluation of the prior literature but rather a well written discussion of prior MDA and logit studies, coupled with an insight into another modelling technique, recursive partitioning (RP) which I will discuss later within this section. The issues of the

[90] Zavgren (1982) is an unpublished predecessor of Zavgren (1985) which utilises both logit and probit estimation techniques.

Discriminant Analysis models are covered in greater detail in this thesis than is in both Jones (1987), and Zavgren (1983), but the lack of an independent empirical evaluation of prior studies is one of the main inspirations for this thesis. The majority of papers in this area, including Zavgren (1983), focus upon finding floors in previous studies, leading to the introduction of a new and improved model derived from a new set of proprietary data. What I deemed to be lacking from the literature was a completely independent review and empirical evaluation of prior models with no ulterior motive[91]. This motivation and others are covered in greater detail in section 6: Objectives and research questions.

The principle aim of Zavgren (1985) was to further develop logit and probit insolvency prediction techniques by introducing significance tests for model variables and coefficients, as well as evaluating the information content contained within. Zavgren's model itself was based upon the ratios contained within Pinches et al. (1973) in which these authors identified seven ratios representing seven key factors attributable to firm solvency. Zavgren notes that the following ratios "*can make unprejudiced, accurate distinctions between failing and healthy firms*" and are detailed as follows: total income/total capital – representing return on investment, sales/net plant – representing capital turnover, inventory/sales – representing inventory turnover, debt/total capital – representing financial leverage, receivables/inventory – representing receivables turnover, quick assets/current liabilities – representing short-term liquidity[92], and cash/total assets – representing cash position. These seven factors are similar in nature to those used by several prior studies already detailed in this thesis, with profitability, liquidity, turnover, and leverage being widely regarded as the key constituents of a firm's solvency position.

Zavgren uses a matched sample of 45 failed and 45 non-failed US manufacturing firms from the period 1972-1978, a typically small sample for this time period, to

[91] One such study is that of McGurr and DeVaney (1998) who empirically evaluate the MDA and logit models of Deakin (1977), Gentry et al. (1985), Gombola et al. (1987), Ohlson (1980) and Zavgren (1985). This study was conducted upon a small sample of 112 companies within the retail industry and showed that the models performed considerably less well with the new ex-ante data, and significantly different to the results obtained by the original authors.

[92] Zavgren uses the acid test ratio instead of the current ratio as used by Pinches et al. (1973) because it eliminates accumulating stock levels which might be apparent and counter-active in the case of a failing firm.

develop logit and probit models for the prediction of failure. An in-sample test of this model using the standard "find the optimal cut-off point which minimises the total error rate" approach, yielded error rates of 18%, 17%, 28%, 27%, and 20% for years one to five prior to insolvency. These somewhat unremarkable results are barely worth a mention (noting the similarity between the one year misclassification rate to that of Ohlson (1980)), with further holdout tests making little progression in academic evolution.[93]

Despite this somewhat standardised introduction I have provided, Zavgren (1985) does lead us into a more innovative testing procedure, that of entropy theory. Based upon the work of Shannon (1948) and his theory for analysing the transmission of information through a communication network, and subsequently applied to accounting by Lev (1969). Entropy theory measures the information content gained or lost through successive predictions of firm failure, with entropy being defined as the degree of uncertainty of the occurrence of the event.

With regards to failure probability p, in a predictive context and if the actual failure of a firm is unknown, the entropy h would be defined as $h(p)$ and $h(1\text{-}p)$, for failed and non-failed event probabilities respectively. Therefore the expected entropy H can be weighted by its probability of occurrence and represents the uncertainty remaining over the occurrence of failure:

$$\begin{aligned} H &= ph(p) + (1-p)h(1-p) \\ &= p\ln(1/p) + (1-p)\ln[1/(1-p)] \end{aligned} \tag{5.18}$$

Logit and probit models are able to provide successive estimates of failure probability over a number of years as more information (messages) becomes available (usually in the form of annual accounts). Using these messages it is possible to measure the positive or negative increments in information contained within the model after the receipt of each new message, which can be represented by:

[93] 16 failed and 16 non-failed firms were used as a holdout sample from a subsequent time period, unlike previous studies e.g. Beaver (1966) whose holdout sample is derived from the same period as the training sample. Seeking inter-temporal generalisation for the first time, Zavgren's results showed an unimpressive 31% error rate.

$$\ln(1/p_2) - \ln(1/p_1) = \ln(p_1/p_2) \tag{5.19}$$

Where p_1 is the probability derived from the first set of information, and p_2 is the probability calculated from the new set of information. Theil (1967) presents this more formally; given two states and two consecutive messages the expected information is:

$$\ln(p_2/p_1) = p_1(s_1)\ln\left[\frac{p_1(s_1)}{p_2(s_1)}\right] + p_2(s_2)\ln\left[\frac{p_1(s_2)}{p_2(s_2)}\right] \tag{5.20}$$

Zavgren uses this theory to determine the increase/decrease in entropy, as her model provides failure probabilities, or messages, over a five year period. The results indicated that for the failed firms, eighteen percent more information was made available by the model over the five year period on average, measured by a decrease in entropy of 0.11 nits, increasing the certainty of failure. This result is comparable to that of Lev (1971) and his balance sheet decomposition model. For the non-failed sample of firms, the uncertainty of survival shrinks by sixteen percent over five years. Using Entropy theory, Zavgren provides alternative empirical evidence of failure detection being more easily determined closer to the event horizon, complementing the findings of previous studies which only used predictive accuracies as their basis.

Keasey and McGuiness (1990) offer an extension to the work of Zavgren (1985), and based upon UK data from 1976-1984, they sought to determine whether this entropy measure would be a reliable indicator to decision makers. Their results found similar to Zavgren in that their logit functions showed increasing information content as failure approaches over a five year period. They note that although the results were concordant with those of Zavgren, these ex-post studies do not realistically simulate a real-work decision making process as the survival state of the firm population is already known. In addition, Keasey and McGuiness (1990) apply an ex-ante application of the entropy information gained to a cross-sectional holdout sample, and found that the entropy measure did not show a general increase in information as the failure date approached. The conclude that "…*for the UK at least, it is questionable whether it would be sensible* *to conclude that we can rely on better use of information as the event of failure approaches*".

Despite the similarity in the nature of the results of Zavgren's logit models and those found within previous MDA-based models, the entropy theory cannot provide a directly comparable link between the two modelling techniques as its requirement for probabilistic estimates is not supported by the MDA approach. A method for converting the continuous score outputs into probabilities is discussed later in this thesis which provides a comparable link between the two methodologies, and indeed allows direct comparisons between multiple models of differing methodology.

In section 5.2.6 I discussed some of the pitfalls associated with the MDA modelling approach, and as the introduction of this section suggests, the adoption of conditional probability models by academics in the 1980s indicates a conscious attempt to avoid some of these pitfalls; in particular, the logit and probit models neither require multivariate normally distributed variables nor do they rely on equal variance-covariance matrices of the two classification groups. Given this information, one might expect that the conditional probability model would be able to more accurately classify failed and non-failed firms than the MDA approach. Despite this intuitive notion, there are actually very few empirical tests that try to determine which type of technique is actually the more accurate as a predictor.

One of the first studies of this kind was that of Press and Wilson (1978), whose breast cancer study directly compared the logit and MDA approaches. In this direct comparison, the authors provided two holdout sample tests for breast cancer prediction and found that the logit methodology provided a slightly better predictive accuracy with respect to MDA upon an identical data set. The first test accuracies were reported as 62% and 59% for logit and MDA respectively, and the second test yielded accuracies of 72% and 68%, showing that logit was the superior modelling technique. A similar result was obtained by Collins and Green (1982) who also compared the two modelling techniques upon failed and non-failed credit unions, a more relevant exercise with regards to this particular thesis. They found in holdout tests that the logit modelling technique provided a higher overall predictive accuracy (94%) than its MDA counterpart (91%). A more concise comparative exploration of the classification accuracies of the logit and MDA approaches can be found in Hamer (1983). Using the chi-square test statistic, Hamer evaluated the differences between the accuracies of each of the techniques, based upon four different variable sets over the five years before firm failure, and concluded that there was no significant difference between the two approaches. These results are consistent with those of Press and Wilson (1978) and Collins and Green (1982) who find similar explanatory power when comparing logit and MDA approaches. Lo (1986) also finds that under certain distributional assumptions, both procedures yield consistent estimates and the MDA estimator is asymptotically efficient.

This evidence advocates that, despite a small number of tests demonstrating the superiority of the logit model over MDA, in reality there is likely to be little difference in terms of predictive accuracy between the two technologies. Lennox (1999) sought to determine the reasons for these similarities in predictive abilities, and suggested that both logit and probit models could be improved in terms of accuracy if they are adequately and correctly specified. Lennox (1999) noted that previous studies do not report tests for misspecification, in terms of both omitted variable bias and heteroskedasticity; and that these factors may cause bias in both coefficient estimates and their standard errors. An application of Lagrange Multiplier tests as developed by Davidson and MacKinnon (1984) demonstrated that homoskedasticity could be rejected in a large sample of UK firms with respect to a linear probit model. Homoskedasticy could not, however, be rejected when tests were applied to both non-linear logit and probit models. Lennox (1999) found that non-linear effects were present in his cashflow and leverage variables and that once these effects were taken into consideration, a probit model's accuracy could be marginally improved. This improvement and overall accuracy of the "well specified" model however, was only significantly different from the MDA model over a very small cut-off range and, for the majority of cut-off points, the accuracy for each of the six model variants which were tested were not significantly different from one another.

5.3.4 Notable applications to UK data

Peel and Peel (1987) were among the first to apply logit analysis in the UK. In an attempt to refine the 'classic' financial ratio-based failure model, they added a number of non-conventional ratios and variables (Charitou et al, 2004). The major purpose of the paper was to explicitly allow for loss-making companies in the non-failed sample as they suggest that "A *common feature of previous work on failure prediction models of UK companies is that the non-failed samples are restricted to include only sound or healthy companies*". The focus of their criticism is aimed at Taffler (Taffler, 1984; Taffler and Tisshaw, 1976) in which they note that both studies exclude firms appearing less than "healthy" from the non-failed samples[94]. Peel (1987) discovered that when a sample of loss-making companies was included in his non-failed estimation sample, both the explanatory power and the classification accuracy of his reported (logit) models declined significantly (Peel and Peel, 1987).

The aim of Peel and Peel (1987) was to investigate the loss-making class of non-failed firms which prior studies such as Altman (1968) would often deem to fall within the "gray area". The data consisted of 146 firms comprising of 56 failures, 56 non-failed profit-making firms, and 34 non-failed loss-making firms and a three-group multilogit model was employed and compared to a three-group multi-discriminant model. The results indicated that the logit model correctly classified 66.7% of firms out of sample whilst the discriminant model correctly classified only 62.5%. The authors suggest that although the main study was only partly successful, there were a number of interesting side-results: the first being that a going concern qualification acted positively in favour of the firm surviving, in concordance with finding by Taffler and Tseung (1984), and the opposite to what one might expect. Secondly, the authors tried to determine whether it was possible for a model to distinguish between the different insolvency procedures (receivership vs non-receivership), but unfortunately could not find any variables which discriminated between the two categories.

Keasey and Watson (1987) investigate small companies and whether adjustments for Current Cost Accounting as recommended by SSAP16 would have any impact on insolvency prediction modelling. They find that there is little or no difference in classification accuracies between models using adjusted and non-adjusted financial data, noting that the CCA adjustment effectively reduces all the companies profit and other ratios by a constant amount. Platt and Platt (1990) investigated the effect of industry-

[94] See Taffler (1984) footnote 38, which describes the errors of a Datastream Z-score model being attributable to a sample not being restricted to only healthy (profit making) firms.

relative ratios on the likelihood of corporate failure and found that a model built upon industry-relative financial ratios and measures of industry growth was superior to a more traditional model using unadjusted variables. Platt and Platt (1994) develop an industry-relative "early warning" model utilising data of financially distress and non-financially distressed firms (no legally defined failures). They suggest that it may be possible to identify pending insolvency and take corrective actions to "*ameliorate financial distress before it disrupts production*". The authors also compare their model to a traditional insolvency prediction version with statistical tests rejecting the hypothesis that financial distress and insolvency are the same process.

"*It appears that financial distress happens to companies whose operating decisions perform below expectations. Bankruptcy, in contrast, seems to be a result of a decision that companies make to relieve themselves of problems such as excess debt levels*". Platt and Platt (1994, p.155).

The issue further treated in Lennox (1999) who demonstrated that the industry sector, company size and the economic cycle have important effects on the likelihood of corporate failure. He posits that corporate failure is expected to increase when the company in question is unprofitable, is highly leveraged, and it has embedded liquidity problems.

5.4 The rise of the artificially intelligent systems in the 1990s

5.4.1 An Introduction to neural networks

The Artificial Neural Network (NN) has been applied extensively to an increasingly wide variety of business areas, including financial forecasting, credit analysis, bond ratings, bankruptcy prediction and fraud detection (Charitou et. al., 2004).

NN's were originally conceived in the 1940s and 1950s (McCulloch and Pitts, 1943; Hebb, 1949; Rosenblat, 1957) as a way of mimicking the function of the human thought process, suggesting that the human brain exists as a series of interacting processes whose function depends on the interaction and connection between them. These processing parts, or neurons, tend to be grouped into cell assemblies with each assembly being self-organised to include information and impulses appropriate for certain behaviour. It was theorised that this self-organisation leads to unique and appropriate interconnections between each of the neurons.

Widrow and Hoff (1960) provides the first practical example of a computer-based neural network design in which a traffic control model was developed, and Hopfield (1982) developed this theory further identifying that the neural network procedure would provide "*rapid solutions to some special classes of computational problems*".

The NN consists of a number computational functions, neural nodes, organised in particular structure with particular inter-connections which are dependent on the designated application. The typical structure of a NN is that of the feedforward network. The basic feedforward network comprises three layers of neurons: the input layer, hidden layer, and output layer. The connections between these layers flow in one direction from the input node through the network in a forward *direction of activation propagation*. When activated, the input node(s) sends information via weighted parallel connections to each connected node in the hidden layer. Each node in the hidden layer sums the weighted inputs and forms a weighted linear combination of the values (z) before applying a non-linear transformation to produce its own output (y) which is then sent to the output layer. This non-linear transformation may follow, for example, the sigmoid function, an increasing function with asymptotic properties. The output layer receives weighted contributions from each of the hidden nodes, transforms these secondary inputs by again, forming a weighted linear combination of the inputs and applying the non-linear transformation which results in a single output (see Figure 5.7). Mathematically the operation executed by the neuron can be described by equations (5.21) and (5.22) given a particular training observation j with i variables.

$$z = \sum_{i=0}^{n} w_i \, x_i \tag{5.21}$$

$$y = f(z) \tag{5.22}$$

The transformations and outputs which occur within the hidden layer and the network itself are based on *'supervised' learning* and are done so by "training" the network using data from a training sample. The learning process within the network is determined by a training algorithm. The training algorithm is used to identify patterns within the data and uses the training sample as a template with the most popular of these being the back-propagation algorithm, based on the principle of continuous error feedback into the network[95].

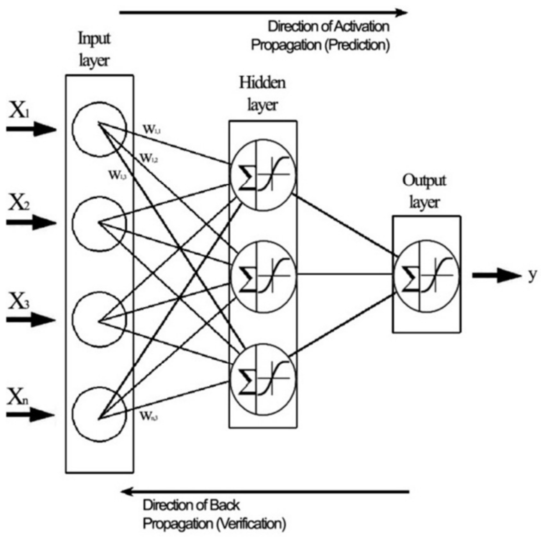


Figure 5.7: A feedforward neural network with backpropagation.

[95] The back-propagation training algorithm was originally developed by Rumelhart et al. (1986) and is discussed in Appendix I of this thesis.

Unlike the MDA and logit approaches, neural networks have the advantage that they do not rely on any pre-specification of a functional form, nor are they restricted in its assumptions regarding the characteristics and statistical distributions of the variables. Further, Neural networks are more able to quickly analyse complex patterns within the data and with a high degree of accuracy (Shachmurove, 2002), and that they are able to progressively learn from examples, given no prior knowledge (Back et al, 1996).

Neural networks do however have their downsides. Firstly neural networks will have a greater tendency to over-fitting (Trigueiros and Taffler, 1996) meaning that the model constructed during the training process, may very well fit the training data almost perfectly but when applied to a holdout sample of different firms the relative weightings may not be so appropriate. This problem, which leads to poor generalisation, is due to the large number of parameters usually involved in neural models (Becerra et al, 2005). The second problem is related to the "black box" nature of the neural network. With regards to this problem, and particularly within the context of failure prediction:

"*a neural network does not reveal the significance of each of the variables in the final classification and the derived weights cannot be interpreted. We have completely no understanding or knowledge concerning how the network classifies companies into the failing and non-failing group. It is impossible to understand how the relations in the layer-structure are estimated and hence, the network cannot be "applied" as such in order to classify new cases. In other words, the black-box problem makes it impossible to use the NN in practice, in a decision context"*. (Balcaen and Ooghe, 2004).

A further problem pertaining to the neural network can be attributable to its very construction. When faced with creating a neural network, the researcher has a number of options which are largely selected or applied on a somewhat arbitrary basis, namely: the size of the input layer (number of variables), the number of hidden layers, the number of nodes/neurons in each of the hidden layers, the number of epochs (training cycles), and the type of training algorithm to be used. Given that the construction of the network is likely to be critical in its application, it is strange that there is no consensus on any of the answers. The size of the input layer is a problem faced by any insolvency prediction methodology, and the optimal choice of variables, as this thesis would suggest, has been hotly debated since the conception of the art. Unless a model can be attached to some form of theoretical under-pinning, the arbitrary selection of variables with inevitably persist.

With respect to the construction of the hidden layers, in the majority of situations the is no clear-cut method for determining the optimal number of hidden layers and the number of nodes within them, without training several networks and estimating error rates for each. Too few hidden units will result in a high training error and high classification errors when applied to further samples because the model is under-fitted. On the other hand, if you have too many hidden units you might get a low training error but might still get large errors when applied to further samples because of over-fitting. Tetko et al. (1995) provide a discussion as to the effects of this trade-off, whilst others offer certain "rules of thumb" for selecting the number of hidden layers and nodes:

"*How large should the hidden layer be? One rule of thumb is that it should never be more than twice as large as the input layer*". (Berry and Linoff, 1997, p. 323).

"*To calculate the number of hidden nodes we use a general rule of*: *(Number of inputs + outputs) * (2/3).*" (www.neuralware.com).

"*A rule of thumb is for the size of this [hidden] layer to be somewhere between the input layer size ... and the output layer size ..."* (Blum, 1992, p. 60).

"*you will never require more than twice the number of hidden units as you have inputs in an MLP with one hidden layer*." (Swingler, 1996, p. 53).

"*Typically, we specify as many hidden nodes as dimensions needed to capture 70-90% of the variance of the input data set*." (Boger and Guterman, 1997)

With no recognised "best" method for the selection or construction of the hidden section of the neural network, a number of constructive algorithms have been proposed in an attempt to provide assistance to the researcher when defining the neural network structure (Becerra et al, 2005). Pruning algorithms start with a large number of neurons and subsequently eliminate those that are identified as redundant (Kun et al., 1990; Hassibi et al., 1993). Growing algorithms, on the other hand, start with a small network, which receives new neurons until the performance of the model is considered satisfactory (Ash, 1989; Setiono and Hui, 1995). A third possibility is the use of evolutionary algorithms (Yao, 1999), which aim at optimizing both the structure and the parameter values of the network by using global search methods inspired in the mechanisms of natural evolution. It is worth noting that the majority of these constructive methods employ search algorithms which are sensitive to the set of initial conditions (Becerra et al, 2005).

5.4.2 Application of NN's to the prediction of insolvency

By the late 1980s and early 1990s, the application of the neural network to insolvency and bankruptcy prediction was gaining popularity within academia, largely arising from the increase in computer processing power and off-the-shelf software which was becoming increasingly available for both academic and commercial usage. Couple this with the less restrictive assumptions of the modelling process, and we find that artificially intelligent systems, in particular the neural network, dominate the literature throughout the 1990s.[96]

The first application of the neural network to failure prediction can be attributable to Odom and Sharda (1990). Using a simple three-layered (1 x input, 1 x hidden, 1 x output) network with a back-propagation training algorithm they compared the accuracy of the network to a discriminant analysis model. The input variables for both models were the same five variables attributable to Altman's (1968) Z-score model, and in both cases these variables were provided by the annual accounting data of 74 companies from 1975-82. In comparing the accuracies of the two types of modelling technologies, Odom and Sharda found that the less restrictive neural network provided better results than that of the discriminant model, achieving a perfect classification within the training sample. Further application of the resulting models to a validation/holdout sample consisting of 55 firms, 27 of which were failures, showed that the neural network correctly classified 22 (81.5%) of the failed firms, whereas the discriminant analysis model only classified 16 (59.3%) correctly. An interesting point to make here is to the comparison between these results and those gained by Altman (1968). The original model achieved a successful classification rate of 96% in holdout sample tests of similar

[96] Daubie and Meskens (2002) provide an excellent meta-analysis of the literature of the 1990s.

size, much higher than either of the models provided by Odom and Sharda. What may have been beneficial to this study would be the direct application and comparison of Altman's original coefficients to Odom and Sharda's data set, in order to determine whether these original coefficients still provided such a high discriminant power, and how they compared to the new coefficients of Odom and Sharda's discriminant model, which were also not provided in the paper.

Odom and Sharda (1990) also make several observations, possibly criticisms, of the neural network modelling procedure. They note that the number of iterations needed to "learn (train) the data" was vast, converging after some 191,400 iterations and 24 hours on average. The complexity of the continuous feedback provided by the back-propagation algorithm and the limited processing power of late 1980s computers, indicates that although results are favourable for the neural network, the exercise and application in practice could be somewhat tedious.[97]

Tam (1991)[98] applied neural network architecture to Texas bank failures between 1985 and 1987. Two neural networks, one two-layered (no hidden layer), and one three-layered (one hidden layer consisting of ten nodes), were trained upon a sample of 59 failed and 59 non-failed banks utilising 19 variables. Similar to Odom and Sharda (1990) the three-layer network was able to successfully classify all of the failed firms correctly within sample. A holdout sample of 22 failed and 22 non-failed banks was used to compare the accuracies of these two networks with other modelling technologies (discriminant analysis, factor logistic, k nearest neighbour (kNN), and ID3 which is a recursive partitioning decision tree methodology[99]). The results of the holdout test demonstrated that both of the neural networks provided greater classification accuracies than any of the other modelling techniques, 18.2% and 14.8% overall misclassification for the two and three-layered networks respectively. The best discriminant model achieved a misclassification rate of 25% one-year prior to failure. Tam concludes that neural networks can easily be extended to other financial applications such as credit scoring and loan evaluation.

---

[97] This study reappears under the guise of Wilson and Sharda (1994) several years later. The study uses the same empirical results with different phraseology. One questions what happened to Mr Odom.

[98] See also Tam and Kiang (1992).

[99] For logistic factor analysis see West (1985); for kNN see Cover and Hart (1967); for ID3 please see Messier and Hansen (1988).

The comparison between neural networks and more traditional methodologies such as MDA and logit becomes a dominant theme throughout the nineteen nineties with the majority of papers championing the use of this “new” technology, which given its relative lack of constraints, continued to outperform these traditional counterparts. Salchenberger et al. (1992), for example, compared neural networks to logistic regression and found that neural networks out-performed logistic regression considerably with regards to their classification accuracy. With an 18-month lead-time, logistic regression provided a 86.4% classification accuracy while the neural network achieved an accuracy of 91.7%. Similarly Coats and Fant (1993) compared the discriminant analysis methodology to that of the neural network. Particularly, and similarly to Odom and Sharda (1990), the discriminant analysis model(s) that they compare is based upon the variables of Altman’s Z-score. Their results showed that classification accuracy of neural networks was between 81.9% and 95.0%, based upon on particular lead-times of four and one year’s respectively. The results of the discriminant analysis ranged between 83.7 and 87.9% between the same lead-times. Coats and Fant conclude that the NN approach is more effective than MDA for pattern classification and although the MDA approach produces consistently higher type II hit rates, the type I hit rates of the models are notably less reliable being as low as 63.8% in the year prior to failure.

A similar study utilising Altman’s variables was that of Zhang, Hu and Patuwo (1999). Zhang et al. compared the performance of logistic regression and neural networks without any prior specification of time to failure. Their sample consisted of production companies, and they applied a five-layer neural network with multiple interactions. They used Altman’s financial ratios for the input variables of the neural networks, adding current assets/short-term liabilities as an additional variable. The neural network performed considerably better than the logistic regression model, with a classification accuracy of 88.2% compared to 78.9%.

Fletcher and Goss (1993) conduct a similar comparison to Salchenberger et al. (1992) upon neural network and logit methodologies. Based on a small sample of eighteen failed and eighteen non-failed firms, matched by industry and size, three ratios were used as input variables into each of the models. The three ratios used were the current ratio (CR), the quick ratio (QR), and the income ratio (IR)[100]. Their small study showed that neural networks provided a higher predictive accuracy than that of a logit regression based upon identical variables and conclude in concordance with other prior studies that neural networks "*are a viable alternative to more traditional methods of estimating casual relationships in data*".

Kerling and Poddig (1994) performed a study upon French companies in order to compare neural networks and discriminant analysis performance over a three-year lead-time. Neural network gave 85.3–87.7% accuracy, whereas the accuracy of the discriminant analysis gave 85.7%. Kerling and Poddig also tried a number of interaction studies and an early stopping algorithm, without any ground-breaking results which might have led to a change in the direction of the literature.

Another neural network and MDA comparison was conducted by Altman, Marco and Varetto (1994). This time the study was based upon 1,000 Italian companies, and compared results based-upon a one-year lead-time. Although their study did not definitively demonstrate a clear "winner" when it came to classification accuracies, both over 90% accurate on holdout sample tests, it was surprisingly discriminant analysis which yielded marginally better results – contrary to the findings of prior studies. Altman et al. suggested that neural networks be used with caution and recognise the "black box" problem along with sometimes illogical weightings and behaviour patterns. The authors suggest that discriminant analysis "*compares rather well*" to the neural network, and although the "fine tuning" of the discriminant function takes longer than the programming of a network, the faster estimation speed of the model once it is up and running, makes it more cost efficient. They also note that discriminant analysis has the advantage in that it is possible to interpret the model's operating logic on the basis of the coefficients. Altman et al. (1994) concludes by suggesting that perhaps a combined approach should be adopted for "predictive reinforcement". The study was conducted in conjunction with the Centrale dei Bilanci, Turin, Italy, who aimed to incorporate the results of the study into a system which provides banks with a tool that quickly identifies companies in distress. Today, Centrale dei Bilanci are part of the European Committee of Central Balance and offer credit rating and risk assessment

[100] CR defined as current assets / current liabilities; QR defined as cash plus near -cash assets / current liabilities; IR defined as net income / working capital.

services to the Italian banking system. It is entirely possible that the Altman et al. study provided a foundation for the proprietary systems which they use today, although it has not been possible to uncover any direct link.

Alici (1995) was the first study to apply neural networks to UK data. The author used principal component analysis and self-organising maps to create a neural network to which he compared the results to those gained from MDA and logit methodologies. The data for the study was from the period 1987-92 and constituted 46 failed and 46 non-failed firms. The results showed that, neural networks gave 73.7% accuracy, as against that of discriminant analysis (65.6%) and logistic regression (66.0%); in line with the majority of prior studies of this nature but its application to UK data makes it particularly relevant to this thesis.

Leshno and Spector (1996) investigated a novel neural network comprising of cross conditions and consinus relationships. Several forms of the network were experimented with and depending on the form of the network, the resulting classification accuracy ranges from between 74.2% and 76.4% over a two-year lead time. These results were compared to more "simple" linear perceptron network which achieved a slightly lower accuracy of 72%. These results suggest that a more complex network adaptation may be able to improve classification accuracy, albeit only slightly.

Back et al. (1996) provide us with another comparative study. An interesting part of this study, however, is the comparison between the neural networks and discriminant and logit models. The key difference of this study compared to prior comparative studies is that the choice of input variables for each type of model was not identical throughout, but each modelling technique was free to "choose its own" variables. Back et al. (1996) used a new approach; that of a genetic algorithm to select variables which would be most apt to act as input variables to a neural network[101]. The authors note that:

"The variables for the input vectors can be chosen by an exhaustive search from the available variables, but this becomes very time consuming when the choice is to be done among several variables. Another method to choose the variables for the networks is to use a genetic algorithm". Back et al. (1996, p.4).

[101] The genetic algorithm process is described in more detail in Appendix J.

The genetic algorithm process selected a number of variables from a potential pool of 31, based on popular prior studies, for three neural network models, one, two and three-years prior to failure, consisting of 6, 7 and 15 variables respectively.

With regards to the discriminant analysis models, a step-wise selection process was used, similar to that of prior MDA studies such as Altman (1968) and Taffler (1984). Variables were chosen to enter or leave the model based upon the significance level 0.05 from an F test of covariance analysis. From the same 31 potential variables, the MDA process selected 4, 6 and 3 variables for the models one, two and three-years prior to failure respectively. A similar procedure was used to select the variables for the logit model although the number of variables selected as inputs was less than that of MDA at 3, 4 and 2 for one, two and three-years prior to failure. A summary of the variable population and to which model they are prescribed as input variables is presented in Table 5.12.

Table 5.12: Variable pool and their assignation to the different models

Ratios	Study
R1 Cash/Current Liabilities	E, D
R2 Cash Flow/Current Liabilities	E
R3 Cash Flow/Total Assets	E-M
R4 Cash Flow/Total Debt	Bl, B, D
R5 Cash/Net Sales	D
R6 Cash/Total Assets	D
R7 Current Assets/Current Liabilities	M, B, D, A-H-N
R8 Current Assets/Net Sales	D
R9 Current Assets/Total Assets	D,E-M
R10 Current Liabilities/Equity	E
R11 Equity/Fixed Assets	F
R12 Equity/Net Sales	R-F, E
R13 Inventory/Net Sales	E
R14 Long Term Debt/Equity	E-M
R15 MV of Equity/Book Value of Debt	A, A-H-N
R16 Total Debt/Equity	M
R17 Net Income/Total Assets	B, D
R18 Net Quick Assets/Inventory	Bl
R19 Net Sales/Total Assets	R-F, A
R20 Operating Income/Total Assets	A, T, A-H-N
R21 EBIT/Total Interest Payments	A-H-N
R22 Quick Assets/Current Liabilities	D, E-M
R23 Quick Assets/Net Sales	D
R24 Quick Assets/Total Assets	D, T, E-M
R25 Rate of Return to Common Stock	Bl
R26 Retained Earnings/Total Assets	A, A-H-N
R27 Return on Stock	F, T
R28 Total Debt/Total Assets	B, D
R29 Working Capital/Net sales	E, D
R30 Working Capital/Equity	T
R31 Working Capital/Total Assets	W-S,M,B,A,D

Legend:

A	Altman 1968
A-H-N	Altman, Haldeman, and Narayanan 1977
B	Beaver 1966
Bl	Blum 1974
D	Deakin 1972
E	Edminster 1972
E-M	El Hennawy and Morris 1983
F	Fitzpatrick 1932
M	Merwin 1942
R-F	Ramser and Foster 1931
W-S	Winakor and Smith 1935

Source: facsimile from Back et al. (1996, p.6).

Table 5.13: Variables selected for discriminant analysis

1 year prior to failure	2 years prior to failure	3 years prior to failure
R16	R14	R4
R4	R27	R5
R19	R5	R28
R24	R25	
	R11	
	R12	

Source: facsimile from Back et al. (1996, p.7).

Table 5.14: Variables selected for logit analysis

1 year prior to failure	2 years prior to failure	3 years prior to failure
R4	R14	R4
R24	R27	R5
R28	R5	
	R25	

Source: facsimile from Back et al. (1996, p.8).

Table 5.15: Variables selected for neural networks using genetic algorithm

1 year prior to failure	2 years prior to failure	3 years prior to failure
R1	R6	R1
R15	R9	R3
R17	R11	R5
R19	R23	R7
R24	R25	R8
R30	R27	R10
	R30	R14
		R15
		R18
		R19
		R20
		R21
		R22
		R27
		R29

Source: facsimile from Back et al. (1996, p.9).

What can be deduced from these variable selection processes is the large difference between the number of variables which are selected, being between 2 and 15. There is also a wide selection of variables being chosen by each process, showing little sign of a coherent or definitive decision as to what a "perfect" set of variables might consist of. R24, quick assets/total assets is the only variable which is included in all three models one-year prior to failure. R25 and R27 are also selected by all three models – this time two-years prior to failure. R5, cash/net sales, is the only variable which is selected by all three models three-years prior to failure. Back et al. also note that the variables with all of the logit models are a subset of those within the discriminant models and that only three variables R2, R26 and R31 are not selected at all.

Back et al. suggest the following reasons for these differences:

"*There can be real and significant different characteristics in different firms that can be measured by different financial ratios. Hence, alternative empirical methods use this information in alternative ways and thus all three models differ from each other. The second possible explanation to this divergence is that because we have selected a large amount of ratios in our original data that was used in model construction, it is quite obvious that there are measures which are highly correlated. If the ratios are measuring same economic dimension, high correlation can interfere the results or the difference in ratios can be so small, that the selection between two or more ratios into the model can be more or less random*". Back et al. (1996, p.10)

When comparing the predictive accuracies of each model, cross-validation was used instead of a true ex-ante holdout sample. The results that are presented, however, largely correspond to previous comparative studies in that the neural network presents a great predictive accuracy compared to logit and discriminant analysis. With an accuracy of 97.3% one-year prior to failure the neural network out-performed logit (96.5%) and MDA (85.1%). Interestingly for the models two-years prior to failure it is MDA which performs the best with an accuracy of 78.4%, with neural network and logit achieving 73% and 71.6% respectively. For models three-years prior to failure the neural network reclaimed the top accuracy classification with 83.8%.

Olmeda and Fernandez (1997) supplied further evidence that neural networks can provide higher classification accuracy than other traditional methodologies. Neural were compared to MDA, logit and two types of decision trees. The study upon Spanish banks suggested that yet again neural networks were the superior choice of modelling technique with an accuracy of 82.4%

Piramuthu, et al. (1998) developed a novel technique which allocated symbols to the inputs of multi-layer neural networks. They used their method on a sample of Belgian companies, and similar to that of Zhang et al. (1999) made no assumptions regarding lead-times. This symbol method yielded 82.9% accuracy, whilst the procedure which did not utilise the symbols was accurate at only 76.1%. The symbol method was also conducted upon American bank solvency positions over a period of two years. Again, this new procedure yielded far better results than its traditional counterpart.

A large study of Tan (1999) compared probit and a three-layer feed-forward back-propagation network using a sample of 2144 American credit institutions, of which 66 were insolvent. He conducted the analysis upon 13 financial ratios and 4 dummy variables (in order to avoid seasonal changes in the variables). The classification accuracy of the probit model was 92.5%, while that of the neural network having had a extremely large 3000 training epochs was 92.2%. The study at least brings some balance into the decade in which most papers of this nature posit in favour of the neural network in terms of classification accuracy. One does wonder however if this particular network was over-trained, given the large number of epochs.

Yang, Platt and Platt (1999) also offer some sobering thoughts perhaps in an attempt to stem the flow of positive neural network literature. Despite the popularity of the back-propagation algorithm Yang et al. point out that (a) it has a high computational intensity, (b) is unable to explain conclusions, and (c) lacks a formal theory which imposes a need for expertise on the user, something which will be discussed within the next section of this thesis.

Yang (2001) furthered his neural network research by developing an "early warning system" with probability-based neural networks using Bayes' classification theory and the method of maximum likelihood, comparing the classification accuracy with those of the usual traditional methodologies. Based on the data of 2408 UK construction companies, the probability-based neural network yielded a classification accuracy of 95.3%, greater than those of the backpropagation network (90.9%), logistic regression (88.9%) and discriminant analysis (81.3%).

A study by Charitou et al. (2004), added to the comparative literature with an aim to develop reliable failure prediction models for UK public industrial firms, and to explore the incremental information content of operating cash-flows in explaining and predicting financial collapse. They compare several methodologies and draw on some of the issues pertaining to UK insolvency prection:

"Why not use a model developed from US financial data to predict a UK company failure? There exists evidence that: (a) there are significant financial reporting differences between the two countries (Nobes and Parker, 1999), and (b) the UK and US also have different insolvency codes (Franks et al., 1996). As far as UK research in corporate insolvency prediction is concerned, evidence shows that such research was undertaken mainly in the 1980s and early 1990s.5 It can thus be argued that the models developed in those studies may not be currently applicable, given that various economic changes have occurred in the UK since then. This study employs a more recent UK dataset to develop and test a number of alternative insolvency classification models. In addition, since the majority of the existing UK failure classification models were based on the MDA approach, a method that was shown to suffer from certain limitations, the research is extended by applying more contemporary approaches; namely, logit analysis and the neural network methodology". Charitou et al. (2004, p.473).

The data used within their data consisted of 51 failed companies from the decade 1988 to 1997 which were paired with 51 companies which remained solvent in a matched pair analysis. A frugal model was developed to include only three variables (out of a possible 26), representing operating cash-flow, profitability and financial leverage. These variables were the result of a step-wise logistical regression and respectively representing the afore-mentioned categories were cash-flows from operations to total liabilities (CFFOTL), earnings before interest and taxes to total liabilities (EBITTL), and total liabilities to total assets (TLAT). These three variables were used as the input variables to a feed-forward neural network which unitised a conjugate gradient training algorithm (Charlambous, 1992), and compared to a logit model with the same three variables as inputs. Models were derived for one, two and three-years prior to failure for each methodology and for overall classification accuracy, were tested upon a holdout sample from the period 1995-97. Neural networks achieved the highest overall classification results for all three years prior to insolvency, with classification accuracies of 83.33%, 76.19%, and 75%. The logit model, although it achieved a lower percentage of overall correct classification accuracies (80.95%, 73.81%, 72.92%) produced slightly lower type I error rates with an average of 16% against 17% in case of neural networks.

Further analysis by Charitou et al. (2004) comprised of a comparison between Altman's (1968) model and those models aforementioned in the preceding paragraph. The five variables used within this model were applied with updated coefficients based upon a logit analysis rather than a discriminant one because of the limitations and restrictive assumptions associated with MDA. The resulting classification accuracy of the Altman variable model is comparable to the logit and neaural network in the first year prior to failure (82.5%) but is substantially less in subsequent years (62.5% and 68%). The results suggest that the transferability of variables from one side of the Atlantic to the other when it comes to insolvency prediction models is not particularly interchangeable, and that variables which seem to be well-fitting in one economy may not necessarily be applicable to another.

The final paper that I wish to discuss in this section is Becerra et al. (2005) a neural network study utilising UK data, subsequent to that used in Charitou et al. (2004), using failed (29 firms) and non-failed (31) firms from the period 1997-2000. A form of wavelet network[102] is developed and compared to a traditional multi-layer perceptron neural network and a discriminant model. The sample was split into training (21 failed and 21 non-failed firms) and validation (8 failed firms and 10 non-failed firms) sub-samples. The input variables to all of the models were that of Altman (1968)[103]. The resulting models when applied to the validation sample showed that the discriminant model had an accuracy of 78%, the multi-layer perceptron network had an accuracy of 83% and their version of the wavelet network had an accuracy of 89%.

[102] For an in-depth discussion on wavelet networks lease refer to Zhang and Benveniste (1992) who pioneer this methodology.

[103] Book value of equity was used instead of the market value of equity in the variable X_4 (VETL).

Although the neural network and its variants differ widely in their construction and statistical make-up, one notion seems apparent from this review of the literature. With few exceptions it is the neural network which consistently provides classification accuracies higher than that of the more traditional modelling techniques such as discriminant analysis and logistical regression. The reason for increase in accuracies of these types of model, is largely because they do not rely on any pre-specification of a functional form, nor are they restricted in its assumptions regarding the characteristics and statistical distributions of the variables. Their sometime time-consuming complexity, arbitrary construction, and the associated “black box” problem means, however, that practical application to future data is somewhat limited, and perhaps not as “attractive” to the general user as a clearly defined multiple regression model.

Neural networks, associated training algorithms, variants such as wavelet networks and genetic algorithms, have been discussed here within the context of corporate insolvency prediction. These methodologies are part of a class of technologies termed “Artificially Intelligent Expert Systems” (AIES), of which many more methodological examples can be found. Other constituents are methods such as survival analysis, decision trees, recursive partitioning, fuzzy rules, rough sets, and case-based reasoning. Balcaen and Ooghe (2004) discuss some of these alternative methods with respect to insolvency prediction, but I will limit this thesis to those already discussed. Although time factors do not allow a discussion of every single methodology employed in the field of corporate insolvency prediction since its conception, a meta-analysis of the literature can be found in section 8.1 which provides the frequency of each of the different methodologies within my large selection of literature.

5.5 Issues arising from the traditional approaches to insolvency prediction

So far, this thesis has discussed the most popular methodologies and technologies employed to predict corporate insolvency. With a limited number of exceptions, these methodologies have predominantly been statistically and mathematically driven, with no fruitful attempts to attach established theoretical underpinning to the problem of prediction.

"Business researchers employing discriminant analysis and classification have typically started out with a large number of variables and reduced them to a smaller number (m) for inclusion in the final model. There are virtually no studies in which the variables employed were determined on the basis of a theoretical model that leads logically to the specification of the variables." (Pinches, 1980).

All of these traditional approaches are essentially goal seeking exercises in the form of supervised classification which has become termed the *classical paradigm*.

"given a set of firms with known descriptor variables and known outcome class membership, a rule is constructed which allows other companies to be assigned to an outcome class on the basis of their descriptor variables". (Hand, 2004).

Balcaen and Ooghe (2006) note that this paradigm "*clearly fails to take account of some important aspects of the real problem of corporate failure prediction*", and offer discussions upon four problems associated with the classical paradigm. These problems are discussed in short summary below, but for a full discussion of the associated problems I would recommend Balcaen Oogue (2006) as an excellent and concise point of reference.

5.5.1 Problems related to the classical paradigm

The first problem associated with the classical paradigm is the arbitrary definition of failure which by convention, (for most statistical processes this will be a necessity), will assume a dichotomous dependant variable. The dependant variable separates the population of firms into failed and non-failed groups, but the definition of failure itself is an arbitrary choice made by the researcher. Usually failure is defined as the firm entering some form of legal process such as insolvency in the UK or bankruptcy in the US, which provides a simple statutory objective criterion on which the group separation can be based. A handful of studies however, have used alternate definitions of failure for example financial distress (e.g. Platt and Platt, 2002), where the distress definition is defined as being either several years of negative net operating income, suspension of dividend payments, or major corporate restructuring. Other failure definitions have also been used such as cash insolvency (Laitinen, 1994), loan defaults (Ward and Foster, 1997), and loan covenant renegotiations (Agarwal and Taffler, 2003).

It seems apparent that failure is not a clear or precise dichotomy, but rather a complex process which may develop and evolve over many years. Legal insolvency itself is often seen as a last resort where the firm may have been technically insolvent or financially distressed for time previous, and, even then, insolvency is only one possible exit for a financially struggling firm. Other possible routes could include mergers, acquisition, or dissolution through the sale of large parts of the business. These exit routes for the financially distressed firm are not easily observable on the back of annual accounting data, but may be preferable, or indeed more common, than insolvency itself. Whereas the omission of these instances as actual failures within the context of an empirical study will undoubtedly limit and reduce the population of firms within this class, it is not possible to attribute a clear "failed" status to a particular firm merger or takeover for example, as the true nature or scale of distress is likely to be known only to the management and directors. Therefore the use of clearly defined instance of insolvency as a proxy for failure within the context of failure prediction is seen as an inevitability.

Financial distress seems to be the symptom detected by many insolvency prediction models, indeed if correctly applied in a true ex-ante nature then this will of course be the case. What researchers should be interested in, are the final triggers that see a company's position change from being financial distressed due to any myriad of factors, into a declaration of a formal insolvency process. For this reason alone it is clear that a legal and formal definition of insolvency should be used within this type of modelling. Balcaen and Ooghe (2006, p.73) suggest that "*bankruptcy prediction models may be contaminated by firms that are declared bankrupt despite the fact that they do not show* *any real signs of failure*". In the UK under the Insolvency Act (2000) there are several forms of insolvency that may be entered into by a company which have already been discussed in section 4. With the exception of Members Voluntary Liquidation (MVL), each process must be met by some form of financial failure, be that technical insolvency, inability to repay debt, or other. Removing MVL from inclusion to the failure defined group ensures that voluntary winding-up does not contaminate the models in this instance.

A handful of studies have sought to address this binary classification problem by introducing additional failure groups. Lau (1987) investigated five different states; state 0 - financial stability; state 1 – omitting or reducing dividend payments; state 2 – technical default or default on loan payments; state 3 – protection under Chapter X or XI of the US Bankruptcy Act; and state 4 – bankruptcy and liquidation. Using a holdout sample for testing, Lau's model forecasted the probability of a firm entering each one of the five states. The results showed that she was able to predict entry into the non-failed state (state 0) with an accuracy of 85%. Classification of firms into the four states of failure however, was less fruitful with correct classifications of 50%, 67%, 20% and 20% for states one through four. Jones and Hensher (2004) use an innovative mixed logit methodology in an attempt to correctly classify firms into three failure groups; nonfailure, outright failure (defined as entering administration, receivership, or liquidation), and insolvency (defined as failure to pay listing fees, loan default, restructuring in order to meet debt payments, and capital raising to finance continuing operations). The firm classification by their model boasted an impressive 99.16% accuracy, suggesting that the mixed logit approach is one which should be investigated in further research.

The second problem with the classical paradigm relates to the instability of data and its non-stationarity over time. The traditional statistical forms of insolvency prediction models of which I have discussed so far, require that the relationships between the variables remain stable over time and will continue to hold true over time, particularly within the predictive context in which the majority of models are associated. Stationarity implies that there is a stable relationship between the dependant and independent variables, and also stable inter-correlations between each of the independent variables (Edminster, 1972; Zavgren, 1983). Wood and Piesse (1987) suggest that this is closely related to "population drift" (or "data stability").

Regrettably however, many studies have shown that data is not stable over long periods of time; in fact there is much evidence of data instability and non-stationarity (Barnes, 1982; Zmijewski, 1984). Data instability may be the result of many factors including inflation or interest rate changes, changes in economic conditions, changes in reporting practices, phases of the business cycle, or strategic or competitive changes within the marketplace. It has also been shown that data instability for financial ratios is usually at its greatest for firms that are about to fail (Dambolena and Khoury, 1980). It can therefore be argued that insolvency and failure prediction models will generally suffer from some form of stationarity problem (Charitou et al, 2004).

If we are to presume therefore that all traditional forms of insolvency prediction model suffer from the stationarity problem, then it is not possible to assume that a model will have robust or efficient failure classification accuracy when applied to future samples of data. Therefore these traditional models may need constant redevelopment or updating over time (Joy and Tollefson, 1975) resulting in new coefficients which will undoubtedly vary from sample to sample.

The third associated problem of the classical paradigm lies with sampling selectivity. Ooghe and Joos (1990) suggest that if a classic statistical failure prediction model is eventually to be used in a predictive context, the estimation samples of failing and non-failing firms should be representative of the whole population of firms. It is indeed evident from the earlier literature previously discussed within this section (Beaver, 1966; Altman, 1968; Deakin, 1972; Blum, 1974; Altman et al, 1977; Taffler and Tisshaw, 1977, Taffler, 1983; Keasey and Watson, 1987; Altman et al, 1995) that many models have been derived using the matched pair sampling approach where failed firms are matched by size, industry, or other to a non-failed mate. These samples are often hand-picked as it enables the researcher to control for some variables (e.g. size) that are thought to attribute some predictive power to the model but are not included in the final model's set of predictor variables (Balcaen and Ooghe, 2006; Ooghe and Verbaere, 1985; Jones, 1987; Keasey and Watson, 1991). Also, if the estimation samples are non-random it may be anticipated that the coefficient estimates are biased (Zmijewski).

To overcome this selection problem, studies such as Ohlson (1980) and later studies which I am yet to discuss (for example, Agarwal and Taffler, 2008; Hillegiest et al, 2004; Bharath and Shumway, 2008), use a much wider sample of firms within their empirical analysis, more representative of the true population, with the selections often containing data for all firms which are available at the time of research. Claims that non-random estimation samples may have serious consequences for the resulting traditional failure prediction model (Balcaen and Ooghe, 2006) can only be put to the test with a true ex-ante study, and is one of many claims and questions which my empirical study seeks to provide evidence.

The final problem associated with the classical paradigm is the choice of optimisation criteria, or put another way, the way in which the model's accuracy is assessed or measured. Many of the studies that I have reviewed so far use a variety of "goodness of fit" measures as their optimisation criteria including the percentage of (mis)classifications, often separable into both type I and type II classification errors, R^2-type measures, measures based on entropy, likelihood and discrete hazard models. In recent years the Receiver Operating Characteristic (ROC) curve, tests of relative information content and tests for economic value have more frequently been used. The variety of tests available to assess model performance, different sample constructions, and inconsistency of method within the literature, make it very difficult to directly compare one model to the next. With regards to the classical paradigm, by arbitrarily choosing a particular performance measure, the paradigm fails to take into account the true nature of insolvency prediction.

5.5.2 Neglecting of the time dimension of failure

Problems with corporate insolvency prediction also arise from the neglect of the time dimension of failure.

"The static classic statistical failure prediction models ignore the fact that companies change over time, which causes various problems and limitations". (Balcaen and Ooghe, 2006).

Many of the traditional insolvency prediction models are estimated and validated using a single observation set for each firm within the sample – one set of annual accounting data contained within one annual report. This relies heavily on the implicit assumption that annual reports pertaining to one particular firm are independent to one another, which is extremely unlikely to be the case. Several problems relate to this single observation or "snapshot" approach. Firstly, Shumway (2001) suggests that the choice, or timing, of when to observe a firm is likely to cause selection bias in the resulting model. For instance, if a company has a particularly poor financial year, those accounts may show the firm as being failed, whereas the adjacent accounts do not. Surely, however, such models are constructed in this way in order to detect these poor financial performances and warn stakeholders that certain care/precautions should be undertaken in the following year to put the company back on track.

There have been, to my knowledge, no studies that incorporate multivariate, multidimensional data into a single model which is easy to replicate and apply in practice. In the real world it is entirely conceivable that a loan officer would review each loan application on an individual basis, involving the perusal of many past years of accounting data. To incorporate many years of lagged data into an all-encompassing multi-firm model, would require so many variables it would unlikely to be practical. Several studies have investigated trend analysis and note its importance (Tamari, 1966; Edminster, 1972; Chalos, 1985), and other studies such as Beaver (1966) show that the trends of individual ratios vary over time, but none incorporate multiple data lags.

In addition, there is a "signal inconsistency problem" (Keasey et al, 1990; Luoma and Laitinen, 1991) pertaining to the use of a single year of observations. In an ex-ante predictive context, the repeated application of an insolvency prediction model to consecutive annual accounts for one particular company may result in a whole range of potentially conflicting predictions.

5.5.3 Problems related to the application focus

Perhaps the most characteristic problem associated with the traditional insolvency prediction models is there inherent arbitrary nature. In constructing a particular model, the researcher has an arbitrary choice of predictor variable, an arbitrary choice of companies, an arbitrary choice of timeframe, and an arbitrary choice of modelling technology and method, thus are the result of "putting the cart before the horse" (Cybinski, 2001).

Scott (1981) explained the problem with application focus as follows: "*Most bankruptcy models are derived using a paired-sample technique ... A number of plausible and traditional financial ratios are calculated from financial statements that were published before failure. Next, the researcher searches for a formula based either on a single ratio or a combination of ratios that best discriminates between firms that eventually failed and firms that remained solvent*" (p. 320).

Balcaen and Ooghe (2006) augment this notion by stating that "*The models are simply the outcome of a statistical search through a number of plausible financial indicators in order to empirically find some characteristics that distinguish between failing and non-failing firms. There seems to be no consensus on the superior predictor variables or on the superior modelling method*" (p.79).

With regards to the arbitrary selection of predictor variables, the lion's share of the traditional insolvency prediction models begin with some form of empirically-based variable selection method which reduces a large group of possible variables to a small chosen few. The initial assemblage of variables is usually hand-picked on the basis of their popularity in prior literature or/and their reported predictive ability in previous research (possibly creating a snowball effect as successful variables will be more likely chosen in future research, whereas less successful variables may not even be considered). Scott (1981) states that, "In searching for the best set of financial variables to predict bankruptcy, researchers are neither guided nor restricted by theory. They face an almost unlimited number of possible variables, and since many financial ratios are highly correlated [e.g., Foster (1978, ch. 6.2)], the choice of variables is often made on the basis of slight differences in predictive power" (p.325).

Beaver (1966) suggests that including variables based solely on popularity may be problematic, "The ratios advocated in the literature are perceived by many to reflect crucial relationships. It will be interesting to see the extent to which popularity will be self-defeating – that is, the most popular ratios will become those most manipulated by management (an activity known as window dressing) in a manner that destroys their utility" (pp.78-79).

Following the selection of potential predictor variables, they usually become subject to certain empirical and statistical tests such as their individual discriminant ability, the significant of the estimation parameters, predictive accuracy (classification results) gained through different combinations of variables, stepwise (discriminant analysis) procedures, principle components and factor analysis. The results of these tests advocate a reduced number of variables (an arbitrary number chosen by the researcher, most popularly between four and seven) for inclusion in the final model. One conclusion of Scott (1981) was that "While such tests are useful, they do not change the fact that the model itself resulted from an empirical search. The best evidence that some over-fitting has occurred is the existence of so many dissimilar prediction models. The danger is that the models so derived will not predict well when confronted with new data".

Despite not conforming to any theoretical underpinning, and the fact that there is very little agreement between researchers as to which variables are the best (Zavgren, 1983; Altman et al, 1977), several predictor ratios have proven to be more popular in the literature than their counterparts. Daubie and Meskens (2002) show that the most frequently used financial ratios are: current assets / current liabilities, working capital / total assets, EBIT / total assets, quick assets / current liabilities, and net income / total assets. Keasey and Watson (1991) suggest that empirically-based variable selection has some serious drawbacks, namely, the selection of variables is usually sample specific (Edminster, 1972; Zavgren, 1985), and that any empirical finding may not be suitable for generalisation (Gentry et al, 1897).

Furthermore, several studies (for example Moses and Liao, 1987; Ooghe and Balcaen, 2002) show that a high correlation between individual ratios can often explain some counter-intuitive signs for some coefficients, pertinent to many prediction models such as Ohlson (1980), Zavgren (1985), Keasey and McGuinness (1990), which contradicts the common perspective that variables representing factors such as size, liquidity and profitability should each act in their intuitive directions with respect to the model (Ooghe and Balcaen, 2002).

An additional criticism regarding the selection of variables within these traditional prediction models is the fact that they are nearly always accounting-number based, with the accounting numbers applied in the form of financial ratios. These financial ratios are gleamed from corporate accounts, which, although can be argued to be reliable representation of a firm's financial position, this position will always be of one in the past. Section 4 argued that the majority of insolvency prediction studies in the past have been conducted upon large, publicly owned companies due, in part to the reliability and availability of accounting data for these types of firms which consequently limits the research conducted upon other smaller firm types. Although larger firm accounts are subject to a higher level of scrutiny than their smaller counterparts, a second problem with the use of accounting numbers is whether or not this assumption holds that they do indeed represent a true and fair view.

Ooghe and Joos (1990) suggest and provide evidence that both non-failing and failing firms in particular, have many incentives to manipulate their annual accounting numbers, such as attracting further investment, appeasing shareholders, and "taking a bath" following a significant change of management. With respect to failing firms, they are likely to adjust their earnings upwards to portray a more positive view of their financial situation (Ooghe and Joos, 1990; Rosner, 2003; Charitou and Lambertides, 2003).

A further problem regarding the use of accounting numbers lies with the accuracy and reliability of the numbers provided by different financial databases and delivery platforms. Personal experience with data collection tentatively indicates that smaller providers such as Hemscott Company Guru provide inadequate data due to large rounding errors and numerous missing values, but this may be (to some extent) expected as smaller/embryonic databases will not have the capability, resources or financial backing of some of the larger database providers.

Several studies have evaluated and compared accounting and financial data provided by the larger, more established data providers such as Datastream, Thomson Financial, and Worldscope, although these studies are primarily focus on US application. The most recent study to this effect is that of Ince and Porter (2006) who determined that although there is a large inconsistency (over 35%) regarding the misclassification of equity securities, and differences regarding firm coverage, returns data is generally consistent between Thomson Datastream and the Centre for Research in Security Prices (CRSP).

Evidence regarding the accuracy and consistency of UK accounting and financial data provision is somewhat limited. Board et al. (1991) provide a descriptive overview of several databases but do not provide any empirical comparative evidence regarding either coverage or consistency. Lara et al. (2006) do provide empirical evidence that there is considerable variation in firm coverage between different databases and find that their valuation model was highly sensitive to the database choice, although the data gathered for this study was from fourteen European countries and not specifically UK targeted. Alves et al. (2007) concentrate solely on data gathered for UK companies upon six leading financial databases. They provide evidence that both firm and item coverage vary from one database to another. More importantly perhaps, is the evidence provided which suggests that the properties of standard income statement, cash-flow statement, balance sheet items, and associated ratios, also differ between each database. The authors attribute these differences to the way each database defines and constructs key variables. The choice of database for my own empirical study is considered in section 7.

The last regard to this section concerns the selection of modelling method/technology, an arbitrary choice made by the researcher. Limited theoretically based studies (e.g. the gamblers ruin approach: Wilcox, 1971) are apparent in the literature, with the vast majority of studies adopting various statistical techniques. Although studies such as Scott (1981) attempt to provide links between the variables found to be useful in predicting insolvency, and theoretical underpinnings, each of these traditional statistical approaches neglect to adhere to any sound theory; the most prominent theoretical explanations being attached to Walter (1957) and his loosely defined liquid reservoir theory. As I move into the following section and the turn of the century, there is, however, a definite shift towards a new group of modelling technique which is based on a sound theoretical underpinning: that of the option pricing formulae as derived by Black and Scholes (1973) and Merton (1973). This shift, as we shall see, also realises a distinct move away from the term “insolvency prediction” or “bankruptcy prediction” towards the somewhat more formative “credit risk” or “credit risk theory”.

5.6 Theoretically driven models of the 21st century

Despite the voluminous research, insolvency prediction suffered much criticism because of a lack of theoretical underpinning – giving rise to problems associated with the classical paradigm (choice of definition of firm failure, non-stationarity and instability of data, and sample selection) along with the arbitrary selection of variables and modelling method. These problems have been readily discussed within the previous section.

More recent modelling developments are the contingent claims models which are based on option pricing theory as set out in Black and Scholes (1973) and Merton (1974). The most popular forms of the model are derived from the European call option and the down-and-out call barrier option. In its simplest form, the model assumes that the shareholders of a firm hold a European call option (EC) on the firm, the exercise price being the amount required to discharge its debt liabilities. The time to maturity of the debt is, therefore, taken to be the option period; and the expiry date is taken as the point at which insolvency might occur. At expiry, the shareholders will either mandate discharge of the debt repayment obligation (if firm assets exceed liabilities) or mandate default on the debt and insolvency process (if firm assets less than liabilities). The method assigns a default (failure) probability independently to each firm. Academic examples of the EC approach can be found in Vassalou and Xing (2004), Bharath and Shumway (2008), and Hillegeist et al. (2004).

Several studies, including Brockman and Turtle (2002) and Reisz and Purlich (2007), derive default probabilities by extending the EC model into a barrier options framework. The model values the option as a down-and-out call (DOC). Debt holders are deemed own a portfolio of risk-free debt and a DOC option on the firm's assets which can be exercised should the value of the firm fall below a predetermined legally binding barrier.

Moody's were the first credit rating agency to publish details of their ratings methodologies. Prior to the acquisition of KMV by Moody's Corp in April 2002, very little was known publicly about KMV's proprietary credit risk appraisal. After the acquisition, a selection of papers written by KMV practitioners became publicly available to download from the Moody's KMV website, and contingent claims models have since attracted a deal of interest from academics.[104]

[104] KMV (Kealhofer, McQuown and Vasicek) developed quantitative risk management methodologies based upon option pricing theory. Moody's credit assessment as we see it today is developed from Vasicek (1984).

Two papers which were among the first to be released into the public domain by Moody's described its "RiskCalc" for both public and private companies; Falkenstein et al. (2000) describes Moody's RiskCalc for private firms whilst Sobehart et al. (2000) details Moody's public firm model. Merton's original construction of the options pricing model (1973), treats debt as having an explicit maturity (one year, for example), with the value of the option calculated upon this expiry date. In its simplest form all liabilities are also assumed to be zero-coupon bonds following Black and Scholes (1973) and Merton (1974)[105]. Firms will default on their debt when the market value of its assets fall below a certain level, which again, in its simplest form, will be equal to the value of the company's debt. The firm will therefore survive and pay its debt obligations if the market value of assets exceeds that of the debt value. The Merton model constitutes that the firm's market value of assets at time T when the debt matures, has a probability distribution defined by its expected value and standard deviation. What the model aims to provide is two-fold. Firstly the 'distance to default' is the number of standard deviations the future value of assets is away from the value of its debt, or default point. Secondly, associated formulae are able to calculate the probability of firm default, the greater the value of the firm, and the lesser its volatility, the lower the probability that the firm will default.

In practice there are a number of variations of the Merton model, both in construction and methodological application augmented by both practitioners and academics. The following is intended to generalise the model and formula that underlie these differing approaches. To calculate the default probability using the Merton model, the market value of equity and its volatility, as well as the value of debt, are required and observable. Using the options approach, the market value of equity is the consequence of the Black-Scholes formula:

$$V_E = Ve^{-\delta T}N(d_1) - Xe^{-rT}N(d_2) + \left(1 - e^{-\delta T}\right)V \tag{5.21}$$

[105] Hillegeist *et al.* (2004) point out that further to Merton (1974), "Subsequent studies have incorporated more realistic assumptions, such as allowing for debt covenants (e.g., Black and Cox, 1976) and multiple classes of debt (e.g., Geske, 1977)".

where $N(d_1)$ and $N(d_2)$ are the standard cumulative normal function of d_1 and d_2 respectively, and where:

$$d_1 = \frac{\ln\left[\frac{V}{X}\right] + \left(r + \frac{1}{2}\sigma_A^2\right)T}{\sigma_A\sqrt{T}}$$

and

$$d_2 = d_1 - \sigma_A\sqrt{T} = \frac{\ln\left[\frac{V}{X}\right] + \left(r - \frac{1}{2}\sigma_A^2\right)T}{\sigma_A\sqrt{T}}$$

The optimal hedge equation is defined as:

$$\sigma_E = \frac{[Ve^{-T}N(d_1)\sigma_A]}{V_E} \tag{5.22}$$

where:

V_E= market value of equity

X = book value of liabilities

V= market value of assets

T = time to maturity

r = risk free rate of return

σ_A = volatility of asset value

σ_E =volatility of equity value

From above we can see that there are two equations and two unknown variables: the market value of assets (V_A) and the volatility of assets (σ_A). Each of the remaining variables within the equations are market, accounting, and economic data which can be readily observed. Falkenstein et al. (2000) suggest the iterative Newton-Raphson technique to solve for V_A and σ_A noting convergance normally within a couple of iterations. Other authors such as Hillegeist el al (2004) use specific computer packages to solve the equations.

Falkenstein et al. (2000) and Sobehart et al. (2000) detail Moody's implementation of the Merton model in which the credit rating agency makes some beneficial adjustments to its formulation. The first of these adjustments relates to the "trigger point" of default (the debt amount), since in real life, a company will have many different liabilities with differing maturities, and hence the Merton model is ambiguous in its practical application. It may also be noted that a firm can continue to make payments on its debt obligations even if, technically, the firm is insolvent (liabilities are greater than assets). There may also be circumstances where delayed payments may be negotiated with one, or a number of debt holders, and through careful management legal insolvency may be avoided.

To allow for these particular characteristics, Moody's use half the book value of long term liabilities plus current liabilities as a proxy for the 1-year default point, which, according to Falkenstein et al. (2000), is "*a formulation based on empirical analysis*", and therefore differing to the total liabilities figure used in the conventional Merton model[106]. The authors argue that this adjustment is consistent with the distribution of recovery rates on defaulted bonds.

"*As opposed to having the highest recovery rates close to 100% and declining exponentially to zero, as implied by a strict interpretation of the Merton model, the mode recovery rate is more likely to be in the 50-60% range rather than the 90-100% range. That is, if the trigger point was total debt, the most probable recovery rate would be in the highest range; it is not, which suggests the trigger is well below the amount of total debt. Direct bankruptcy costs, such as legal bills, imply that 99% recovery would not be the mode even if the Merton Model were strictly true, yet empirical estimates of these costs are all well below the 40% that corresponds to the recovery rate averages we observe*". Falkenstein et al. (2000, p.18)

A second adjustment to the Merton model is made by mapping the distance from default into that of a probability. For the majority of cases, probabilities calculated from

[106] Vasicek (1984) was the first to introduce a distinction between short and long term debt which was adopted and developed by KMV. Another variant of the Merton model developed by Vasicek and Kealhofer (also KMV) assumes the firm's equity is a perpetual option with the default point acting as the absorbing barrier for the firm's asset value: See Kealhofer (2003). This model has come to be termed the Vasicek-Kealhofer (VK) model.

the Merton model are seemingly much too low, less than 0.1%, which would imply that the majority of firms should be issued with AAA bond ratings because they are extremely unlikely to fail. The probability of default can, according to the original Merton formulation, be derived from standard normal probability tables using the following formulation:

$$\text{Prob} \; = N\left[-\frac{\ln\left[\frac{V}{X}\right] + \left(\mu \, - \, \frac{1}{2}\sigma^2\right)T}{\sigma\sqrt{T}}\right] \tag{5.23}$$

where μ is the expected return on the firm's assets

To get around this "calibration" problem, Moody's map their output from the Merton model into actual defaults using historical data from proprietary data records, something that is of course not possible to achieve by anyone except those with access to Moody's proprietary defaults database. In academia therefore, researchers must be careful not to attach too much meaning to the distribution or scale of the outputted default probabilities. Although ordinal rankings of firms will be unaffected, the probabilities may need to be calibrated or placed in groups of percentiles, rather than suggesting that firm *X* has a definitive *Y%* chance of failure at the end of this year. These adjustments therefore suggest that the Merton model is more of a guideline than a concrete rule for estimating a quantitative model. "*The final transformation from standard normal probabilities into empirical probabilities implies that even the strongest proponents of the approach do not take the Merton model literally*". Falkenstein et al. (2000).

The Falkenstien et al. (2000) private firm model itself resides in a comprehensive publication spanning some 88 pages which describes Moody's default model for private, middle-market firms. The model was derived using a large database of 28,104 companies with 1,604 defaults and results in a model utilising ten financial ratios. As part of their analysis they investigate a number of different models and methodologies, finding interestingly that univariate models quite often outperform the earlier multivariate models such as the Altman Z-score. They also found that extra information could be added to the Merton model to produce a higher rate of model accuracy.

"We found that by adding information such as profitability, in conjunction with a 'distance to default' measure based on the Merton approach, we can significantly

improve upon a model that uses a stricter interpretation of the Merton framework". Falkenstein et al. (2000, p.12)

Sobehart and Keenan (1999) offer an excellent discussion as to the reasons why adding financial statement data to the Merton model can improve its predictive accuracy and overall performance, and Sobehart and Stein (2000) argue that the Merton model can easily be improved upon. These claims are investigated further in section 10 of this thesis where I too attempt to add power to that of the Merton model by constructing a best-of-breed model. Kealhofer and Kurbat (2001) retort suggesting that the Merton models capture all of the available information contained within traditional credit agency ratings based upon tried-and-tested accounting variables, a claim supported by Duffie and Wang (2004) who show that the Merton model has significant predictive power over time.

With regards to Moody's public firm model, Sobehart et al. (2000) suggest that neither historical financial statement data, nor information gained through market data and the Merton model, adequately capture all distress and default information.

"*The analysis of historical financial statements may present an incomplete or distorted picture of the company's true financial condition. For a variety of reasons including the intrinsic conservatism of accounting principles, financial statements do not necessarily reflect the complete economic reality of the firm.*"

"Merton's original contingent claims model, and most subsequent refinements of it, does not contemplate cases in which firms default on their debt obligations due to severe liquidity problems. In addition, even in situations where market equity contains relatively complete information about a firm's credit quality, this does not guarantee that a structural model will correctly capture and reflect that information. Knowledge of the market information alone does not directly inform an investor as to a borrower's creditworthiness."

The exact methodology that Moody's use to calculated their credit ratings and default probabilities therefore, incorporates components of both the Merton model approach and an empirically derived statistical model based upon accounting and other data, resulting in a hybrid of the two forms. Sobehart et al. (2000) describe this hybrid model as having five key inputs:

- Agency rating when available,
- Modified version of the Merton model (expressed as a distance to default),
- Company financial statement information,
- Additional equity market information; and
- Macroeconomic variables that represent snapshots of the state of the economy or of specific industries which are used for pre-processing model inputs.

More precisely, the nine variables which are used as inputs for Moody's risk model for public firms are listed below, although the precise model derivation, methodology and resulting coefficients for the model are not disclosed due to its proprietary nature.

- Credit Quality (previous rating)
- Return on Assets (ROA)
- Firm Size
- Operating Liquidity
- Leverage
- Market Sensitivity (stock price volatility)
- Equity Growth
- Return on Equity
- Distance to Default (from Merton model)

The model is compared to several other credit risk models using Receiver Operating Characteristic curves (see section 8 of this thesis) and the associated accuracy ratio. The comparative models which are particularly of interest to this thesis are the Merton model itself, and Altman's original Z-score. The authors also compare the public firm model to a Hazard model, a reduced Z-score model, and the single variable ROA. The comparative analysis is conducted using proprietary data covering some 1,406 defaulting firms and 13,041 non-defaulting firms (the largest dataset cited within this thesis).

What results is that the public firm model achieved the highest classification accuracy[107] of all the models (86.5%), with the Merton model a close second with an accuracy of 83.5%. these results compare favourably to those of the original Altman Z-score which only achieved and accuracy of 78%. It is clear therefore that although the contingent claims Merton model out-performed the somewhat antiquated Altman Z-score, the combination of both market and accounting variables within the hybrid model, provides substantially more information regarding the solvency of US public firms.

Academic studies that evaluate the performance of the Merton model become apparent in the literature, subsequent to these publications by Moody's, and are offered in two generalised forms. Vassalou and Xing (2004), Tudela and Young (2003), Hillegeist et al. (2004), and Bharath and Shumway (2008) investigate variants and comparative studies based upon the simplified Merton model (single period European call option), whereas Brockman and Turtle (2003) and Reisz and Purlich (2007) extend the formulation into a barrier options approach.

Vassalou and Xing (2004) examine the impact that default risk has on equity returns. Small firms which also have a high risk of default are likely to have significantly higher returns on equity than larger companies who have sound solvency credentials. They provide empirical evidence to suggest that in general, firms with a high risk of default will provide higher returns than firms with low classes of default risk, providing that they are either small in size or have a high book to market ratio.

Tudela and Young (2003) represent the Bank of England and investigate the Merton model's ability to produce a reliable ordinal ranking of companies based upon UK company data. Their dataset consisted of 65 companies which entered receivership between 1990 and 2001, and 7,459 firm year observations for those which did not fail. For one-year-ahead forecasts the Merton models does a sound job in correctly classifying 62 out of the 65 failures, although the model misclassified 20% of the non-

[107] Based upon the area under the ROC curve (AUC). Not disclosed by the paper but can easily be calculated from the provided Accuracy Ratio. AR = 2*(AUC+0.5).

failed group. The authors suggest that based on this evidence, "*the Merton model appears to be useful tool for identifying companies at risk of receivership*".

Brockman and Turtle (2003) provide a framework for extending the traditional form of the Merton model (standard call option), into one which is expressed through a closed-form, down-and-out-call, barrier option approach (DOC).

The primary feature distinguishing a DOC from a standard call is the existence of a barrier which, when breached, causes the termination of the option (i.e., failure). The barrier level can be set at, above, or below the exercise price of the option, although it must be set initially below the underlying asset value to be viable. A DOC is classified as path-dependent because its payoff is not simply a function of the terminal value of the underlying asset, but is dependent on the specific path taken by the asset over the life of the option. If the barrier is hit at any time before maturity, the DOC expires worthless even if the underlying asset value subsequently recovers. In place of a zero payout upon hitting the barrier, DOCs can pay out a positive value referred to as the rebate… The DOC framework explicitly recognizes the consequences of bankruptcy whenever asset values fall below a pre-specified barrier value. Asset ownership is transferred from shareholders to creditors and any subsequent rise in asset values will accrue to creditors since equity holders' residual claims have been permanently extinguished. In this way, equity value behaves as a DOC option on the underlying assets of the firm. (Brockamn and Turtle (2003, p.514-5).

The probability of firm default using the barrier option approach is calculated similarly to the Merton model albeit with a more expansive formulation as per Equation 5.24.[108]

$$V_E = DOC = V\mathrm{N}(d_1) - X\mathrm{e}^{-r\tau}\mathrm{N}\big(d_1 - \sigma\sqrt{\tau}\big)$$
$$- \left[V\left(\frac{B}{V}\right)^{\frac{2(r)}{\sigma^2}+1}\mathrm{N}(d_1^B) - X\mathrm{e}^{-r\tau}\left(\frac{B}{V}\right)^{\frac{2(r)}{\sigma^2}-1}\mathrm{N}\big(d_1^B - \sigma\sqrt{\tau}\big)\right]$$

[108] There are several versions of this formulation. This particular version is risk neutral and negates to include the optional rebate addition, and does not include dividends which accrue to the equity holders. The latter is included in the formulation within section 8.2. Brockman and Turtle (2003) offer several variants of this formulation.

$$d_1 = \frac{\ln(V/X) + \left(r + (\sigma^2/2)\right)\tau}{\sigma\sqrt{\tau}}$$

$$d_1^B = \frac{\ln(B^2/VX) + \left(r + (\sigma^2/2)\right)\tau}{\sigma\sqrt{T-\tau}}$$

(5.24)

The definitions of the terms within this model are identical to those contained within Equation 5.21 but with the addition of the firm-specific barrier B, T, the date of expiry, and τ, the current date ($(T-\tau)$, therefore equating to the time until expiry). Similarly to the Traditional model, the unobservable variables V (market value of assets) and σ (asset volatility) can be solved simultaneously by iterative or computative methods using equations (5.21) and (5.23). The barrier B can then be solved from Equation (5.24). The risk-neutral probability of default can then be calculated thusly:

$$\text{Prob} = 1 - \text{N}\left(\frac{\ln\left(\frac{V_\tau}{X}\right) + \left(\mu - \frac{1}{2}\sigma_A^2\right)(T-\tau)}{\sigma_A\sqrt{T-\tau}}\right) + \left(\frac{B}{V_t}\right)^{\frac{2(\mu)}{\sigma_A^2}-1} \cdot N\left(\frac{\ln\left(\frac{B^2}{XV_t}\right) + \left(\mu - \frac{1}{2}\sigma_A^2\right)(T-\tau)}{\sigma_A\sqrt{T-\tau}}\right)$$

(5.25)

Similar probabilities can also be found in Cox and Millar (1965), Ingersoll (1987), Rich (1994), and Reisz and Purlich (2007).

Although this new wave of academic and proprietary adoption of the theoretically-driven credit risk model has become popularised in recent years, we must remember that other theoretical models have been pursued in the past such as Wilcox (1973) gambler's ruin theory. For whatever reasons, however, these alternative theories have not gained the popularity within academia as compared to that of the contingent claims models[109]. What is evident, however, is that the release of these proprietary publications sparked a deal of interest within academia, with a selection of papers keen to prove that the

[109] Blums (2003) provides an interesting study which offers a model combining both the gambler's ruin theory and option pricing theory. The model was able to correctly classify 100% of failed firms but only 43% of non-failed companies.

traditional models based upon accounting data could still compare to these more modern techniques.

Comparisons of the performance of contingent claims or ‘market’ models with that of traditional accounting-number based models forms the backbone of much of the empirical analysis found in the recent research; and such comparisons have yielded a variety of results. Tudela and Young (2003) find that their market model out-performs models which use accounting data, and that when the output of their market model is combined with accounting data accuracy can only slightly be improved. Hillegeist et al. (2004) compare their EC model with two accounting number-based methods (being the Altman 1968 ‘Z-score’ model and the Ohlson 1980 model), and find that the market model carries more information than the accounting number-based comparators. They argue that since accounting statements are prepared on a going-concern basis, they are, by design, of limited use in predicting bankruptcy.

Bharath and Shumway (2008) construct a naïve version of the EC model in order to avoid some of the estimation complexity found in Hillegeist et al. (2004) who employ simultaneous equation solving of the unknown Merton model variables. Bharath and Shumway (2008) find little deterioration in model performance using this simplified method and suggest that the KMV-Merton model probability of default is a marginally useful default forecaster, but not a sufficient statistic for default. Reisz and Purlich (2007) compare both EC and DOC models with the Altman Z-score model; and find, in contrast to the findings of Hillegeist et al. (2004), that the Z-score model out-performs both contingent claims approaches in terms of receiver operating characteristic curves for a one-year ahead forecast. Brockamn and Turtle (2003) compare the DOC model to Altman’s Z-score and find that the DOC model dominates the Z-score in most cases in terms of predictive ability. For the UK, Agarwal and Taffler (2008) find similarly to Reisz and Purlich (2007), comparing the Taffler (1983) accounting number-based model (the Taffler Z-score) with EC models of both Hillegeist et al. (2004) and Bharath and Shumway (2008). In their analysis, the Taffler Z-score achieved the highest accuracy (89%), with the models of Hillegeist et al. (2004) achieving 84% and Bharath and Shumway (2008) achieving 87%. Book to market and size variables were also examined, with the resulting accuracies being 68% and 75% respectively. Agarwal and Taffler (2008) suggest that:

"...traditional accounting ratio-based bankruptcy risk models are, in fact, not inferior to KMV-type option-based models for credit risk assessment purposes ... The apparent superiority of the market-based model approach claimed by Hillegeist et al. (2004) reflects the poor performance of their comparator models, not a particularly strong performance by their option-pricing model". Agarwal and Taffler (2008, p. 1,542)

Gharghori et al. (2006) provide another comparative study as to the options verses accounting-number model debate. They compare the traditional Merton model to a barrier option variation, the book to market ratio, size (market value of equity), and a model based upon accounting numbers. Using Australian data, they find that the barrier option model is superior (when the barrier was set to below the value of liabilities), with a fractionally higher accuracy than the Merton model, the accuracies being 86.25% and 86.24% respectively. Book to market ratio provided and accuracy of 76.78% whereas the variable size only achieved an accuracy of 64.3%. The accounting number-based model achieved and accuracy of 76%, far less than the options-based models.

With conflicting academic evidence as to whether the contingent claims approach to insolvency prediction and credit risk is indeed superior to the more traditional forms of insolvency prediction modelling based upon accounting data, a large empirical study is undertaken within this thesis to provide further (and updated) evidence to this debate.

6 Objectives and research questions

6.1 Objectives

From the extensive review of prior literature which has been discussed within the previous section, it is apparent that the volume of previous research within the field has been vast. Although I do not claim to have covered every single piece of research that has previously been published, I am confident that all the key developments and enhancements have been conversed, which is hoped to give the reader a clear overview of the evolution of the art that is corporate insolvency prediction and credit risk modelling. Given this voluminous prior research, it is surprising that the art still attracts a great deal of research to this very day. Many current studies are still determined to hunt down a perfect prediction model but ultimately (and sadly) fail to acknowledge the key methodological issues and detractions arising from the prior research, often employing a somewhat arbitrary and gung-ho approach to arrive at yet another insolvency prediction model.

In a later section of this thesis, I aim to show that creating a new model, given an arbitrary choice of variable and modelling technology, is a fairly simple process which I really have no interest in promoting. One common and potential beguiling question is whether or not a best of breed model may be constructed, based upon the best performing variables from each of the models tested. I have always provided the same answer to this comment in that this kind of research has little or no interest to me. I do feel however that this exercise is one which should be included within this thesis, if only to demonstrate how easily a model can be fitted and augmented around a particular data set, pass it off as being the most accurate model ever produced, and then realise that it is extremely unlikely that it would ever deliver the same efficiency if it were applied to a brand new set of data.

The main objective of my thesis therefore, is not to succumb to this pitfall, but to provide an independent empirical analysis of extant corporate insolvency prediction technologies, in an attempt to discover if there is a single technology, methodology, or set of variables which outperforms all of its counterparts. In doing so, I can gain an impartial empirical study of extant models using up-to-date data, and determine their relevance, efficiency, and applicability to the uncertain financial climate which is the early twenty first century and therefor determine if any one model/method is superior to its counterparts. It is also hoped that this study will be able to identify any potential room for improvement of the existing methods and technologies, with particular reference to UK firms within this current economic cycle. Choosing models which represent key developments in the literature, it should be possible to provide a truly independent empirical overview of the evolution of the art and provide an academic catalogue and reference upon insolvency prediction and credit risk technologies.

The conclusions which result from this study could have great importance, and have a dramatic impact upon the whole of the financial fraternity. The up-to-date information it will provide to investors (particularly debt lenders) and regulators on the appropriate application of insolvency prediction and credit risk technologies, could prove to be invaluable and potentially save the financial sector not only monetary assets but the publicly paraded embarrassment which has recently and frequently been associated with loan defaults. With regards to the recent financial crisis there has been an obvious and well document drying up of credit where banks and other businesses are reluctant to lend monies, and will only tend to do so to the most credit worthy of companies. Whatever the reality maybe however, it is not in the interest of the bank *not* to lend money, it is after all how the institution makes the majority of its profit. There is therefore an inherent need in today's climate not only in the identification of good credit, but an increased important upon the accuracy of any loan decision or credit rating systems, in order that loan applicants can be adequately ascribed the correct or appropriate rate of return.

It is of course somewhat difficult to determine exactly what technologies and methodologies are used by individual debt lenders as each will have their own proprietary systems. Evidence within the literature suggests that the major credit rating agencies such as Moody's rely on models based upon and adapted from the Black-Scholes and Merton option pricing formulae. In order to assess the credit risk of larger public corporations, many other debt lenders such as banks will each have their own databases, proprietary systems, and algorithms, each being developed and adapted over some time. It is therefore possible that this thesis could help any number of stakeholders identify some of the pitfalls associated with their respective systems and help make apparent any possible improvements.

6.2 Research questions

In mind of the research objectives detailed above, it is clear that in order to fulfil them, that there are several underlying questions which arise. The most important perhaps, certainly with respect to the recent financial crisis and the increased importance to adequately asses the credit worthiness of potential loan applicants, or to appraise the solvency of a potential investment, are the related questions: do insolvency prediction and credit risk models still provide the accuracy that has been claimed in the past?; do they still work, and are they still relevant in today's financial climate?; and, can they be adequately relied upon today for decision making purposes? In order to answer this type of question it is fair to say that some amount of empirical research must be undertaken. This in turn opens up a new set of research questions which must also be answered in order to fulfil the objectives of the previous section.

The first, and possibly the most obvious question which arises is which technologies and models warrant empirical testing? I have previously said that the objective of the thesis is to evaluate models that are representative of the evolution of the literature. The choice of models is an extremely important one and therefore one which requires some careful consideration. There are potentially hundreds of individual models which have been published in the prior literature, some of which I have discussed in my literature review, but this is no means an exhaustive list[110]. I have, of course, by guiding the readers through the literature, focused on what I believe are the key papers within the field from my own quest and lengthy search through the archives. It is safe to say therefore that we might expect that the models which are selected for the final empirical analysis will come from a much larger population which has already been pruned in my own pre-conceived way, and by the evolution of the literature itself[111]. Section 8 of this thesis provides the final method that was employed in order to determine the models

[110] From a simple "Google Scholar" search it is notable that Beaver (1966) "Financial ratios as predictors of failure" and Altman (1968) "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy", have 1,904 and 4,511 citations respectively within the journal archive JSTOR alone. Withstanding other archives, working papers, dissertations and theses from around the globe, the potential population of research papers within this field is vast indeed.

[111] The organic growth and self-evolution of the literature will of course determine what are the" important" or key papers, with said papers inspiring a new generation of work. Those which are not so remarkable will inevitably pass us by without much after-thought or reflection.

which will appear in my empirical analysis, utilising both descriptive reasoning and statistical evidence in support.

The second question which arises instinctively follows the first. Assuming I have made an adequate selection of models for testing, what exactly is (are) the test(s) which should be carried out, and what is (are) the most appropriate or best method(s) for efficacy comparison. There are several alternatives which must be considered, each with their own assumptions and restrictions. Similarly to the model selection itself, it would be desirable if the choice of comparative technique would also, in some way, follow the evolution of the art, taking into account all recent developments and enhancements. It would also be unwise to assume at this point, that a single method of comparison is likely to suffice, due to the amount of information that I would wish to capture and portray. Section 8 also discusses the choice of test possibilities along with any pertinent advantages and disadvantages for each of the candidates.

Question three is a simple one. Are we significantly closer to a 'perfect' prediction model than was Beaver 45 years ago? One thing I hope that the reader has gathered from thesis thus far is that the wealth of research in this subject area has been vast. Given this voluminous research, the countless man-hours of reading, examining, number-crunching, postulating and concluding, the amount of money all this research would have cost the individuals, academia and financial institutions, we, or at least I, am interested to discover whether it has all been worth it? That is to say, using empirical evidence, in today's climate, how does a single ratio such as Beaver's cash-flow to total debt, compare to the latest developments in insolvency prediction modelling.

Finally I am interested to determine if any of the extant modelling technologies be enhanced or adapted to produce a better performing model and is there some logical progression towards the next generation of corporate insolvency prediction or credit risk model. I have already divulged my disinterest in simply bolting together a horde of variables in some ad-hoc fashion, so any progression towards a new form of model would have to be backed up by some theoretical underpinning.[112]

[112] Section 10.2 provides a deliberate example of an arbitrary ad-hoc model. Section 8.2 provides an example of a theoretically based model extension as a natural progression to the literature.

7 Data

7.1 Specification of data items required and data sets sought

It is paramount that the empirical undertakings of this thesis should be conducted upon data which is relevant to the nature of the research questions, consistent and reliable, and is up-to-date and recent enough in order to provide meaningful results for current applications beyond this thesis. To this end there are several considerations:

7.1.1 First consideration

The first consideration is to understand what timeframe (or window) should be adopted within which, the necessary data items ought to be collected. I have previously stated the necessity that data items and the study itself should be conducted upon the most up-to-date data currently available at a time where an empirical analysis can be satisfactorily conducted within the completion deadline of this thesis. This consideration is not problematic and will be discussed in due course. One question which does arise, however, is: how far back in time should the data be collected in order that it is still meaningful today? In order that a sound empirical analysis can be undertaken, it is desirable that the dataset is large in nature, containing as many firms (both failed and non-failed) and firm-year observations as possible. On the other hand, the age of the data which is collected, must also be taken into account. Economic conditions, accounting standards, practices and reporting requirements, and other company legislation will change and evolve over time, questioning the comparability of both data collected from company accounts and data gathered from market sources should the timeframe be too large. It is therefore anticipated (both through previous empirical exercises undertaken by myself, and from my understanding of the literature) that a sample period of just a few years is likely to yield a sample of too few data items (particularly that of the desired failed firm population) to justify a sound empirical analysis. The reverse is also true in that a sample period which is too wide, may provide an adequate population in respect to the failed firm population, but is unlikely to be consistence with respect to the individual data items over long periods of time. Therefore, a balance must be struck of an intermediate length in sample timeframe that best circumvents these possible problems.

7.1.2 Second consideration

The second consideration notes the types of company and macro population from which the sample should be drawn. Section 4 dealt with this consideration at some length, detailing the potential research avenues and limitations for individuals, and both private and publicly traded companies. The conclusions gained through this discussion were that (a) large datasets on individuals is largely (if not wholly) proprietary in nature and therefore an empirical analysis upon this population is somewhat problematic. (b) For privately owned companies, accounting data is available but is somewhat limited as to what can be readily accessed in a cost and time effective manner. Questions also pertain as to the reliability of such accounts given that they do not require 3rd part audit. (c) Publicly traded companies offer the most potential for research, and indeed make up the vast majority of prior studies. This is primarily driven of course by the stricter reporting regulations faced by these firms and the increased transparency required to attract potential investors. Many financial databases exist containing vast amounts of information which can be readily accessed by the researcher. (d) The economic significance of public company becoming insolvent is also far greater than that of any other type of firm.

It is therefore clear that an empirical study based upon publicly traded companies is preferred. It should also be evident from previous sections within this thesis that it is my wish to concentrate on matters within the UK and therefore selecting firms from outside the UK population would be nonsensical.

7.1.3 Third consideration

The third consideration relates to the overall size of the sample itself. More precisely the sample warrants two populations, a population of failed firms, and a population of non-failed firms; the size of each should be carefully considered.

Insolvency prediction literature has traditionally relied on matched samples of failed and non-failed firms, as in, for example, Beaver (1966), Altman (1968), Libby (1975), Taffler (1983), Keasey and McGuinness (1990), and Charitou et al. (2004). The paired sampling approach is most certainly a legacy of the discriminant analysis methodology frequently used in the 1960s and 1970s, where failed firms were matched on size and industry with a non-failed mate. Stein (2002) posits that the variable accuracy of any prediction model is *"primarily driven by the number of defaults, rather than the total number of observations"*. This is supported by Falkenstein et al. (2000), who suggest that the hazard model of Shumway (2001) outperformed others in their evaluation because of the large number of defaults (300) in the sample.

Current insolvency prediction and credit risk papers utilize the abundance of financial databases and related search tools to include a far greater number of non-failed firms in relation to the failed ones, more realistically representing the percentage of failures in the actual population. These technological advancements means that all listed firms have the potential to be analysed in greater detail. Moody's "RiskCalc for Public Companies", for example, uses a database of 1,406 failed and 13,041 non-failed firms for their proprietary version of the contingent claims model in the US. In the UK, the number of insolvencies is far less than that in the US, since the overall population of UK firms is much smaller. Agarwal and Taffler (2008) use data from 2,006 total firms and 103 failures over a fifteen year period in the UK, far less than Hillegeist (2004) for example, who studied 756 failures and 14,303 non-failed firms over an almost identical period. Table 7.1 compares sample sizes from a number of the studies cited in this paper.

Table 7.1: Summary of the sample sizes analysed in a selection of prior studies

Author(s) (year)	Model type	Country	Failed	Non-failed
Beaver (1966)	Single Ratio	US	79	79
Altman (1968)	MDA	US	33	33
Ohlson (1980)	Logit	US	105	2058
Taffler (1983)	MDA	UK	46	46
Sobehart and Stein (2000)[a]	Options	US	1,406	13,041
Charitou et al. (2004)	Neural network	UK	51	51
Bharath and Shumway (2004/8)	Options	US	1,449	*
Hillegeist et al. (2004)	Options	US	756	14,303
Reisz and Purlich (2007)	Options	US	799	4985
Agarwal and Taffler (2008)	MDA/Options	UK	103	2,006

* Not Disclosed.

[a] RiskCalc for public companies

The modern trend therefore seems to be geared towards using the maximum amount of data which is possibly available with recent studies employing sample proportions relatively similar to that of the real world (which is of course bound to be true if almost all firms are considered). It is my intention therefore to follow these recent studies and also use the maximum number of observations conforming to the sample criterion.

7.1.4 Fourth consideration

The fourth consideration continues the discussion upon population sizes, but more directly discusses any possible firm size and industry restrictions. It is desirable that this study is all-encompassing, that is to say, that firms should not be discarded or eliminated from the empirical study simply due to their size or specific industry. With regards to size, it is apparent from the prior literature that the size of a firm (whether measured in total assets, market value, number of employees, or other), could play an important part in the ultimate survival of a firm as a large equity reserves or asset bases act as buffer for debt servicing, should a company suddenly lose liquidity and/or profitability. It is therefore important that an arbitrary elimination of potential firms from the sample population, based upon some predetermined maximum and minimum values does not take place. With regards to industry however, many previous studies (e.g. Altman, 1968; Taffler, 1983) have concentrated solely upon the industrial/manufacturing sector, where a small paired sample approach is used. Recent studies (e.g. Agarwal and Taffler, 2007; Hillgeist et al, 2004; Reisz and Purlich, 2007) offer up far larger sample populations, unlimited by industry sector, with just one exception. Firms with financial industry codes are often excluded from sample populations because of their differing capital structure, earnings profile and reporting requirements. To this end, I intend also to eliminate financial firms from my empirical analysis.

7.1.5 Fifth consideration

The Fifth consideration regards the individual data items themselves. A large selection of models are to be considered for the empirical section of this thesis, with each of the models requiring a different (but not necessarily exclusive) set of variables to be sought. Therefore the individual data items are model specific and should be collected accordingly. Considerations also need to be made as to the actual collection of these variables, as many financial databases offer similar financial information without necessarily the same reliability. The most important factor is therefore the selection of a reputable source of information and to ensure that all data items are collected in a consistent manner from firm to firm.

7.1.6 Sixth consideration

Finally, consideration six relates to the possible division of the total sample into two smaller samples for the purposes of training and validation. It is evident from the majority of prior literature that in order to test/validate a new model proposal that the model must first be estimated upon one sample and then tested for its accuracy upon another. This procedure is commonly referred to as holdout sample validation. This holdout procedure, however, is only a desired necessity for models which need to be “trained” in some way in order to estimate the model’s parameters, e.g. MDA, logit, NN analyses or indeed every insolvency prediction model which is based upon a statistical or artificially intelligent system. Only models which are based on pure theory such as the option pricing formulae can be excluded from this testing and validation procedure as no parameters or coefficients require estimation.

The purpose of holdout sample testing is to provide a true ex-ante assessment of a model’s predictive accuracy, although this has not always been applied in an appropriate manner, as I have previously discussed in section 5. Many of the recent studies in the field of insolvency prediction either concentrate on theoretically based models or the assessment of popular extant models derived in an earlier period in time, therefore the use of the holdout sample in recent years has been somewhat limited. It is only when a particular model requires updating (in the coefficients) or re-trained as per a neural network for example, that the training and validation procedure warrants employment.

So the main question here is, do I need a holdout sample for my study? I am to test models representing the key developments within the literature, and as I have previously discussed it makes no sense to assume that the coefficients of earlier models should still be relevant/applicable today given the changes to the reporting environment over the previous half century. It is therefore practical to assume that with respects to any statistically driven models that a training and holdout sample setup will be desirable. Any theoretically driven models will not require training. Therefore, we must be in mind that in order to aid comparability of the models, each must be validated upon the same sample of observations regardless of whether training is required or not.

Various alternatives to model validation and to the holdout sample approach are evident in the literature. Early papers (e.g. Taffler, 1983) employ the Lachenbruch (1973) Jack-knife technique, usually where sample sizes are small and inefficient to divide. The Jack-knife technique reports to offer an almost unbiased measure of a models true ex-ante predictive accuracy and as previously discussed in section 5. In essence the Jack-knife technique is similar to that of "bootstrapping" where one or more observations are left out of the sample for validation based upon the result gained through training on the remaining observations. The process is repeated systematically until all observations have been validated from which estimates of the bias and variance of the models accuracy can be computed.

A similar technique is that of cross-validation, where subsets of the data sample are held out with the remainder used for training. The size of the subset and the number of repetitions is an arbitrary choice by the researcher. The results for each repetition are then averaged to find the models accuracy. More precisely K-fold cross validation splits the sample into K subsets where each is held out in turn for validation.

The problem with these alternative forms of validation however, is that they do not provide a true ex-ante determination of the models predictive ability. Section 8.3 discusses these problems in more detail and provides descriptions of the validation tests selected for use within this thesis.

7.2 Possible alternatives for the collection of data and associated problems

A number of financial databases are provided by the University of Exeter for general use by its students and researchers. The list of possible databases available at this time was as follows: Datastream, Wharton Research Data Service (WRDS), Thomson Reuters One Banker, Compustat, Bloomberg, Reuters 3000xtra, Morningstar, Hemscott Company Guru, and the London Share Price Database (LSPD).

The selection of database(s) from which the data for this study was to be collected, took some careful consideration. Some previous experience of data collection obtained whilst studying for my master's degree indicated that different databases offer a wide variety of data items, information, ease of use, and intuitiveness. The main requirements for the database utilised in this study were therefore not only to be able to necessitate the specifications and considerations as described in the previous section, but also to satisfy personal preferences when it come to the overall practicality, design, layout and functionality of the system. It was clear from the inception of this thesis that the data items required would include both historic accounting data and historic market data for (if possible) all UK public companies listed on the London Stock Exchange (LSE) for at least the previous decade.

Several of the databases were eliminated from consideration from the very beginning due to prior experience of the systems via both "hands-on" workshops, and through my own smaller prior studies. Hemscott Company Guru was eliminated due to a lack of company inclusions, and where data for companies did exist; the data items were often incomplete, missing or rounded to an inappropriate level. WRDS was eliminated due to the lack of data and information upon UK companies, again based on experience of the system some 12-18 months previous. This was considered somewhat unfortunate as the WRDS system itself offers a very usable interface with connectivity through its website, hence accessibility through any PC whether at the university or at home, something particularly desirable to myself. Compustat access is provided to the university via WRDS and therefore suffers the same inherent problems. Morningstar was a relatively new database to the university and was subject to a six month trial. The uncertainty as to the data collection timeframe, and the necessity to have a "re-collection of data" option at some point in the future (should additional companies or data items be unpredictably or suddenly required), meant that Morningstar could not be relied upon and therefore could not be considered. The Bloomberg database was also dismissed and was done so because of two factors. Firstly, the Bloomberg terminal interface is not very intuitive (at least not to myself) and involves the added complexity of a different keyboard layout and instructional language. Although these difficulties are not insurmountable, when coupled with the second factor, in that there is only one such Bloomberg terminal in the university, (which may be booked in one hour slots and has proven to be quite popular), guaranteeing usage of the system prior to booking several weeks in advance was impractical and it was felt that my time would be better applied elsewhere. Reuters

3000xtra was also eliminated, this time due to its somewhat awkward interface. I have previous completed and passed the Thomson Reuters Markets Academy course on the usage of this database but regardless, I did not find the interface particularly appealing or suitable for the needs of this study. The system has a lot of good points but several attempts to successfully locate the relevant data items pertinent to this study were not particularly fruitful.

The choice of database and the collection of data items was now reduced to just three options; Thomson One Banker, Datastream, and LSPD. The LSPD only holds market data and therefore accounting number-based items are not available on this database and was therefore not considered for this purpose. LSPD does, however, have some very useful key information pertaining to the identification of failed companies; something from previous experience has been particularly difficult to obtain for UK listed companies. LSPD provides a comprehensive list of every company that that has been listed on the LSE since January 1955 with data collected from a number of recognised sources including the Stock Exchange Daily Official List, the Financial Times, and Extel's EXSHARE service. The database is maintained and supported by the Institute of Finance and Accounting at London Business School. The list of companies (denoted by names and SEDOL numbers, is accompanied by various data items such as General Descriptive records (including deaths, de-listings, and types of death), Capital Changes, Dividends, Units, Prices, Share Capital. Information within the General Descriptive records has proven to be invaluable, and is something that I have found lacking from other databases in such a readily identifiable fashion. Previous smaller studies undertaken by myself have relied on company failure (or death) information gathered from the London Gazette, the "Official Newspaper of Record for the UK" as it is legal requirement by all insolvency processes to advertise here as part of their procedure (see Appendix B). The London Gazette is a publication which specialises in recording and disseminating official, regulatory and legal information and has an historical archive of over 350 years[113]. Using this resource it is possible to search the archive for notices on companies which match any of the insolvency criteria as discussed in some depth in section 4.3. However, it was clear that an obligatory manual search through the website archives was a particularly laborious process, even for my smaller pilot studies which

[113] http://www.london-gazette.co.uk/.

only incorporated one or two years of company data. Hence, to empirically evaluate a wider selection of companies over a longer timeframe, an alternative record of company failures and insolvencies had to be sought. Fortunately I was able to source a copy of the LSPD which not only provided all the information in one easily searchable database, but was also easily exportable into Microsoft Excel format to be manipulated as desired.

Finally I must discuss the data items themselves and both Datastream and Thomson One Banker remained for consideration, and if am to be honest, this was a fairly easy decision to make. It was evident from the outset that Datastream was only available on two computers within the university, and like the Bloomberg terminal suffered from a very busy booking policy. Although this is something that can be worked around, it is not ideal as I like to be able to access information as and when I desire, dependant on workload, family commitments, etc. More importantly I find my work habits, idea generation, and dare I say enthusiasm, tend to be quite irregular and often quite spontaneous. Therefore if I suddenly find myself in a position where I have a great idea, and have a sudden urge to download and play with some data, I generally like to involve myself with it almost immediately, and therefore waiting to book a computer at the university in two weeks' time, is something I would tend to avoid, in favour of a system that I can readily access through any computer at any time of the day.

Thomson One Banker (Thomson Financial) satisfies nearly all of my requirements from a database and is adequately suited to my needs and style of working. The database itself allows access to quotes, earnings estimates, financial fundamentals, transaction data, corporate filings and many other data items for over 60,000 global companies. More importantly for this study, there is an abundance of UK listed firms contained in the database with access to both accounting and market-based data with large historical archives. The data itself is sourced from many industry leading sources such as Worldscope, Reuters, Extel, Datastream, and Investext. A study by Lara et al (2006) examines the effects of database choice on accounting research and provides evidence that Thomson Financial delivers more complete data items than a selection of six alternate sources with regards to UK companies. These results empirically justify my choice of database and the collection of data.

The database can be accessed via two routes. Firstly access can be gained through any web browser, although this proved to be inefficient when trying to gather multiple company records and seems to be focused upon data at individual or firm specific level. The most attractive feature of this database however, is the second method of access, which comes in the form of a Microsoft Excel plugin.

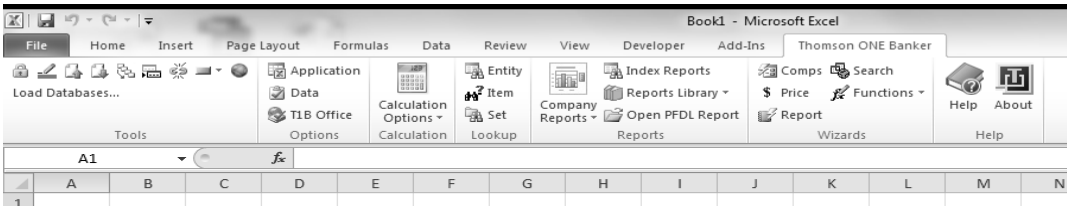


Figure 7.1: Thomson One Banker command ribbon.

The plugin integrates itself fairly seamlessly into an existing Excel package, adding a comprehensive set of functions to a new tab within the command ribbon as per Figure 7.1. What follows therefore is a brief discussion of the methodology employed in gathering the required information as well as a description of the final data set employed in the empirical section of this thesis.

7.3 Methodology and composition of the final dataset

The first task was to realise the timeframe (or window) in which this study was to be based upon. From my research objectives it was evident that I needed to use the most up-to-date data which was available, but the question pertained as to how far back in time data should be collected in order that it should still be meaningful and relevant. The timeframe required is largely dependent upon the number failed firms to be used within the empirical analysis, as past studies have shown (e.g. Stein, 2002) that the accuracy of any insolvency prediction model is primarily driven by the number of firms within the failed/insolvent group. In order to collect a sufficient number of failed firms for a sound empirical analysis, an average failure rate of around fifteen public companies per year was identified and therefore (somewhat arbitrarily) a decade of firm failures with an expected population of 150 was deemed to be sufficient.

The initial goal therefore was that the study utilised accounting and market data upon a sample of LSE-listed non-financial failed and non-failed firms, the sample period for such data being the decade 1st January 2000 to the 31th December 2009. All firm specific information being obtained through LSPD and Thomson One Banker.[114]

The next step was to create a list of companies that were both non-failed (alive) during this period and separate these from those which had failed (died). LSPD provides a General Descriptive record for every firm listed on the LSE since 1955. One of the fields in the General Descriptive record is G10 (Type of Death), which provides an indication of the reason why the company ceased to be quoted in the SEDOL. Several codes within this field are relevant to this study with the death codes matching my description of forms of failure and insolvency as per section 4; 7, the company is valueless after a creditor liquidation; 16, the company is valueless after the appointment of an administrator; 20, the company is in administration / administrative receivership; and 0, indicates that the firm is still alive and trading on the LSE.

As regards failed firms, this study takes the population of all London Stock Exchange (LSE) listed non-financial firms recorded on the London Share Price Database as failing between 30th September 2000 to 30th September 2009, reduced only by the exclusion of firms for which insufficient data is available to allow application of the insolvency prediction models under consideration. Failed firms were identified from the LSPD general descriptive record (G10) by codes 7, 16 and 20.[115] Firms with financial industry codes[116] are excluded from the population because of their differing capital structure, earnings profile and reporting. The final number of failed firms considered in the study is 178.

The sample of non-failed firms was selected from a non-financially coded population which was deemed to be "alive" at 30th September 2009 with LSPD code G10 "Type of Death" equal to zero. For this study a non-failed sample containing 2244 eligible firms was identified for potential analysis.

[114] One possible problem associated with this specific timeframe is the adoption by the UK of IFRS reporting standards in 2005. A change in reporting such as this may have a dramatic effect on empirical results if not dealt with in a correct manor. The treatment of this accounting regulation change is dealt with later in this section. Whereas this is not ideal, to gather an adequate number of failed firms, this was unfortunately a necessity.

[115] Representing, respectively: *creditor liquidation*, *receiver appointed*, and *in administration*.

[116] LSPD (G17) codes 8X, 8XX and 8XXX representing firms from the financial sector.

Once the name and SEDOL for each firm within the failed and non-failed groups had been identified through LSPD, the next step was to begin the construction of a master spreadsheet (or database). Careful consideration was required as to the layout, format, and usability and to the requirement that unforeseen/additional data items may be added with relative ease if desired at some point post collection. It was clear from the outset that none of the database report templates provided by Thomson would be adequate for my requirements and therefore I needed to design my own.

Within the chosen sample period it was desired that for each company, a number of firm-year observations would be required following recent prior literature. Therefore a maximum of ten years' worth of data was potentially available for each firm identified through LSPD. The inclusion of multiple firm years of observations has several impactors. Firstly it enables the observation of trends in ratios/model scores over time, particularly in the periods preceding the onset of insolvency. Secondly it enables a direct comparison of firms which fail within a certain year with accounting numbers and financials from the non-failed population within that same year. If only one year of observations (i.e. from the last set of accounts and likely to be circa 2009) was collected from the non-failed population, then these cannot be directly compared to a failed company's data circa 2000 due to changes in economic and reporting conditions. By collecting multiple years of data, the total sample can be split and contrasted over each year of the decade, a further and more detailed discussion of this can be found in the forthcoming section entitled the "structure of annual samples".

It was possible to import the list of firms (or entities) identified through LSPD into Thomson One Banker, although this proved to be somewhat monotonous. Although LSPD provides company name and SEDOL reference, Thomson likes to use its own reference key for populating its datasheets[117]. Once the Thomson key had been established for each firm, using a purposely written excel macro, it was possible to manipulate the layout of the datasheet so that each firm had ten identical rows (one for each year) each containing the company's descriptive information from the LSPD, along with the Thomson key identifier.

[117] It would have been simpler (and saved a lot of time) to match the SEDOL numbers provided by LSPD to those of Thomson. A number of attempts (for unknown reasons) did not yield satisfactory results and I was forced to match each company individually.

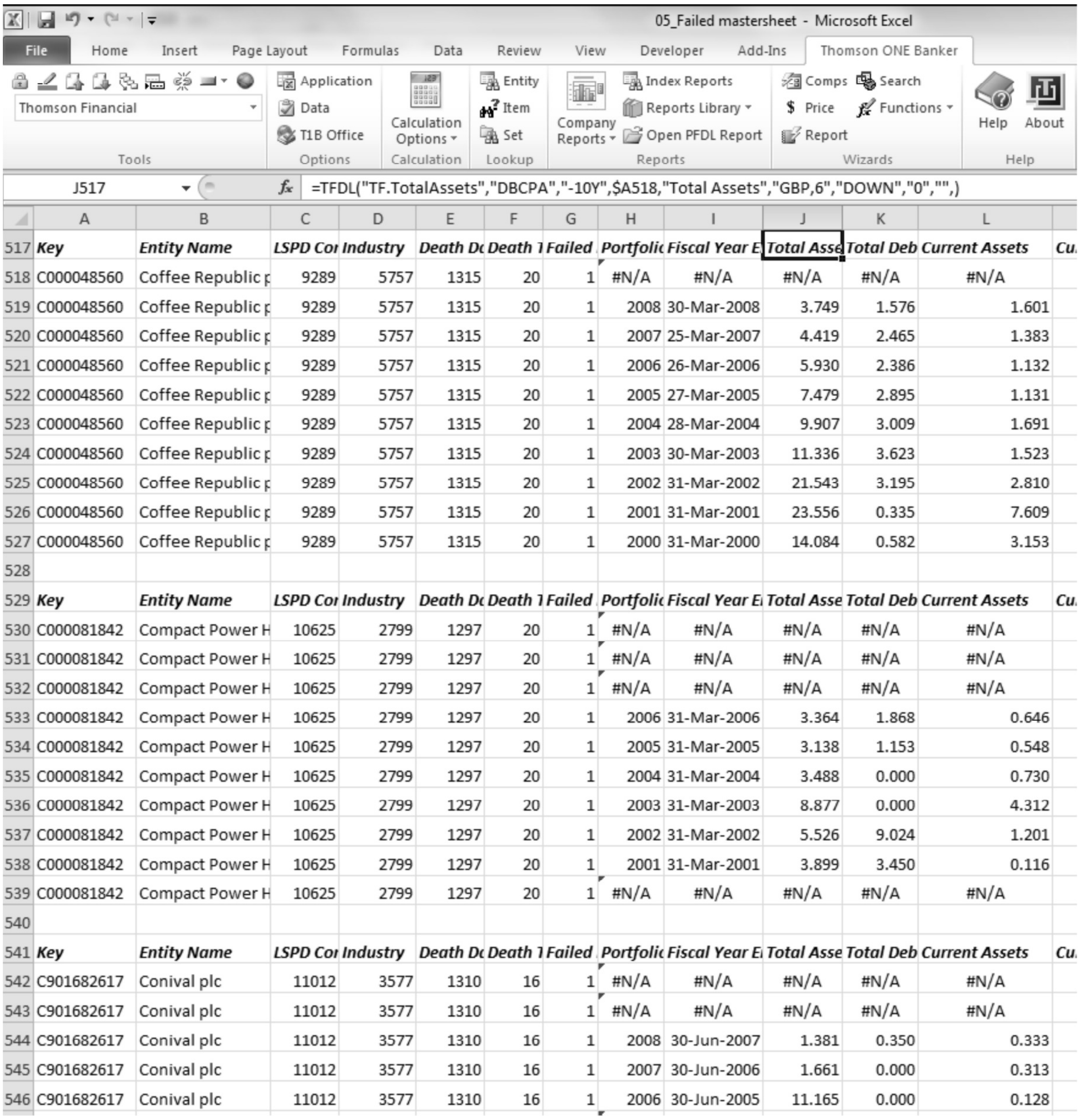
The next stage was to populate the datasheet with the data items themselves, but this led to the inevitable question of which items should be included. From the prior literature it is evident that ratios representing liquidity, profitability and size are commonplace and that variations of these ratios are often composed of a number of key variables. Also taken into consideration were possible prediction model candidates for inclusion in the empirical analysis although at this point the choice of models was not conclusive. The initial number data items required, not including firm specific descriptors, was just seventeen[118] and were as follows: total assets, total debt, current assets, current liabilities, cash, retained earnings, EBIT, sales, inventories, operating expenses, depreciation, net income, funds from operations, ROE, dividends, market capital and price volatility.

To populate the data items, it was necessary to enter the functions specific to Thomson One Banker into the Excel spreadsheet. For example, to populate the sheet with the previous ten years of total assets in descending rows, for the entity in cell A2 I used the function:

```
=TFDL("TF.TotalAssets","DBCPA","-10Y",$A2,"Total Assets","GBP,6","DOWN","0","",)
```

As by convention, each column of the datasheet represents a different data item, and therefore the function (which acts as a header for each firm set) is repeated across all columns with only the data item reference (for the case of total assets this is “TF.TotalAssets”) and the column header (Total Assets) requiring amendment according to the desired item. Figure 7.2 shows a section of the resulting datasheet, noting that if no data items are available for a particular year then the function returns the value #N/A.

[118] The number of possible two-variable ratio combinations in this instance is 289 (17^2). There are of course many more possibilities if more than two variables are combined.



	A	B	C	D	E	F	G	H	I	J	K	L	
517	Key	Entity Name	LSPD Coı	Industry	Death Dc	Death 1	Failed	Portfolic	Fiscal Year E	Total Asse	Total Deb	Current Assets	Cu.
518	C000048560	Coffee Republic p	9289	5757	1315	20	1	#N/A	#N/A	#N/A	#N/A	#N/A	
519	C000048560	Coffee Republic p	9289	5757	1315	20	1	2008	30-Mar-2008	3.749	1.576	1.601	
520	C000048560	Coffee Republic p	9289	5757	1315	20	1	2007	25-Mar-2007	4.419	2.465	1.383	
521	C000048560	Coffee Republic p	9289	5757	1315	20	1	2006	26-Mar-2006	5.930	2.386	1.132	
522	C000048560	Coffee Republic p	9289	5757	1315	20	1	2005	27-Mar-2005	7.479	2.895	1.131	
523	C000048560	Coffee Republic p	9289	5757	1315	20	1	2004	28-Mar-2004	9.907	3.009	1.691	
524	C000048560	Coffee Republic p	9289	5757	1315	20	1	2003	30-Mar-2003	11.336	3.623	1.523	
525	C000048560	Coffee Republic p	9289	5757	1315	20	1	2002	31-Mar-2002	21.543	3.195	2.810	
526	C000048560	Coffee Republic p	9289	5757	1315	20	1	2001	31-Mar-2001	23.556	0.335	7.609	
527	C000048560	Coffee Republic p	9289	5757	1315	20	1	2000	31-Mar-2000	14.084	0.582	3.153	
528													
529	Key	Entity Name	LSPD Coı	Industry	Death Dc	Death 1	Failed	Portfolic	Fiscal Year Ei	Total Asse	Total Deb	Current Assets	Cu.
530	C000081842	Compact Power H	10625	2799	1297	20	1	#N/A	#N/A	#N/A	#N/A	#N/A	
531	C000081842	Compact Power H	10625	2799	1297	20	1	#N/A	#N/A	#N/A	#N/A	#N/A	
532	C000081842	Compact Power H	10625	2799	1297	20	1	#N/A	#N/A	#N/A	#N/A	#N/A	
533	C000081842	Compact Power H	10625	2799	1297	20	1	2006	31-Mar-2006	3.364	1.868	0.646	
534	C000081842	Compact Power H	10625	2799	1297	20	1	2005	31-Mar-2005	3.138	1.153	0.548	
535	C000081842	Compact Power H	10625	2799	1297	20	1	2004	31-Mar-2004	3.488	0.000	0.730	
536	C000081842	Compact Power H	10625	2799	1297	20	1	2003	31-Mar-2003	8.877	0.000	4.312	
537	C000081842	Compact Power H	10625	2799	1297	20	1	2002	31-Mar-2002	5.526	9.024	1.201	
538	C000081842	Compact Power H	10625	2799	1297	20	1	2001	31-Mar-2001	3.899	3.450	0.116	
539	C000081842	Compact Power H	10625	2799	1297	20	1	#N/A	#N/A	#N/A	#N/A	#N/A	
540													
541	Key	Entity Name	LSPD Coı	Industry	Death Dc	Death 1	Failed	Portfolic	Fiscal Year Ei	Total Asse	Total Deb	Current Assets	Cu.
542	C901682617	Conival plc	11012	3577	1310	16	1	#N/A	#N/A	#N/A	#N/A	#N/A	
543	C901682617	Conival plc	11012	3577	1310	16	1	#N/A	#N/A	#N/A	#N/A	#N/A	
544	C901682617	Conival plc	11012	3577	1310	16	1	2008	30-Jun-2007	1.381	0.350	0.333	
545	C901682617	Conival plc	11012	3577	1310	16	1	2007	30-Jun-2006	1.661	0.000	0.313	
546	C901682617	Conival plc	11012	3577	1310	16	1	2006	30-Jun-2005	11.165	0.000	0.128	

Figure 7.2: Sectional extract of the resulting initial datasheet.

The final step of the collection procedure was to eliminate the rows which contain the header information for each firm along with the blank row separator and those rows for years with no data. This was simply achieved by running another custom macro after copying the values of the “live” spreadsheet into a new blank sheet. The resulting sheet contained all firms (failed and non-failed) with all anticipated data items, for all years of available data between 2000 and 2010, as per Figure 7.3. This resulting dataset contained 820 firm year observations for the 178 failed firms, whereas the non-failed group provides 12,855 firm-year observations from the 2244 available firms.

	A	B	C	D	E	F	G	J	K	L	M	N	O
1	*Key*	*Entity Nar*	*LSPD Com*	*Industry*	*Death Dat*	*Death Typ*	*Failec*	*Portfolio*	*Portfolio da*	*Fiscal Year*	*Total Asse*	*Total Deb*	*Current A*
27	C00003241	Planestati	5625	2777	1267	20	1	2000	29/09/2000	31/03/2000	87.604	22.127	#N/A
28	C0000878	Plasmon p	9498	9572	1306	20	1	2008	30/09/2008	31/03/2008	46.716	11.108	17.33
29	C0000878	Plasmon p	9498	9572	1306	20	1	2007	28/09/2007	31/03/2007	51.278	11.381	19.19
30	C0000878	Plasmon p	9498	9572	1306	20	1	2006	29/09/2006	31/03/2006	60.094	9.052	24.23
31	C0000878	Plasmon p	9498	9572	1306	20	1	2005	30/09/2005	31/03/2005	61.265	12.707	26.80
32	C0000878	Plasmon p	9498	9572	1306	20	1	2004	30/09/2004	31/03/2004	58.346	13.823	25.49
33	C0000878	Plasmon p	9498	9572	1306	20	1	2003	30/09/2003	31/03/2003	65.902	14.132	32.58
34	C0000878	Plasmon p	9498	9572	1306	20	1	2002	30/09/2002	31/03/2002	70.013	11.565	36.38
35	C0000878	Plasmon p	9498	9572	1306	20	1	2001	28/09/2001	31/03/2001	59.521	10.826	29.05
36	C0000878	Plasmon p	9498	9572	1306	20	1	2000	29/09/2000	31/03/2000	41.13	2.173	20.8
37	C9016784	Playgolf (I	10975	5755	1312	20	1	2008	30/09/2008	31/12/2007	17.96	10.573	1.64
38	C9016784	Playgolf (I	10975	5755	1312	20	1	2007	28/09/2007	31/12/2006	22.335	10.993	0.44
39	C9016784	Playgolf (I	10975	5755	1312	20	1	2006	29/09/2006	31/12/2005	22.817	10.073	1.10
40	C9016784	Playgolf (I	10975	5755	1312	20	1	2005	30/09/2005	31/12/2004	11.994	4.712	1.50
41	C9016784	Playgolf (I	10975	5755	1312	20	1	2004	30/09/2004	31/12/2003	7.763	2.047	0.36
42	C0000730	Po Na Na	10264	539	1240	20	1	2002	30/09/2002	31/03/2002	31.64	8.038	5.86
43	C0000730	Po Na Na	10264	539	1240	20	1	2001	28/09/2001	25/03/2001	38.457	9.469	4.00
44	C0000730	Po Na Na	10264	539	1240	20	1	2000	29/09/2000	26/03/2000	27.824	7.681	2.38
45	C0000796	Public Net	10544	974	1243	16	1	2002	30/09/2002	31/03/2002	1.262022	0.427548	0.19195
46	C0000796	Public Net	10544	974	1243	16	1	2001	28/09/2001	31/03/2001	1.130149	0.256772	0.10848
47	C0000527	Pubs 'N' B	9243	5757	1320	20	1	2009	30/09/2009	31/12/2008	48.88184	35.55356	3.04093
48	C0000527	Pubs 'N' B	9243	5757	1320	20	1	2008	30/09/2008	31/12/2007	60.5519	34.62485	3.04608
49	C0000527	Pubs 'N' B	9243	5757	1320	20	1	2007	28/09/2007	31/12/2006	39.59532	20.29364	2.6709
50	C0000527	Pubs 'N' B	9243	5757	1320	20	1	2006	29/09/2006	31/12/2005	39.08291	20.61188	3.20025
51	C0000527	Pubs 'N' B	9243	5757	1320	20	1	2005	30/09/2005	31/12/2004	28.0156	11.20756	2.23081
52	C0000527	Pubs 'N' B	9243	5757	1320	20	1	2004	30/09/2004	31/12/2003	30.14923	13.22012	3.59838
53	C0000527	Pubs 'N' B	9243	5757	1320	20	1	2003	30/09/2003	31/12/2002	28.38372	12.549	3.20152
54	C0000527	Pubs 'N' B	9243	5757	1320	20	1	2002	30/09/2002	31/12/2001	25.27642	12.838	2.13177
55	C0000527	Pubs 'N' B	9243	5757	1320	20	1	2001	28/09/2001	31/12/2000	24.50349	12.257	3.37157
56	C0000488	Qualceran	9649	3726	1311	16	1	2008	30/09/2008	31/12/2007	54.39704	12.68395	44.3218
57	C0000488	Qualceran	9649	3726	1311	16	1	2007	28/09/2007	31/12/2006	46.21521	13.39617	28.5191
58	C0000488	Qualceran	9649	3726	1311	16	1	2006	29/09/2006	31/12/2005	45.65436	14.19892	27.2325
59	C0000488	Qualceran	9649	3726	1311	16	1	2005	30/09/2005	31/12/2004	48.52148	10.09183	28.5990
60	C0000488	Qualceran	9649	3726	1311	16	1	2004	30/09/2004	31/12/2003	50.777	13.70588	30.0962
61	C0000488	Qualceran	9649	3726	1311	16	1	2003	30/09/2003	31/12/2002	66.10165	33.45233	27.9726
62	C0000488	Qualceran	9649	3726	1311	16	1	2002	30/09/2002	31/12/2001	72.96021	35.91144	30.3379
63	C0000488	Qualceran	9649	3726	1311	16	1	2001	28/09/2001	31/12/2000	76.93531	36.62144	33.0289

Figure 7.3: Sectional extract of the final dataset.

An additional data item, the risk-free rate of return was also anticipated as being required for some models, particularly the evaluation of the latest form of credit risk assessment models based on option pricing formulae. The risk-free rate of return is calculated annually using the nominal yield on a 12-month UK treasury bill as at 30th September each year in order to satisfy the requirements of the portfolio structure detailed below.

7.4 Structure of annual samples

The empirical analysis involves the consideration on the same date each year of a set of all firms from my sample for which pertinent data is available in respect of that year, some of which fail within twelve months of the portfolio date (failed firms), and the remainder which do not (non-failed firms). In the prior literature, such sets have been referred to as annual ‘portfolios’.

The predictive power of any model depends upon when the information (financial report) is assumed to be available. Some previous studies have not been careful in this regard. (Ohlson, 1980).

The formation of annual portfolios is a necessity for my empirical analysis due to the inclusion of both market and accounting data. To aid a realistic and correct comparison between the models it is essential that the market data is taken from the same date on each year to withstand seasonal changes in economic and financial conditions, and that the accounting data is given a sufficient lead-time to allow for preparation and publication after the fiscal year-end.

A portfolio is created for each year from 2000 to 2009 on 30th September, following Agarwal and Taffler (2008) to aid both comparability with that study and between the different models which I test. Each portfolio is constructed on the 30th September using market data up to and including this date. Accounting data, however, is taken only from annual reports made up to 30th April of the same year or before, to duplicate the information conditions which would have been available to a real-world user of the models on 30th September of the year concerned.

The choice of the portfolio date is somewhat arbitrary. Hillegeist et al. (2004) use a four month lag of portfolio creation presumably due to the SEC requirement of report publishing within four months of the accounting year end. In the UK the statutory requirement is that public limited company reports must be filed and published within seven months of the accounting year end. Agarwal and Taffler (2008) use the 30th September of each year (at least five months from the fiscal year end) to construct their portfolios and I do the same. The resulting portfolios provide an average lead-time of 9.5 months quite a bit shorter than Olson’s (1980) lead-time of 13 months. Interestingly whereas Ohlson found that the lead-times were on average larger in length for failed companies in the year prior to insolvency, from data it is evident that no such anomaly exists, with failed and non-failed companies exhibiting remarkably similar average lead-times. Figure 7.4 shows the frequency of lead-times pertinent to this study along with the matching distribution of fiscal year-ends.

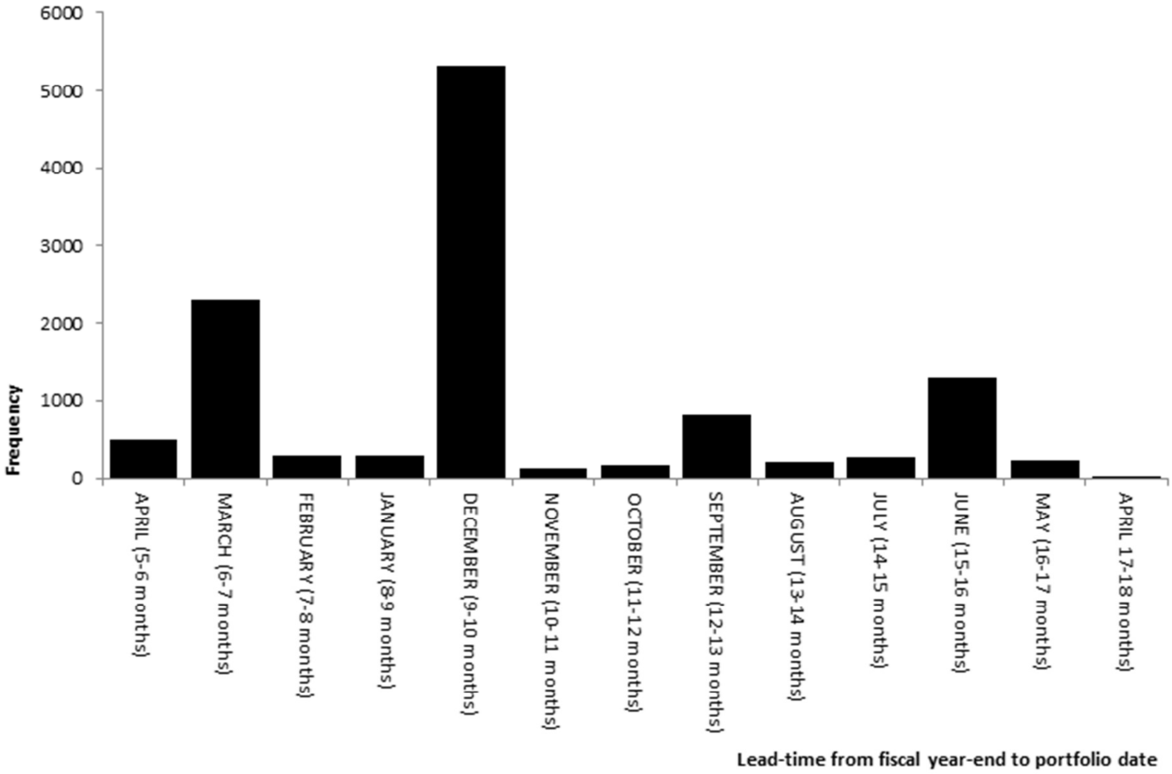


Figure 7.4: Lead-time between portfolio date and availability of accounting data

7.5 Reduction, training and validation

In order that I conduct an empirically sound study upon the collected data, two factors need to be considered; with the first these factors ultimately leading to a reduction of the potential 12,855 firm year observations as described previously to approximately half.

It is critical for an empirical comparison of any nature that the same sample (in this case the sample of firms) remains of the same with respect to both size, and of population, i.e. the exact same number of firms, as well as each individual firm, must be constant throughout all of the tests. Particular to this study, and true to many statistical tests, the number of “events” for which one is trying to classify, is critical to the results which are obtained. The events in this case are corporate insolvencies, and from previous studies mentioned in this thesis along with the potential data sample as described in the previous sub-sections, the reader will be aware that from a whole population, these events are somewhat rare. Stein (2002) demonstrates that the number of these rare events is critical in determining the classification accuracy of insolvency prediction models through the use of Receiver Operating Characteristic (ROC) curves (see section 8.3), and is therefore apparent that each model and each test should be conducted upon the exact same sample of firms[119].

With this data restraint it is therefore paramount that each firm-year observation will contain all of the variables and data items attributable to each of the different models which are sought to be tested and these models are described in detail subsequently within section 8. As much as this reduction is a necessity, it does potentially attract some bias into the test results. Zmijewski (1984) investigates the bias in predictive accuracy of conditional probability models when samples are selected (or hand-picked) to ensure data is complete, such as the intentions of this study. He notes that:

The sample selection bias is examined by comparing probit estimates of a financial distress model conditional on complete data to estimates from a bivariate probit assessment of the model which incorporates the probability of an observation having complete data into the estimation of the financial distress model parameters. The bivariate probit model consists of two probit equations, the financial distress equation and the complete data equation. The data constraint when using bivariate probit is that all observations must have complete data for the complete data equation. The results are qualitatively similar to the choice-based sample results in that a bias is clearly shown to exist, but, in general, it does not appear to affect the statistical inferences or overall classification rates.

For the confines of this study therefore, it would suggest that this bias should not influence the results, but of course, a great deal of information is inevitably lost through the omission and deletion of firms who do not provide 100% of the data items sought.

[119] King and Zeng (2001) suggest that these rare events can also dramatically underestimate probabilities within popular statistical methodologies such as logistic regression, and propose an adjustment methodology to counteract this bias.

Lastly I would like to discuss training and validation. For some of the prediction models of which I test (notably any statistical model which requires some sort of parameter or coefficient estimation), both training and validation samples are required in order to assess the model's predictive accuracy. There are various alternative forms of approach to validation (e.g., random sample cross-validation, *k*-fold cross validation, Lachenbruch jack-knife), each designed to portray a true ex-ante depiction of the models ability to distinguish firms which fail from those which do not. With regards to these alternative methodologies, it is thought that many are rather antiquated and have previously been adopted only where sample have been limited in size. A true measure of a models ex-ante predictive ability is based upon the holdout sample. Many studies previously described within this thesis have utilised holdout samples but are often selected from a population covering the same time period as that of the training sample and therefore disregard the issues pertaining to application focus which I discussed in section 5.5.

For this study therefore, in the first instance, and in order to reflect the practices of construction and use of the models by real-world users, I split the sample through time rather than by some random assignment. Training data comes from portfolios 2000-2005 inclusively, and the validation data (on which all models, irrespective of whether or not they require training, are tested) is derived from portfolio years 2006-2009.[120]

Table 7.2 presents the number of firm-year observations employed in the training and validation samples within this study, noting the reduction in total number of non-failed firm-years from a potential 12,855 (as described in section 7.3) to 6,595 in order to satisfy the condition of completeness.

.

[120] Accounting data required for validation/testing, therefore, is from the post-IFRS implementation era. It would have been more satisfactory to conduct the whole study using IFRS data. Obtaining an acceptable sample size (in respect of number of failed firms) to incorporate both training *and* validation samples post-IFRS is, however, not yet possible.

Table 7.2: The Number of firm-year observations employed in the study

	Training	Validation	
Period	2000-05	2006-09	
Failed	39	62	**101**
Non-Failed	2964	3530	**6494**
Total	**3003**	**3592**	**6595**

Additionally, I generate 10,000 random splits between training and validation samples. In each case the number of both failed firm-year observations and non-failed firm year observations in the training and validation samples are as in Table 7.2; but the allocations now being made on a random basis, rather than according to chronology. This provides me with the opportunity to investigate variation in efficacy of the prediction models resulting from variation in allocation of observations between the training and validation samples.

8 Method

8.1 Model selection

The academic literature contains a vast array of techniques which have been posited over the last five decades to 'predict' corporate insolvency. The selection choice of technologies for comparative study in this paper is based on three factors: (i) the frequency of study and reported efficacy in the prior academic literature; (ii) relevance to the UK; and (iii) the desire to incorporate and compare market-based contingent claims models and their traditional accounting number-based counterparts.

My review of the literature has identified 24 different methodologies,[121] which are shown as the abscissa labels in Figure 8.1. These range from the univariate and multivariate discriminant analysis of the 1960s and 1970s, through the logit and probits of the eighties and the artificial learning models of the 1990s, to the recent crop of contingent claims models in the 2000s.

A number of papers in the literature compare and contrast these various methodologies and techniques, e.g., Scott (1981), Zavgren (1983), Laitinen and Kankaanpaa (1999), and Balcaen and Ooghe (2006). A meta-analysis by Aziz and Dar (2006) considers a number of methods and techniques from a sample of 89 empirical studies (although there is little representation of market-based contingent claims type models). Their paper compares the frequency of usage (by academics) and efficacy of 16 different methodologies which they identified as being used to predict insolvency and assign each of them to one of three categories, 'statistical models'(represented by black bars in the following charts), 'artificially intelligent expert systems (AIES)' (represented by grey bars), and 'theoretical models' (represented by white bars). The Aziz and Dar results show MDA and logit to be the most popular for study amongst academics; and AIES, such as neural networks and genetic algorithms, to have, marginally, the best reported efficacy.

[121] Many more, if variations within method are counted.

Figure 8.1 shows the frequency of my academic investigation of each of the 24 corporate insolvency prediction technologies found in my literature review, grouped according to the Aziz and Dar (2006) categories. Having surveyed over 350 papers, a larger sample of studies than was considered by these authors, the relative frequency of academic investigation of techniques that I find is in accordance with their results.[122]

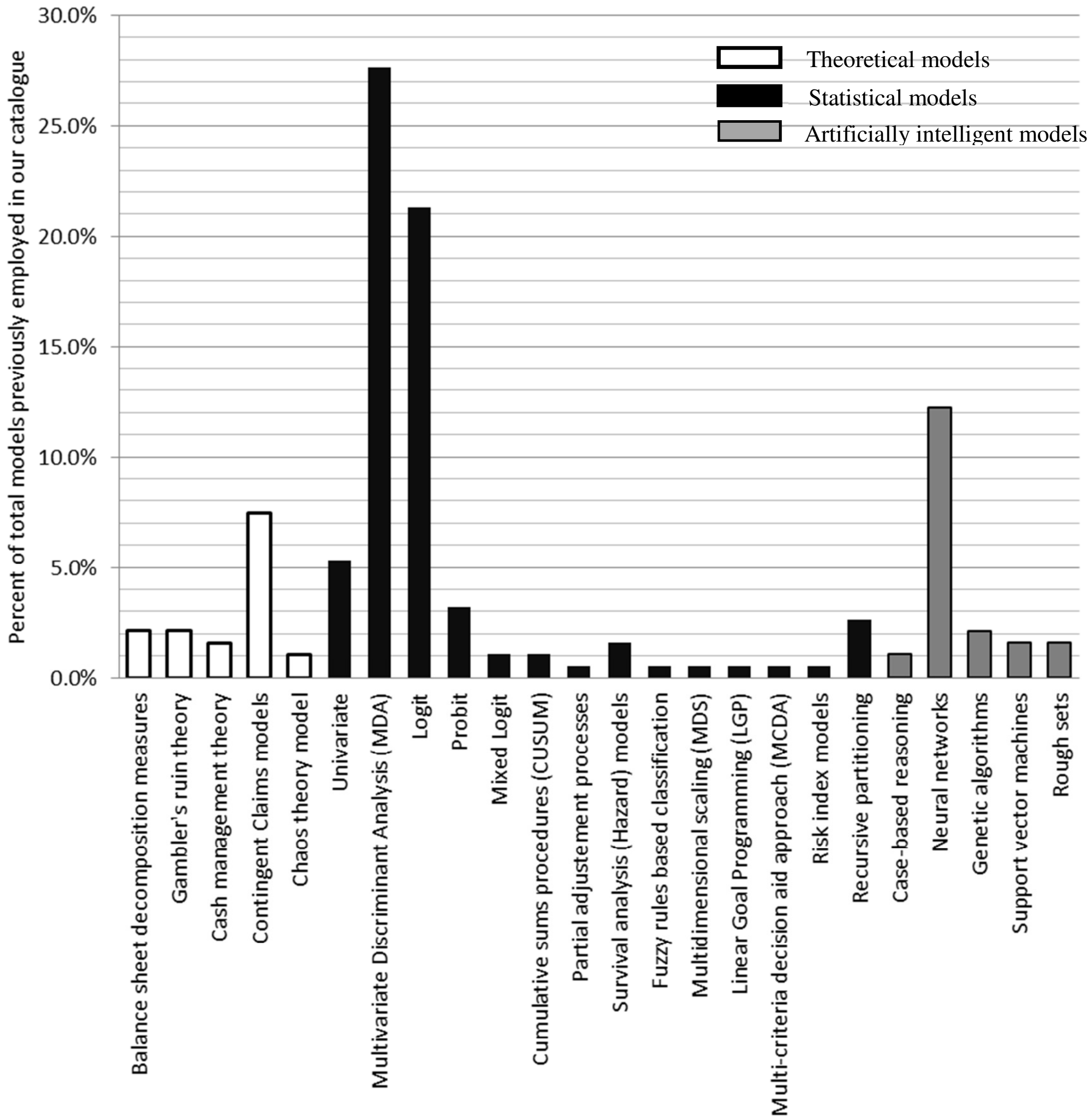


Figure 8.1: Frequency of models employed in previous studies

[122] Although approximately 350 papers have been surveyed, the differing nature of the literature means that there are not 350 different models. Some papers offer a new model (1 entry), some are purely discursive (0 entries), and some papers are comparative (a paper empirically evaluating 4 types of models would therefore have 4 entries).

Further to this frequency analysis, the next stage was to determine whether any of the prediction technologies demonstrate proportionally higher prediction accuracy over its counterparts. In order to determine the accuracies of each individual technology, reported accuracies of each model within the groupings needed to be collated from the individual papers within my literature catalogue.

Some difficulty arose when trying to determine what constitutes “accuracy” as the literature contains varied methods on how to report the efficacy of these types of models (differing sample types, holdout verses validation tests, one-to-five year ahead forecasts, confusion matrices, ROC curves etc). As there is no general acceptable rule for reporting accuracy, directly comparing empirical results from one paper to the next, making inferences within this regard is somewhat dangerous. What this investigation can do however is act as a meta-generalisation of the literature which may or may not provide an insight into classification accuracies across different modelling methodologies.

To ensure as much comparability as possible, the accuracies of the methodologies are, where possible, are taken from individual model tests one-year prior to failure. In addition, those accuracies reported are preferably from holdout sample tests, whether that be a cut-off and overall error approach, a Jack-knife validation approach, ROC curves or k-fold validation. This analysis resulted in 188 observations across the 24 different modelling technologies.

This process was quite lengthy and somewhat tedious, but was a necessary part of the review process. It was hoped that the results would demonstrate whether one particular prediction technology on average provided consistently higher accuracies and therefore warrant inclusion in the empirical work within this thesis. The results of this meta-analysis are shown in Figure 8.2 with summary statistics presented in Table 8.1.

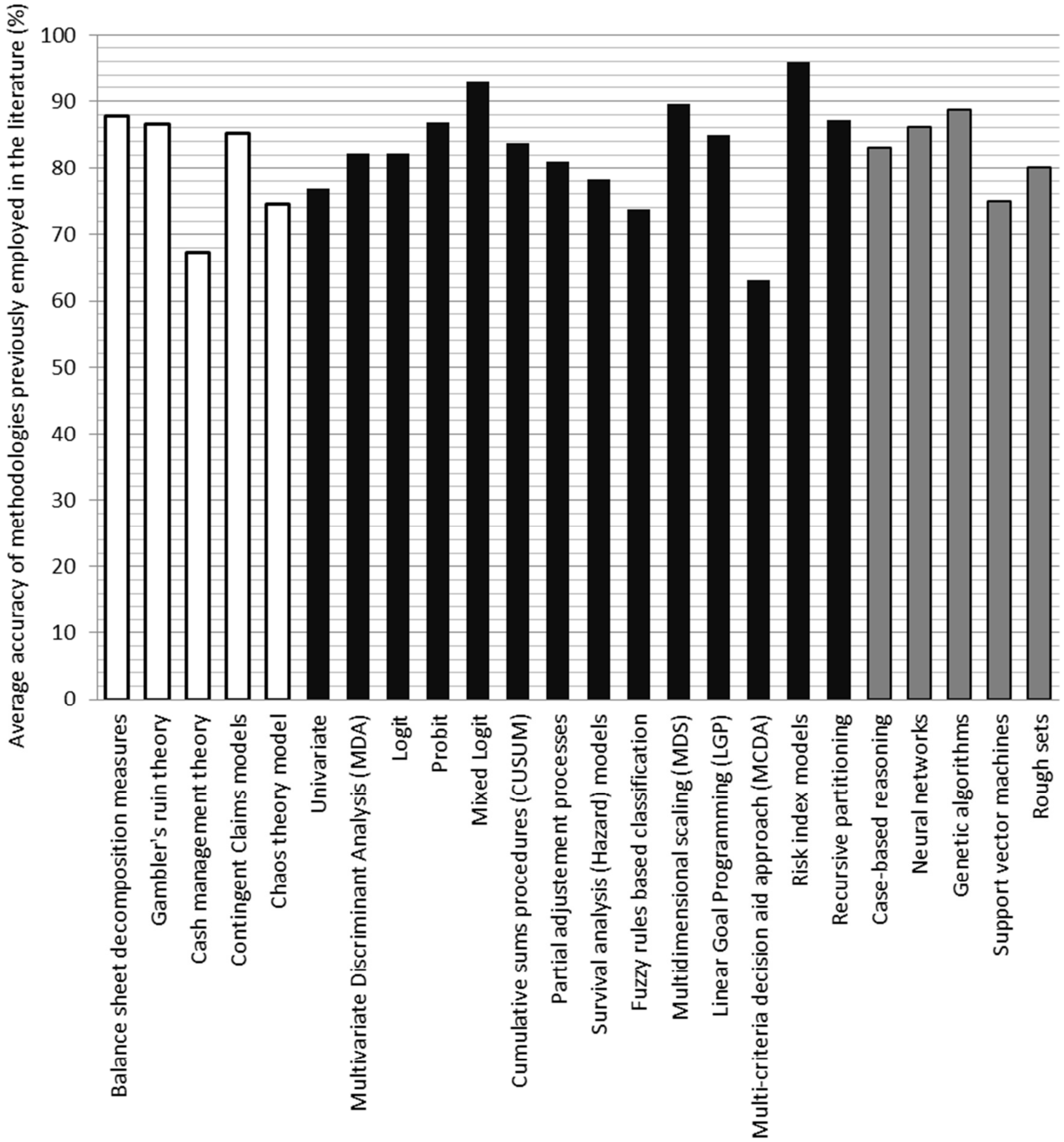


Figure 8.2: Accuracy of insolvency prediction models employed in previous studies

The technologies show accuracies of between 63% and 93% across the sample of literature, although several of the technologies do not offer a sufficient number of papers to provide empirical robustness. Only seven of the technologies provide five observations or more rendering the accuracies of the less-used technologies difficult to interpret. From these seven technologies, the best performer on average is the Probit model (86.8%) closely followed by neural networks (86.1%). The lowest ranked from these seven are the univariate studies (76.9%).

Table 8.1: Summary statistics for different model types and their accuracy

Model type	N	Mean accuracy (%)	SD
Balance sheet decomposition measures	4	87.7	3.3
Gambler's ruin theory	4	86.5	8.7
Cash management theory	3	67.3	8.3
Contingent Claims models	14	85.1	5.2
Chaos theory model	2	74.6	13.5
Univariate	10	76.9	8.0
Multivariate Discriminant Analysis (MDA)	52	82.2	11.7
Logit	40	82.2	16.0
Probit	6	86.8	10.4
Mixed Logit	2	93.0	2.8
Cumulative sums procedures (CUSUM)	2	83.7	1.7
Partial adjustment processes	1	81.0	n/a
Survival analysis (Hazard) models	3	78.3	15.7
Fuzzy rules based classification	1	73.8	n/a
Multidimensional scaling (MDS)	1	89.5	n/a
Linear Goal Programming (LGP)	1	85.0	n/a
Multi-criteria decision aid approach (MCDA)	1	63.2	n/a
Risk index models	1	96.0	n/a
Recursive partitioning	5	87.2	11.3
Case-based reasoning	2	83.0	2.1
Neural networks	23	86.1	8.4
Genetic algorithms	4	88.7	8.8
Support vector machines	3	75.0	8.5
Rough sets	3	80.0	13.1
TOTAL	188		

If I combine these scores and present them in their three main groups, theoretical, statistical, and artificially intelligent models, one can observe that the accuracies become less disperse. Statistical models are the weakest (82%), theoretical models are second (83%), and the best performing group are the artificially intelligent systems (85%). The overall average reported accuracy for all models representing all technologies and all studies throughout the evolution of the literature is approximately 83%.

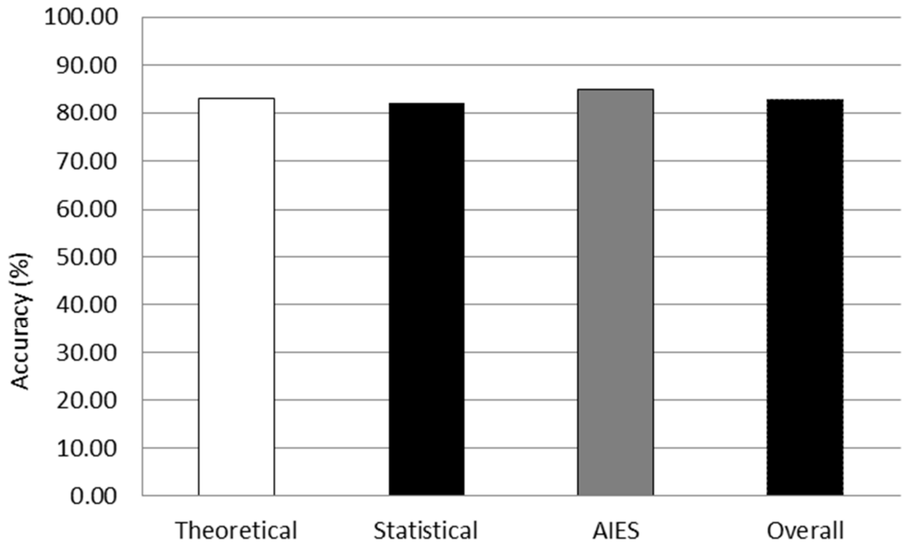


Figure 8.3: Aggregated predictive accuracies across the literature

The inclusion (exclusion) of models in (from) the empirical examination content of this thesis is therefore based upon the finding of the meta-analyses highlighted above. Univariate, MDA, logit, NN and Contingent Claims models warrant inclusion due to their popularity in the literature and representative predictive accuracies in comparison to the mean average. These five model types also allow for representation of the three broader groupings of model types – Theoretical, Statistical, and Artificially Intelligent. Technologies which provide small proportions of the catalogue, despite some instances of above average predictive accuracy, are reluctantly excluded from the study as they do not represent the literature as a whole. Individual model selection(s) representing the chosen technologies are described in the following section along with the reasons for inclusion.

8.2 Model specification

8.2.1 Single variable models

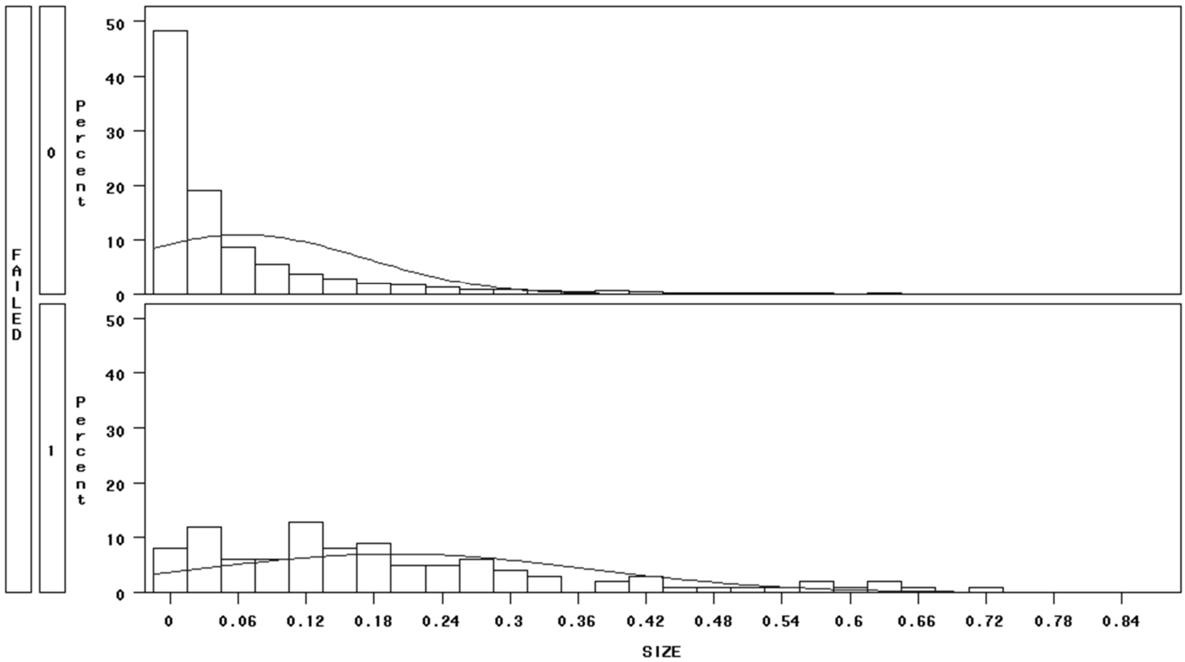
I include three single variable 'models' in the comparative analysis: cash-flow to total debt (which I designate model BV); firm size (model SIZE); and book-to-market value ratio (model B/M).

Vassalou and Xing (2004) find empirical evidence that 'both size and book-to-market exhibit a strong link with default risk'; and Agarwal and Taffler (2008) use these variables as benchmarks against which to compare multivariate models. Size is intuitively appealing as a factor in a firm's chances of surviving financial setbacks, since high equity reserves or asset bases act as buffer to enable continued debt servicing, should a company suddenly lose liquidity and/or profitability. This can be seen in the case of BP PLC. In 2010 BP, one of the largest companies listed on the FTSE, saw its share-price collapse as a result of both ecologic and political blunders surrounding the Deepwater Horizon disaster. The sheer size of BP has ultimately ensured its survival (despite the $20 billion spill response fund it was ordered to create by US President Obama. In contrast, one the largest companies to suffer insolvency proceedings in the last five years was Woolworths PLC, one of the UK's largest and established high street chains, although with an asset base of just 1% of BP's. Woolworths could not weather the onset of recession, and bad management and the failure to modernise are attributed to its demise. I use the natural logarithm of GDP deflated Market value of equity as a proxy (SIZE) to determine its influence on failure probability. The histogram for this, and proceeding models, represents the ratio/score when converted into a probability using the formula $p = e^{score}/(1 + e^{score})$. The requirement for conversion is dealt with in section 9.3 but in essence it acts as a control in order that each of the models which I test are comparable. The plots are broken down into failed and non-failed firms, normalised and over-laid with a normal curve approximating the distribution using the SAS procedure UNIVARIATE with the option histogram / normal. This provides a consistent view of the data within each of the models which I test to further aid comparability in a visual sense.

Table 8.2: Summary statistics for SIZE variable

Model		Mean	Median	SD	Var	Kurt.	Skew	Min	Max	N
SIZE	FAILED	0.197	0.1565	0.172	0.029	0.838	1.176	0.001	0.717	101
	NON-FAILED	0.062	0.016	0.109	0.012	10.447	2.970	0.000	0.870	6494
	t-stat	7.856*								

* Denotes significant difference in means at 5% level
Histogram maps the difference in distribution between failed and non-failed firms after converting scores into probabilities.

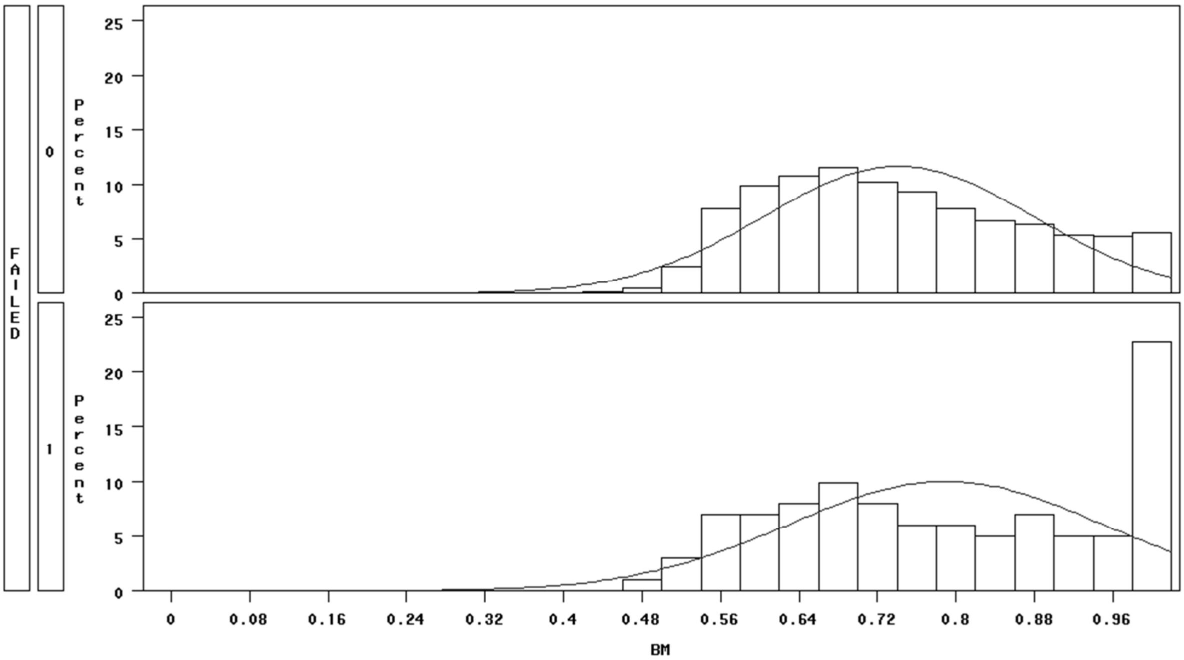


The second variable I include is the book to market ratio (B/M) defined as the book value of a firm (total assets less total liabilities) divided by the market value of the firm (market capital). I use this ratio to investigate whether failed firms are typically over-valued compared to the non-failed firms and whether a lower ratio score suggests a higher probability of failure. The book to market ratio allows, in essence, investigation of the extent of which market capitalisation of future earnings is useful alone in predicting failure. The book-to-market ratio is converted into a failure probability using the formula $p = e^{score}/(1 + e^{score})$.

Table 8.3: Summary statistics for B/M variable

Model		Mean	Median	SD	Var	Kurt.	Skew	Min	Max	N
B/M	FAILED	0.790	0.782	0.160	0.026	-1.33	-0.077	0.467	1.0	101
	NON-FAILED	0.741	0.725	0.137	0.019	-0.03	0.064	0.000	1.0	6494
	t-stat	3.008*								

* Denotes significant difference in means at 5% level
Histogram maps the difference in distribution between failed and non-failed firms after converting scores into probabilities.



The third single variable that I incorporate is cash-flow to total debt as determined by Beaver (1966) to be the best performing ratio for the prediction of failure. Re-examinations of this ratio are infrequent in recent literature[123] as most studies focus on an array of multivariate and non-linear approaches. I include Beaver and his best performing ratio (BV), to investigate how one of the simplest and earliest posited methods of insolvency prediction compares with the latest efforts of academics, practitioners and modern processing power. That is, to see if developments over the last

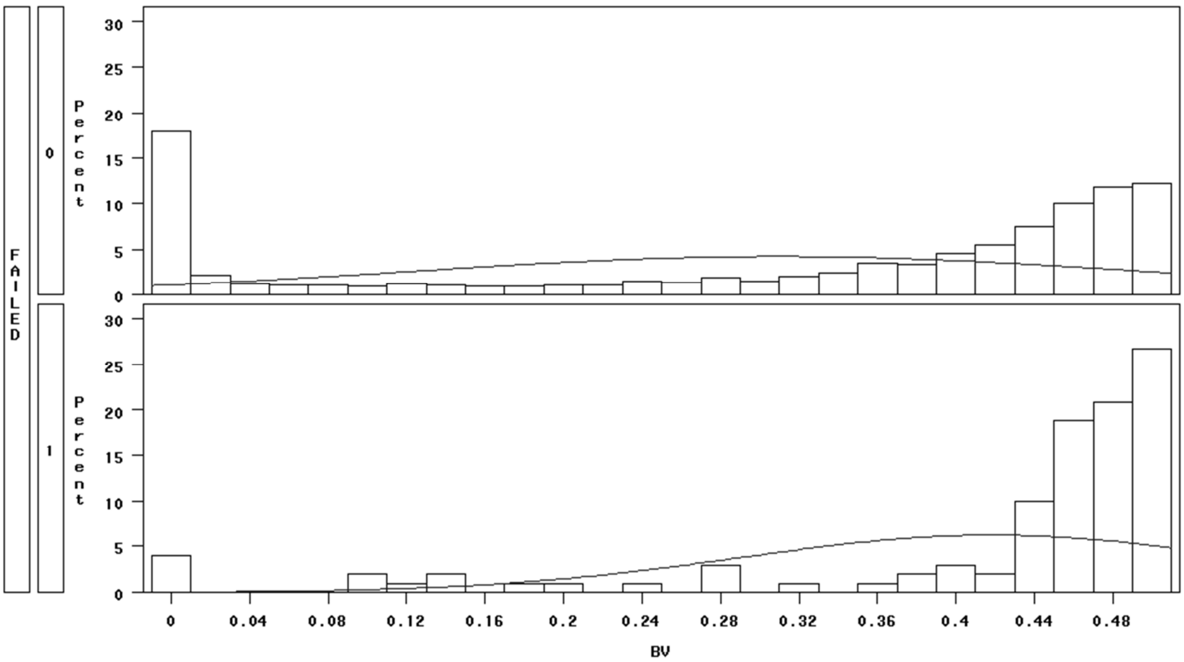
123 Although Zavgren (1983), for example, includes Beaver's ratio in her comparative study.

44 years have significantly advanced our ability to separate the failing from the non-failing firms.

Table 8.4: Summary statistics for BV model

Model		Mean	Med	SD	Var	Kurt.	Skew	Min	Max	N
BV	FAILED	2.794	0.125	20.39	415.71	95.4	9.664	0	202.9	101
	NON-FAILED	40.90	0.407	737.9	544540	3461	53.890	0	50493	6494
	t-stat	-4.063*								

* Denotes significant difference in means at 5% level
Histogram maps the difference in distribution between failed and non-failed firms after converting scores into probabilities.



8.2.2 Discriminant Analysis (Z-Score Models)

I examine two forms of the Z-score model, the Altman (1968) Z-score and the Taffler (1983) Z-score. [124] These are included in my study owing to their popularity in prior literature, past reported efficacy and for the latter, UK relevance. For the initial test where I split the sample through time and thus replicating a real-world setting, I test each of these the models with both their original coefficients and with updated

[124] The latter model coefficients being later disclosed in Agarwal and Taffler (2007).

coefficients as estimated from my more recent data set (the training sample). It is intuitive to expect that the models, for which the coefficients have been updated, would outperform their original counterparts given that these models were derived using data from the 1960s and 1970s. Moyer (1977) was able to reduce misclassification from 25% to 10% by updating the coefficients and I would expect similar results. Table 8.7 sets out the models and their original and updated coefficients. Model re-estimations are performed using the training sample (approximately half of my data set and non-IFRS accounting data), and a validation sample (again, approximately half of my data set but using IFRS accounting data) is used to assess each model's performance. As previously described, I do this in order to replicate real-world information conditions whilst ensuring two large samples of firms. The identities of the variables in the models are as follows: *WC/TA* is working capital divided by total assets; *RE/TA* is retained earnings divided by total assets; *EBIT/TA* is earnings before interest and taxes divided by total assets; V_E/TL is market value of equity divided by total liabilities; *S/TA* is sales divided by total assets; *PBT/CL* is profit before tax divided by current liabilities; *CA/TL* is current assets divided by total liabilities; *CL/TA* is current liabilities divided by total assets; and *NCI* is the 'no-credit interval', being the period a company can continue its operations using its immediate assets if all other forms of short term finance are cut off. More directly, NCI is defined as immediate assets (current assets – stock) less current liabilities, divided by daily operating costs excluding depreciation. All variables are calculated from data items as obtained from Thomson One Banker. The summary statistics for each of the variables and Z-scores, along with the normalised distributions of the resulting models are provided in Tables 8.5 and 8.6.[125]

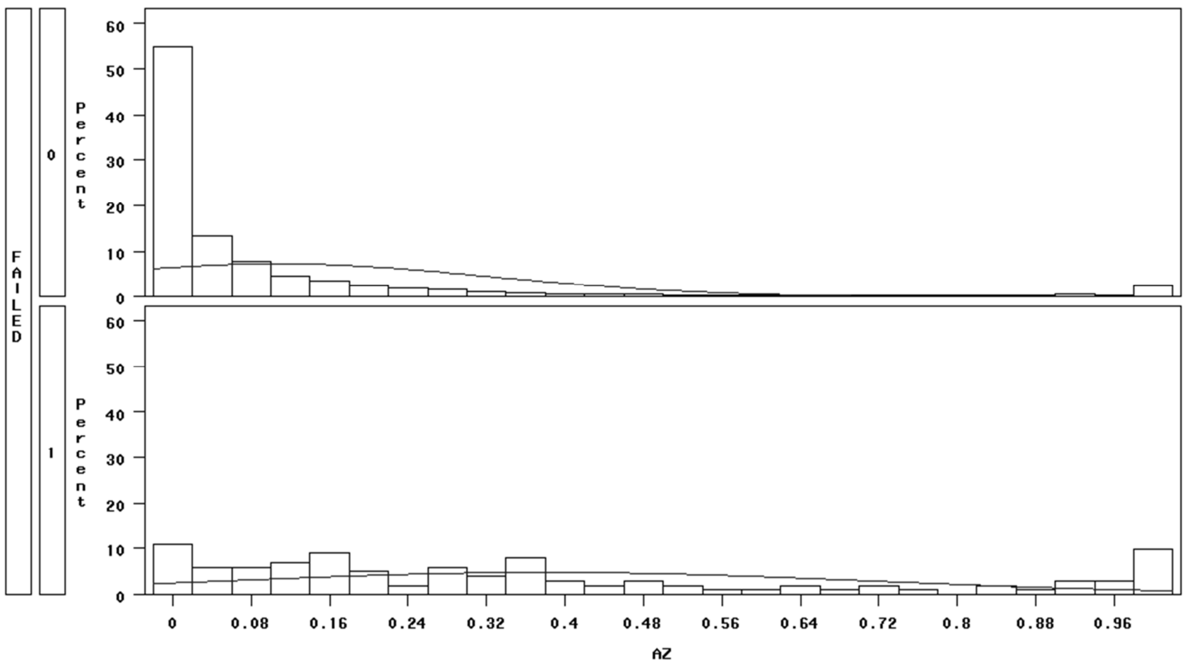
[125] Although the updated models AZU and TZU are trained on the training sample and tested for accuracy via the validation sample, here the coefficient estimates are applied to the whole sample to generate the summary statistics.

Table 8.5: Summary statistics for Altman's Z-score and its variables

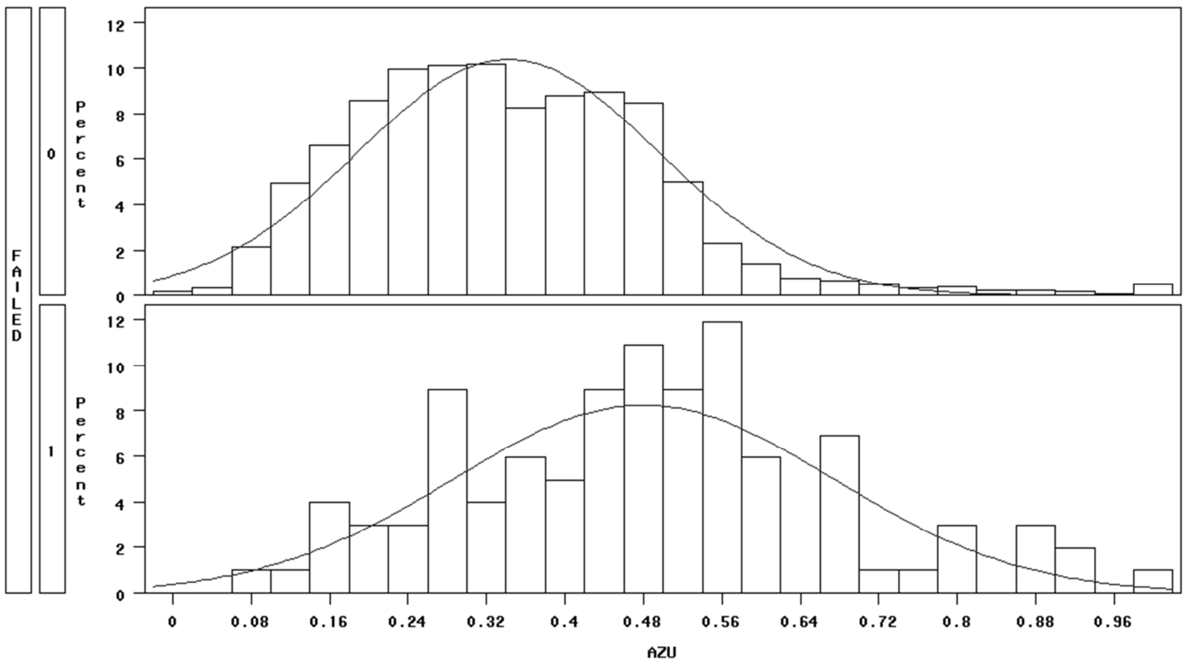
Variable	Group	Mean	Med	SD	Var	Kurt	Skew	Min	Max
WC/TA	Failed	-0.100 (-0.061)	-0.049	0.499	0.249	39.970	-5.07	-4.066	0.743
	Non-failed	0.099 (0.414)	0.098	0.490	0.240	1309.598	-27.7	-25.881	0.992
	t-stat	-3.990*							
RE/TA	Failed	-1.156 (-0.626)	-0.169	3.962	15.698	62.899	-7.42	-36.173	0.496
	Non-failed	-1.654 (0.355)	0.091	92.190	8499	6469	-80.3	-7422	5.364
	t-stat	0.411							
EBIT/TA	Failed	-0.270 (-0.318)	-0.116	0.501	0.251	10.334	-2.88	-2.946	0.318
	Non-failed	-0.043 (0.153)	0.058	0.539	0.291	321.321	-14.7	-16.663	1.718
	t-stat	-4.508*							
VE/TL	Failed	5.539 (0.401)	1.056	17.331	300.361	42.997	6.02	0.025	144.45
	Non-failed	265.775 (2.477)	4.732	4297	18465972	3072.79	50.79	0.004	282458
	t-stat	-4.878*							
S/TA	Failed	0.956 (1.500)	0.799	0.731	0.534	2.639	1.250	0.000	3.996
	Non-failed	1.175 (1.900)	0.973	1.500	2.250	477.685	17.08	0.000	51.145
	t-stat	-2.926*							

Model	Group	Mean	Med	SD	Var	Kurt	Skew	Min	Max
AZ	Failed	-1.64	-0.87	12.49	155.99	25.05	-2.18	-84.3	59.85
	Non-failed	-158	-4.36	2527.5	6388456	3244.9	-51.9	-169474	230.05
	t-stat	4.99*							
AZU	Failed	0.018	-0.06	1.505	2.265	40.067	5.15	-2.225	11.998
	Non-failed	-0.69	-0.71	1.423	2.024	962.23	20.96	-17.929	69.562
	t-stat	4.72*							

* Denotes significant difference in means at 5% level
The values in parenthesis represent the values presented within the original formulation of the model to aid comparison between the updated and original models.
** The signs of the coefficients have been reversed so that the resulting score is increasing with the probability of failure. Histogram maps the difference in distribution between failed and non-failed firms after converting scores into probabilities.



FAILED
0
1
Percent
0
10
20
30
40
50
60
0
0.08
0.16
0.24
0.32
0.4
0.48
0.56
0.64
0.72
0.8
0.88
0.96
AZ



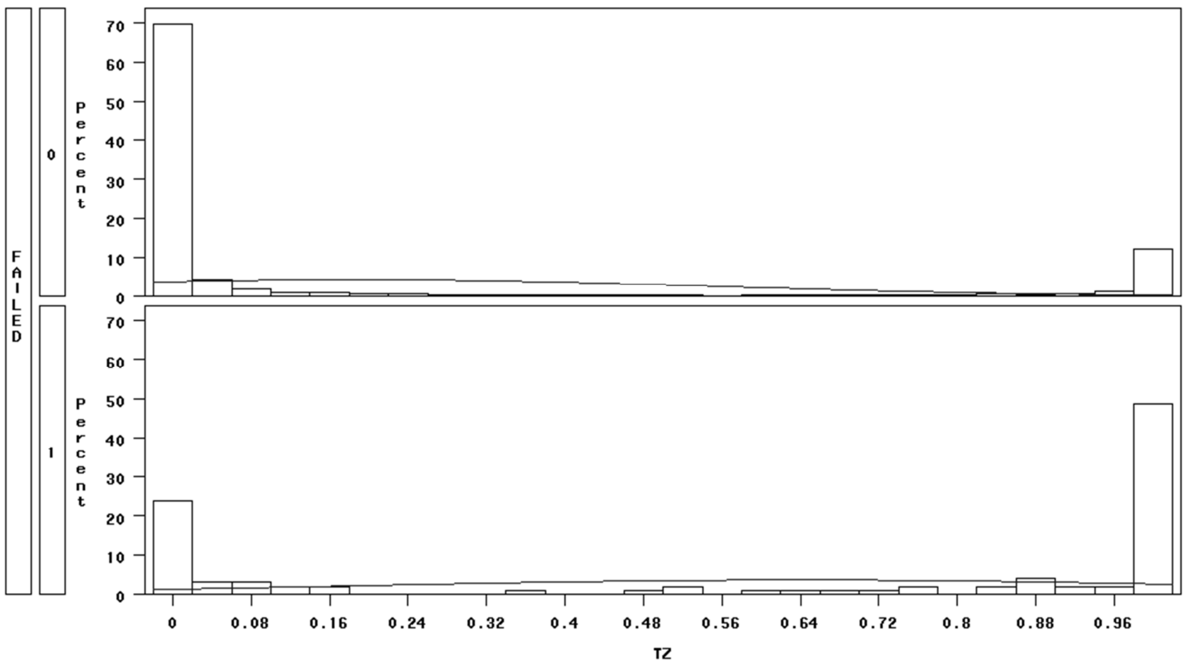
FAILED
0
1
Percent
0
2
4
6
8
10
12
0
0.08
0.16
0.24
0.32
0.4
0.48
0.56
0.64
0.72
0.8
0.88
0.96
AZU

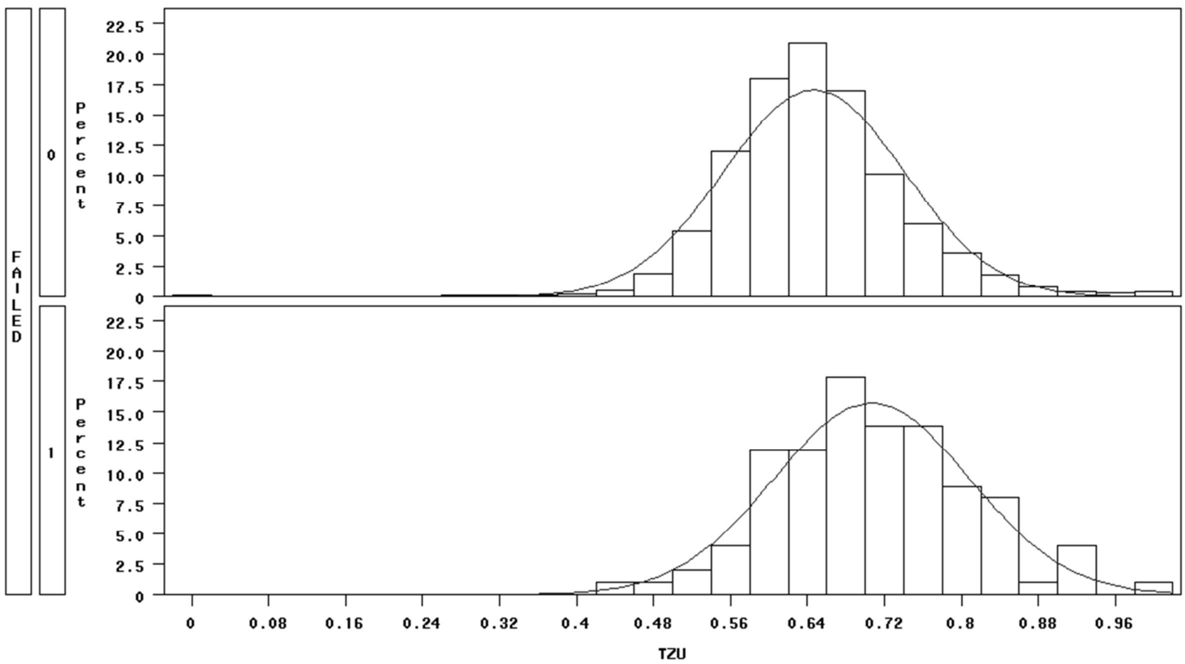
Table 8.6: Summary statistics for the Taffler Z-score and its variables

Variable	Group	Mean	Med	SD	Var	Kurt	Skew	Min	Max
PBT/TL	Failed	-0.695	-0.199	1.484	2.203	9.900	-2.975	-7.628	1.304
	Non-failed	-0.116	0.168	2.622	6.873	2150	32.938	-37.984	159.762
	t-stat	-3.830							
CA/TL	Failed	6.273	1.247	34.413	1184.283	91.431	9.386	0.052	340.493
	Non-failed	67.399	2.485	834.744	696798	2593	44.836	0	53027
	t-stat	-5.603							
CL/TA	Failed	0.511	0.405	0.535	0.286	51.520	6.239	0.032	5.036
	Non-failed	0.370	0.323	0.489	0.239	1442	30.310	0.003	26.881
	t-stat	2.636							
NCI	Failed	-13.696	-55.893	564.508	318670	85.990	8.896	-704.388	5387.986
	Non-failed	132.745	-3.544	3525	12426818	1657.84	39.263	-16891	155125
	t-stat	-2.057							

Model	Group	Mean	Med	SD	Var	Kurt	Skew	Min	Max
TZ**	Failed	-1.3	6.912	90.14	8125	73.396	-8.07	-830	91
	Non-failed	-167	-5.58	2088	4363385	2581	-44.6	-132545	542.4
	t-stat	6.04*							
TZU**	Failed	1.001	0.852	0.964	0.928	49.72	6.045	-0.316	9.088
	Non-failed	0.653	0.579	1.021	1.043	855.7	16.32	-20.93	48.4
	t-stat	3.59*							

* Denotes significant difference in means at 5% level
** The signs of the coefficients have been reversed so that the resulting score is increasing with the probability of failure. Histogram maps the difference in distribution between failed and non-failed firms after converting scores into probabilities.





The models updated via the initial chronological split, show that not all of the original variables are significant or add much information to the model. For the Altman model AZU I can observe that V_E/TL returns a very small value adding no information to the model at all, likewise with the variable RE/TA. For TZU I find that only PBT/CL and CL/TA are significant (both containing the CL variable, intrinsically related to both liquidity and short-term profitability. The signs of each coefficient are mostly acting in the same direction as their original counterpart, particularly within the Taffler model, and notably all of the significant coefficients have identical signs to their original counterparts. This suggests that each variable's directional influence upon insolvency is the same as it was when the original model was formulated which I would expect to be the case. The changes in magnitude reflect the population and period in which the training samples were constructed.

Table 8.7: Z-score models and their initial parameter estimates

Altman (1968) Model	WC/TA	RE/TA	EBIT/TA	V_E/TL	S/TA
Original Coefficients (AZ)[a]	-1.2	-1.4	-3.3	-0.6	-0.99
Updated Coefficients (AZU)[b]	-2.67***	0.001	-0.423***	0.000	-0.38**
Taffler (1984)	PBT/CL	CA/TL	CL/TA	NCI	
Original Coefficients (TZ)[a]	-12.18	-2.5	10.68	-0.029	
Updated Coefficients (TZU)[b]	-0.131***	-0.0001	1.78***	-0.0001	

Significance of estimates of updated coefficients: ***, **, * denote two-tailed significant at the 1%, 5%, 10% levels.
[a] The signs of the original coefficients for each model have been reversed so that the resulting score is increasing with the probability of failure.
[b] The updated coefficients are raw canonical coefficients estimated using the SAS procedure PROC CANDISC in an attempt to mimic the discriminant approach of the original authors. The base data for re-estimation of the Altman model and Taffler models (the "training sample") is described in section 7.

The original Altman model for example was derived from US data from the period 1946-1965, compared to my study which uses UK data from 2000-2005 (in this instance, the chronologic training sample). It is therefore of little surprise if there is significant change in the coefficient estimates. The Taffler Z-score is based on data from the same country to my study and is more recent (1968-1976) than the Altman model. The small magnitude of two (out of four) variables, however, suggests that these indicators are not as relevant to UK distress prediction as perhaps they once were (at least relative to PBT/CL and CL/TA).

One further cause for the change in magnitude of these coefficients may lay in the discriminant process itself and its heavy reliance on both variance and covariance of the underlying variables. Both of these original models were formulated using small, carefully designed datasets with restrictions placed on company attributes such as size and sector. As previously discussed in section 7 of this thesis, with the exception of the removal of all financially coded companies, no such limitations have been placed upon my empirical study. Section 9.3 investigates this further by studying the effects of Winsorisation upon both the accuracy and coefficients estimates of these two discriminant models. In short the results show that under a certain amount of Winsorisation, both the original and updated models can achieve identical accuracies and under this circumstance demonstrate very similar values for their coefficient estimates.

These coefficients which are estimated using this time-split approach are just one part of this evaluation. Further within this thesis I generate 10,000 new sets of coefficients based upon random sample splits, for both of the Z-score models. I do this to investigate the distributional properties of the models predictive accuracy over the last decade and to determine whether splitting the data sample by chronology gives a fair representation of the models predictive ability. Section 9.4 provides the results of this investigation.

8.2.3 Ohlson's logit model

I test one particular version of the logit model, that of Ohlson (1980). The reason for its inclusion is based on its popularity and influence in the prior literature (e.g., Zavgren, 1983; Begley et al, 1996; Hillegeist, 2004). I test the model with both its original coefficients (OL) and with updated coefficients (OLU) as estimated using my training sample. Table 8.3 sets out the model with its original and updated coefficients.

The identities of the variables in the models are as follows: *SIZE* is the natural logarithm of GDP-deflated total assets; *TL/TA* is total liabilities divided by total assets; *WC/TA* is working capital divided by total assets; *CL/CA* is current liabilities divided by current assets; *OENEG* is a dummy variable equal to one if total liabilities exceed total assets, and zero otherwise; *NI/TA* is net income divided by total assets; *FUTL* is funds from operations (pre-tax income plus depreciation and amortization) divided by total liabilities; *INTWO* is a dummy variable equal to one if net income was negative over previous two years, and zero otherwise; *CHIN* is the scaled change in net income calculated as $(NI_t-NI_{t-1})/(|NI_t|+|NI_{t-1}|)$, where NI_t is the net income for the most recent period. All variables are calculated from data items as obtained from Thomson One Banker.

Table 8.8 presents the specifications for the original Ohlson logit model, and a model in which the coefficients have been updated using the chronological training sample. Similar to the Z-score model updates I can see that the majority of coefficient estimates (seven out of ten) have the same sign. Only three variables WC/TA, INTWO, and CHIN (measures of short term liquidity and profitability) are statistically significant, whereas FUTL provides almost nothing to the model. Similarly to the MDA models, the changes in coefficient magnitude reflects the population and period in which the training samples were constructed, suggesting that their influence has changed over time and/or is distorted by inter-country differences. The original Ohlson model was derived from US data from the period 1970-1976 and is therefore unrealistic to expect too many similarities to my study, although there is nothing to suggest that any of the variables used in the model(s) would not be an intuitive predictor.

Table 8.8: Ohlson's logit model and parameter estimates

Ohlson (1980) Model	Constant	SIZE	TL/TA	WC/TA	CL/CA	OENEG	NI/TA	FUTL	INTWO	CHIN
Original Coefficients	-1.32	-0.41	6.03	-1.43	0.08	-2.37	-1.83	0.285	-1.72	-0.52
Updated Coefficients[a]	-1.91***	-0.217*	2.492***	-0.946*	-0.007	-12.208***	0.321	0.003	1.434**	-0.639

Significance of estimates of updated coefficients: ***, **, * denote two-tailed significant at the 1%, 5%, 10% levels.
[a] The updated coefficients are estimated by logistical regression using the SAS procedure LOGISTIC. The base data for the re-estimation (the training sample) is in accordance with that which is described in section 7.

The summary statistics for each of the variables and Z-scores, along with the normalised distributions of the resulting models are provided in Table 8.9. On an individual basis, seven of the nine variables have means which are significantly different between the failed and non-failed groups of firms. The magnitudes of the values in all but one case are comparable to the values presented in Ohlson (1980). The reason for this scaling difference is due to a number of observations in which the total debt figure is considerably low in comparison to the funds from operations.[126]

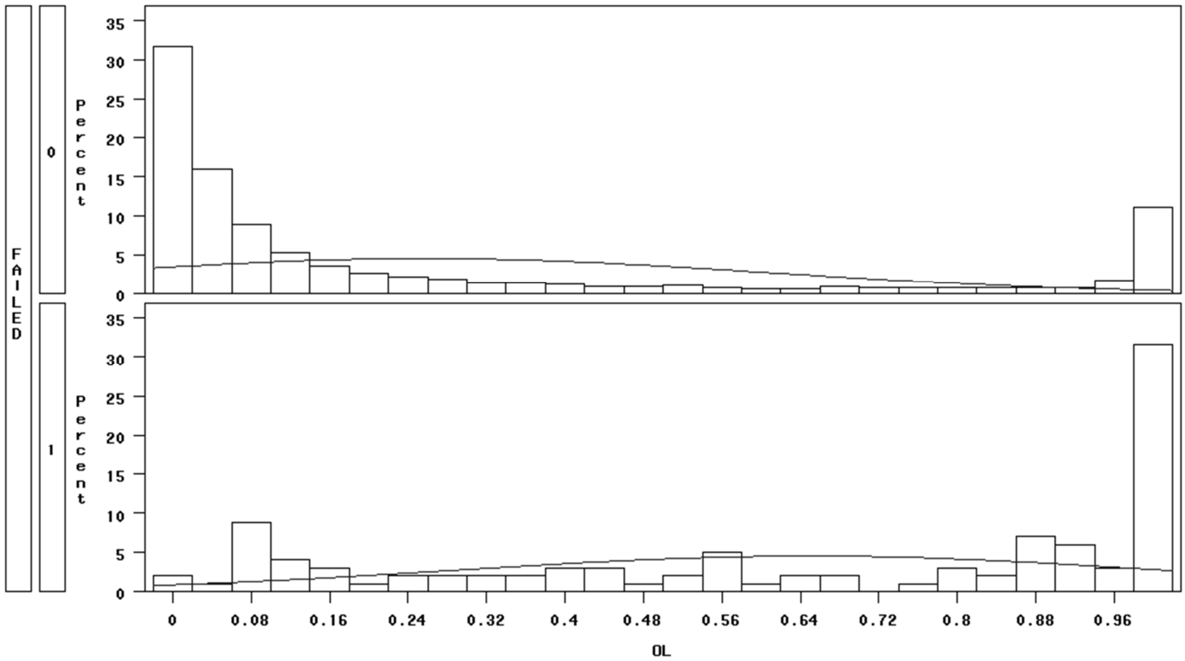
[126] Although the updated models AZU and TZU are trained on the training sample and tested for accuracy via the validation sample, here the coefficient estimates are applied to the whole sample to generate the summary statistics.

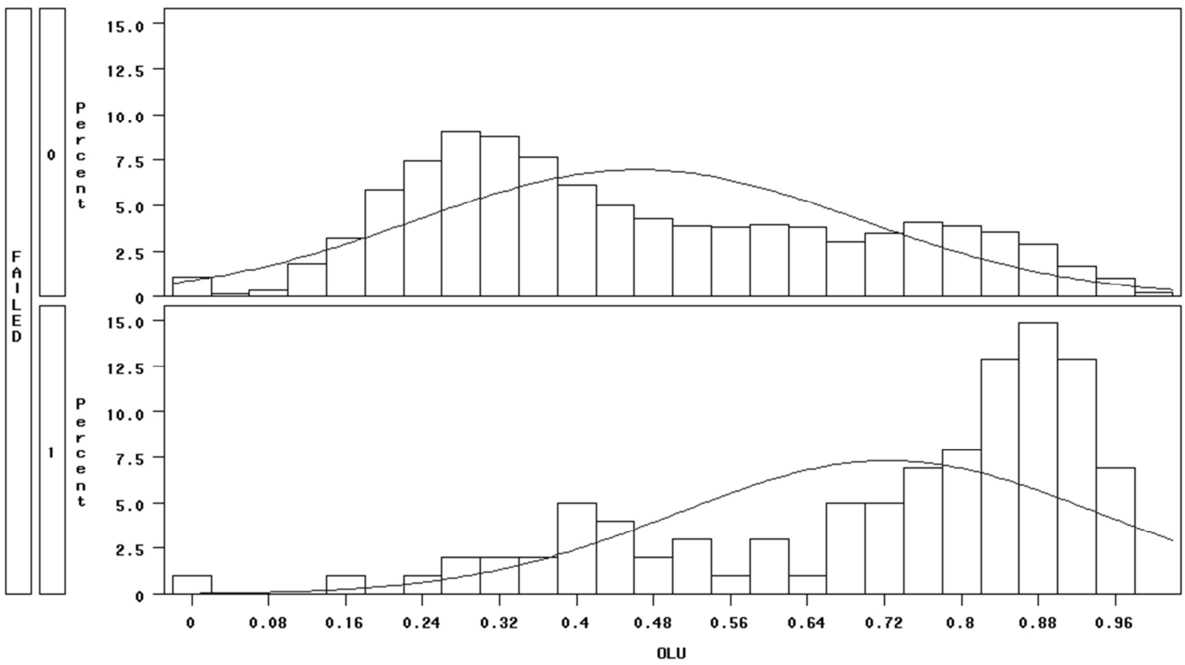
Table 8.9: Summary statistics for the Ohlson logit model and its variables

Variable	Group	Mean	Med	SD	Var	Kurt	Skew	Min	Max
SIZE	Failed	17.004	16.914	1.718	2.950	0.248	0.235	11.750	21.535
		(12.134)		(1.38)					
	Non-failed	18.330	18.117	2.271	5.155	-0.056	0.321	7.844	26.039
		(13.26)		(1.570)					
	t-stat	-7.656*							
TL/TA	Failed	0.362	0.326	0.289	0.084	23.044	3.560	0.003	2.388
		(0.905)		(0.637)					
	Non-failed	0.224	0.178	0.257	0.066	145	8.042	0.000	7.072
		(0.488)		(0.181)					
	t-stat	4.791*							
WC/TA	Failed	-0.100	-0.049	0.499	0.249	39.970	-5.073	-4.066	0.743
		(0.041)		(0.608)					
	Non-failed	0.099	0.098	0.490	0.240	1309.598	-27.783	-25.881	0.992
		(0.310)		(0.182)					
	t-stat	-3.990*							
CL/CA	Failed	1.705	1.217	1.919	3.684	18.476	3.918	0.046	12.740
		(1.32)		(2.52)					
	Non-failed	1.268	0.771	13	171	3244.264	54.414	0.008	864
		(0.525)		(0.740)					
	t-stat	1.746							
OENEG	Failed	0.010	0.000	0.100	0.010	101.000	10.050	0	1
		(0.18)		(0.385)					
	Non-failed	0.010	0.000	0.098	0.010	98.166	10.007	0	1
		(0.0044)		(0.0660)					
	t-stat	0.020							
NI/TA	Failed	-0.298	-0.137	0.502	0.252	10.091	-2.852	-2.954	0.380
		(-0.208)		(0.411)					
	Non-failed	-0.075	0.030	0.538	0.290	320	-14.354	-17	1.678
		(0.052)		(0.075)					
	t-stat	-4.426*							
FUTL	Failed	-2.413	-0.178	14.396	207.259	94.902	-9.616	-143.513	1.109
		(-0.117)		(0.421)					
	Non-failed	-7.654	0.311	161.205	25987	1154.631	-29.835	-7684	381.83
		(0.28)		(0.36)					
	t-stat	2.130*							
INTWO	Failed	0.743	1.000	0.439	0.193	-0.746	-1.126	0	1
		(0.39)		(0.488)					
	Non-failed	0.367	0.000	0.482	0.232	-1.696	0.552	0	1
		(0.043)		(0.203)					
	t-stat	8.508*							
CHIN	Failed	-0.203	-0.214	0.605	0.367	-0.791	0.348	-1	1
		(-0.322)		(0.644)					
	Non-failed	0.017	0.053	0.559	0.313	-0.431	-0.117	-1	1
		(0.0379)		(0.458)					
	t-stat	-3.635*							

Model	Group	Mean	Med	SD	Var	Kurt	Skew	Min	Max
OL**	FAILED	0.65	0.813	0.356	0.126	-1.331	-0.519	0.013	1
	NON-FAILED	0.256	0.069	0.349	0.122	0.015	1.267	0	1
	t-stat	11.0*							
OLU**	FAILED	0.723	0.814	0.218	0.047	0.274	-1.07	0.016	0.972
	NON-FAILED	0.464	0.408	0.228	0.052	-0.867	0.406	0	1
	t-stat	11.8*							

* Denotes significant difference in means at 5% level
** The signs of the coefficients have been reversed so that the resulting score is increasing with the probability of failure. Histogram maps the difference in distribution between failed and non-failed firms. The values in parenthesis represent the values presented within the original formulation of the model to aid comparison between the updated and original models.





8.2.4 Artificial Neural Networks

For the neural network analysis I use a particular training algorithm called the backpropagation algorithm. Backpropagation verifies and adjusts the weighted signals produced by each node in the hidden layer according to the amount in which they contribute to the overall least squares error function

$$E = \frac{1}{2}\sum_{j=1}^{n} |o_j - t_j|^2 \quad (8.1)$$

where o_j is the output produced by the training observation j and t_j is the target output for the same observation. The algorithm looks for the minimum of this error function in weight space using the method of gradient descent. The signals are fed back through the network and the combination of weights which minimizes the error function is considered to be a solution of the learning problem. Figure 8.4 visualises the three layers of the NN along with the inter-connections. A full derivation of the backpropagation algorithm can be found in Rojas (1996).[127]

[127] It was desirable that this derivation was included within the Appendix to this thesis. The length of the derivation and the time constraints put upon me has meant that this has not been possible. However Appendix I does offer a small discussion of the leading principles involved within the development of the training algorithm..

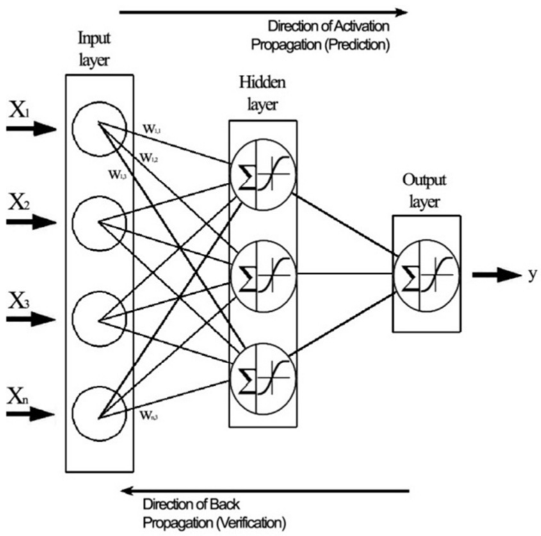


Figure 8.4: A feedforward neural network with backpropagation

I apply the NN methodology described above to form three different models based upon the variables employed by Altman (1968), Ohlson (1980), and Taffler (1983). By comparing the original model, the model with updated coefficients, and the non-linearly constrained NN model, I hope to determine and evaluate each models performance from a different perspective, i.e. from the perspective of the technology employed when the inputs remain constant, rather than that of the model itself.

With respect to the construction of the hidden layers, referring back to section 5.4 and with reference to the majority of situations, there is no clear-cut method for determining the optimal number of hidden layers and the number of nodes within them, without training several networks and estimating error rates for each. Too few hidden units will result in a high training error and high classification errors when applied to further samples because the model is under-fitted. On the other hand, if one has too many hidden units one might get a low training error but might still get large errors when applied to further samples because of over-fitting. Tetko et al. (1995) provide a discussion as to the effects of this trade-off, whilst others offer certain “rules of thumb” for selecting the number of hidden layers and nodes (Berry and Linoff, 1997; Blum, 1992; Swingler, 1996; Boger and Guterman, 1997).

For my networks therefore, there was a certain amount of trial and error while trying to determine the optimal level of hidden layer items, training epochs, and momentum. The optimal structure consisted of n inputs in the input layer (model dependant), n+1 hidden nodes, one output node, zero momentum and twenty-five raining cycles (epochs) using the backpropagation algorithm; this was applied to each of the three models to aid comparability.

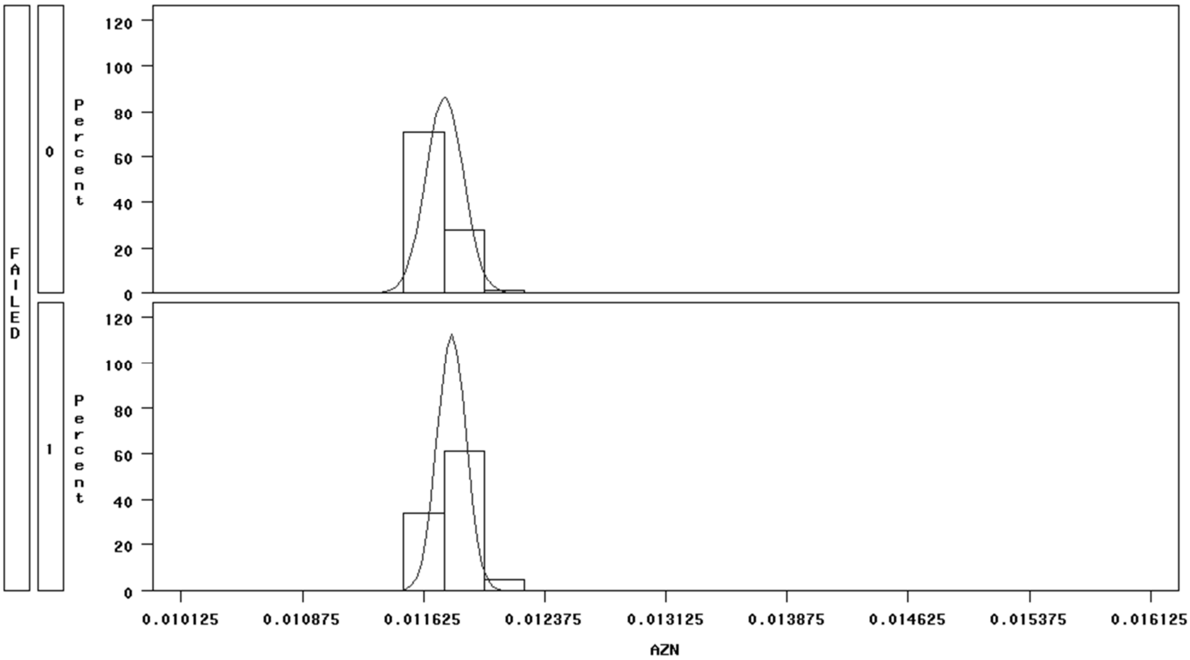
For a description of the variables used for each model please refer back to Sections (8.2.2) and (8.2.3). Table 8.10 provides the summary statistics for each of the three models.[128]

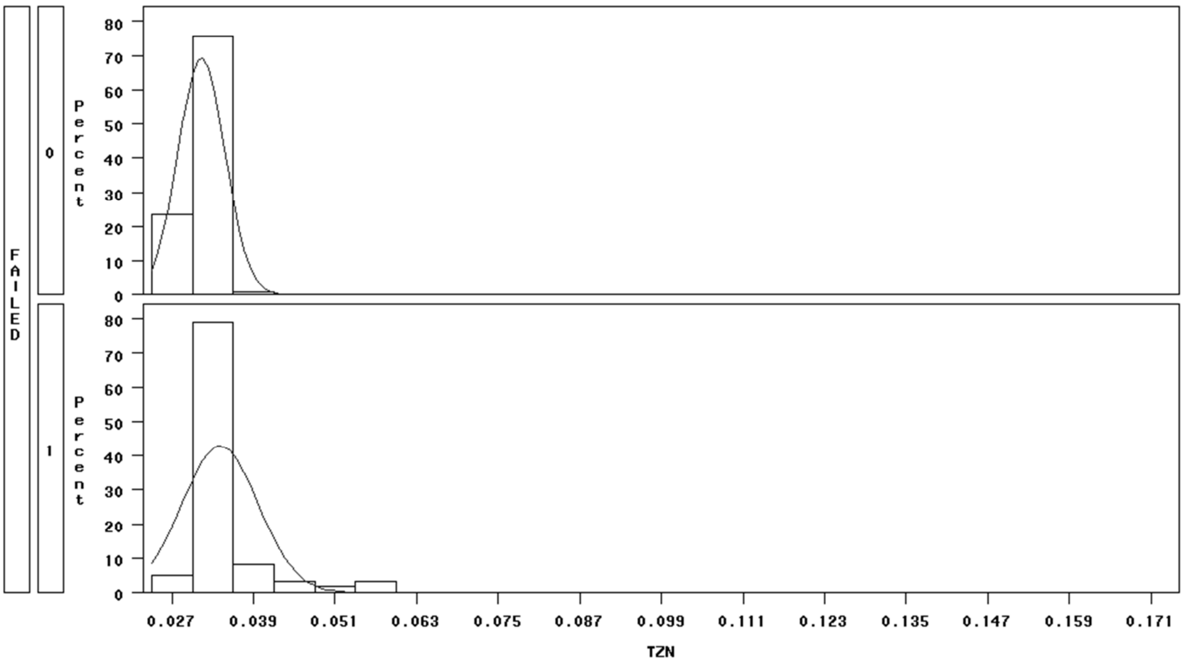
Table 8.10: Summary statistics for the neural network models

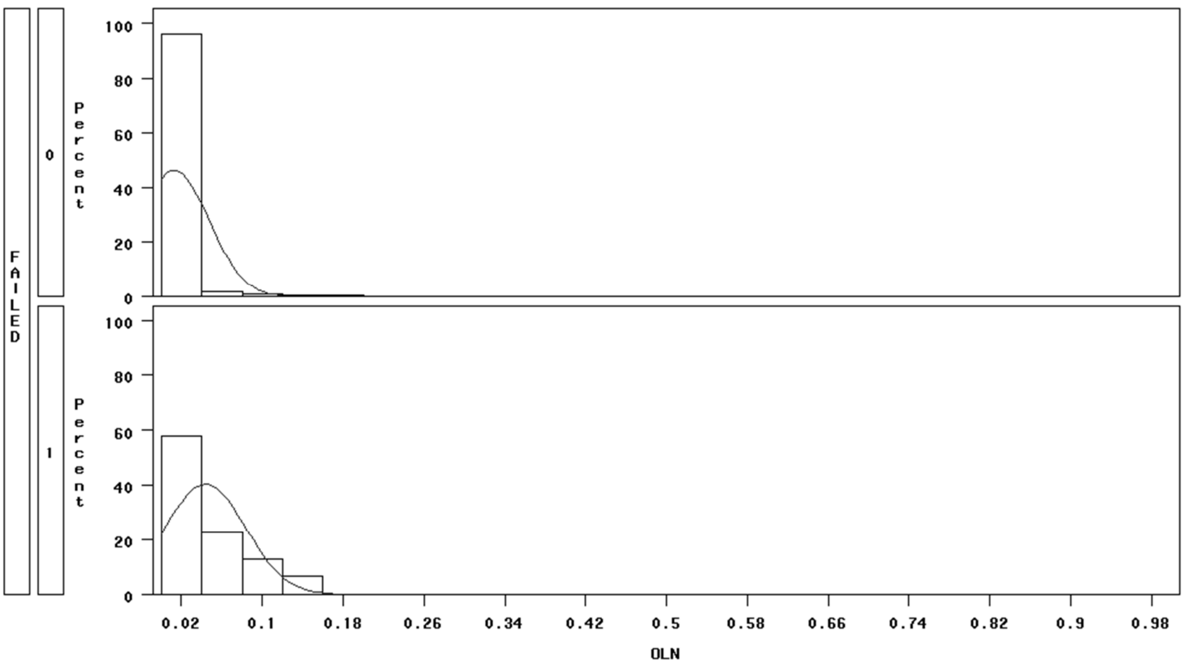
		Mean	Med	SD	Var	Kurt	Skew	Min	Max	N
AZN	FAILED	0.011	0.011	0.000	7.85E-09	5.738	2.233	0.0116	0.0121	62
	NON-FAILED	0.011	0.011	0.000	1.34E-08	645.1	18.906	0.010	0.0161	3530
	t-stat	3.77*								
TZN	FAILED	0.034	0.032	0.005	3.14E-05	8.557	2.880	0.029	0.057	62
	NON-FAILED	0.031	0.031	0.003	1.2E-05	827.27	23.331	0.024	0.171	3530
	t-stat	3.85*								
OLN	FAILED	0.043	0.032	0.0396	0.0015	0.161	0.995	0.0033	0.1560	62
	NON-FAILED	0.012	0.007	0.034	0.0012	319.6	15.199	0.001	0.996	3530
	t-stat	6.11*								

* Denotes significant difference in means at 5% level.
Histograms map the difference in distribution between failed and non-failed firms after converting scores into probabilities.

[128] Due to the black-box nature of the neural network, the scores were not available for the training sample, The summary stats are therefore only available for the validation sample.


FAILED
0
1
Percent
120
100
80
60
40
20
0
0.010125
0.010875
0.011625
0.012375
0.013125
0.013875
0.014625
0.015375
0.016125
AZN


FAILED
0
1
Percent
80
70
60
50
40
30
20
10
0
0.027
0.039
0.051
0.063
0.075
0.087
0.099
0.111
0.123
0.135
0.147
0.159
0.171
TZN



8.2.5 The European call contingent claims model

The methodology and theoretical underpinning to the contingent claims models have previously been discussed within section 5.6 of this thesis; therefor this first paragraph merely acts as a reminder of the basic principles of the model. What follows is a discussion of two different approaches to the pertaining parameter estimation process.

The EC model is used to measure the default probability of a firm and is based on the work of Black and Scholes (1973), Merton (1974) and adopted and modified by Moody's Rating Methodology as described in Sobehart and Stein (2000) and Crosbie and Bohn (2003).

The model views equity as a European call option on a firm's assets with a strike price equal to the face value of its liabilities. All liabilities are assumed to be zero-coupon bonds following Black and Scholes (1973) and Merton (1974). The option expires at time T when the debt matures. At this point the equity holders either; exercise their option and pay off the debt (if the firm's assets (V) are of greater worth than the face value of its liabilities (X)) or, let the option expire (if the assets are not sufficient to cover the cost of the maturing debt). If the option is left to expire then the debt holders will file for liquidation and the assets will be transferred to them without cost. The residual claim for the equity holders is zero. The contingent claims model determines the probability of these two outcomes. McDonald (2002, p. 604), shows that the probability of failure V (T) < X can be calculated as follows:

$$\text{Prob} = \mathrm{N}\left[-\frac{\ln\left[\frac{V}{X}\right] + \left(\mu - \delta - \frac{1}{2}\sigma^2\right)T}{\sigma\sqrt{T}}\right] \tag{8.2}$$

Equation (8.2) contains variables which are both observable and those that are not. The observable variables of Equation (8.2) are X (face value of long-term liabilities), T (time to expiry – taken to be one year), and δ (annual dividend rate). V, σ and μ are not directly observable and, therefore, must be estimated. V represents the market value of the firm's assets, σ is the asset volatility and μ is the expected return of the firm. The next section summarises two versions of the EC model which are evaluated within my study, that contain differing approaches to the estimation of the unobservable parameters V, σ and μ, and to treatment of the X value.

8.2.5.1 Hillegeist et al. (2004)

The value of equity as a call option on the value of the firm's assets V_E is:[129]

$$V_E = V\mathrm{e}^{-\delta T}\mathrm{N}(d_1) - Xe^{-rT}\mathrm{N}\left(d_1 - \sigma\sqrt{T}\right) + \left(1 - \mathrm{e}^{-\delta T}\right)V \tag{8.3}$$

where

$$d_1 = \frac{\ln\left[\frac{V}{X}\right] + \left(r - \delta + \frac{1}{2}\sigma^2\right)T}{\sigma\sqrt{T}}$$

Here, the optimal hedge equation is defined as:

$$\sigma_E = \frac{\left[Ve^{-\delta T}N(d_1)\sigma\right]}{V_E} \tag{8.4}$$

The values V and σ can be found by solving Equations (8.3) and (8.4) simultaneously. The solution methods employed by Moody's are described by Crosbie

[129] The call equation is modified to include dividends, reflecting that the dividend stream flows to equity holders.

and Bohn (2003) as a *"complex iterative procedure"* and Hillegeist et al. use SAS implementation to back-out these unknowns. In this study I use a modified version of the SAS code as provided by Hillegeist el al. (2004, pp 30-31) with initial values of V and σ computed according to Bharath and Shumway (2008). The values of X are proxied by total liabilities, departing from the Moody's method that uses current liabilities plus half of long-term debt. Hillegeist et al. suggest that lowering the value of liabilities mechanically reduces the probability of failure estimates and slightly weakens the model.

The expected return on the firm's assets μ is proxied by the actual return on assets for the previous year thus:

$$\mu_t = \max \left[\frac{V_t + Dividends - V_{t-1}}{V_{t-1}}, r\right] \tag{8.5}$$

where r represents the risk free rate and $Dividends$ are the sum of the common and preferred dividends declared during the year. The expected return from Equation (8.5) is bounded below by the risk-free rate (as shown), and also bounded above by 100%.

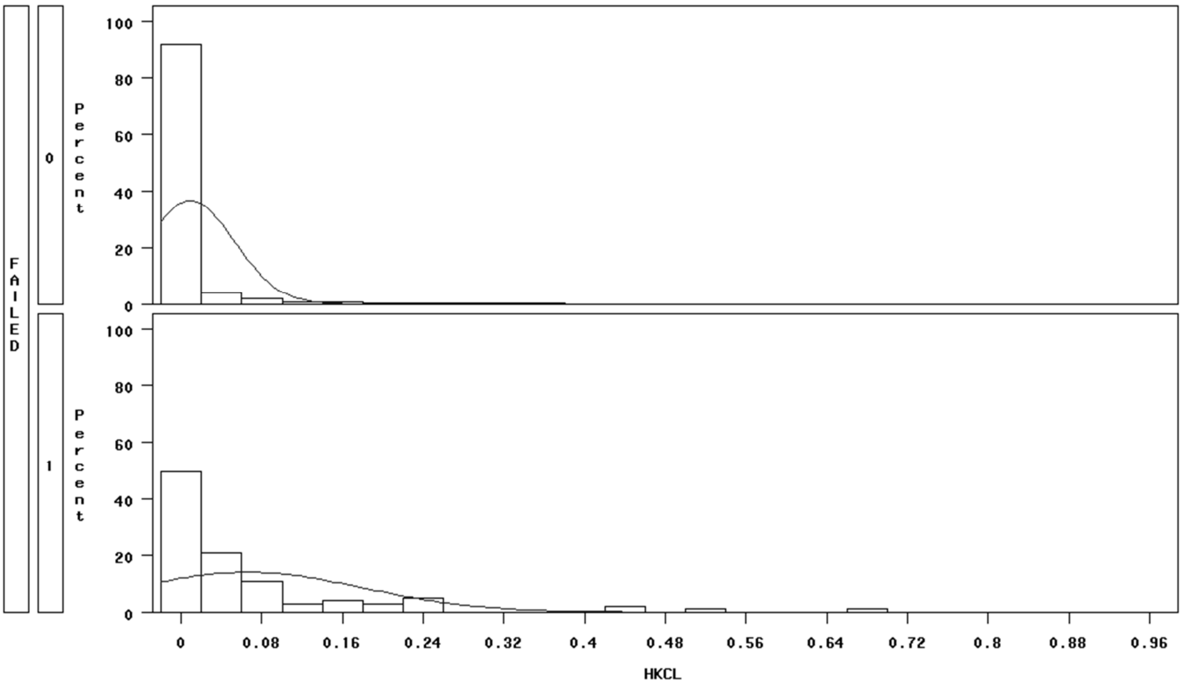
Inserting the values derived from expressions (8.3) through (8.5) into expression (8.2) provides the Hillegeist et al. (HKCL) probability of default.[130]

[130] Moody's methodology differs here in that the defaults are not taken from an assumed normal distribution, but instead mapped to historical default frequencies from their large proprietary database.

Table 8.11: Summary statistics for HKCL model and its variables

		Mean	Med	SD	Var	Kurt	Skew	Min	Max	N
HKCL	FAILED	0.065	0.021	0.114	0.013	11.280	3.081	0.000	0.686	101
	NON-FAILED	0.009	0.000	0.044	0.002	153.42	10.27	0.000	0.978	6494
	t-stat	$\boxed{4.90*}$								

* Denotes significant difference in means at 5% level
Histogram maps the difference in distribution between failed and non-failed firms.



8.2.5.2 The Bharath and Shumway (2008) "naive" approach

The naive EC model presented Bharath and Shumway (2008) simplified the methodology examined by Hillegeist et al. (2004) – in which where the values of V and σ are estimated by solving two simultaneous non-linear equations. Bharath and Shumway instead use the following formulations:

$$V = V_E + X \tag{8.6}$$

$$\sigma = \frac{V_E}{V}\sigma_E + \frac{X}{V}\sigma_D \tag{8.7}$$

$$\sigma_D = 0.05 + 0.25\sigma_E \tag{8.8}$$

Where σ_D represents the volatility of the firms debt – naively deduced by a linear transformation of the volatility of equity, σ_E.

Bharath and Shumway approximate the value of liabilities, X, to be current liabilities plus one half of long term debt, following Moodys' methodology. Reisz and Purlich (2007) present that within a barrier option framework, the barrier that triggers insolvency is more likely to be lower than the face value of long-term liabilities, therefore, current liabilities plus one half of long term debt may be a more realistic approach for a call option model.

The expected return on the firm's assets μ is estimated simply by using the prior year stock return bounded between the risk-free rate and 100%. According to Agarwal and Taffler (2008), a problem with this bounded variable is that

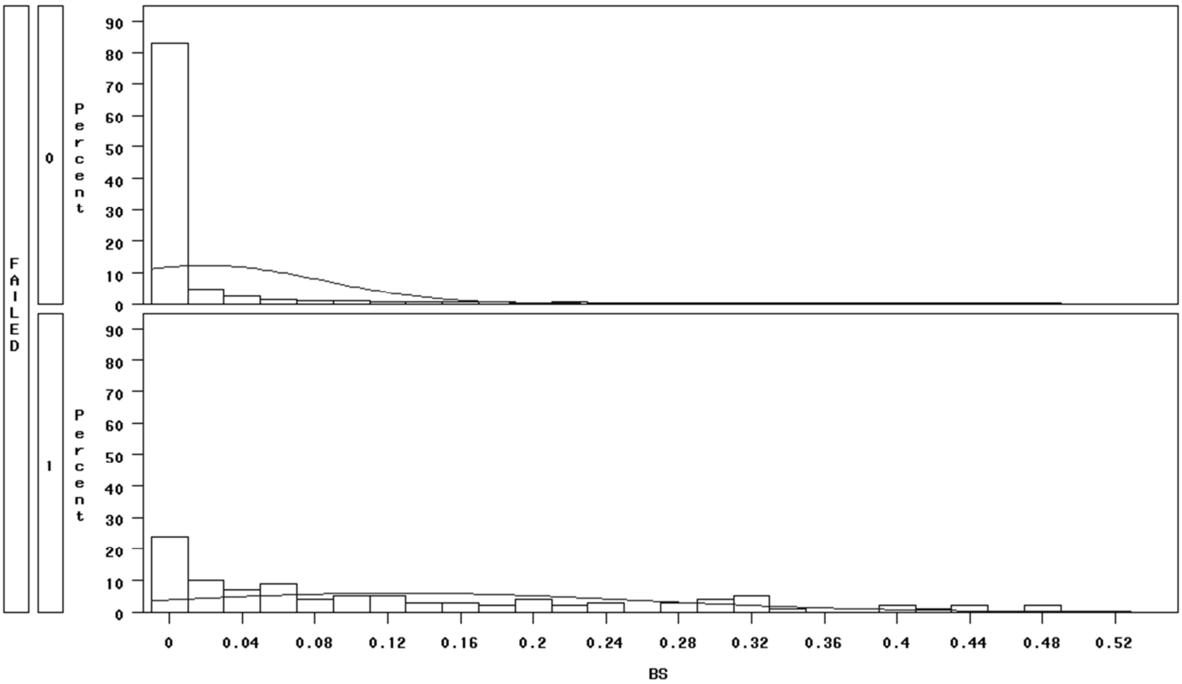
"…even if realized return on equity is a good proxy for expected return on equity, it will be a good proxy for expected return on assets only if the expected return on debt is the same for all the firms".

More problematic may be that, in a real world setting, the prior year stock return is likely to be negative particularly within a sample period enveloping a period of economic downturn, thus setting a floor at the risk-free rate does not reflect the possibility of equity degradation. Agarwal and Taffler (2008) show, however, that the expected return generating model does not significantly affect the outcome of the probability estimates.

Table 8.12: Summary statistics for BS model and its variables

		Mean	Med	SD	Var	Kurt	Skew	Min	Max	N
BS	FAILED	0.126	0.074	0.134	0.018	-0.057	1.007	0.000	0.482	101
	NON-FAILED	0.020	0.000	0.064	0.004	20.681	4.348	0.000	0.548	6494
	t-stat	7.93*								

* Denotes significant difference in means at 5% level
Histogram maps the difference in distribution between failed and non-failed firms.



8.2.6 The down-and-out call barrier option

The barrier option framework extends the EC model and treats the underlying value of firm equity to be treated as a DOC option. Debt holders own a portfolio of risk-free debt and a DOC option on the firm's assets which can be exercised should the value of the firm falls below a predetermined barrier. The implied firm-specific barrier level is estimated by using market values of firm traded equities (or an appropriate proxy) which can be achieved by adopting Equation (8.9) and solving for the barrier (B). Equation (8.9) shows the DOC formula under the condition that the barrier is set at, or below, the strike-price/liability value.[131]

[131] See Brockman and Turtle (2002) and Reisz and Purlich (2007) for details of possible variations, and an excellent discussion and formulation of barrier options framework. The DOC formula is often present as containing an included term to take into consideration a possible rebate which may be attributable to the

$$V_E = DOC = V\mathrm{e}^{-\delta\tau}\mathrm{N}(d_1) - X\mathrm{e}^{-r\tau}\mathrm{N}\left(d_1 - \sigma\sqrt{\tau}\right)$$
$$- \left[V\mathrm{e}^{-\delta\tau}\left(\frac{B}{V}\right)^{\frac{2(r-\delta)}{\sigma^2}+1}\mathrm{N}(d_1^B) - X\mathrm{e}^{-r\tau}\left(\frac{B}{V}\right)^{\frac{2(r-\delta)}{\sigma^2}-1}\mathrm{N}\left(d_1^B - \sigma\sqrt{\tau}\right)\right]$$

$$d_1 = \frac{\ln(V/X) + \left(r - \delta + (\sigma^2/2)\right)\tau}{\sigma\sqrt{\tau}}$$
$$d_1^B = \frac{\ln(B^2/VX) + \left(r - \delta + (\sigma^2/2)\right)\tau}{\sigma\sqrt{T-t}}$$

(8.9)

The setting of the barrier (higher than, lower than, or at, the liability value is a matter of contention, of which a good discussion is provided by Reisz and Purlich (2007). For example, companies which exhibit a low credit rating, high monitoring costs or long term illiquid assets, may see debt-holders set a barrier above the liability value. However, given the definition of insolvency as being V (T) **<** X, I would not anticipate that a barrier set above liabilities would trigger any insolvency procedure in the UK. I would expect that the position of the barrier would be set at, or below, the value of debt to be serviced. Moody's practitioners Crosbie and Bohn (2003) find that:

"...in general firms do not default when their asset value reaches the book value of their total liabilities. While some firms certainly default at this point, many continue to trade and service their debts. The long-term nature of some of their liabilities provides these firms with some breathing space. We have found that the default point, the asset value at which the firm will default, generally lies somewhere between total liabilities and current, or short-term, liabilities".

The definitions contained within Equation (8.9) are similar to the EC option formula, V_E represents the market value of equity, V is the market value of assets, σ is the asset volatility, r equals the risk-free rate of return and $\tau = (T - t)$ and is the time until the option expires. As with the section 4.5.1 the values V and σ are determined by solving

shareholders should the option expire. As with similar studies, we assume a zero rebate so the term collapses. Note that if both the rebate and the Barrier are set to zero, then the equation further collapses into the non-dividend EC option model.

the call option equation (8.4) and the optimal hedge equation (8.5). Several papers[132] follow Moody's and set liabilities (X) equal to current liabilities plus half long-term debt. Others stick more closely to the BSM theory and use total liabilities for X[133] and I do the same.

Once *B, V* and σ have been backed out, the probability of default can be calculated using Equation (8.10) which represents the actual probability of default rather than the risk neutral probability.[134]

$$\text{Prob} = 1 - \mathrm{N}\left(\frac{\ln\left(\frac{V_t}{X}\right) + \left(\mu - \delta - \frac{1}{2}\sigma_A^2\right)(T-\tau)}{\sigma_A\sqrt{T-\tau}}\right) + \left(\frac{B}{V_t}\right)^{\frac{2(\mu-\delta)}{\sigma_A^2}-1} \cdot N\left(\frac{\ln\left(\frac{B^2}{XV_t}\right) + \left(\mu - \delta - \frac{1}{2}\sigma_A^2\right)(T-\tau)}{\sigma_A\sqrt{T-\tau}}\right) \tag{8.10}$$

For my study, I employ two different versions of the DOC model. The first model (DOC) is constructed according to the method described above. The known input variables for the model are the market value of equity, volatility of equity (standard deviation of daily log returns), Time until expiry (one year), rebate amount (zero), risk-free rate of return (proxied by 12 month UK Treasury bond yields), and annual dividend rate. I then solve for the unknowns *B, V* and σ. Firstly I solve Equations (8.3) and (8.4) to derive my estimates of *V* and σ. Return on assets μ can then be calculated as per Equation (8.5). Second, I determine the size of the barrier as per Equation (8.9). I then apply equation (8.10) to obtain the rick neutral probability of default within one year.

---

[132] For example: Vassalou and Xing (2004), Bharath and Shumway (2008), Reisz and Purlich (2007).

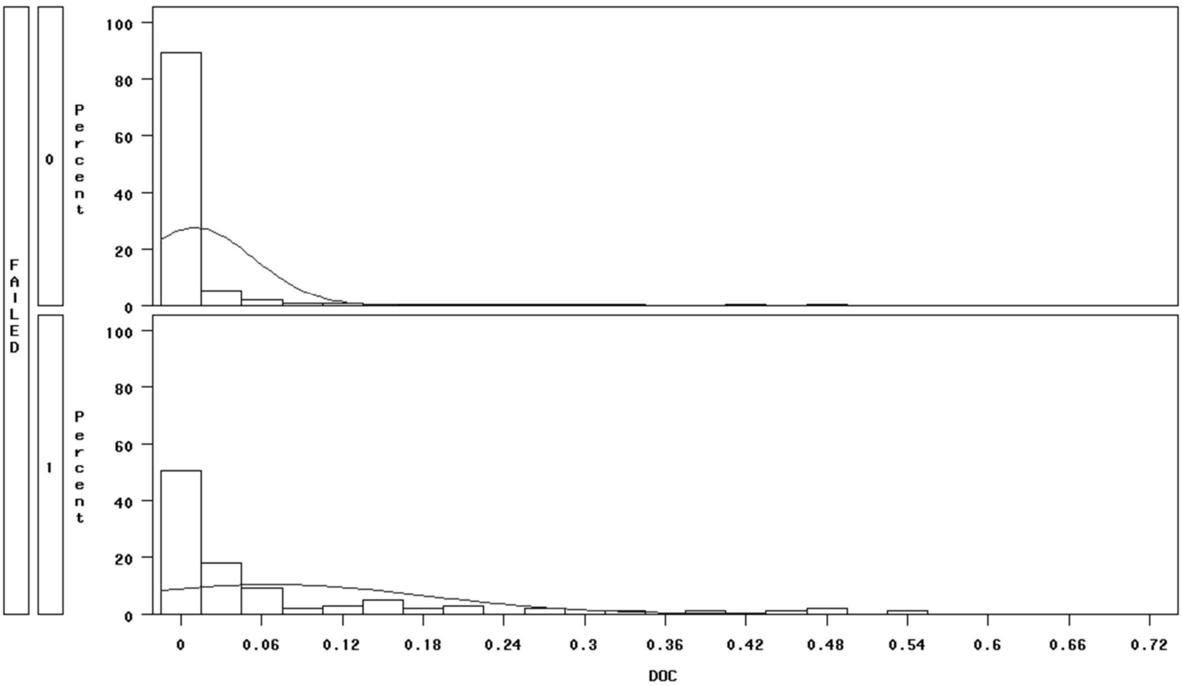
[133] E.g. Hillegeist et.al. (2004).

[134] A concise derivation can be found in Appendix 16.8 which follows Reisz and Purlich (2007) (Appendix A. pp123-129) which in turn follows that of Merton (1974).

Table 8.13: Summary statistics for DOC and its variables

		Mean	Med	SD	Var	Kurt	Skew	Min	Max	N
DOC	FAILED	0.066	0.015	0.116	0.014	6.074	2.490	0.000	0.550	101
	NON-FAILED	0.010	0.000	0.044	0.002	65.772	7.289	0.000	0.731	6494
	t-stat	4.84*								

* Denotes significant difference in means at 5% level
Histogram maps the difference in distribution between failed and non-failed.



8.2.7 A new naive DOC model

It may be argued that a large part of the success/interest which has surrounded the Altman Z-score throughout its existence, as well as its implied predictive ability, is its sheer simplicity. Five variables, a simple linear model, and a proven track record, has established the Altman Z-score as the insolvency prediction model which first comes to mind for the average student or academic. I therefore attempt to simplify the DOC model so that it might more easily be applied across varying practices and disciplines, as a theoretically driven, convenient and practical model which determines a within-one-year probability of failure.

Just as Bharath and Shumway (2008) propose a naive version of the EC option model, I attempt to do the same for the DOC model. I first follow Bharath and Shumway in eliminating the backing out of V and σ by estimating their values as per

Equations (9) and (10)[135]. I then calculate the firm specific barrier, not by solving equation (12), but setting the barrier to the same level as the firm's total liabilities which I feel to be an acceptable proxy given the legal balance sheet insolvency definition as being $V(T) < X$. I also assume a no dividend policy, zero rebate, costless insolvency proceedings, and set the return on assets equal to the risk-free rate.

The naive risk-neutral DOC (JW_DOC) is shown below in Equation (8.11).

$$\text{JW\_DOC} = \mathrm{N}\left[\frac{\ln\left(\frac{X}{V}\right)-\left(r-\frac{1}{2}\sigma^2\right)}{\sigma}\right] + \left(\frac{X}{V}\right)^{\frac{2r}{\sigma^2}-1} N\left[\frac{\ln\left(\frac{X}{V}\right)+\left(r-\frac{1}{2}\sigma^2\right)}{\sigma}\right] \qquad (8.11)$$

where,

X = Book value of total debt

$V = V_E + X$

$\sigma = \frac{V_E}{V}\sigma_E + \frac{X}{V}(0.05 + 0.25\sigma_E)$ (σ_E = SD of daily log returns)

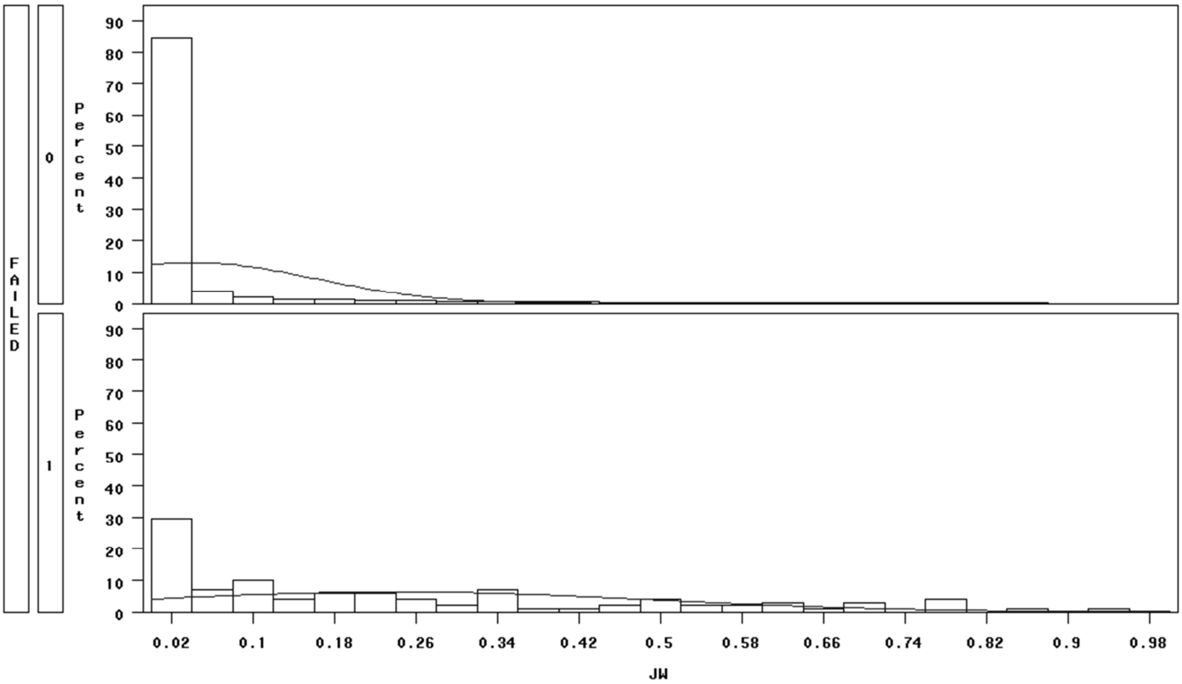
r = risk free rate of return

[135] X is proxied by book value of total debt, V_E is the market value of equity and σ_E is the volatility of equity taken as standard deviation of daily log returns over the previous twelve months.

Table 8.14: Summary statistics for JW_DOC and its variables

		Mean	Med	SD	Var	Kurt	Skew	Min	Max	N
JW_DOC	FAILED	0.240	0.157	0.251	0.063	-0.093	0.992	0.000	0.938	101
	NON-FAILED	0.040	0.000	0.121	0.015	18.486	4.109	0.000	0.986	6494
	t-stat	7.95*								

* Denotes significant difference in means at 5% level
Histogram maps the difference in distribution between failed and non-failed firms.



8.3 Comparing the models

This section provides a discussion of the test procedures used to compare and contrast each of the sixteen models. From the literature it is evident that many different techniques have been used to evaluate model accuracy and performance since the conception of the art, and to describe, discuss and apply each and every technique is not an efficient use of time resources. Some early techniques for model evaluation for example are out-dated and somewhat irrelevant to this study and therefore would make no sense in their application. The literature review of section 5 discussed in some detail several early classification procedures such as the confusion matrix, the Lachenbruch (1967) jack-knife technique, and k-fold cross validation. Although these techniques are still valid and could be correctly applied to this study, they have been improved upon and superseded by the techniques which are used in this study and are now discussed.

8.3.1 Converting scores into probabilities

In order that the continuous Z-score and single ratio models may be directly compared to both the logit and contingent claims probability measures, I follow both Hillegeist et al. (2004) and Agarwal and Taffler (2008) by using the logistic cumulative function $p = e^{z-score}/(1 + e^{z-score})$.[136]

This transformation is not a panacea, as noted by each of the aforementioned studies. Maddala (1983) shows that discriminant scores should be converted into probabilities using a linear probability model after suitable transformation through regression sum of squares. McFadden (1973) shows, however, that the MDA and logit approaches are closely related under normality assumptions and therefore the use of the logistic cumulative transformation is valid.

8.3.2 Receiver operating characteristic (ROC) curves

For the first test of model comparativeness, I assess each model's predictive ability using the ROC curve. Traditional tests for model accuracy have relied on the minimising of type I and type II errors based on an arbitrary cut-off point chosen by the researcher. Table 8.4 shows the form of a typical classification table, similar to that first used by Altman (1968). The problem with this classification method is the single cut-off point. The ROC curve addresses this issue by evaluating the model's performance over the whole range of possible cut-off points.[137]

---

[136] Having reversed the signs for the coefficients in the original MDA models, we do the same for BV and SIZE so that the ratios increase with the probability of failure.

[137] ROC curves have traditionally and most commonly been used in the field of medicine for assessing various diagnostic procedures and treatment techniques.

Table 8.15: The confusion matrix

	ACTUAL		
PREDICTION	Failed	Non-failed	TOTAL
Failed	TP	FP [a]	TP+FP
Non-failed	FN [b]	TN	FN+FP
TOTAL	TP+FN	TN+FP	TP+FP+TN+FN

[a] Type II error
[b] Type I error
Notes: TP (True Positive) is the correct classification of a failed firm. TN (True Negative) is the correct classification of a non-failed firm. A false negative outcome (FN) is predicting a firm to survive when it actually fails (type I error). A false positive outcome (FN) is predicting a firm to fail when it actually survives (type II error).

My construction of ROC curves is as in Gönen (2006). Firm-year observations from each portfolio are ranked from the highest probability of failure to the lowest. For each of 101 possible cut-offs in probability, ranging from 1.00 to 0.00 on steps of 0.01, the number of TP, FP, TN and FN outcomes are be observed. Then for each of the 101 possible cut-offs, the true positive rate (TPR) and the true negative rate (TNR) are deduced. The TPR is calculated as TP / (TP+FN) and is often referred to as the "sensitivity". The TNR is calculated as TN / (TN+FP) and is referred to as the "specificity". The ROC curve is then plotted as sensitivity against 1-specificity (also the false positive rate) as cut off probability varies. The model's overall accuracy is then deemed to be represented by the area under the ROC curve, θ[138]. Three examples of the ROC curve can be found in Figure 8.5 below. The figure depicts two models which have strong and medium accuracies, along with one of a random choice model demonstrated by the 45^0 line. As demonstrated by the example, a model with high or near perfect accuracy will have each of its measurements (for each cut-off) clustered near to the position (0,1) where there is 100% sensitivity and 100% specificity.

[138] Stein (2000) provides an excellent examination and interpretation of the meaning of the area under a ROC curve.

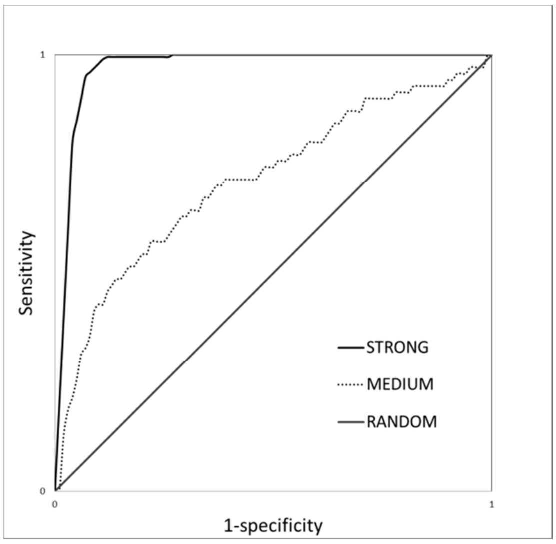


Figure 8.5: ROC examples

The area under the ROC curve is estimated using the trapezoid rule for which Hanley and McNeil (1982) show is “virtually identical” to the Wilcoxon test statistic and an unbiased estimator of the true smoothed area under the curve. As my ROC curves are constructed using percentile values of specificity and, my probability estimates are on a continuous scale, the trapezoid rule for calculating the area is empirically sound. The standard error of the area under the ROC curve $SE(\hat{\theta})$ is calculated following Hanley and McNeil (1982) as per Equation (8.12).

$$SE(\hat{\theta}) = \sqrt{\frac{\hat{\theta}(1-\hat{\theta}) + (n_F - 1)(Q_1 - \hat{\theta}^2) + (n_N - 1)(Q_2 - \hat{\theta}^2)}{n_F n_N}} \tag{8.12}$$

where n_F and n_N are the number of failed and non-failed firms in the sample respectively; Q_1 is the probability that two randomly selected failed firms will both be classified as having a higher probability of failure than a randomly selected non-failed firm ($Q_1 = \hat{\theta} \div (2 - \hat{\theta})$); and Q_2 is the probability that one randomly selected failed firm will be classified as having a higher probability of failure than two randomly selected non-failed firms ($Q_2 = 2\hat{\theta}^2 \div (1 + \hat{\theta})$). The standard error is used to determine the *t*-statistic for any significant difference in the model’s accuracy over a model with no predictive ability.[139]

[139] A model with no predictive accuracy, a random choice model, is represented by a 45degree line on the ROC diagram and therefore has an area of 0.5.

The accuracy ratio, AR, of each model, following Englemann et al. (2003), is a linear transformation of the area under the ROC curve: AR = 2*(θ - 0.5).

8.3.3 Test of economic value

The ROC curve described in the previous section, although a common tool for model comparison, suffers from the problem of ascribing equal error costs which, although convenient for econometric modelling, does not provide a true picture similar to what one might find in a real-world situation. Within a loan market, the cost to a lender of misclassifying a failing firm as being non-failing, granting a loan, and potentially losing 100% of the loan amount, is potentially far greater than the reverse in which a non-failing firm is deemed to be failing and hence the lender forgoes potential interest repayments. Many researchers have neglected to take into account these differing error costs, treat them as being equal, and simply aim to minimize the combined total error rate presumably because in practice, the specification of error costs seems to be difficult and very subjective (Steele, 1995). Therefore if I presume that the misclassification costs are not equal, how does one go about determining a value for these costs?

Differing misclassification costs have been used within insolvency prediction ever since its conception, but not necessarily used with respect to assessing a model's performance. Anderson (1962), Edminster (1972), and Altman (1977) for example used these differing misclassification costs in order to determine an optimal cut-off score for their predictive models. Altman (1993) provides an example of how to determine these costs, a revised approach to that which is found within Altman (1977); used to calculate the optimal cut-off score for the ZETA model as previously described in section 5.2. With regards to costing the type I error, Altman (1977) investigated a small sample of 33 regional banks and 25 large city banks (1969-1975) and determined that the loss incurred by banks with regards to defaulting loans was approximately 69% of the loan value for small banks and a loss of 72% by the major banks upon the loan value. For "computational purposes" Altman took the cost of misclassifying a failed company as being non-failed (type I error) as being 70% of the loan value. A revised value of the type I error appears in Altman (1993) as being 62% of the loan value, which is based on questionnaires returned from 55 commercial banks from the period 1975-1977 with information on 377 loans.

We can observe that over a small number of years, differing information and from slight changes to calculation methodology, the cost of a type I error can indeed be very subjective; shifting 8% in value for the Altman example. More recent studies such as Stein (2005), Blochlinger and Leippold (2006), and Agarwal and Taffler (2008), determine the type I error cost to be much lower than that of Altman's 62% which is not surprising given the decades between the studies. Stein (2005) uses a base loss rate of 50% in his discussion of ROC analysis and loan pricing; Blochlinger and Leippold (2006) use a loss given default (LGD) rate of 40% based on a weighted average of collateralised and unsecured credits gained through their positions with Credit Suisse and the Swiss Banking Institute; Agarwal and Taffler (2008) use an LGD of 45% under the assumption that the banks within their test of economic value follow the Basel II Foundation Internal Ratings-based approach, and that all loans are unsecured senior debt. For comparability, relevance, and desire to follow internationally accepted treatments of LGD, for my analysis of economic value of insolvency prediction models, I also treat the LGD as being 45%.

The cost of classifying a non-failed firm as being failed (type II error) is also open to a degree of subjectivity, although it is perhaps easier to measure in real monetary terms than its type I counterpart. The cost of not making a loan to a non-failing, payment making company, because your model determined the company to fail, can be based on the concept of opportunity costing. Within this concept, the lender will negate to collect the loan repayments from the credit worthy company because it has failed to lend, but the opportunity cost is the rate of interest the lender might have earned if the loan amount was invested elsewhere. Therefore the type II error cost in this instance is the forgone rate of return on the rejected loan, minus the opportunity cost that could be made through investing. The opportunity cost in this situation is likely to be somewhere around the base rate of interest (currently 0.5%) that will fluctuate over time, whereas the forgone loan rate will be firm specific, making an absolute value for a type II error difficult to assess. Altman (1993) approximates this cost at 2%, although I would suggest that in today's financial climate this would perhaps be too high, and a more appropriate value might be around the 1% mark. The model that I use for testing economic value when misclassification costs differ, which I will now go on to describe, does not, however, specifically require a fixed type II error cost. Instead the model, as we shall discover, treats the cost of a type II error as pure loss of business, affecting the lender's profit and return figures in an artificial economy.

The economic value model which I use is based upon the framework defined in Blochlinger and Leippold (2006) (B&L) in which they provide a method to link the accuracy of credit risk models in distinguishing good and bad loans, the pricing of said loans, and the decision making process of both lender and borrower. B&L note that the traditional cut-off regime is "*rather conservative since the bank uses the credit score at an exogenously given risk premium*", i.e. in reality a bank will offer a spread of interest rates – firms with higher perceived default risk will be charged a higher rate of interest than a firm which has a lower perceived default rate. Additionally, firms which are determined to be of too high a risk will not be offered a loan at all. B&L derive a pricing regime which determines the credit risk spread as a function of a firm's credit score (as determined by a particular model). Assuming that the lender will only offer loans which have a positive net present value, and ignoring any strategic value which might be gained through relationship banking, the chargeable rate of interest can be calculated as follows:

$$R(t) = \frac{P(Y = 1|S = t)}{P(Y = 0|S = t)} LGD + k \tag{8.13}$$

Where $R(t)$ is the chargeable interest rate given a credit score t, $P(Y = 1|S = t)$ is the probability that the borrower will become insolvent and default on its loan given a credit score of t, $P(Y = 0|S = t)$ is the probability that the borrower will not default given a credit score t, LGD is the percentage of the loan value lost given borrower default, and k is the interest rate charged to those borrowers perceived to be of the highest quality. B&L round the value of $R(t)$ up to the nearest quarter of one per cent.

With respect to the test itself, I follow B&L and imagine an artificial economy, with a number of banks competing for business. Each bank within the economy prices its loans based upon a particular credit scoring method (e.g. Altman's Z-score) and presents its offer to the potential borrower. A rational borrower will accept the loan from the bank which offers the lowest rate of interest which, with the omission of any relationship banking, provides the most favourable terms. B&L note that: "*When the* *bank derives the price from the credit score, a low-power model results in a non-competitive pricing system* (does not get the customer's business). *Hence, a marginal power improvement may lead to a sizeable profitability increase* (grants more loans because it is more attractive to borrowers)".

The test, therefore, is constructed thusly: I assume that the total value of the loan market is worth £100 billion and that each of the *n* firms (borrowers) within the loan market are seeking an equal loan of value £100bn / n. The economy has within it *m* banks competing for business, each using a different credit scoring model[140]. Each bank refuses loans to those companies which fall within the bottom 5% of their respective credit spread but otherwise will offer a loan at interest rate $R(t)$. The potential borrow will evaluate all the loan offers that are presented, and accept the offer which provides the lowest rate of interest. In the event that two (or more) bank offers are of equal terms, the value of the loan is divided equally between these banks. The scenario may also arise where one (or more) firms are refused loans by all banks, and therefore it is possible that the market share gained by the banks may not sum to one.

Once all of the loan applications have been processed I can determine the profitability and return on assets (ROA) figures generated by each bank, based upon total loan value, interest charged, and loan amounts lost through default. The banks performance will be ranked according to their resulting ROA and thus provide an insight into the economic value for of the insolvency prediction and credit risk models investigated within this thesis, for when misclassification costs differ.

[140] Several scenarios are played out. The first scenario is based upon the original validation sample split through time (2006-2009) and includes all sixteen models. The second scenario incorporates the whole decade (2000-2009) and includes ten models: AZU, AZN, TZU, TZN, OLU, and OLN are dropped from the original collection of sixteen models as to not to allow any bias to enter the test (these six models being estimated using data from the period 2000-2005).

9 Results and findings

I present first, in Sections 9.1 and 9.2, results based on my chronological split between training and validation/testing samples (with training data from portfolios 2000-2005 inclusive, and validation data from portfolio years 2006-2009). This split was performed in order to reflect the practices of construction and use of the models by real-world users. I then proceed into section 9.4 and present findings as to the distributional properties of θ for each model, derived by a 10,000-iteration Monte Carlo approach to training and validation sample splits. Section 9.5 presents the test results for economic value when misclassification costs differ.

9.1 ROC Analysis: summary statistics

The mean predicted probability of firm failure as between failed on non-failed firms from each model is presented in Table 9.1. The table summarises the failure probabilities as defined formerly within the model specification, and are transformed into probabilities where appropriate. All the models give a higher mean predicted probability of failure for the group of actually failed firms that they do for the non-failed firms[141]; and this significant at the 1% level for twelve out of sixteen models - the exceptions being OLU, AZN, TZN and OLN, where there is no significance at generally acceptable levels.

The measure of how closely a model's predicted probability of failure conforms to the actual failure rate is known as calibration. The actual failure rate within my sample is 1.7%, being 62 of 3,592 firm-years in the validation sample. This percentage is much lower than that predicted by many of the accounting number-based models tested here (excluding neural network versions), indicating that these models are not particularly well calibrated.

Far better calibrated in general are the market-based models, with models HKCL and DOC predicting failure in 1.1% and 1.2%, respectively, across all firms; and their naïve counterparts, models BS and JW_DOC, predicting 2.4% and 4.7% respectively.[142]

[141] Albeit this is only apparent in the fifth decimal place for model AZN.

[142] Also, model OLU predicts a 1.5% failure rate across the sample.

The failure probabilities provided by the European call models HKCL and BS are close to (yet marginally higher than) the 0.96% and 2.12%, respectively, reported by Agarwal and Taffler (2008). The single variable SIZE model (itself based on market price data) is reasonably, but no so, well calibrated, predicting failure of 6.6% of firms. Therefore, the financial turbulence and economic recession during my validation sample period does not seem to have had any dramatic impact on the market-based model's failure probability calibration. Stein (2002, pg. 7) suggests that the calibration of a model can be achieved by mapping its score to an empirical probability of default using historical data and adjusting for the difference; without Moody's proprietary database however, this is somewhat problematic. The calibration of the models is not a necessity here, however, because it does not impact on the overall accuracy of the individual model to discriminate between failed and non-failed firms, the ordinal ranking is unaffected, and hence the overall power of a model is unaffected by its calibration.[143]

Table 9.1: Mean predicted probabilities of failure for failed versus non-failed groups

Model	All	Non-Failed	Failed	t-stat[a]
SIZE	0.066	0.063	0.198	6.400*
Book to Market (B/M)	0.738	0.737	0.781	2.385*
Beaver (1966) best performing ratio CF/TD (BV)	0.307	0.305	0.404	3.734*
Altman (1968) Z-score (AZ)	0.116	0.111	0.366	7.956*
Altman Z-score with updated coefficients (AZU)	0.401	0.399	0.520	5.334*
Altman Z-score estimated using NN (AZN) [b]	0.012	0.012	0.012	0.003
Taffler (1983) UK Z-score (TZ)	0.193	0.186	0.571	8.038*
Taffler UK Z-score with updated coefficients (TZU)	0.433	0.431	0.497	3.937*
Taffler UK Z-score estimated using NN (TZN) [b]	0.032	0.031	0.034	0.171
Ohlson (1980) Logit Model (OL)	0.267	0.261	0.615	7.665*
Ohlson Logit model with updated coefficients (OLU)	0.015	0.014	0.040	1.500
Ohlson Logit model estimated using NN (OLN) [b]	0.013	0.013	0.044	1.830
Hillgeist et al. (2004) market model (HKCL)	0.011	0.009	0.067	3.317*
Bharath and Shumway (2008) naive market model (BS)	0.024	0.022	0.115	5.019*
Barrier option as Down-and-Out-Call (DOC)	0.012	0.011	0.069	3.334*
New naïve down-and-out call (JW_DOC)	0.047	0.044	0.221	7.800*

Notes:
* denotes one-tailed significance at the 5% level.
[a] Test of difference in mean predicted probability of failure, non-failed firms versus failed firms.
This table reports the probabilities of failure for each firm observation in my sample at the end of September each year in accordance with the designed portfolio structure. Those firms which fail in the 12 months after this date are deemed failed and those which do not are the non-failed firms.
[b] The relative magnitude of the neural network estimates are not directly comparable to the other models as previously noted.

[143] For a more detailed discussion of power, accuracy, and calibration refer to Stein (2002).

Somewhat surprisingly, the largest significant difference between the two groups of firms comes from the original Taffler Z-score model (TZ), followed closely by the original Altman Z-score (AZ), JW_DOC and Ohlson (OL) models. With respect to these original accounting number-based models, the separation is much wider than that of their updated counterparts (although still highly significant), which is not what one might expect, and as the next section demonstrates, does not necessarily translate into a higher predictive accuracy.

9.2 ROC Analysis: predictive power

The ROC results are presented numerically in Table 9.2 and graphically in Figure 9.1, highlighting several interesting points for discussion. The accuracy ratio of the single variable B/M model is a low outlier at 0.15; with the accuracy ratios of the rest of the models being in the range 0.31 to 0.68. Again, I observe the market-based models performing well in comparison with the accounting number-based models, with models HKCL and DOC at 0.65 and 0.64 respectively; and their naïve versions at 0.68 and 0.66 respectively. The single variable SIZE model, with an accuracy ratio of 0.61, stands above the accounting number-based models, with the exception of OLN whose accuracy ratio is 0.62.

Focusing on the area under the ROC curve, the single variable B/M model is the weakest, having $\theta = 57.7\%$, which is significantly different from 50% at only the 5% level. This is some distance from 68% θ reported for B/M by Agarwal and Taffler (2008), based on earlier UK data. All the other models which I test, however, have θ significant at the 1% level; indicating, broadly, success in my setting of the variables and technologies employed in the prediction of firm failure. The multivariate accounting number-based models, AZU, TZU and OLU (implemented using original variables and technologies, re-estimated on my own training data set) have θ of 68.9%, 65.3% and 78.1% respectively; that is, AZU and TZU have similar predictive accuracy, while OLU outperforms them both. Employing a simple neural network technology allows us to obtain more predictive ability out of the Altman, Taffler and Ohlson variables, which is unsurprising, given that neural networks are less constrained than MDA and logit. The θ values achieved by the multivariate models in the current setting are, however, lower than the θ values reported in earlier studies: Reisz and Purlich (2007) report θ at 78% for the original Altman model, using US data from 1988-2002; Agarwal and Taffler (2008) report θ at 89% for the original Taffler model, using UK data from 1985-2001; and Duan and Shrestha (2011) report θ at 87.6% for the original Altman model and 88.9% for the original Ohlson model.

Ahead of the multivariate accounting number-based models, the best performing of the models in this study are the market-based models. The European call-based model HKCL achieves θ of 82.7%, with its naïve variant, BS, achieving 83.9%. The barrier option-based models achieve similar levels of accuracy, with DOC achieving a θ of 81.8% and its naïve variant, JW_DOC, achieving $\theta = 83.0\%$. My result for the HKCL model is similar to the 84% reported by Agarwal and Taffler (2008); but is much higher than the 71% reported by Reisz and Purlich (2007). As regards model BS, Agarwal and Taffler (2008) report a θ of 87%, which, consistent with my findings, places it superior to the more sophisticated HKCL. Of the remaining single variable models, SIZE, with $\theta = 80.6\%$, outperforms the multivariate accounting number based models (indeed, it is in the league of the contingent claims-based models). Beaver's best performing ratio, BV, at 45 years old, has $\theta = 67.0\%$ - which places it alongside the multivariate accounting number-based models AZU and TZU in terms of predictive accuracy.

With regards to the original accounting models, the original Altman model provides the highest accuracy of all the original accounting-number based models at approximately 79%, closely followed by the original Ohlson model (76.3%) and the original Taffler model (74.9%), the only model here which was originally formulated from UK data. Beaver's single ratio CF/TD (67%) provides a reasonable accuracy compared to several of the other models, but is towards the low end of the spectrum. These numbers are in contrast to the original author's claims of 87% (Beaver), 83% (Altman), 85% (Taffler), and 85% (Ohlson).[144]

[144] With the exception of the Ohlson model, each percentage is taken from out-of-sample (holdout) tests and is the overall error percentage given a particular cut-off point. Ohlson does not provide an out-of-sample test as he was not concerned with providing a "precise evaluation of a predictive model". The accuracy of the Taffler model was taken from Agarwal and Taffler (2007) due to its larger sample size.

With reference to the discriminant models of Altman and Taffler, one can observe that the original models outperform their updated counterparts. This result is not what I would expect given the age of the original formulations, and one would rather expect the updated models to be superior in accuracy. This counter-intuitive result can also be found in Altman et al. (1977) who attribute the anomaly to smaller within-group variances of the later sample and hence, making the discriminant separation less efficient. Section 9.3 deals with this in more detail and attributes this anomaly to outliers and high variances within the training sample of data, something detrimental to the MDA process and not prevalent in the carefully constructed sample selection of the original models. This anomaly, as I will later explain in more detail, is not one which causes a great deal of concern.

Table 9.2: Area under curve (θ) and accuracy ratio (AR)

Model	AUC (θ)	SE(θ)	Z	AR
SIZE	80.6%	0.034	9.097	0.61
Book to Market (B/M)	57.7%	0.038	2.004	0.15
Beaver (1966) best performing ratio CF/TD (BV)	67.0%	0.038	4.461	0.34
Altman (1968) Z-score (AZ)	78.7%	0.035	8.291	0.57
Altman Z-score with updated coefficients (AZU)	68.9%	0.038	4.994	0.38
Altman Z-score estimated using NN (AZN)	73.1%	0.037	6.293	0.46
Taffler (1983) UK Z-score (TZ)	74.9%	0.036	6.882	0.50
Taffler UK Z-score with updated coefficients (TZU)	65.3%	0.038	3.992	0.31
Taffler UK Z-score estimated using NN (TZN)	72.4%	0.037	6.046	0.45
Ohlson (1980) Logit Model (OL)	76.3%	0.036	7.364	0.53
Ohlson Logit model with updated coefficients (OLU)	78.1%	0.035	8.048	0.56
Ohlson Logit model estimated using NN (OLN)	80.9%	0.033	9.242	0.62
Hillgeist et al. (2004) market model (HKCL)	82.7%	0.032	10.076	0.65
Bharath and Shumway (2008) naive market model (BS)	83.9%	0.032	10.733	0.68
Barrier option as Down-and-Out-Call (DOC)	81.8%	0.033	9.655	0.64
New naïve down-and-out call (JW_DOC)	83.0%	0.032	10.267	0.66

The probability of firm failure is estimated using ten accounting number-based models (BV, AZ, AZU, AZN, TZ, TZU, TZN, OL, OLU and OLN) and six market based models (SIZE, B/M, HKCL, BS, DOC and JW_DOC). The area under the ROC curve, AUC(θ), represents a model's overall accuracy. AUC(θ) is calculated using the trapezoid method which Hanley and McNeil (1982) show to be equal to the Wilcoxon test statistic. The standard error, SE(θ), is also calculated following Hanley and McNeil (1982). The accuracy ratio (AR), following Englemann et al. (2003), is a linear transformation of the area under the ROC curve: AR = 2*(θ - 0.5). The Z column contains shows the test statistic on deviation from a null hypothesis that the area under the curve = 0.5, i.e., no predictive ability.

An issue pertaining to the accounting number-based models was the adoption of International Financial Reporting Standards (IFRSs) in the UK from January 2005. Accounting standard changes will certainly impact on the predictive ability of an accounting number-based failure prediction model. As discussed above, this sub-section has reported results based on testing/validation of each of the models using post-IFRS data, but with training/estimation of certain models performed using predominantly pre-IFRS accounting data – the empirical method being designed to: (i) use up-to-date data; (ii) include sufficient failed firms to allow for sound analysis; and (iii) replicate use in practice, by estimating models (where required) on a sample period preceding model application. Avoidance of the transition to IFRS is, for the time being, not realistic; and a study based wholly in the post- IFRS era might be an area for future research. In the following sub-section, however, I consider the distributional of properties of θ for each model, derived by a 10,000-iteration Monte Carlo approach to training and validation sample splits – without any chronological constraint on the membership of the training and validation samples, i.e., allowing pre- and post-IFRS data in both training and validation samples.

With regards to the neural networks, when I run the less restrictive feedforward neural network with backpropagation training algorithm using the variables described in the original model as the input variables to the neural network, additional accuracy is achieved over the original technology. The increased accuracy is between 3 and 7 percentage points overall, depending on the model type. Given the additional mathematical complexity of the models, this result is unsurprising, but it is interesting to see that the MDA models are able to be improved a great deal more than the logit model.

The results obtained by using single variables as predictors are strikingly varied. B/M ratio does not provide any predictive ability at all with an AUC of just 57.7%, a very slight improvement on a random model. In contrast to this I find that using the SIZE variable I am able to predict insolvency quite well with an accuracy of 80.6%. Interestingly this accuracy is higher than that of almost all the accounting-number based models and is only out-performed by the Ohlson-based model (OLN) , and also each of the contingent claims models. The 45 year-old single ratio of Beaver shows a reasonable predictive accuracy of 67%, a score towards the bottom end of the sample, indicating that insolvency prediction has been as substantial as the literature might imply.

Finally, I provide evidence that using contingent claims models for insolvency prediction in the current UK climate would be preferred to traditional accounting models. Each of the four models has a higher accuracy than their accounting-number counterparts.

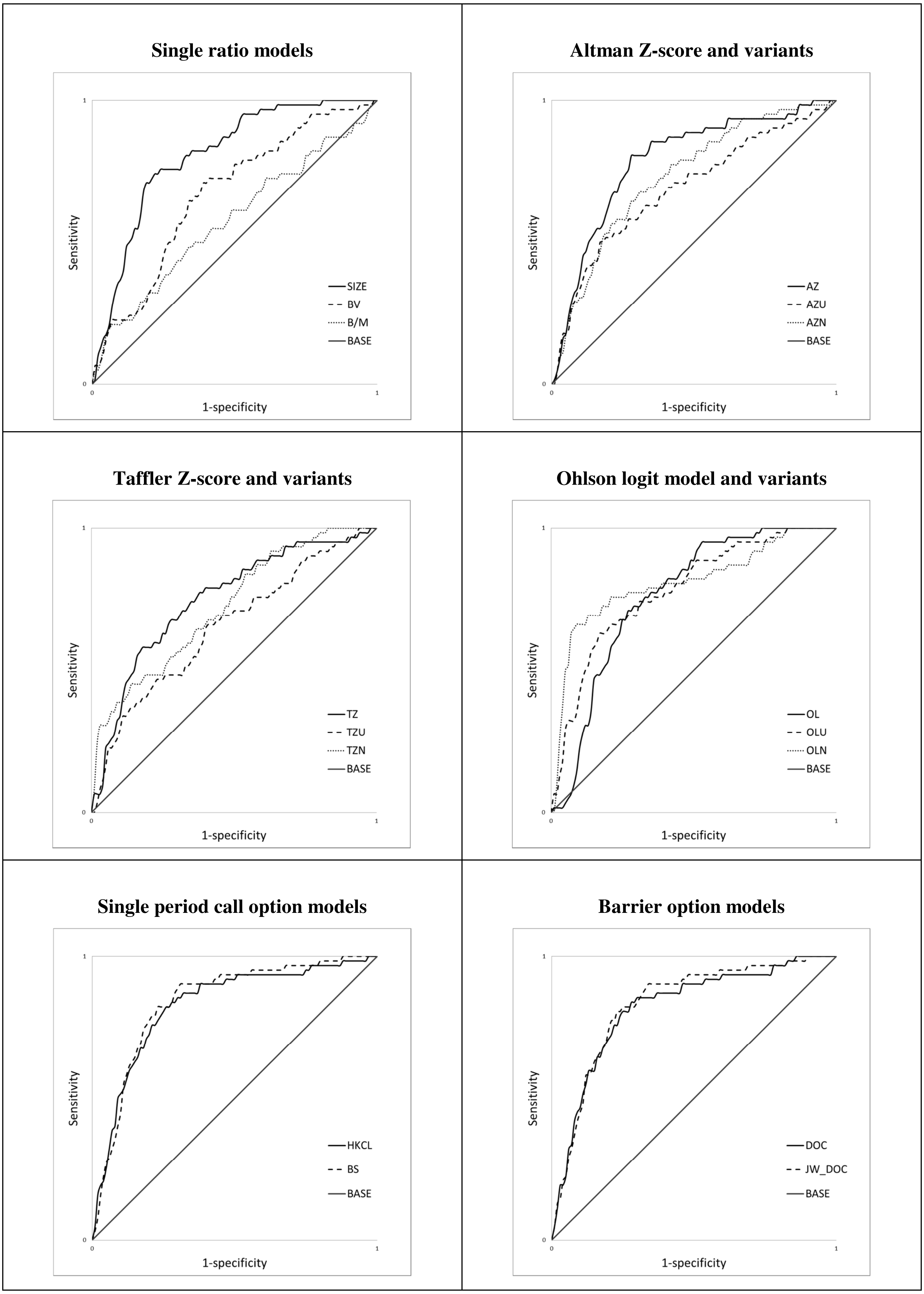


Figure 9.1: Receiver operating characteristic (ROC) curves.

The 45-degree line (BASE) represents a model with no predictive accuracy, i.e., a random choice model with enveloped area equal to 50%.

9.3 The effects of outliers on discriminant function accuracy

Under ROC analysis, updating the coefficients with the original technologies (MDA and logit) I am surprisingly only able to increase the accuracy of Ohlson's logit model; I am unable to increase the accuracy for the Altman and Taffler models using discriminant analysis to re-estimate the coefficients. This result is counter-intuitive and contrary to what I would expect, given the age of the data employed to construct their original forms, and the number of firms within the original samples.

An initial explanation could simply be that that accounting data is no longer sufficient (at least employing these particular variables) to be able to discriminate between failed and non-failed firms, or at least not as apparent as perhaps they once were. Insolvency in today's climate and the onset of corporate failure may be a swifter process than in the past, due perhaps to the less complex administration process which is available in the UK today. If this were the case however, it does not explain why the original Altman and Taffler models *are* able to make this distinction to greater effect, and therefore this explanation falls flat.

A further explanation may lie in my sample design itself, and perhaps the mix of IFRS and non-IFRS populations has had a greater impact than I might have anticipated. This problem is circumnavigated in the next section by randomly selecting training and validation samples and applying the technologies 10,000 times.

There are several assumptions/design fundamentals of the MDA process which are also potential causes of this strange phenomenon, the first relating to sample size. The original models were constructed using small samples of a matched-pair nature, carefully selected by the authors. Within my sample of firms, because failure is such a relatively rare event, a matched sample approach would severely reduce the sample size of this study, eliminating the majority of non-failed firms. It is generally accepted that conducting discriminant analysis with unequal sample sizes is acceptable (Poulson and French, 2004; Tabachnick, 1989), and therefore unequal sample sizes is unlikely to be the cause of the poor performance of the MDA process. The only constraints with respect to sample sizes are that the smallest group needs to exceed the number of predictor variables, and that the smallest sample size should be at least 20 but preferably at least 5 times that of the number of predictor variables. The number of variables is 5 for the Altman model, and 4 for the Taffler model, therefore my training sample of 39 failures is sufficient.

The assumption of multivariate normality has been discussed at length in section 5 of this thesis. Lachenbruch et al. (1973), and Altman et al. (1977), find that linear discriminant models are not compromised in terms of their predictive ability when the assumption of multivariate normality is overlooked and compare favourably to their quadratic counterparts. Poulson and French (2004) note that violations of the normality assumption are not "fatal" and the resultant significance tests are still reliable as long as non-normality is caused by skewness and not outliers (Tabachnick and Fidell, 1996).

This therefore leads me to what I suspect is the cause of the poor performance of the models updated using the MDA process (AZU and TZU) as compared to the original models with the original coefficients (AZ and TZ) – outliers. The overall significance tests within the MDA process are based upon pooled variances (the average variance across both the failed and non-failed groups). Therefore, the significance tests of the variables which provide the largest means along with the largest variances would be based upon and calculated using the variances of the other variables with the relatively smaller variances. This would therefore result in poor estimates and inferences regarding the significance and value of the coefficients variables with larger distributions of data.

Table 9.3: Descriptive statistics for AZ and TZ

Altman		Mean	Med	SD	Var	Kurt	Skew	Min	Max	N
Variable	WC/TA	0.10	0.10	0.49	0.24	1282	-27.3	-25.88	0.99	6595
	RE/TA	-1.65	0.09	91.48	$\boxed{8369}$	6568	-80.9	-7422.3	5.36	6595
	EBIT/TA	-0.05	0.06	0.54	0.29	315	-14.0	-16.66	1.72	6595
	VE/TL	261.7	4.64	4264.2	$\boxed{18184155}$	3120	51.18	0.00	282457	6595
	S/TA	1.17	0.97	1.49	2.22	481	17.13	0.00	51.15	6595
Taffler		**Mean**	**Med**	**SD**	**Var**	**Kurt**	**Skew**	**Min**	**Max**	**N**
Variable	PBT/CL	-0.12	0.164	2.609	6.806	2159	32.91	-37.984	159.76	6595
	CA/TL	66.46	2.453	828.37	$\boxed{686199}$	2633	45.18	0.001	53027	6595
	CL/TA	0.372	0.324	0.490	0.240	1408	29.79	0.003	26.88	6595
	NCI	130.5	-3.94	3498	$\boxed{12241633}$	1682	39.54	-16891	155125	6595

Table 9.3 provides summary statistics for the variables used within the two discriminant models. It can be seen that severe outliers in four out of the nine variables create exceptionally high variances within the data set. This is therefore a possible explanation for the poor performance of the updated discriminant models AZU and TZU.

To investigate this further, I designed some code which would enable me to compare AZ to AZU, and likewise TZ to TZU, as the number of outliers are gradually reduced. The code was designed to automatically Winsorise the data with 0.2% decrements from 100% Winsorisation (no change in data) to 50% Winorisation (the top 25% of scores are set to the 25th percentile, and the bottom 25% of score are set to the 25th percentile)[145]. At each of these stages the coefficients were updated for AZU and TZU and then AUC was calculated for each of the four models. The Winsorisation was conducted upon the training sample of firms and the validation sample was left unchanged. By lowering the variances of the variables this way, I am able to determine what effect (if any) outliers have on the predictive accuracy of the resulting discriminant function. Figures 9.2 and 9.3 show the plots of how AUC changes as the outliers are gradually reduced through Winsorisation.

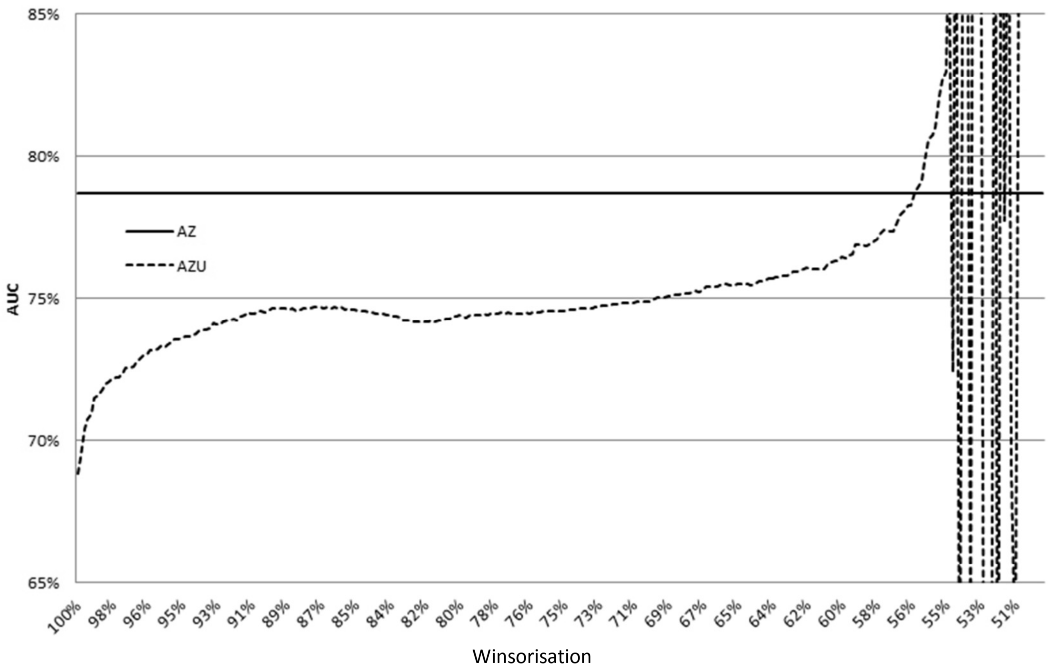


Figure 9.2: The effects of Winsorisation on the Altman model

[145] Beyond 50% Winsorisation the models begin to fall apart.

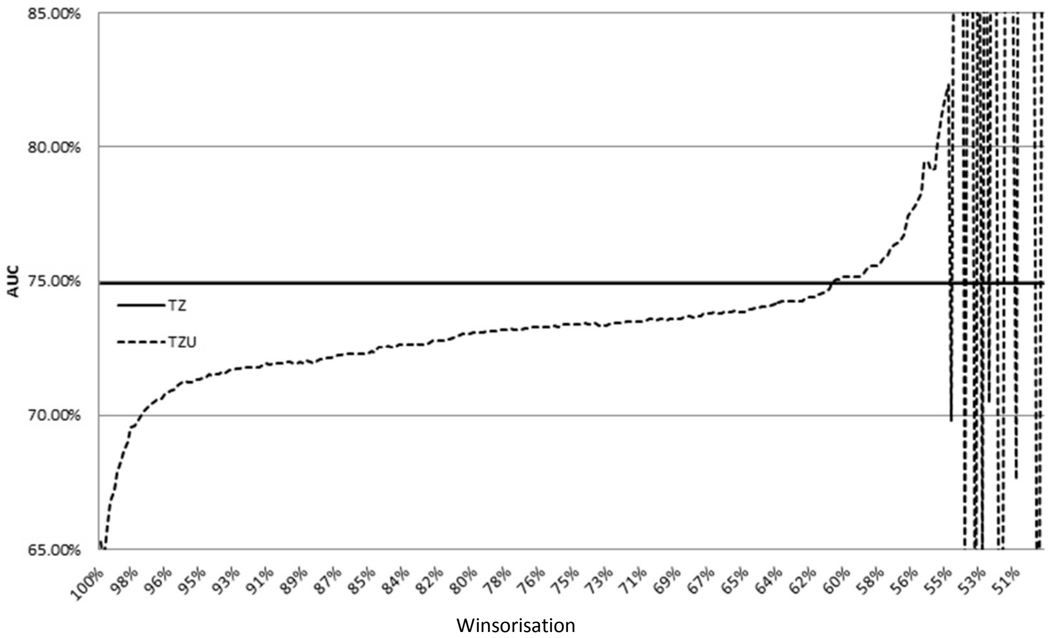


Figure 9.3: The effects of Winsorisation on the Taffler model

The solid lines in each Figure represent the models with their original coefficients, and because the validation sample remains unchanged, so do the predictive accuracies of the models. The AZU and TZU plots begin at 100% Winsorisation (the original AZU and TZU models as derived through the chronological splitting of firms into training and validation, i.e. with no removal of outliers). As I move across the graphs from left to right, the data within the training sample is progressively Winsorised using 0.2% intervals. The graphs show initially, for both models, that there is a sharp increase in model accuracy over the first 3% of Windorisation. This rapid increase in accuracy is what I would expect given the removal of the most extreme outliers, which consequently reduces the variance of the independent variables and allows the discriminant function to be calculated more efficiently.

After this initial increase in predictive power, the scale of this increase deteriorates but in both cases the accuracies steadily approach that of the original coefficients. For the Altman model this occurs at approximately 56% Winsorisation, and for the Taffler model this occurs at approximately 60% (i.e. 60% of the data remains unchanged). After this crossover point, accuracy is accelerated for several percent until both of the models break down at around 54% Winsorisation, as the data becomes too uniform for the discriminant process to operate effectively.

Figures 9.2 and 9.3, clearly demonstrate what effect outliers can have on the discriminant process and to what extent the careful selection of firms within the samples has on the overall accuracy of the predictive model. Both of the original models were formulated using very small numbers of carefully selected firms with variables falling within pre-specified ranges. For example Altman (1968) limited his sample to 66 manufacturing firms with asset sizes below $25 million, whilst Taffler (1984) excludes firms which appeared to be "less than healthy" from the non-failed sample of 23 firms. Peel and Peel (1987) note that "*This type of a priori screening process appears to go somewhat further than merely excluding loss-making firms*", with reference to Taffler (1984) eliminating weaker firms from the non-failed sample. Therefore the removal of outliers has the potential to substantially increase the overall accuracy of the predictive discriminant model, and a carefully selected sample is likely to generate more positive results than one chosen randomly from the population.

An interesting observation within this sub-study is the coefficient values of the updated models at the point in which the original and updated models provide equal AUC's. Table 9.4 shows the comparison of model coefficients at the point of intersection.

Table 9.4: Comparison of coefficients at the point of intersection

Altman (1968) Model	θ (%)	WC/TA	RE/TA	EBIT/TA	V_E/TL	S/TA
Original Coefficients (AZ)	78.7	1.2	1.4	3.3	0.6	0.99
Updated Coefficients (AZU)	78.7	5.74	2.43	13.80	0.09	-0.32
Taffler (1984)		PBT/CL	CA/TL	CL/TA	NCI	
Original Coefficients (TZ)	74.9	12.18	2.5	-10.68	0.029	
Updated Coefficients (TZU)	74.9	18.44	1.29	-14.31	0.037	

With respect to the Taffler model, it may be observed that the coefficient estimates for updated model are remarkably similar to those of the original model formulated upon 40-year-old UK data with a much smaller sample. The magnitudes of the coefficient estimates are similar and their signs are identical. It is, however, difficult to interpret this result, given that the same exacting cohesion cannot be seen within the Altman model. The coefficient estimates at the point of intersection for the Altman models (AZ and AZU) are fairly close in terms of magnitude and all of the signs are identical except for the variable S/TA in which the sign is reversed. For this particular variable, increasing the extent of Winsorisation lowers the value of the coefficient. If I refer back to the formulation of AZU with no changes to the data, this coefficient estimate was 0.39. This value falls steadily as I increase the number of outliers which are Winsorised, switching signs in the process.

To conclude this sub-section I have shown that the removal of outliers can have a dramatic effect on the accuracy and power of a prediction model. By systematically removing outliers from the training data, there is a convergence between the accuracy of the original model and that of the updated version. There are wide differences between my data set and those of the original authors, both in terms of sample size, type of firm, the decades between them, and the changes in reporting practices which have occurred in between. At the point of convergence, the coefficients of the original and updated models are remarkably similar, which is surprising, but difficult to interpret this result. Research upon different samples of firms should be open to future analysis to determine whether my results go some way to defining an optimal set of coefficients for these variables (in which the original authors may have already defined), or whether this result is mere coincidence. I leave this investigation to future research.

9.4 Distributional properties of models using repeated random sub-sampling.

The previous ROC analysis was based upon a real-world scenario where a lender would have access to information in the past (the training sample) and apply it to loan decisions in the future (the validation sample). Although this may make intuitive sense, it may not provide a true or fair evaluation of the predictive accuracy of each model over the entire sample period. Therefore I investigate whether the results obtained in the previous section are representative and significantly indifferent to that which I might obtain through differing sample splits. In order to determine whether or not these previously obtained results are consistent throughout the sample period, it seemed appropriate to apply Monte Carlo simulation over different sample splits to determine the distributional properties of each of the models ROC accuracy.

This approach is perhaps not too different to cross validation (where the training and validation samples are “switched” and the average of the two results are taken), or k-fold cross validation (where an average of k number of training samples are used to generate results from the remainder of the sample). What does differ here is the scale of the operation (10,000 repetitions) and its application to ROC accuracies, something that has not been seen in the literature before.

In order to achieve this, each firm-year observation is assigned an evenly distributed random real number greater than or equal to 0 and less than 1. The observations are ordered lowest to highest. The first 50% of failed firms (with the lowest randomly generated numbers) and the first 50% of the non-failed firms are selected and transferred to the training sample with the remaining failed firms and non-failed firms being transferred to the validation sample.

Where applicable (AZU, TZU, OLU), model estimation is performed on the training sample which providing parameter estimates for each variable used in each of the three models. These parameter estimates are then applied to the firms within the validation sample resulting in a failure probability for each firm year observation. For the remaining models (those which do not require parameter estimation), ROC analysis is simple conducted upon the randomly generated validation sample and the model's accuracy is then determined by the area under a receiver operating characteristics curve (AUC) as described in the previous section. This process is repeated multiple (10,000) times in order to investigate the distributional properties of the AUC for each of thirteen models. I also plot the 10,000 ROC curves for each model, to obtain a visual interpretation of its accuracy for which I have termed distributional feathers.

The reader may notice the reduction in model numbers at this point, dropping from sixteen to thirteen. The reason for this reduction is the loss (or non-inclusion) of the three Neural Network models from the sample which requires some explanation. The primary reason for the loss is time limitation; each of the 10,000 repetitions, and the whole of this analysis, is based upon custom Visual Basic macros for which I have had to write myself. I could not find a way of generating the required results using SAS (as my preferred statistical package), and the time spent investigating and developing a process, which might not even be possible, was deemed to be better spent writing the procedure in Visual Basic for which I am more comfortable, being certain that the goal could be achieved. The resulting macro procedures incorporated the random sample generation, the Discriminant Analysis and Logit coefficient generation (where appropriate), the construction of the ROC curve, and the calculation of the various statistics required for each run.

Each run varied in its duration of calculation depending upon each model's requirement for estimation, and the computer upon which it would run. For models that did not require any re-estimation (SIZE, BV, B/M, AZ, TZ, OL, BS, HKCL, DOC, and JW_DOC) the typical run would take anywhere between 7-10 seconds, meaning that each model would usually require somewhere between 24 and 28 hours to generate 10,000 repetitions.

For models that *did* require re-estimation, the relative timeframe required to generate 10,000 runs was dramatically increased. The estimation of OLU required the necessity to invoke the solver function to complete the maximum likelihood estimation, and generate a set of new coefficients. One complete logit run cycle would typically take between 2-3 minutes on my Dell Latitude E6400 laptop[146] which was left to number-crunch in the corner of the room. For 10,000 repetitions this process took around 20 days. The MDA estimation of AZU and TZU increased this timescale even further. Due to the nature of the estimation process using matrix algebra upon a large data set, the aforementioned laptop could not manage the estimation process on a dataset larger than 2000 firms due to its limitation in specifications, and even a limited sample size would result in a 10 minute run window. Fortunately one of my home computers has a little more performance and was able to manage the task more effectively[147] although the average run time was still a large 4 minutes. These two sets of results took approximately 2 months to generate. In hindsight it would be fair to say that an alternative method would have been advantageous, and that a more advanced knowledge of statistical software may have produced these results in a shorter time period.

This then leads me back to the original explanation of why the NN models have not been included in this analysis. The first problem is incorporating an existing (or designing from scratch) a NN algorithm macro which would fit neatly into my existing VBA program. Although not outside the realms of possibility, early attempts have not

[146] 2GB RAM, 160GB HDD and Intel Core Duo P8400 running Windows Vista 32bit OS.

[147] 16GB RAM 1TB HDD and AMD PhenomII X6 1055T running Windows 7 64bit OS.

been successful and time constraints have seen my time spent on other areas of this thesis. I also envisage that even if I had the NN/ROC macro working at this present moment in time, it would not be any faster than the other models which required re-estimation and therefore would be entirely possible that it might take another 3 months to generate another 3 sets of results for AZN, TZN, and OLN, which is time I currently do not have. It would therefore, I think, be more beneficial to leave these 3 models for future research, where a more efficient generation process could be explored and developed. I now return the attention to that of the results which I have generated for the remaining thirteen models.

Summary statistics regarding θ for the remaining thirteen insolvency prediction models are set out in Table 9.5 along with distribution histograms[148] and ROC "feathers" depicted in Figure 9.4; the relative frequency distributions of θ for each of these models are superimposed graphically in Figure 9.5. Table 9.6 compares the average accuracies of the distribution exercise to those results gained through the original training and validation samples (time-split).

Results as to the efficacy of the models as measured by the mean of θ are in agreement with the results presented in the previous section. The ordering of model efficacy by mean θ from Table 9.5 is almost identical to that according to the original θ from Table 9.2. Further, each θ from Table 9.2 is close to the corresponding mean θ also depicted within Table 9.65.

[148] Each histogram is overlaid with a normal curve based upon the models mean and standard deviation. Tests for normality suggest that the actual distributions for the models are not significantly normal, although the overlaid plot demonstrates that this is very difficult to distinguish by eye.

Table 9.5: Distributional properties[a] of model predictive accuracy, as measure by area under receiver operating characteristic curves (θ).

Model [model designation]	Mean	Median	SD	Skewness	Excess kurtosis[b]	Minimum	Maximum	Jarque-Bera ~ $\chi^2(2)$
Single variable approaches								
SIZE	80.5%	80.5%	0.0149	0.0403	-0.1327	75.1%	86.3%	10.04**
Book to Market [B/M]	58.3%	58.3%	0.0260	0.0391	-0.0314	49.1%	68.8%	2.96
Beaver (1966) best performing ratio CF/TD [BV]	69.5%	69.5%	0.0195	0.0472	-0.0823	62.9%	76.7%	6.53*
Altman Z-score [AZ]	80.7%	80.7%	0.0171	0.0788	-0.0277	75.0%	87.6%	10.67**
Altman Z-score with updated coefficients [AZU]	73.5%	73.6%	0.0256	-0.1206	0.0979	60.4%	83.0%	28.25**
Taffler UK Z-score [TZ]	77.9%	77.9%	0.0184	0.0819	-0.0728	71.8%	85.7%	13.39**
Taffler UK Z-score with updated coefficients [TZU]	69.4%	69.7%	0.0346	-0.3681	-0.1861	54.2%	79.3%	240.24**
Ohlson logit model [OL]	78.4%	78.4%	0.0125	0.0652	-0.0533	73.7%	82.9%	8.27*
Ohlson logit model with updated coefficients [OLU]	78.6%	78.7%	0.0207	-0.2298	0.2614	68.6%	87.1%	116.51**
Hillegeist et al. (2004) market model [HKCL]	83.2%	83.1%	0.0165	0.1116	-0.1266	77.8%	89.0%	27.43**
Bharath and Shumway (2008) naïve market model [BS]	86.1%	86.0%	0.0131	0.1103	-0.1995	81.9%	91.0%	36.84**
Barrier option as down-and-out call [DOC]	81.5%	81.5%	0.0167	0.0648	-0.0605	75.4%	87.8%	8.52*
New naïve down-and-out call [JW_DOC]	85.0%	85.0%	0.0138	0.1319	-0.1511	80.5%	89.9%	38.52**

[a] Based on 10,000 random splits between training and validation samples. In each case the number of both failed firm-year observations and non-failed firm year observations in the training and validation samples are as in Table 7.2; but the allocations now being made on a random basis, rather than according to chronology.
[b] Excess kurtosis is the unbiased measure of deviation of kurtosis from 3 (3 being the kurtosis of a normal distribution).
**, * denote one-tailed significance at the 1% level and 5% level respectively

Also it is shown that there are no differences between Table 9.2 and corresponding Table 9.5 mean θ significant at the 5% level as presented in Table 9.6. The results of section 9.2 are not, therefore, outliers or atypical in the context of the distributions of possible θs.

The θ distributions of all of the models show a reasonable degree of regularity and symmetry; but Jarque-Bera testing of the distributions finds that a null of normality may be rejected for nine of the thirteen models at the 1% level of significance. The exceptions are BV, OL and DOC, where normality may be rejected only at the 5% level of significance; and B/M, where normality may not be rejected at any generally acceptable level of significance. It is interesting to note that model B/M, which has the most 'normal-like' of the distributions, is by far the least efficient of the models tested.

Concerning the ordering of the different models by mean θ, each model's mean θ is different from the mean(s) of the model(s) of neighbouring efficacy with significance at the 1% level. Whilst the means of the distributions ordered by efficacy are significantly different one to the next, it is clear from Figure 9.4 that there is considerable overlap in the distributions – and that a particular sample split into training and validation samples might suggest an alternative ordering. This leads to discussion of the spread of the distributions.

The less spread in a distribution, the less sensitive the model's efficacy to variation in the composition of training and validation/test samples. Model TZU's θ distribution shows the highest spread, with a standard deviation of 0.0346. Models B/M, AZU, OLU and BV have distributions with the next highest standard deviations (being 0.0260, 0.0256, 0.0207 and 0.0195 respectively). In this respect, superior again are the market-based models: BS, JW_DOC, SIZE, HKCL and DOC, with their distributions of θ having standard deviations of 0.0131, 0.0138, 0.0149, 0.0165 and 0.0167 respectively. The relatively high standard deviations of θ distributions for the accounting number-based multivariate models AZU, TZU and OLU is consistent with, indeed may explain, the variety of efficiency claims for these models as between different studies.

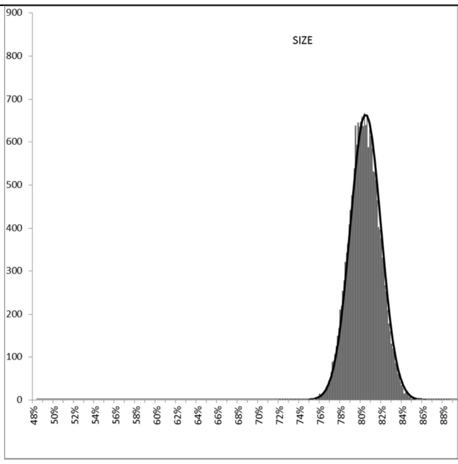
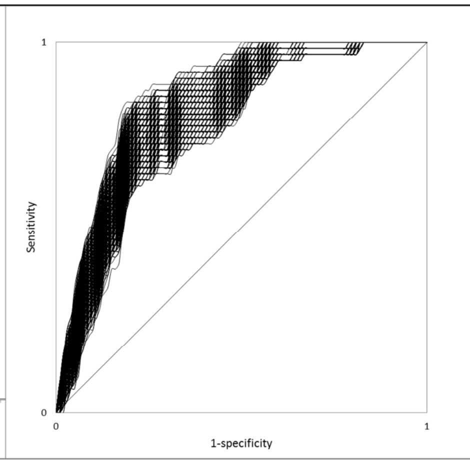
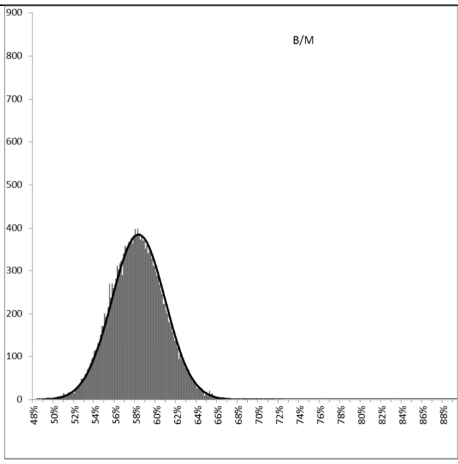
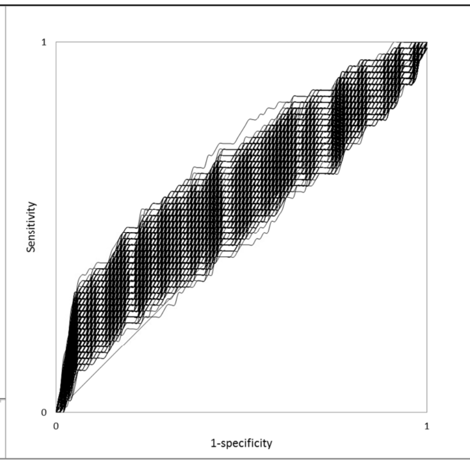
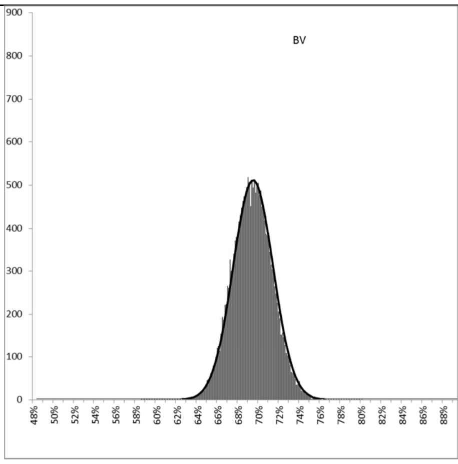
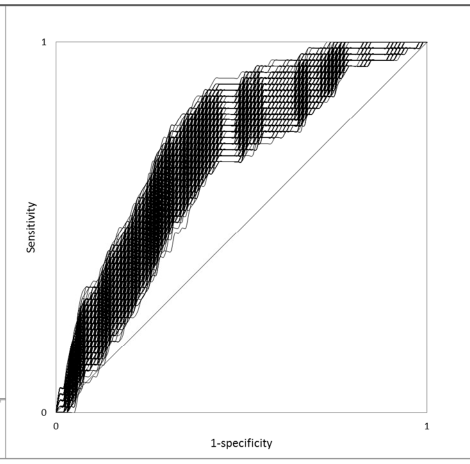
Model	Distribution	ROC plots
SIZE		
B/M		
BV		

Figure 9.4: Distributional properties of model predictive accuracy, as measured by area under receiver operating characteristic curves (θ)

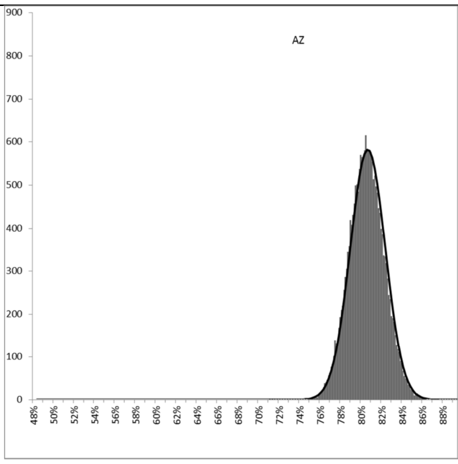
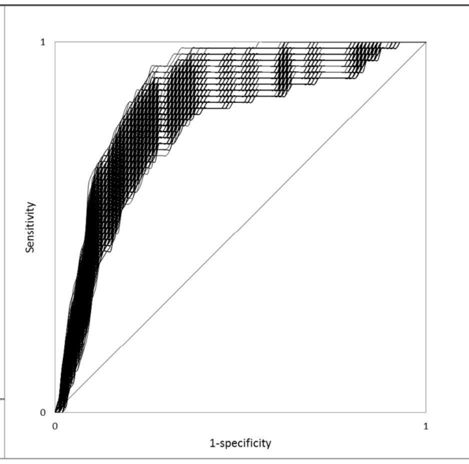
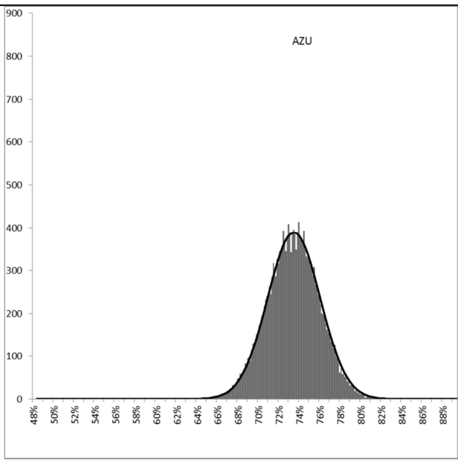
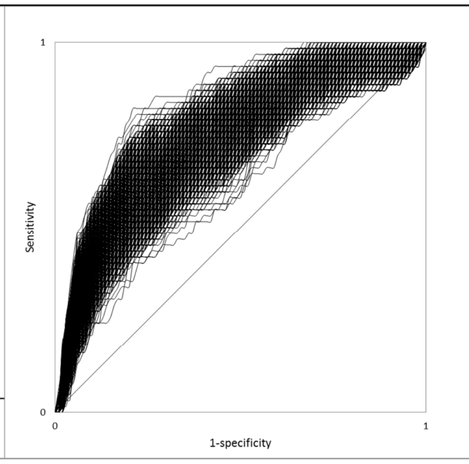
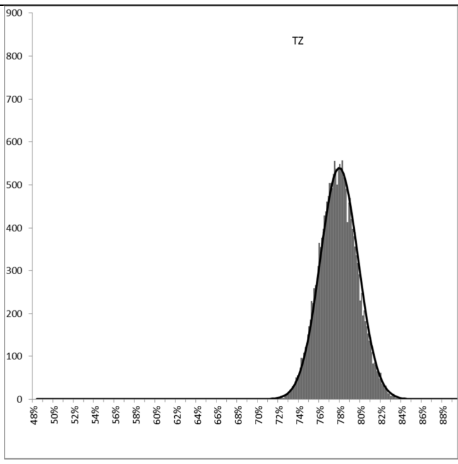
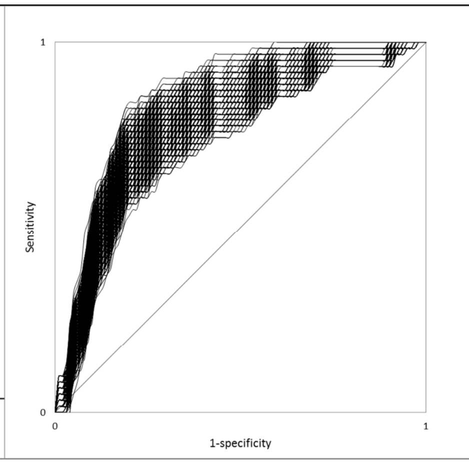
Model	Distribution	ROC plots
AZ		
AZU		
TZ		

Figure 9.4 Continued: Distributional properties of model predictive accuracy, as measured by area under receiver operating characteristic curves (θ)

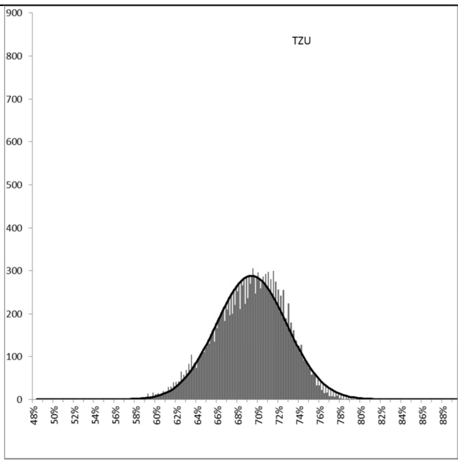
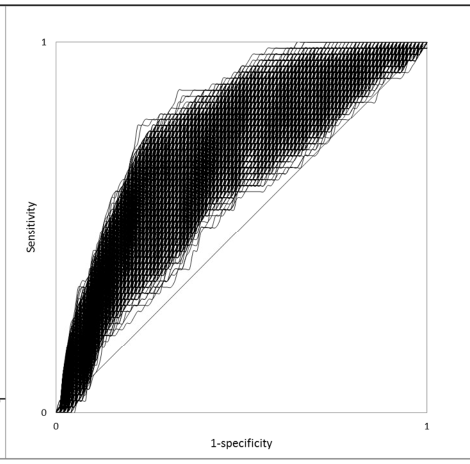
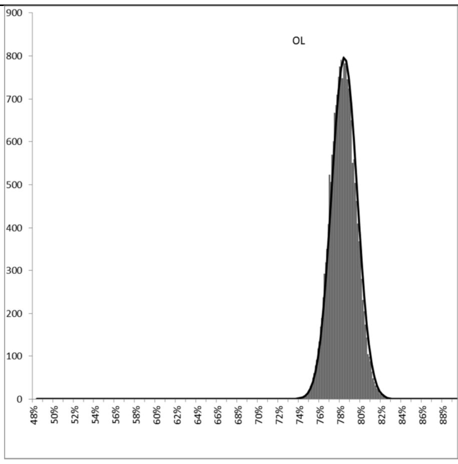
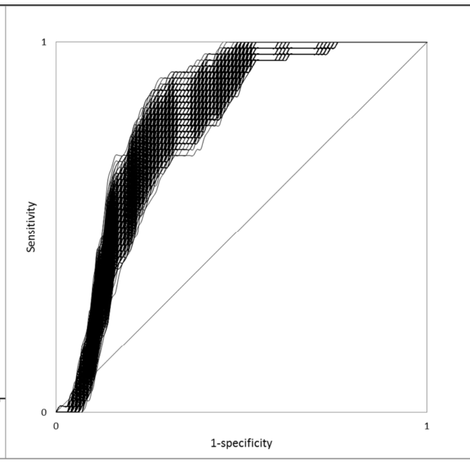
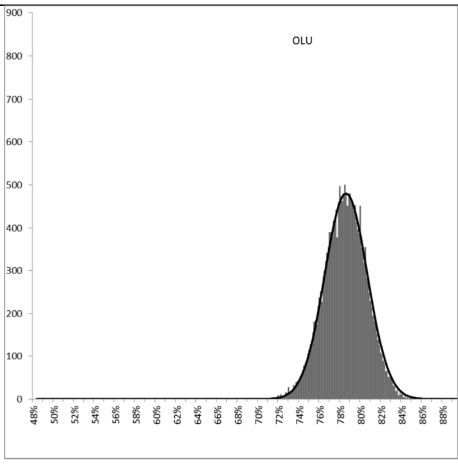
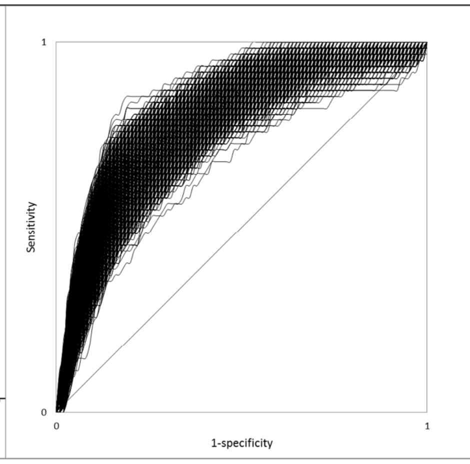
Model	Distribution	ROC plots
TZU		
OL		
OLU		

Figure 9.4 Continued: Distributional properties of model predictive accuracy, as measured by area under receiver operating characteristic curves (θ)

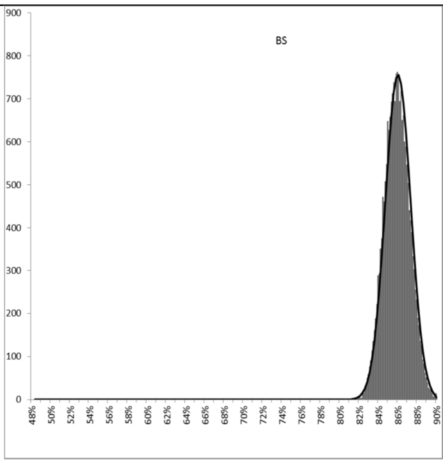
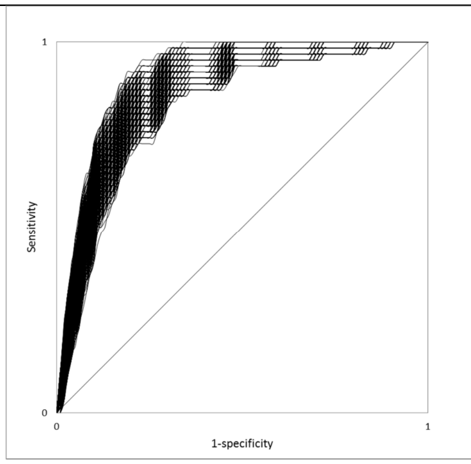
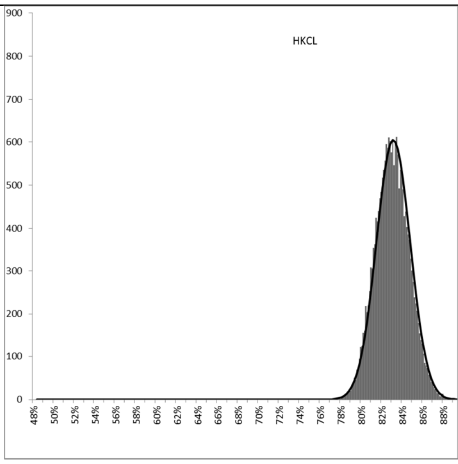
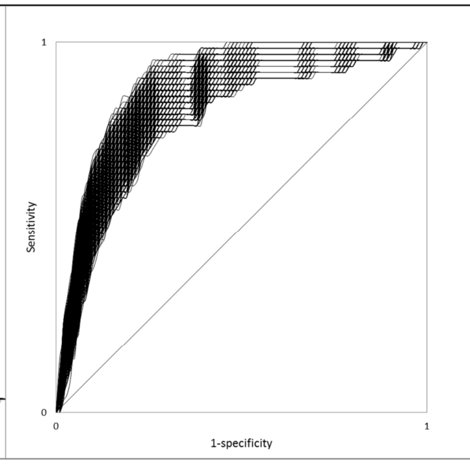
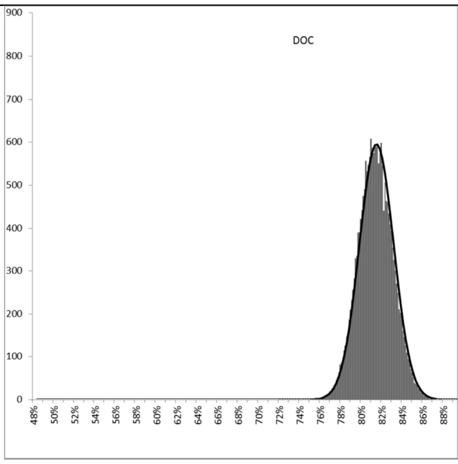
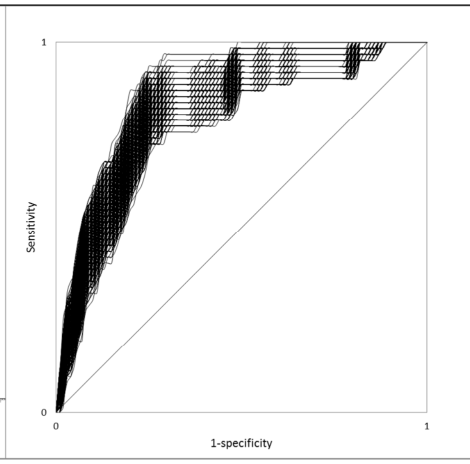
<table>
<tr><th>Model</th><th>Distribution</th><th>ROC plots</th></tr>
<tr><td>BS</td><td>
</td><td>
</td></tr>
<tr><td>HKCL</td><td>
</td><td>
</td></tr>
<tr><td>DOC</td><td>
</td><td>
</td></tr>
</table>

Figure 9.4 Continued: Distributional properties of model predictive accuracy, as measured by area under receiver operating characteristic curves (θ)

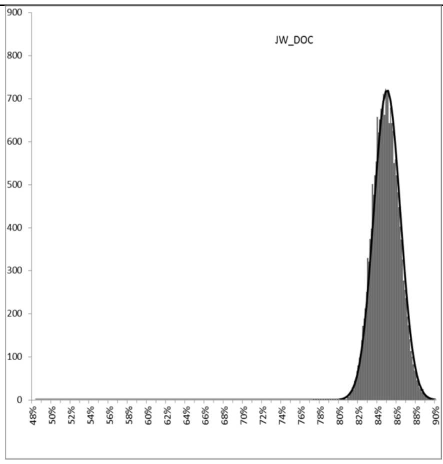
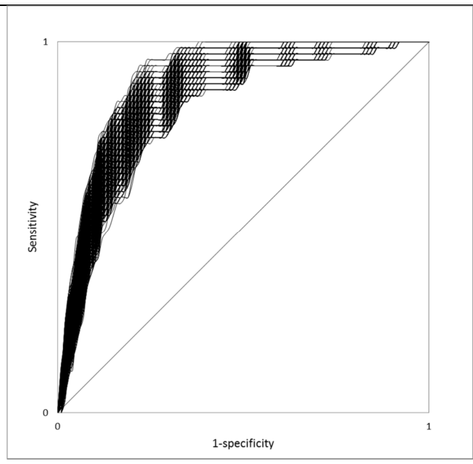
Model	Distribution	ROC plots
JW_DOC		

Figure 9.4 Continued: Distributional properties of model predictive accuracy, as measured by area under receiver operating characteristic curves (θ)

Table 9.6: Comparison of ROC accuracies between original time-split sample and Monte Carlo simulation

MODEL	Time-split	Monte Carlo	SD	Z[1]
SIZE	80.6%	80.5%	1.5%	0.09
BV	67.0%	69.5%	1.9%	-1.31
B/M	57.7%	58.3%	2.6%	-0.24
AZ	78.7%	80.7%	1.7%	-1.16
AZU	68.9%	73.5%	2.6%	-1.82
AZN	73.1%	N/A	N/A	N/A
TZ	74.9%	77.9%	1.8%	-1.65
TZU	65.3%	69.4%	3.5%	-1.19
TZN	72.4%	N/A	N/A	N/A
OL	76.3%	78.4%	1.2%	-1.74
OLU	78.1%	78.6%	2.1%	-0.25
OLN	80.9%	N/A	N/A	N/A
HKCL	82.7%	83.2%	1.6%	-0.30
BS	83.9%	86.1%	1.3%	-1.63
DOC	81.8%	81.5%	1.7%	0.17
JW_DOC	83.0%	85.0%	1.4%	-1.42

[1] Significance of the difference between the original time-split samples to that of the Monte Carlo simulation.
* Significantly different at 95% confidence level (none).

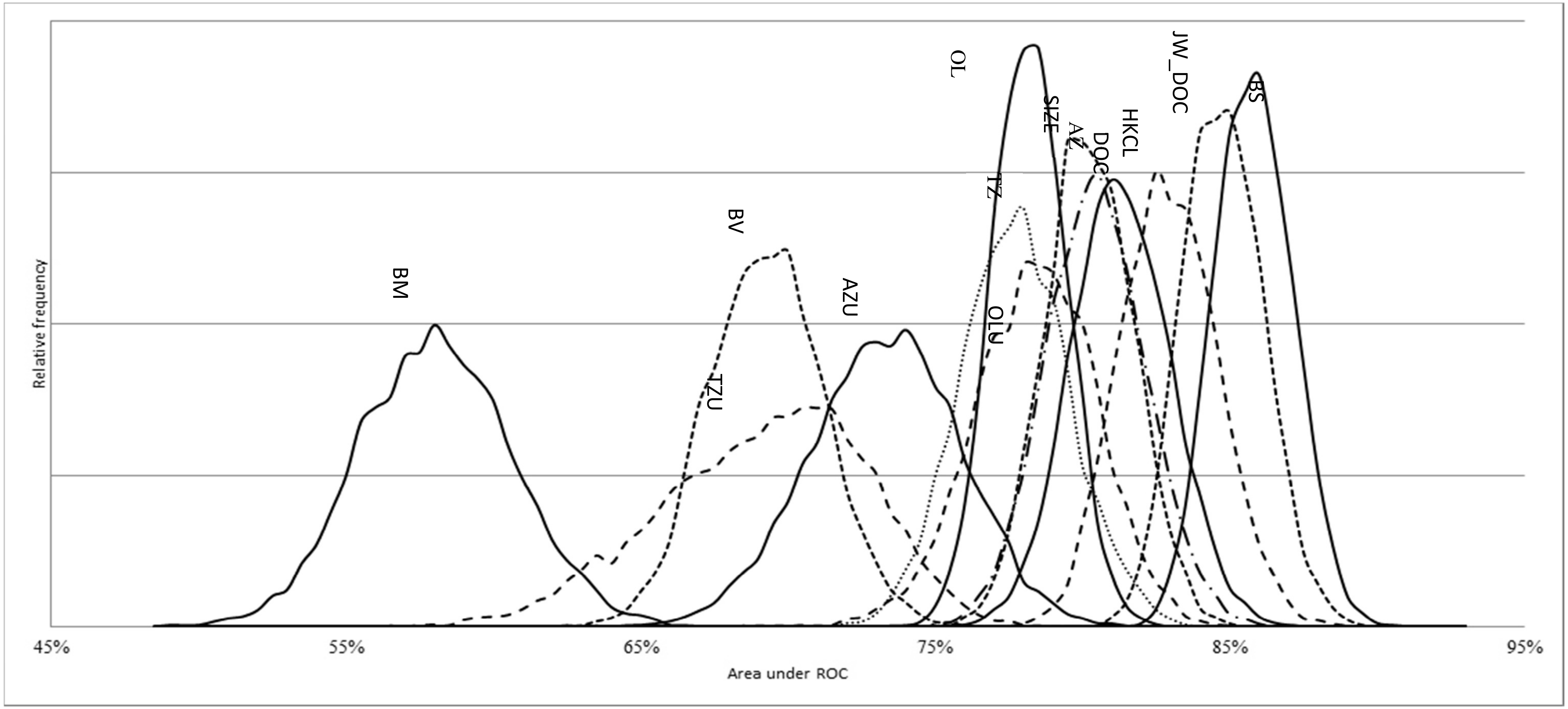


Figure 9.5: Relative frequency distributions of area under receiver operating characteristic curves (θ) for thirteen insolvency prediction models. Based on 10,000 random splits between training and validation samples

It is debatable whether the Monte Carlo simulation provides a truer depiction of performance over that of the original time-split analysis, but it can go some way to determine how the accuracy of the models may vary over time, depicted by their standard deviation. The simulation is however, not necessarily a good conveyor of the real-life decision making faced by lenders with respect to each of the models which require training, as each repetition produces a different set of coefficients to which future judgements and decision making must be made.

The two models which are repeatedly re-estimated using discriminant analysis (AZU and TZU) provide average accuracies of approximately ten percentage points lower than their original counterparts, of which the coefficients were estimated using much smaller data samples, many decades earlier. Why the persistent re-estimation of these models provides results which are less favourable then their original counterparts can be explained by the outlier effect as described in the previous sub-section, but it is somewhat surprising as to the extent of this effect over 10,000 repetitions. Logistical regression in contrast is more robust with the updated model providing a higher average accuracy than its original counterpart over the simulation.

For the remaining ten models (the models which do not require coefficient estimation) some generalised observations can be made. Although, as previously suggested, this particular examination of the models does not act as a true ex-ante study of the models accuracy, it is likely to be beneficial to any potential lending decisions made in the future. The distribution of the model accuracy is based upon randomly selecting half of the total sample and applying ROC analysis to determine the models predictive power, and therefore the 10,000 repetitions of this process allow us to observe how much each model is dependent upon sample selectivity.

I can observe that the higher scoring models have lower standard deviation suggesting better consistency as well as accuracy and therefor are certainly more reliable than their underperforming counterparts. The lower scoring models tend to have larger variances and thus one snapshot may not give a true picture for example; B/M ranges from between forty-nine and sixty-nine per cent, and BV ranges from between sixty-three and seventy-seven per cent, indicating that the choice of sample make a huge difference to the resulting model power with power-spreads of up to 20%. The lowest power-spread comes from the models OL, JW_DOC and BS with spreads of around ten per cent (standard deviations of 1.3%), with JW_DOC and BS also providing the highest average accuracies from all of the models of which I test.

The inclusion of the distributional feather plots aids us further in visualising the performance of the models. It is clear to observe that the worse performing model B/M shows no sign of being a curve at all, as suggested by its 58% average accuracy. The plot simple straddles the 45^0 random model line in a somewhat parallelogramesque manner. For the other models each feather is what one might expect, with the lower scoring models providing fatter, flatter feathers compared to the thinner, curvaceous feathers associated with the higher power models. One point of interest to note regarding the feathers is their position within the graph as the different models increase in accuracy. In each case the feather gravitates towards the top right-hand of the plot area (i.e. where sensitivity is 100%) rather than the opposite bottom left-hand corner of the plot area where specificity is at 100%. The phenomenon is readily explained by the sample sizes of the failed and non-failed groups of firms. With approximately 70 times more non-failed firms in the sample than failed firms, it is far easier for the model to capture all of the failed firms than it is to capture all of the non-failed firms. Other than analysing equal group sizes (which will of course detract from the empirical validity of the any study, given the relatively small number of failed firms in the population), the slightly skewed feather shape will inevitably persist.

In sum, my results show that the market-based, contingent claims models, in all of the guises I adopt, performed better than the accounting-number based models; and that, within the family of accounting number-based models, the Altman Z-score is persistently more accurate than all of the other models, and the single variable approach of Beaver (1966), does not perform as well as I would have hoped, being one of the poorest models as compared with the more complex multivariate approaches. Reversely using a single variable representing size can provide highly comparable predictive accuracies compared to any other methodology.

9.5 Economic value

This section applies the economic value model described in section 8.3 within several scenarios. The first scenario envisages an economy consisting of sixteen banks, each bank making lending decisions based upon a different insolvency prediction or credit risk model. The data for this economy is provided by my original validation sample in order that six of the bank's models can be trained in accordance to what has been previously described in this thesis.

Similar to Agarwal and Taffler (2007), because each of the models to which I test are miscalibrated as noted in section 9.1, it is evident that it would not be appropriate to employ the loan rate formulae as detailed in section 8.3. As this formula relies upon the default probability as determined a particular model, and the average probability of default varies enormously (as detailed in Table 9.1 to range between 72% for the B/M model and 1.2% for the DOC model), in this example the DOC model would always charge a lower rate of interest than the B/M model and therefore a truly biased economy would exist.

To eliminate the problem of miscalibration, a single credit spread is used for each model. For each model the firms are ranked from lowest probability of default to highest probability of default and then separated into twenty separate groups. The 5% of firms with the lowest probability of default are ascribed a risk premium *k* of 0.3% following both Blochlinger and Leippold (2006) and Agarwal and Taffler (2007). At the other end of the spectrum, the 5% of firms with the highest probability of default are not offered a loan at all by the bank. The remaining 90% of firms are ascribed loan offer rates according to Table 9.8 with each class of loan increasing by 0.25% as the probability of default increases.

The remaining specifications of this test and the model construction are now detailed and recalled. The artificial economy which is created contains 3499 firms, 61 of which are known to have failed. Each firm requires an equal share of a loan market worth £100bn (approximately £28.5m each). Within the economy are sixteen banks competing for business, each using a different model to assess credit worthiness of each firm. All of the banks reject firms within the bottom 5% of their credit spread[149], but make offers to each of the others based upon their position in the spread (Table 9.7). For the customer, or firm perspective, each firm will accept an offer from a bank based upon the lowest offered rate. If one or more banks offer the same low rate then the value of the loan is divided equally among this number.

Once all lending and borrowing decisions have been made, I am then able to calculate revenue, profitability, loss given default, market share, and ROI for each of the banks to determine which model has the greater economic value. Loss given default (LGD) is calculated as being 45% of the loan value, following the Basel II internal ratings approach.[150]

Table 9.7: Credit spread for artificial economy

LOAN TYPE	GROUP	Premium
HIGHEST QUALITY	1	0.30%
	2	0.55%
	3	0.80%
	4	1.05%
	5	1.30%
GOOD	6	1.55%
	7	1.80%
	8	2.05%
	9	2.30%
	10	2.55%
MEDIUM	11	2.80%
	12	3.05%
	13	3.30%
	14	3.55%
	15	3.80%
POOR	16	4.05%
	17	4.30%
	18	4.55%
	19	4.80%
DO NOT LEND	20	N/A

[149] As a consequence, the market share may not sum to 1 if all banks refuse to loan to one or more firms.

[150] Agarwal and Taffler (2007) also use this value on their 2-bank analysis. Blochlinger and Leippold use 40% in their study based upon their findings at Credit Suisse and the Swiss Banking Institute.

Table 9.8: Economic value tests upon sixteen credit scoring models

BANK	SIZE	B/M	BV	AZ	AZU	AZN	TZ	TZU	TZN	OL	OLU	OLN	BS	HKCL	DOC	JW	TOTAL
Value of Good Loans	13280	9553	5384	3803	10048	6163	3465	7802	9408	3186	8016	6592	2417	2610	3970	2561	98257
Value of Defaulting Loans (£m)	0.00	652.57	42.87	57.16	242.93	23.82	57.16	100.03	209.58	42.87	71.45	128.61	0.00	28.58	85.74	0.00	1743
TOTAL LOAN VALUE (£m)	**13280**	**10205**	**5427**	**3860**	**10290**	**6187**	**3522**	**7902**	**9618**	**3229**	**8087**	**6721**	**2417**	**2638**	**4055**	**2561**	**100000**
Market Share	13.28%	10.21%	5.43%	3.86%	10.29%	6.19%	3.52%	7.90%	9.62%	3.23%	8.09%	6.72%	2.42%	2.64%	4.06%	2.56%	100.00%
Average lending rate	0.98%	1.04%	0.85%	0.65%	0.98%	0.70%	0.60%	0.79%	0.93%	0.73%	0.75%	0.83%	0.59%	0.64%	0.77%	0.61%	
Share of Good Loans	13.52%	9.72%	5.48%	3.87%	10.23%	6.27%	3.53%	7.94%	9.58%	3.24%	8.16%	6.71%	2.46%	2.66%	4.04%	2.61%	100.00%
REVENUE (£m)	**130**	**103**	**59**	**32**	**109**	**45**	**27**	**64**	**90**	**31**	**61**	**61**	**17**	**21**	**37**	**19**	
Share of Defaults	0.00%	37.43%	2.46%	3.28%	13.93%	1.37%	3.28%	5.74%	12.02%	2.46%	4.10%	7.38%	0.00%	1.64%	4.92%	0.00%	100.00%
LOSS (£m)	**0**	**294**	**19**	**26**	**109**	**11**	**26**	**45**	**94**	**19**	**32**	**58**	**0**	**13**	**39**	**0**	
PROFIT (£m)	**130**	**-190**	**39**	**6**	**0**	**34**	**1**	**19**	**-5**	**11**	**29**	**3**	**17**	**8**	**-1**	**19**	
RETURN ON ASSETS	**0.98%**	**-1.87%**	**0.73%**	**0.17%**	**0.00%**	**0.56%**	**0.04%**	**0.23%**	**-0.05%**	**0.35%**	**0.36%**	**0.04%**	**0.69%**	**0.30%**	**-0.03%**	**0.75%**	

Table 9.8 presents the result of this first test of economic value based upon all sixteen banks/models with several points of interest emerging. Firstly, the return on investment for each of the banks is consistent with prior studies, being less than 1% in all instances, suggesting that aggregated lending profits for banking institutions are not as substantial as one might expect given the volume of potential transactions. The results from this sixteen bank economy show that ROI ranges from 0.98% (SIZE) to -1.87% (B/M). Whereas it is not unsurprising that B/M has the least economic value given its poor performance in prior ROC analysis, it is interesting that the SIZE model is the top performer at almost one quarter of a percent higher ROI than its nearest rivals JW_DOC (0.75%) and BV (0.73%), which is also quite interesting given the relatively poor performance of BV in the distributional test.

Three out of the top four models, SIZE, JW_DOC, and BS do well to avoid lending to any of the failed firms within the sample. With respect to JW_DOC and BS, this avoidance of losses is perhaps key to their higher ROI given that overall market share for these two firms are among the three lowest (2.6% and 2.4% respectively). With profit figures of £19.28m and £16.71m, these models are mid-ranked but their smaller initial outlay ensures that they are highly ranked with respects to ROI. Profits for SIZE are the largest of all the banking institutes at almost £300m, some £90m ahead of its nearest rival explained by it having the highest market share (13.3%) and one of the highest average lending rates (0.98%).

Several of the models do not perform particularly well in this test for economic value. Four models (B/M, AZU, TZN, and DOC) post negative profits for this test, which translates into ROI's of -1.87%, 0%, -0.05%, and -0.03% respectively. The large losses incurred by B/M are again consistent with the weaknesses of this model discovered in prior accuracy tests. Although B/M gains a high percentage of non-defaulting loans (10%), its value is lost by lending to 37.4% of the failed firms, which leads to overall losses of £190m. One surprising result here is the poor performance of the DOC model, to which one might have expected a relatively much higher ROI given its performance in the accuracy tests. However, a small market share (4.1%) and a 4.6% share of defaulters led to an overall loss of £1.37m.

Further to this sixteen bank study, a second-phase test is now performed. The second-phase test expands the sample timeframe to cover the whole of the decade 2000-

2009 inclusive. To ensure that no bias enters the findings, the six banks using models which require training (AZU, AZN, TZU, TZN, OLU, and OLN) are dropped from test, resulting in an economy consisting of the ten remaining banks. Given the relatively mediocre impact that these models have provided within all prior tests contained within this thesis, it may be more beneficial to eliminate these firms all together in order that the sample size may be increased. As a consequence, comparisons can be made to the previous sixteen-bank economy in terms of both competition and of customer availability.

Table 9.9 shows the results of this second-phase test and several similarities are evident. Firstly, SIZE remains the better bank, posting profits of £233m and an ROI of 0.97% almost identical to the previous test. The main difference in the numbers reported by SIZE is that is has now contracted 7.7% of the defaulting loans compared to its previously healthy 0%. An increased market share of 24% is almost double that which it achieved in the sixteen bank economy and cements its position as the most profitable bank, suggesting that the size of a potential loan customer is important to making profits. JW_DOC and BS which were previously in the top four firms remain so, with an increase in BS from 0.69% to 0.75%, and a decrease in JW_DOC from 0.75% to 0.66%. Although both of their market shares have now increased, this time it is only BS which avoids defaulting loans altogether, and the only bank to do so in this test.

B/M remains the worst performing bank in terms of economic value, posting losses of £92m and an ROI of -0.48. Its share of defaults is almost half at 46% which cannot be compensated by a high market share of 19%, cementing its place at the bottom of the rankings in terms of both profits and ROI. In this test B/M is the only bank to post a negative ROI and with all other firms making profits of between £12.95m (HKCL) and £233m (SIZE).

With the change in economy, the elimination of six of the banks and an increase in customer base, the DOC model has a reversal of fortune. Previously DOC posted losses and a negative ROI, but here the bank posts profits of £32.56m and an ROI of 0.48%. Although its share of defaults has changed only slightly, it has managed to increase its share of good loans from 4% to 6.8%, almost double what it previously achieved.

Table 9.9: Economic value second-phase test upon ten credit scoring models

BANK	SIZE	B/M	BV	AZ	TZ	OL	BS	HKCL	DOC	JW	TOTAL
Value of Good Loans	23874	18587	15072	5556	9115	8668	3650	3490	6686	3759	98458
Value of Defaulting Loans (£m)	118	711	263	64	171	97	0	30	79	8	1542
TOTAL LOAN VALUE (£m)	**23993**	**19298**	**15335**	**5620**	**9286**	**8765**	**3650**	**3520**	**6765**	**3767**	**100000**
Market Share	23.99%	19.30%	15.34%	5.62%	9.29%	8.77%	3.65%	3.52%	6.77%	3.77%	100%
Average lending rate	1.20%	1.22%	1.34%	0.93%	1.14%	1.14%	0.75%	0.76%	1.02%	0.75%	
Share of Good Loans	24.25%	18.88%	15.31%	5.64%	9.26%	8.80%	3.71%	3.54%	6.79%	3.82%	100%
REVENUE (£m)	**286**	**228**	**201**	**52**	**104**	**99**	**27**	**27**	**68**	**28**	
Share of Defaults	7.68%	46.08%	17.08%	4.17%	11.11%	6.29%	0.00%	1.96%	5.15%	0.49%	100%
LOSS (£m)	**53**	**320**	**118**	**29**	**77**	**44**	**0**	**14**	**36**	**3**	
PROFIT (£m)	**233**	**-92**	**83**	**23**	**26**	**55**	**27**	**13**	**33**	**25**	
RETURN ON ASSETS	**0.97%**	**-0.48%**	**0.54%**	**0.41%**	**0.29%**	**0.63%**	**0.75%**	**0.37%**	**0.48%**	**0.66%**	

Some other points to note regarding the change in economy are as follows. One of the better performing models in the previous test was BV with an ROI of 0.73%, but here an increase in loan defaults from 2.5% to 17.1% sees this ROI drop to 0.54% but still ranked 5th on this basis. The remaining models which have not been mentioned up until now have been largely inconspicuous. AZ, TZ, and OL have managed to increase their respective profits and ROI's as compared to the prior sixteen bank test, with AZ and TZ each holding lower to middle rankings. OL shows the biggest improvement in terms of ROI rankings, now posting the third highest ROI at 0.63%; and what of HKCL?

HKCL was ranked third in the distributional properties test in section 9.4 and the third best performer in the ROC analysis of section 9.2. Here, in terms of economic value, the model does not perform so well. ROI has increased from 0.30% to 0.37% as I change the structure of the economy, but neither test posts any note-worthy results, ranking towards the lower end of the spectrum.

These results suggest that lending decisions based on the size (market capitalisation) of a potential customer are more profitable than any other credit rating methodologies that this thesis has investigated, and has quite an intuitive and interesting appeal which can be linked back to several of the causes of financial distress and insolvency as discussed in section 3. It is clear that the size of a company is closely related to its age and profitability; many companies will start off small and will only grow in size over a number of years given sound managerial decision-making and consistent provision of profit. Smaller companies will inevitably find it harder to raise capital, attract good management, and have the capability to reclaim bad debts, noted by the R3 as being key factors in the cause of insolvency. The largest factor and cause of insolvency as stated by the R3 is loss of market which attributes to 29% of all insolvencies. Larger firms traditionally will have age, brand identity, customer loyalty and a substantial share of a particular market. The size of the operation means that these larger firms may better adapt to shrinking or changing market places by being able to invest more heavily in research and development than smaller, less-established firms.

To conclude this section I present a third-phase test of economic value when misclassification costs differ which extends the previous ten-bank economy model thusly. The test of economic value gained in terms of profitability and ROI is now repeated over subsequent "years", in which the competitive nature of the banking industry means that after each test, the bank with the lowest ROI is eliminated from model. The test is then repeated with the remaining banks (upon the same data) until only one bank remains. By performing this test it is hoped that I might better understand the effect that the number of banks has on this type of test, and the interaction between them, given that prior studies have been limited to two or three banks; the "elimination" test will also provide an alternate method of evaluating and ranking the economic value of each of the banking models, to that of the previous two economies. The ROI figure for each bank and for each year is presented in the summary Table 9.10 below. Detailed results of this test can be found in Appendix I.

Table 9.10: Economic value elimination test – resulting ROI

BANK	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8	Year 9
AZ	0.41%	0.68%	0.71%	0.70%	0.71%	0.71%	0.95%	1.15%	1.27%
SIZE	0.97%	0.96%	0.97%	1.00%	1.01%	1.07%	1.11%	1.11%	1.19%
BS	0.75%	0.58%	0.43%	0.54%	0.65%	0.93%	0.92%	1.07%	
OL	0.63%	0.53%	0.53%	0.52%	0.53%	0.55%	0.70%		
BV	0.54%	0.43%	0.43%	0.52%	0.54%	0.46%			
DOC	0.48%	0.48%	0.52%	0.56%	0.51%				
HKCL	0.37%	0.31%	0.38%	0.43%					
TZ	0.29%	0.33%	0.35%						
JW_DOC	0.66%	0.22%							
B/M	-0.48%								

Each bank is order top (best) to bottom (worse) according to economic value elimination test which yield some surprising results. Year 1 contains the ROI's as previously depicted in Table 9.9 with the lowest ranked bank B/M eliminated from year 2 and with the elimination of B/M; year 2 throws a surprising result. JW_DOC, previously ranked in the top three for all of the accuracy, power, and economic value tests, provides the lowest ROI attributable to a large increase in its share of defaults (8.3%). Although generating profits of £13m, as the lowest scoring model it is reluctantly eliminated. The next model to be eliminated in year 3 is TZ followed surprisingly by two more of the contingent claims-based models which performed so well in the ROC and distributional tests. Subsequently BV, OL, and BS are eliminated in that order leaving an economy consisting of just two banks, SIZE and AZ. Unexpectedly it is the Altman Z-score which comes out on top in this head-to-head test, despite SIZE's top performance and AZ's average performance in the previous two tests.

By analysing the ROI and how it changes over time as competition is reduced, we can see that for SIZE it is quite consistent, steadily increasing as the number of banks diminishes from 0.97& in year 1 to 1.19% in year 2. AZ also increases over time but is more irregular increasing from 0.4% in year 1 to 0.68% in year two, remaining at approximately 0.7% through years 3-6 and then rising in the final three years to 1.27%, beating SIZE in the process.

It is difficult to interpret these results given the swing in rankings to both the sixteen and ten bank tests. Two consistencies remain within these comparisons; B/M provides no economic value to lending decisions and seems to be firmly rooted to the bottom of the results table, not only in the economic value tests, but within the ROC and distributional tests discussed previously. SIZE also remains consistent but at the other end of the spectrum, ranking first in two out of the three economic value tests and ranking second in the other. Other results are more perplexing such as the poor performance of the contingent-claims models JW_DOC, DOC and HKCL. HKCL performs poorly in all of the economic value tests and DOC not much better, which is surprising given their good showing throughout sections 9.1, 9.2 and 9.4. Further, the poor ROI of JW_DOC in year 2 of the latter test is in complete contrast to all previous results, to which this model has done particularly well on all accounts.

It is with respect to this particular result that one final test is conducted in which bank JW_DOC re-enters the economy to tend for business against the top two performers AZ and SIZE. The test is conducted purely out of curiosity given JW-DOC's previously excellent results.

Table 9.11: Re-entry of JW_DOC

BANK	*SIZE*	*AZ*	*JW_DOC*	TOTAL
Value of Good Loans	42004	33651	22273	97929
Value of Defaulting Loans (£m)	597	491	348	1436
TOTAL LOAN VALUE (£m)	42602	34142	22621	99365
Market Share	**42.9%**	**34.4%**	**22.8%**	100%
Average lending rate	1.8%	1.8%	1.7%	
Share of Good Loans	42.9%	34.4%	22.7%	100%
REVENUE (£m)	**761**	**620**	**387**	
Share of Defaults	41.6%	34.2%	24.2%	100%
LOSS (£m)	**269**	**221**	**156**	
PROFIT (£m)	**492**	**399**	**230**	
RETURN ON ASSETS	**1.15%**	**1.17%**	**1.02%**	

The re-entry of JW_DOC as shown in Table 9.11 shows some consistency in the results previously gained in that AZ is still the better performing bank (albeit by a much smaller margin) over SIZE, with JW_DOC performing least well out of the three. This shows that despite my eagerness for JW_DOC to re-enter the model, provide the highest ROI, and completely turn the prior results on their head, the better performing banks, with respect to economic value at least, are SIZE and AZ with the more consistent bank being that of SIZE across each of the tests performed concerning economic value. The results therefore suggest that given this simplified economy, lending decisions based upon the size of the borrower alone is the most profitable approach.

10 Thesis extension: a best of breed model

Section 6 posited the notion that it would be desirable to offer a new model – based upon the best components, traits, and variables contained within the selection of extant models that have been analysed within this thesis thus far. However, unless it can be based on some sound underlying theoretical argument, the addition of another statistically derived insolvency prediction model into the literature has little appeal. This being said, it is apparent that for an all-encompassing thesis, it would be of benefit and of some interest to the reader, if this exercise was undertaken.

There are two main considerations which much be addressed when developing a model of this type: (1) what variables are to be used? and (2) which statistical technology is best able to provide both maximum separation of the two groups of firms and also provide a readily understandable model? With regards to the selection of variables, I begin by citing a relevant paragraph by Falkenstein et al. (2000):

"Financial ratios are related to firm failure the way that the speed of a car is related to the probability of crashing: there's a correlation, it's nonlinear, but there's no point at which failure is certain. Failure depends upon other variables, such as the environment and driver skill. When modelling such phenomena, one must, therefore, build in a tolerance for extreme values, as these observations do not necessarily relate to failure - as evidenced by their prevalence.

Yet all is not noise: higher leverage, lower profitability, firm size, and liquidity are all related to default in ways most credit analysts would expect. What is not clear is the shape of the relationship (what default rates correspond to which input ratio levels), which ratios are most powerful (e.g., profitability: net income or EBIT?), and how correlations affect the relative weightings assigned to these various ratios in a multivariate context." Falkenstein et al. (2000, p.27)

The literature review of section 5 refers to a vast array of variables which have been used to model credit risk and predict insolvency over the last half century with most accounting models centred on the size, profitability, and liquidity paradigm, with leverage and variability being addressed within the market models. Studies such as

Keasey and McGuiness (1990), with particular reference to UK data demonstrate that profitability and efficiency ratios are important in explaining failure in the years immediately prior to failure. The model which I develop should therefore incorporate all of these aspects of failure.

Firstly I deconstruct the Altman (1968), Ohlson (1980) and Taffler (1984) models to obtain the variables within, and categorise according to type. This results in 17 different variables out of the 18 variables available (WC/TA appearing in both the Altman and Ohlson models). These variables are OL SIZE[151], TL/TA, WC/TA, CL/CA, OENEG, NITA, FUTL, INTWO, CHIN, RE/TA, EBITTA, VE/TL, S/TA, PBTCL, CA/TL, CL/TA, NCI[152]. I also include the three single variables SIZE, B/M and Beaver's CF/TD (BV). The probabilities of failure as designated by all four of the market models BS, HKCL, DOC and JW_DOC, are also included as input variables to the model, this results in a total of 24 variables for initial inclusion into the best-of-breed model.

With regards to the best technology to employ, not only to best separate the two groups of firms, but to produce a meaningful model with coefficients applicable to future samples, I decided to employ the logit model. The MDA methodology was disregarded due to its vulnerability to outliers as discussed in section 9.3, and the Neural Network methodology negated due to its black box nature as discussed in section 5.4. A probit model was considered to be an option, but findings such as in Zmijewski (1984) relating to the similarity of performance between the logit and probit models suggests that both methodologies would perform equally as well. The logit model was arbitrarily chosen because of its greater popularity within the literature. Should the aim of thesis been focused upon deriving and presenting a brand new arbitrary model, one might have opted to use a mixed logit approach similar to that described by Jones and Hensher (2004), a type for which little research has been undertaken in subsequent years. Given the almost perfect results of the Jones and Hensher study, it is certainly an avenue that would warrant further investigation. For simplicity purposes and given the nature of this thesis extension however, the standard linear logit model is employed.

In order to determine the best combination of the 24 variables within the logit model, and ignoring (for now at least) the fact that there may be correlation between some of

[151] OLSIZE being Ohlson's Size variable (ln GDP deflated total assets), which is not to be confused with SIZE (ln GDP deflated market value of equity).

[152] Please refer back to section 8.2 (model specification) for precise variable definitions.

these variables, I employ a "step-wise" procedure which eliminates one variable at a time from the model. The criterion for elimination is simply the variable which adds the least to the model, i.e. has the least significant coefficient, using the chronologic training sample of data (2000-2005).

As the variables are reduced from 24 down to just 1, at each stage of the process I apply the coefficients (the model) to the validation sample of data (2006-2009) and calculate the area under the receiver operating characteristics curve. This step-wise reduction of variables is detailed in Table 10.1 below. For each model, T24 to T1, the coefficient estimates are shown along with the corresponding Wald Chi-Squared test statistic. The variable with the lowest test statistic is eliminated from the model. Panel A shows the estimates for models from T24 through T13, and Panel B shows the estimates for models T12 through T1.

It can be seen from the results contained within Table 10.1 that this step-wise procedure yields some interesting (possibly strange) results. Firstly, HKCL, one of the best performing variables on an individual basis, is the fourth variable to drop out of the model, and not what I would expect. Reversely, B/M, the poorest performing variable (model) under ROC analysis, does not drop out until T12 or half way through the procedure. BS remains until T7, but I would expect that this variable would be the "last one standing" given that it achieved the highest accuracy in previous ROC tests.

The explanation for these strange results may lie in the fact that there may be some correlation between a number of the variables, and that this correlation may affect the estimation process.

The two most significant variables are SIZE and OLSIZE respectively, which is not too surprising given the previous good performance of SIZE in both ROC and information content tests. These two variables however, are potentially highly correlated which may also affect the estimation procedure. This problem of correlation is addressed further later in this section.

Table 10.1: Panel A. Step-wise reduction showing coefficient estimates with Wald ChiSq in parenthesis

	alpha	SIZE	OLSIZE	STA	WCTA	OENEG	RETA	BS	DOC	JW	INTWO	CLCA	BM	PBTCL	CHIN	CFTD	FUTL	CLTA	VETL	NITA	EBITTA	HKCL	CATL	TLTA	NCI
T24	-1.4	-1.0	0.5	-0.4	-3.0	-5.3	0.1	18.7	-6.7	-6.77	0.82	-0.20	-1.93	0.13	-0.28	-0.07	-0.02	-0.43	0.00	1.02	-0.92	1.27	0.01	0.18	0.00
	(1.1)	(9.1)	(2.8)	(3.6)	(6.5)	(3.8)	(0.7)	(3.1)	(1.8)	(1.67)	(3.08)	(1.33)	(1.56)	(0.49)	(1.02)	(0.72)	(0.93)	(0.16)	(0.13)	(0.12)	(0.09)	(0.05)	(0.61)	(0.04)	(0.02)
T23	-1.4	-1.0	0.5	-0.4	-3.0	-5.3	0.1	18.8	-6.6	-6.80	0.82	-0.20	-1.93	0.13	-0.28	-0.07	-0.02	-0.44	0.00	1.02	-0.93	1.26	0.01	0.18	
	(1.1)	(9.1)	(2.8)	(3.6)	(6.5)	(3.8)	(0.7)	(3.1)	(1.8)	(1.68)	(3.08)	(1.34)	(1.57)	(0.49)	(1.01)	(0.72)	(0.93)	(0.17)	(0.13)	(0.12)	(0.10)	(0.05)	(0.61)	(0.04)	
T22	-1.5	-1.0	0.5	-0.4	-3.1	-5.1	0.1	18.9	-6.6	-6.78	0.81	-0.20	-1.91	0.13	-0.28	-0.05	-0.01	-0.34	0.00	0.82	-0.75	1.10	0.00		
	(1.2)	(10.0)	(3.1)	(4.1)	(7.0)	(3.9)	(1.2)	(3.1)	(1.8)	(1.69)	(3.04)	(1.41)	(1.52)	(0.50)	(1.05)	(0.40)	(0.48)	(0.11)	(0.08)	(0.09)	(0.07	(0.04)	(0.02)		
T21	-1.4	-1.0	0.5	-0.4	-3.1	-5.1	0.1	18.8	-6.6	-6.71	0.81	-0.20	-1.91	0.13	-0.28	-0.05	-0.02	-0.35	0.00	0.84	-0.75	1.15			
	(1.1)	(9.8)	(2.9)	(4.1)	(7.1)	(3.9)	(1.1)	(3.1)	(1.8)	(1.66)	(3.03)	(1.44)	(1.55)	(0.49)	(1.04)	(0.48)	(0.75)	(0.12)	(0.11)	(0.10)	(0.07	(0.04)			
T20	-1.4	-1.0	0.5	-0.4	-3.1	-5.0	0.1	19.8	-5.8	-7.18	0.82	-0.20	-1.92	0.13	-0.28	-0.05	-0.02	-0.37	0.00	0.82	-0.73				
	(1.1)	(9.8)	(2.9)	(4.1)	(7.1)	(3.9)	(1.1)	(4.2)	(3.8)	(2.27)	(3.09)	(1.45)	(1.56)	(0.52)	(1.09)	(0.48)	(0.74)	(0.13)	(0.10)	(0.09)	(0.07				
T19	-1.4	-1.0	0.5	-0.4	-3.1	-5.2	0.1	19.5	-5.8	-7.05	0.83	-0.21	-1.99	0.12	-0.28	-0.05	-0.02	-0.38	0.00	0.12					
	(1.0)	(10.4)	(3.1)	(4.0)	(7.4)	(4.5)	(1.2)	(4.1)	(3.8)	(2.22)	(3.17)	(1.52)	(1.72)	(0.48)	(1.07)	(0.44)	(0.76)	(0.14)	(0.11)	(0.05)					
T18	-1.4	-1.0	0.5	-0.4	-3.2	-5.2	0.1	19.5	-5.8	-7.04	0.82	-0.21	-1.97	0.15	-0.28	-0.05	-0.02	-0.46	0.00						
	(1.0)	(10.2)	(3.0)	(4.2)	(7.8)	(4.6)	(1.9)	(4.1)	(3.9)	(2.20)	(3.14)	(1.58)	(1.69)	(1.25)	(1.02)	(0.37)	(0.75)	(0.22)	(0.12)						
T17	-1.4	-1.0	0.5	-0.4	-3.1	-5.3	0.1	19.4	-5.8	-7.10	0.82	-0.21	-1.98	0.14	-0.28	-0.07	-0.01	-0.36							
	(1.1)	(11.6)	(3.6)	(4.4)	(7.5)	(4.6)	(1.9)	(4.1)	(3.8)	(2.26)	(3.08)	(1.54)	(1.69)	(1.19)	(1.07)	(0.69)	(0.57)	(0.14)							
T16	-1.5	-1.0	0.6	-0.4	-2.8	-5.6	0.1	19.5	-6.0	-7.21	0.81	-0.19	-2.07	0.14	-0.28	-0.07	-0.01								
	(1.4)	(12.6)	(4.5)	(5.8)	(12.1)	(5.7)	(2.0)	(4.1)	(4.3)	(2.35)	(3.05)	(1.43)	(1.88)	(1.15)	(1.06)	(0.71)	(0.61)								
T15	-1.5	-1.0	0.6	-0.4	-2.8	-5.6	0.1	19.5	-6.0	-7.21	0.83	-0.19	-2.11	0.13	-0.28	-0.06									
	(1.3)	(12.6)	(4.5)	(5.9)	(12.1)	(5.8)	(2.0)	(4.1)	(4.3)	(2.34)	(3.19)	(1.43)	(1.96)	(1.11)	(1.07)	(0.57)									
T14	-1.6	-1.1	0.6	-0.4	-3.0	-6.0	0.1	20.1	-6.1	-7.38	0.81	-0.20	-2.11	0.15	-0.29										
	(1.5)	(13.6)	(5.0)	(5.9)	(14.5)	(6.4)	(2.6)	(4.3)	(4.4)	(2.43)	(3.03)	(1.48)	(1.96)	(1.26)	(1.14)										
T13	-1.6	-1.1	0.6	-0.4	-3.0	-6.1	0.1	21.3	-6.4	-7.90	0.80	-0.21	-2.07	0.12											
	(1.5)	(14.0)	(5.3)	(5.7)	(15.1)	(6.9)	(2.9)	(4.9)	(4.9)	(2.82)	(2.95)	(1.57)	(1.88)	(1.09)											

Table 10.1: Panel B. Step-wise reduction showing coefficient estimates with Wald ChiSq in parenthesis

	Alpha	SIZE	OLSIZE	STA	WCTA	OENEG	RETA	BS	DOC	JW	INTWO	CLCA	BM	PBTCL	CHIN	CFTD	FUTL	CLTA	VETL	NITA	EBITTA	HKCL	CATL	TLTA	NCI
T12	-2.0	-1.0	0.6	-0.4	-3.1	-5.8	0.1	20.2	-5.7	-7.61	0.68	-0.19	-1.67												
	(2.5)	(13.5)	(5.3)	(4.6)	(15.8)	(6.1)	(4.2)	(4.5)	(4.3)	(2.63)	(2.24)	(1.37)	(1.26)												
T11	-3.3	-0.9	0.5	-0.4	-2.9	-5.0	0.1	21.4	-5.6	-7.98	0.65	-0.16													
	(28.0)	(12.1)	(3.9)	(4.5)	(14.0)	(4.0)	(4.1)	(4.9)	(4.2)	(2.75)	(2.09)	(1.07)													
T10	-3.6	-0.9	0.5	-0.3	-2.4	-4.0	0.1	21.7	-5.2	-8.31	0.64														
	(38.2)	(12.2)	(3.9)	(3.7)	(18.3)	(4.0)	(3.4)	(5.0)	(3.6)	(2.96)	(2.00)														
T9	-3.0	-0.9	0.4	-0.3	-2.5	-4.3	0.1	21.0	-5.3	-7.72															
	(54.7)	(13.0)	(3.5)	(4.4)	(19.7)	(4.4)	(3.5)	(4.9)	(3.9)	(2.70)															
T8	-3.0	-0.9	0.4	-0.3	-2.4	-4.3	0.1	6.1	-4.0																
	(54.6)	(11.6)	(2.7)	(4.4)	(18.9)	(4.8)	(3.0)	(6.0)	(2.5)																
T7	-3.1	-0.9	0.5	-0.3	-2.5	-4.1	0.1	3.5																	
	(58.3)	(13.5)	(4.2)	(4.1)	(21.1)	(4.0)	(3.9)	(3.3)																	
T6	-3.0	-1.3	0.8	-0.4	-2.8	-4.2	0.1																		
	(56.2)	(57.0)	(24.3)	(6.1)	(27.7)	(4.1)	(4.0)																		
T5	-3.0	-1.3	0.8	-0.5	-2.1	-4.4																			
	(66.0)	(62.6)	(30.3)	(18.9)	(26.6)	(3.9)																			
T4	-3.2	-1.3	0.8	-0.4	-1.4																				
	(71.0)	(61.3)	(29.4)	(21.0)	(16.5)																				
T3	-3.6	-1.3	0.8	-0.6																					
	(97.4)	(59.8)	(27.3)	(8.9)																					
T2	-3.5	-1.2	0.8																						
	(90.3)	(56.1)	(21.9)																						
T1	-2.5	-0.6																							
	(92.0)	(37.6)																							

Before I attempt to address the correlation problem, I must first present the results of the ROC analysis for each of the 24 models based on this initial step-wise procedure. The over-riding aim of this section was to prove that by a simple data-mining exercise it would be possible to formulate a model which, using my training and validation datasets could out-perform all of the sixteen prior insolvency prediction and credit risk models which have been evaluated within this thesis. Figure 10.1 plot the ROC accuracy of the model(s) as the variables are reduced from 24 down to 1.

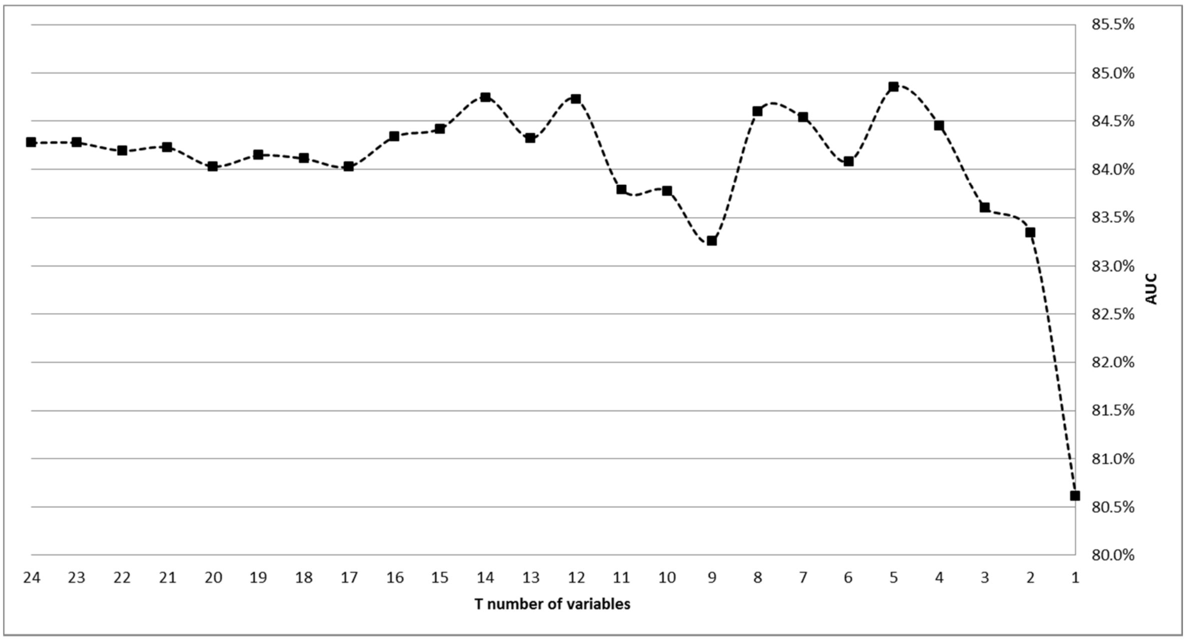


Figure 10.1: ROC accuracies achieved through step-wise reduction of 24 variables

If I refer back to section 9.2, the highest accuracy achieved through ROC analysis based on the chronological training and validation data samples was the naïve market model of Bharath and Shumway (2008), BS, which achieved an accuracy (area under the ROC curve) of 83.9%.

The plot of ROC accuracies achieved through step-wise reduction of 24 variables shows that 18 out of 24 versions of the model have an accuracy higher than 84%. The plot dips under the 84% mark at T11, T10, and T9 but regains "domination" culminating in a maximum ROC accuracy of 84.9% at T5. From this point the model diminishes in accuracy as it approaches T1, the single variable SIZE, which provides an accuracy of 80.6% in concordance with the result within section 9.2.

The highest accuracy is achieved with 5 variables (SIZE, OLSIZE, WC/TA, OENEG, and S/TA) at 84.9%. The resulting best-of-breed model would therefore be presented as follows:

$$PROBABILITY\ OF\ FAILURE = \frac{1}{1 + e^{-Z}}$$

where

$$Z = -3.03 - 1.28(SIZE) + 0.84(OLSIZE) - 2.13(WC/TA) - 4.04(OENEG) - 0.50(S/TA)$$

I have shown within the literature review of section 5 that quite often in these multivariate models, the signs of the coefficients make no intuitive sense although they cannot be interpreted in the same way as a linear regression for example. Here, the variables SIZE and OLSIZE, natural logarithms of GDP deflated market value of equity and total assets respectively, should be expected to have a high correlation, yet the signs of the coefficients are opposite to each other. OENEG (a dummy variable, 1 if total liabilities exceed total assets and 0 otherwise), would be expected to act in the opposite direction to the other variables, but it does not. The variables do however represent key features of the failure process. WC/TA is a liquidity variable, S/TA represents profitability, OENEG attributable to leverage, and SIZE and OLSIZE representing size. This demonstrates to some degree therefore, that the model encapsulates the key components of the failure process.

Previously mentioned in this section was the problem with correlated variables. To investigate this further, the intention is to re-run the step-wise procedure after the removal of highly correlated variables. To do this I perform Pearson rank correlation tests on the 24 variables. Variables are removed in order that correlation coefficients are less than the critical value 0.404 (N(24)-2 degrees of freedom at 5% significance – 2 tail test).[153]

[153] A good explanation of the critical values can be found at http://www.une.edu.au/WebStat/unit_materials/c6_common_statistical_tests/test_signif_pearson.html.

Table 10.2: Correlation between the 24 variables

	SIZE	B/M	BS	HKCL	DOC	JW	OL SIZE	TL/TA	WC/T	CL/CA	OENEG	NI/TA	FUTL	INTW	CHIN	RE/TA	EBIT/T	VE/TL	S/TA	PBT/C	CA/TL	CL/TA	NCI	BV
SIZE	1.000																							
B/M	-0.227	1.00																						
BS	-0.277	0.17	1.00																					
HKCL	-0.214	0.10	0.75	1.00																				
DOC	-0.243	0.08	0.78	0.92	1.00																			
JW	-0.284	0.18	0.96	0.73	0.75	1.00																		
OL SIZE	0.897	0.07	-0.07	-0.07	-0.10	-0.07	1.00																	
TL/TA	0.060	-0.17	0.35	0.24	0.28	0.38	0.10	1.00																
WC/T	-0.032	0.05	-0.14	-0.10	-0.10	-0.15	-0.02	-0.50	1.00															
CL/CA	0.072	0.03	0.00	0.00	0.00	0.00	-0.04	0.01	-0.06	1.00														
OENEG	-0.096	-0.23	0.10	0.11	0.13	0.14	-0.15	0.59	-0.37	0.01	1.00													
NI/TA	0.240	0.10	-0.14	-0.12	-0.14	-0.14	0.29	-0.22	0.33	-0.02	-0.28	1.00												
FUTL	0.043	0.01	0.02	0.01	0.01	0.02	0.05	0.05	-0.07	0.00	0.00	0.09	1.00											
INTW	-0.379	0.07	0.23	0.16	0.18	0.25	-0.37	0.06	-0.05	0.04	0.09	-0.33	-0.10	1.00										
CHIN	0.061	-0.10	-0.14	-0.11	-0.10	-0.13	-0.02	-0.03	-0.01	-0.01	0.02	0.13	0.02	-0.02	1.00									
RE/TA	0.016	0.04	0.00	0.00	0.00	0.00	0.09	0.00	0.24	0.00	-0.01	0.41	0.03	-0.03	-0.03	1.00								
EBIT/T	0.254	0.09	-0.14	-0.12	-0.13	-0.14	0.30	-0.19	0.31	-0.02	-0.27	0.98	0.10	-0.34	0.14	0.39	1.00							
VE/TL	0.000	-0.07	-0.03	-0.02	-0.02	-0.03	-0.13	-0.07	-0.15	0.18	-0.01	-0.38	-0.26	0.07	0.02	-0.38	-0.37	1.00						
S/TA	-0.081	0.01	-0.01	0.01	0.00	-0.02	-0.10	0.04	-0.33	-0.03	0.14	-0.04	0.04	-0.08	0.05	0.01	-0.03	-0.04	1.00					
PBT/C	0.255	0.04	-0.10	-0.07	-0.07	-0.10	0.28	0.03	-0.07	-0.03	-0.05	0.61	0.19	-0.39	0.21	0.03	0.33	-0.10	0.11	1.00				
CA/TL	-0.025	0.00	-0.02	-0.02	-0.02	-0.02	-0.02	-0.07	0.08	0.00	-0.01	-0.02	-0.37	0.05	-0.01	0.00	-0.03	0.21	-0.03	-0.12	1.00			
CL/TA	-0.154	-0.07	0.14	0.13	0.14	0.14	-0.20	0.42	-0.78	0.00	0.45	-0.39	0.03	0.07	0.05	-0.27	-0.37	0.20	0.54	0.02	-0.03	1.00		
NCI	0.058	-0.06	-0.02	-0.01	-0.01	-0.02	0.04	-0.04	0.04	-0.05	-0.01	0.01	0.00	-0.03	0.02	0.00	0.01	-0.01	-0.04	0.00	0.01	-0.04	1.00	
BV	-0.017	0.00	-0.02	-0.01	-0.01	-0.02	-0.02	-0.05	0.07	0.00	0.00	-0.02	-0.38	0.05	-0.01	0.00	-0.02	0.18	-0.03	-0.11	0.28	-0.03	0.01	1.00

The removal of variables can be justified as follows. The high correlations between the market models are unsurprising and because BS outperformed each of the other market models in previous tests, HKCL, DOC and JW_DOC are removed. CL/TA has high correlations with four other variables and therefore it is removed. NITA has high correlations with three other variables and is also removed. SIZE is, as expected, heavily correlated with OLSIZE and I arbitrarily remove OLSIZE. OENEG is significantly correlated with TL/TA and given that OENEG is a dummy variable and TL/TA is not, OENEG is removed. The resulting variable population is therefore reduced to 17 from the original population of 24.

I repeat the previous step-wise reduction, this time only utilising these 17 variables. The results of this reduction are shown in Table 10.2 with the ROC accuracies show in Figure 10.2.

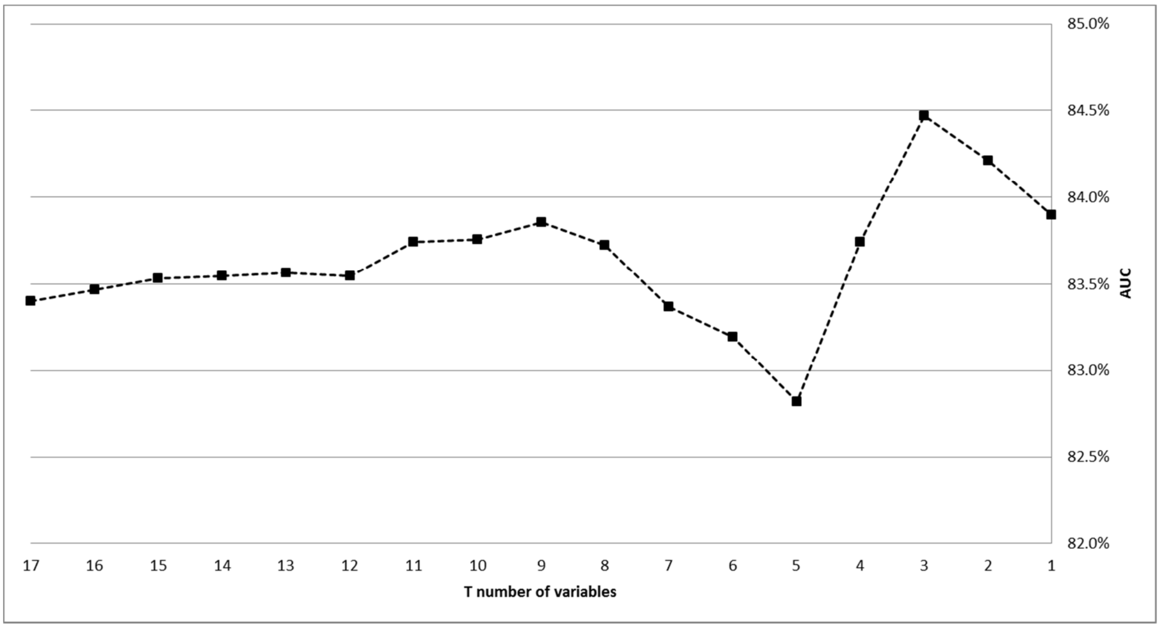


Figure 10.2: ROC accuracies achieved through step-wise reduction of 17 variables

Table 10.3: Panel A. Step-wise reduction of non-correlated variables, showing coefficient estimates with Wald ChiSq in parenthesis

	Intercept	BS	SIZE	WCTA	RETA	TLTA	STA	INTWO	VETL	CHIN	EBITTA	FUTL	CLCA	CFTD	CATL	NCI	PBTCL	BM
T17	-2.71	6.39	-0.39	-2.51	0.08	-1.28	-0.37	0.74	-0.02	-0.42	0.34	-0.03	-0.08	-0.07	0.02	0.00	0.04	-0.10
	(4.9)	(19)	(12.1)	(9.2)	(1.8)	(3.2)	(4.1)	(2.5)	(1.3)	(2.3)	(0.4)	(3.4)	(0.4)	(0.8)	(0.6)	(0.1)	(0)	(0)
T16	-2.75	6.29	-0.39	-2.52	0.08	-1.29	-0.37	0.74	-0.02	-0.42	0.35	-0.03	-0.08	-0.06	0.01	0.00	0.04	
	(17.5)	(18.8)	(12.1)	(9.2)	(1.8)	(3.4)	(4.1)	(2.5)	(1.1)	(2.3)	(0.4)	(2.5)	(0.4)	(0.5)	(0.2)	(0.1)	(0)	
T15	-2.75	6.30	-0.39	-2.55	0.08	-1.28	-0.37	0.73	-0.02	-0.41	0.43	-0.03	-0.09	-0.06	0.01	0.00		
	(17.4)	(18.9)	(12)	(9.5)	(1.8)	(3.4)	(4.3)	(2.5)	(1.2)	(2.3)	(1.1)	(2.6)	(0.4)	(0.6)	(0.2)	(0.1)		
T14	-2.75	6.30	-0.39	-2.53	0.08	-1.28	-0.37	0.73	-0.02	-0.41	0.43	-0.03	-0.09	-0.06	0.01			
	(17.5)	(18.9)	(12)	(9.4)	(1.8)	(3.4)	(4.3)	(2.5)	(1.2)	(2.3)	(1.1)	(2.7)	(0.4)	(0.6)	(0.2)			
T13	-2.79	6.32	-0.40	-2.45	0.08	-1.24	-0.35	0.72	-0.01	-0.41	0.37	-0.02	-0.09	-0.04				
	(18.6)	(19.8)	(12.6)	(9.2)	(1.9)	(3.3)	(4.1)	(2.5)	(0.7)	(2.3)	(0.9)	(1.5)	(0.4)	(0.3)				
T12	-2.71	6.09	-0.39	-2.52	0.09	-1.31	-0.35	0.73	-0.03	-0.40	0.32	-0.01	-0.08					
	(17)	(17.5)	(12.5)	(9.5)	(2.1)	(3.6)	(4.2)	(2.5)	(1.7)	(2.2)	(0.7)	(0.7)	(0.4)					
T11	-2.86	5.92	-0.40	-2.20	0.07	-1.17	-0.32	0.72	-0.03	-0.40	0.32	-0.01						
	(21.6)	(17)	(12.7)	(11.6)	(1.8)	(3.3)	(3.8)	(2.4)	(1.7)	(2.2)	(0.8)	(0.7)						
T10	-2.88	5.94	-0.40	-2.19	0.07	-1.16	-0.32	0.72	-0.03	-0.40	0.32							
	(21.8)	(17.1)	(12.8)	(11.6)	(1.8)	(3.2)	(3.8)	(2.5)	(1.7)	(2.2)	(0.7)							
T9	-2.92	5.95	-0.38	-2.23	0.10	-1.21	-0.29	0.64	-0.03	-0.35								
	(22.5)	(17.2)	(12.2)	(12.3)	(4.1)	(3.6)	(2.9)	(2)	(1.8)	(1.7)								

Table 10.3: Panel B. Step-wise reduction of non-correlated variables, showing coefficient estimates with Wald ChiSq in parenthesis

	Intercept	BS	SIZE	WCTA	RETA	TLTA	STA	INTWO	VETL	CHIN	EBITTA	FUTL	CLCA	CFTD	CATL	NCI	PBTCL	BM
T8	-2.90	6.17	-0.38	-2.19	0.10	-1.20	-0.30	0.66	-0.03									
	(21.7)	(18.4)	(12)	(12.1)	(4)	(3.5)	(2.5)	(2.1)	(1.9)									
T7	-3.23	7.13	-0.39	-2.21	0.11	-1.02	-0.27	0.60										
	(28.4)	(27.1)	(12.2)	(13.8)	(5.2)	(2.9)	(1.9)	(1.8)										
T6	-2.68	7.41	-0.44	-2.33	0.12	-1.09	-0.29											
	(38.3)	(29.6)	(17.2)	(15.1)	(5.4)	(3.2)	(2.3)											
T5	-3.02	7.31	-0.45	-2.18	0.15	-0.86												
	(60.3)	(29.5)	(17.9)	(12.8)	(10.9)	(2.4)												
T4	-3.25	6.55	-0.44	-1.50	0.11													
	(77.9)	(26.1)	(17.5)	(12.8)	(6.9)													
T3	-3.54	7.28	-0.39	-0.27														
	(101.7)	(35.9)	(14.9)	(3.4)														
T2	-3.50	7.34	-0.41															
	(101.4)	(36.6)	(16.3)															
T1	-4.91	9.81																
	(534.9)	(85.5)																

The highest accuracy is achieved with the 3 variable model (SIZE, BS, WCTA) with an accuracy of 84.5%, higher than any model tested within section 9.2 This model however, falls slightly short of the 84.9% achieved previously within the 24-variable step-wise procedure. Despite this small shortfall, the elimination of variables (now that I have removed those which were heavily correlated) is done so in a more logical sequence. B/M is the first variable to be eliminated which is in accordance with its poor prior performance, and the variables within the poor-performing Taffler model are also quick to be eliminated. It also makes sense that BS is the most significant variable given that it achieves the highest accuracy in the ROC tests within section 9.2. The resulting best-of-breed model in this instance is therefore presented as follows:

$$PROBABILITY\ OF\ FAILURE = \frac{1}{1 + e^{-Z}}$$

where

$$Z = -3.54 + 7.28(BS) - 0.39(SIZE) - 0.27(WC/TA)$$

It has been demonstrated therefore that through careful selection of variables, it is possible to provide several models which perform better than any of the sixteen insolvency prediction models plucked from the prior literature, based upon a statistical step-wise reduction of variables. Although this data mining approach provides a model(s) with the highest accuracy in this study there is, of course, no theoretical underpinning, and there is also no guarantee that similar accuracies would be achieved when applied to future datasets. For these reasons alone it not my intention to champion any of these models as being superior to anything presented in the prior literature, but merely act a demonstration as to what can be achieved through an appropriate selection of variables and modelling criterion.

11 Integration of results and conclusions

11.1 Summary of objectives, methods and findings

Within this thesis I have presented a review of the methodologies proposed and employed for the prediction and corporate insolvency and the assessment of credit risk over the last five decades, and catalogued 25 different methods within three broad groupings: theoretical, statistical and artificially intelligent. This thesis, for the first time, brings together these different approaches to insolvency prediction and credit risk analysis, in a single document of reference, utilising both descriptive and quantitative methods for contrast, assessment and evaluation of performance.

Informed by prior frequencies of use and reported efficacies of the various models in combination with an extensive literature review, I select a variety of these models for testing, representing key developments and initiatives spanning the evolution of the literature. The empirical study uses up-to-date UK data, including post-IFRS data, with the sample period encompassing the recent financial crisis to determine whether the reported accuracy of extant models and whether their ability to correctly distinguish between insolvent and healthy firms still holds true in times of economic turmoil. I also develop and test a new naïve version of the down-and-out call barrier option model and find that this model performs favourably in relation to the fifteen extant models which are investigated.

Using ROC curve analysis, not only in a chronological separation of training and validation samples, but using a Monte Carlo random separation procedure based upon 10,000 repetitions, I show that the traditional accounting-based models are outperformed in this setting by theoretically-driven, market-based, contingent claims methods. Of sixteen models tested, the four best performing models are based on European call and barrier option models. The new, naïve version of the down-and-out call option model which I propose is second only the Bharath and Shumway (2008) naïve market model in predictive accuracy; and each of these models outperforms its antecedent. This Monte Carlo approach in combination with ROC analysis is presented as a new and innovative method for assessing the accuracy of prediction models, providing evidence as to the distributional properties of model accuracy over time.

Two single variable accounting number-based models, the book-to-market ratio and Beaver's (1966) best performing ratio, cash-flow to total debt, do not compare well against more sophisticated multivariate accounting number-based approaches using this methodology. Despite its use as focal/benchmark variable in earlier studies, in my setting book-to-market ratio – capturing, in essence, extent of market capitalisation of future earnings – is not found to be useful alone in predicting firm failure. In stark contrast, the single variable SIZE (natural logarithm of GDP deflated market value of equity) performs well compared to the multivariate approaches.

With reference to the accounting number-based models, I am able to obtain only marginally improved predictive ability from the variables of the Altman (1968), Taffler (1983) and Ohlson (1980) models by applying simple Neural Network technology, over that of the original technologies of MDA and logit. An unexpected result was in reference to the original MDA models of Altman (1968), and Taffler (1983) when applying their original coefficients. In ROC tests these models outperformed their updated counterpart which was not what I would expect. The reason for this result could be explained by the heavy impact that outliers have on the MDA technology, relying heavily on the variance and covariance's of the variables. Further results indicated that, as the number of outliers was reduced through winsorisation, the area under the ROC curve increased, and proceeded to dominate the original models, increasing in accuracy up to a point where the models simply collapsed. The use of winsorisation upon the associated Altman and Taffler MDA samples shows that this methodology relies heavily on a carefully selected sample. The original authors were able to use restricted samples of a small number of firms to carefully remove the incidents of outliers, thus ensuring the efficacy of the MDA process. When the MDA is applied to a sample of firms representative of the actual population, the MDA process is found lacking.

Regarding the relatively poor performance of accounting number-based models in the ROC tests, it is not surprising that they are inferior to their market-based counterparts. Since accounting information is historical and does not reflect the current position of a firm, even on its date of publication, the accuracy of multivariate accounting number-based models to predict insolvency is unlikely to be stable over a period of significant and general economic turmoil. Market-based models, incorporating up-to-date market information, are likely to fare better.

This hypothesis was put to the test with the development of a best-of-breed model which utilised a step-wise reduction of 24 accounting and market variables. Several models where proposed, each providing a higher ROC accuracy than any of the sixteen previously-tested extant corporate insolvency and credit risk models. The best performing best-of-breed model contained both market and accounting variables when there was no significant correlation between the initial variable populations. The results of this thesis extension demonstrate that producing an insolvency prediction model which purports to be higher in predictive accuracy than any prior extant model, is a relatively simple exercise. Caution must be used, however, when suggesting that these types of model, statistically fitted to one particular training dataset and tested on a similar holdout sample of data, would demonstrate the same efficacy when applied to further samples in the future. This casing point is seen by the poor performance of the Taffler model for example within my analysis. Previous studies by the author have demonstrated near perfect predictive accuracy, but the same levels of performance are not prevalent in my own empirical study.

The empirical ROC analysis within this thesis, regardless of model methodology or variable selection, indicated that there was a maximum accuracy which could be achieved. This maximum accuracy was 85%, and this barrier could not be broken by any of the models whether extant or proprietary best-of-breed. This is somewhat in-line with the meta-analysis conducted within section 5.8 in which the average accuracy reported from the prior literature was determined to be approximately 83%. To date, my empirical analysis has provided an impartial comparison upon more insolvency and credit risk models, and upon a larger selection of companies, than any study previously found within the literature. The 85% accuracy barrier however, was provided only by one particular test, the ROC curve. The accuracy achieved by such a test does not take into account the differences in economic value of the different associated error types. A type I error (incorrectly classifying a failed firm as not failing) would be expected to incur a larger economic loss within a lending scenario, than A type II error (incorrectly classifying a healthy firm as one which will fail). To allow for these differences in economic value, a further test of predictive accuracy was empirically evaluated.

The second-phase of testing involved the economic value of the insolvency prediction and credit risk models, where an artificial economy was created. The economy contained 3499 firms, 61 of which are known to have failed. Each firm required an equal share of a loan market worth £100bn (approximately £28.5m each), and each insolvency and credit risk model was assigned to a different bank. Banks would base their loan decisions purely upon the failure probability provided by their particular model. The rate at which each bank was willing to lend to a particular company was determined by its position in the credit spread as defined by the banks failure probability default model. Each firm within the economy (the chronologically derived validation sample) acted as a potential loan customer who would ultimately borrow from the bank which offered the best loan deal, denoted by the lowest offered rate of interest. For each loan applicant, the bank would either succeed in attracting the business or it would not. For the loans which were granted there would be two outcomes, either the customer would survive the loan period and make the repayments, or the customer would default. Using this information I was able to determine which of the banks was the most profitable, and therefore which of the models provides the most economic value.

The results of the economic value tests demonstrated similarities to those of the ROC tests but initially there where some stark differences. In the initial test several of the models do not perform particularly well, with four models (B/M, AZU, TZN, and DOC) posting negative profits and therefore negative returns. What is surprising here is the poor performance of DOC and its inability to provide positive economic value, given that it was one of the better performing models under ROC analysis. The remainder of the market models out-perform most cases of the models being based upon accounting numbers.

In order to circumnavigate the training requirement of six of the models (AZU, TZU, OLU, AZN, TZN, and OLN), and to also provide a larger sample of firms for the analysis, the second test dropped the aforementioned models from the economy. The new results showed that the book-to-market ratio performed least well, again posting negative returns for its respective bank, consistent with its prior poor performance under both ROC analysis and the first test of economic value. The remaining nine banks each posted positive returns, with SIZE posting the largest return and the market models BS and JW_DOC ranking second and third respectively.

The third test of this nature involved the removal of one (the least profitable) bank after each year of trading. In this test, SIZE was the penultimate bank standing but was out-survived somewhat surprisingly by the original Altman Z-score, although both banks reported to be profitable. The reason for the good performance of the Altman Z-score can be attributed to its ability to capture the key components attributable to company survival namely profitability, liquidity and size. In this regard the model appears to capture a company's health far more effectively than its competitor accounting-number models, the Taffler Z-score and the Ohlson logit model, of which the former does not perform particularly well in any of the tests undertaken within this thesis.

The single variable SIZE was shown to be a consistently good performer being the dominant bank in two out of the three tests undertaken. Although these economic value tests are based upon a very simplified economy, the results could be seen as mirroring the real loan making decisions faced by banks every day. Size is an important factor in the loan-making process, with companies with large asset bases for security more likely to be granted loans, and also less-likely to default. I have previously stated that the size of a company is closely related to its age and profitability; many companies will start off small and will only grow in size over a number of years given sound managerial decision-making and consistent provision of profit. Smaller companies will inevitably find it harder to raise capital, attract good management, and have the capability to reclaim bad debts, noted by the R3 as being key factors in the cause of insolvency. The largest factor and cause of insolvency as stated by the R3 is loss of market which attributes to 29% of all insolvencies. Larger firms traditionally will have age, brand identity, customer loyalty and a substantial share of a particular market. The size of the operation means that these larger firms may better adapt to shrinking or changing market places by being able to invest more heavily in research and development than smaller, less-established firms. Therefore the incidents of SIZE being the most profitable bank are not that surprising.

What is not so clear within the tests of economic value is why the four market models perform so inconsistently, with banks utilising decision making based upon simple factors, such as the size of a company, gaining comparatively larger returns for their investment. In the first two tests these models perform relatively well and in keeping with my prior findings. In the third (elimination) test, once the bank using book-to-market as its tool was eliminated, three out of the four market models quickly follow, leaving just the Bharath and Shumway (2008) model to survive into third place.

With the exception of the book-to-market model, in terms of economic value, the ability of insolvency and credit risk models to generate positive returns within a period of global financial turmoil is apparent. The scale of these returns and how they might be affected year-on-year is a matter for future research and something I wish to pursue. It would be an interesting exercise to map and measure the performance of each bank over several years of isolated data, and determine whether the return trends would match that which we might see within the real UK economy. This matter is discussed further in section 12.4 (Difficulties, limitations and suggestions for further research).

11.2 Conclusions and achievement of objectives

The overriding objective of this thesis was to provide an independent discussion and empirical analysis of extant corporate insolvency prediction technologies, in an attempt to discover if there is a single technology, methodology, or set of variables which outperforms all of its counterparts both in terms of model accuracy and the economic value which they deliver. In doing so, I hoped to be able to gain an impartial view of extant models and determine their relevance, efficiency, and applicability to the uncertain financial climate which is the early twenty first century, and whether these types of model can be relied upon in times of economic turmoil. As a further consequence it was hoped that this thesis may act as a catalogue of insolvency prediction and credit risk literature, covering the key developments throughout the evolution of the art, and pin-pointing its current state.

To this end, I believe that the objectives and research questions discussed within section 6 have been answered effectively. Given the potential scope and direction in which this thesis could have taken, and the voluminous amount of literature available for review, I have been able to adequately condense the information into a single document. By separating the literature into sub-suctions based upon a chronological timeframe, I have been able to structure the literature and its developments into readily identifiable passages reflecting the changes in the art through the decades, which is something that has not been apparent in the literature before.

This thesis adds to the literature upon the relative efficiency of different models for insolvency prediction, using post-IFRS data from the UK, providing comparisons between broader range of contingent claims and accounting number-based models than are found in earlier extant studies, and using receiver operating characteristic curves and tests of economic value as the basis for those comparisons. This thesis also examines, for the first time, the distributional properties of model accuracy when the training and validation samples vary in formation using Monte Carlo simulation of receiver operating characteristic curve accuracies.

The results of both the ROC analysis and the tests of economic value delivery, provide a range of fortunes for different models which are examined. As a general rule, it is apparent that as the field of insolvency prediction and credit risk modelling evolves over time, the models are better able to distinguish between the failed and non-failed firms. Single variables (with the exception of SIZE) on average perform the least well, with multivariate models based upon simple statistical techniques such as MDA and logit providing further predictive ability. The accuracy of such models is able to be improved upon by applying more complex statistical techniques as represented in this thesis by Neural Networks, although the margin of improvement is not as great as one might expect. When evaluating the performance of the latest credit risk (market model) approaches, it is evident that these types of model are generally superior to those based on accounting numbers alone.

I posited the question whether we are any closer to developing a perfect insolvency prediction model than we were 45 years ago after the publication of Beaver's seminal work on single variables as predictors. The 45 year-old single ratio of Beaver (cash-flow to total debt) shows a reasonable predictive accuracy of 67%, in the chronologically split holdout ROC tests, fairing slightly better in the Monte Carlo simulation. With an average accuracy of 69.5% however, it is a score residing at the lower end of the spectrum of results. These results would therefore suggest that improvements in insolvency prediction over the past few decades has been quite substantial, which is perhaps what one would expect given the importance and economic significance of the field, coupled with the vast amount of literature which accompanies it.

Given my findings overall, I suggest that insolvency prediction based upon accounting numbers alone, and given that the Altman (1968), Ohlson (1980) and Taffler (1983) models are now over a quarter of a century old, give way to techniques which consider market-based information alongside accounting numbers; including research, development and implementation of market-based, contingent claims models. It has been demonstrated that even with complex statistical techniques such as the Neural Network; the use of accounting numbers alone is not enough to adequately distinguish between failed and non-failed companies, with these types of models consistently being out-performed by market models under ROC analysis.

When investigating the economic value of insolvency and credit risk models, the simple economies generated within this study provide evidence that lending decisions based upon a single variable can be just as profitable (and indeed more profitable) than those decisions based upon more complex modelling techniques. This suggests that the size of a company is one of the most important factors when it comes to firm survival.

11.3 Implications

Within the introduction of this thesis I posited the question as to whether these types of models can be relied upon in times of economic turmoil, suggesting a need for accurate and reliable credit scoring models to aid loan decisions. With regards to the recent financial crisis there has been an obvious and well document drying up of credit where banks and other businesses are reluctant to lend monies, and will only tend to do so to the most credit worthy of companies. It is therefore important to understand whether existing insolvency prediction and credit risk models can accurately determine the credit worthiness of their potential customers.

It is somewhat impossible to gauge exactly what proprietary methodologies and systems are put in place by lending institutions to evaluate the credit worthiness of their potential customers. This thesis has demonstrated that there is a large variety of modelling methodologies available to the lender and that the greatest predictive accuracies can be achieved by using a combination of different approaches (e.g. using the default probability obtained through a market-style model as variable in conjunction with measures of size, profitability and liquidity).

Although the results within this thesis show a mixture of fortunes for the different models tested, the credit scoring methodologies based upon the option pricing formulae do a good job in distinguishing between failed and non-failed firms. Given that these methods are readily used by ratings institutions, albeit mapped to proprietary default frequencies instead of the normal approximation (used by the models within this thesis), and combined with other proprietary data, it is reasonable to suggest that credit ratings issued to public companies are a reliable indication of their credit worthiness. My findings therefore can act as some reassurance to both lenders and investors, in that these types of model continue to provide accurate predictions of insolvency. Given that these ratings institutions have come under the spotlight in recent times due to their instrumental influence upon market prices and underlying economic factors, this is an important and reassuring finding as the global economy seeks to promote financial stability and economic growth.

11.4 Difficulties, limitations and suggestions for further research

11.4.1 Addition of technologies

Whilst every effort was undertaken to ensure that this thesis represented an all-encompassing review and analysis of the prior insolvency prediction and credit risk literature, it is evident, given the voluminous prior research, that it would be impossible to cover every aspect of the art in meticulous detail. Section 8 of this thesis detailed some twenty-five different technologies which have been used in prior studies in order to distinguish between failed and non-failed firms. Whilst the empirical analysis within this thesis has covered sixteen different models for comparison, more than any other study to date, these models only represent five of these differing approaches, namely Univariate, MDA, logit, Neural Networks, and Contingent Claims, although these technologies do represent the vast majority of the literature in terms of individual models.

To incorporate and represent each of the twenty-five different technologies within an empirical analysis would inevitably be a considerable undertaking, and one which I do not believe to be possible within a realistic timeframe. The inclusion of a small number of additional technologies may be beneficial, for example, the gambler's ruin approach would be an excellent addition given its sound theoretical underpinning. Similarly alternate AIES technologies could be included such as Genetic Algorithms and Support Vector machines, which might possibly be able to surpass the predictive ability of the simple Neural Networks used within this thesis, given their increased complexity. Such technologies would however require further specialised software and the incorporation of results into both the ROC and economic value models; something which I am sure would be overly time-consuming with respect to this current research project, particularly with regards to the distributional properties contained within section 9.4.

A further consideration might be to include additional extant models to the population which I test. The literature review of section 5 detailed several of these models which might warrant an inclusion (Edminster, 1972; Deakin, 1972; Blum, 1974; Altman et al., 1977, for example). Zmejewski's (1984) probit model is one that in hindsight would have been relatively straight forward to incorporate into the study, and would represent an additional technology. Likewise, a mixed logit model in the vein of Jones and Hensher (2004) would also have been beneficial, although more complex and time-restricted in comparison. Given the vast number of different models that have been presented in the literature and the similarity in both variables and modelling technology, particularly through the 1970s and 1980s, it would, again, require a huge undertaking which may not yield any insightful results. I therefore feel that my inclusion of several of these models, and possibly the best known (i.e. Beaver, 1966; Altman, 1968; Ohlson, 1980; and Taffler, 1984), provides a fair depiction of the art throughout this period of time. Future studies may of course wish to expand this model selection further.

11.4.2 Distributional properties of neural network models

Section 9.4 investigated the distributional properties of insolvency prediction and credit risk models. One regrettable limitation of this analysis was the omission of the three Neural Network models AZN, TZN, and OLN. Within the chronological training and validation sample split of section 9.2, each of these models performed (slightly) better than their original counterpart suggesting that the Neural Network technology was superior to that of both MDA and logit in the separation of failed and non-failed firms. Unfortunately, time constraints meant that the running of 30,000 Neural Networks (10,000 for each of the three models), could not be achieved. If I were to successfully design and implement a statistical program which would randomly split the sample of firms into two, train the Neural Network upon the training sample, apply the network to the validation sample, run a ROC analysis, collect the results and repeat 10,000 times, I would expect this process to take approximately 34 days. This timeframe is calculated by using the average time for one manually-run Neural Network (approximately 5 mins) as used within the chronologically split analysis of section 9.2.

Further investigation is therefore warranted in order to determine a more efficient way of generating multiple Neural Network results within a Monte Carlo simulation. If this were to be achieved, I feel that it would contribute a great deal extra to the results already delivered by this thesis, with regards to the full extent of the increased accuracy gained through Neural Network implementation over that of the original technology.

11.4.3 Data limitations and representativeness of sample

Section 7 of this thesis described the data collection process which was not without its share of difficulties, highlighting several limitations which have been imposed upon this study. Firstly, as this is study of UK listed companies, the population of firms is one which is limited, particularly if compared to the population of firms within the US. Using LSPD I was available to identify 2244 non-failed and 178 failed firms eligible for inclusion within the empirical analysis. When compared to Sobehart and Stein (2000) for example whose study utilises 13,041 non-failed and 1,406 failed companies, this limitation in population size is effectively highlighted, but of course one which is inherently unavoidable.

It was decided that a maximum of 10 years of data for each non-failed firm was to be collected (firm years) within the period January 2000 – December 2009. This resulted in the collection of 12,855 non-failed firm years, approximately 5.7 firm years per company, and questions the completeness of the Thomson One Banker (T1B) database. If we are to naively believe that the T1B database has complete records for all companies listed in the UK, this would suggest that the average age of a PLC is 5.7 reporting years, something I find difficult to comprehend. It would therefore be sensible to suggest that this database is, perhaps, only 57% complete.

Further issues arise when we investigate the firm year data in more detail. For the empirical analysis, it was necessary for a fair comparison of the models, that a firm year was only to be included if it contained all of the required financial information to construct each of the sixteen models. From the 12,855 potential observations, only 6,595 contained all of the required data (51%), which in turn means that only 29% (51% of 59%) of the data sought for the study was adequate enough to be used for empirical analysis. This observation is slightly worrying now that I have had time to reflect on the data collection process, and question whether T1B was the correct choice as a data source. Given that T1B collect data from various sources including Datastream and Worldscope, and purports to be one of the largest providers of UK company data, one also wonders whether an alternate database would provide any additional observations and advantages. Given the finding of studies such as Lara et al (2006), empirical evidence exists to support a notion that T1B is the most complete database, with regards to UK companies, and at the time of writing this thesis therefore, the best selection that could have been made.

The largest reduction in firm year data required for this study was undoubtedly attributable to the data item Total Debt (or total liabilities). Out of the 6,260 firm years which were eliminated due to missing data, approximately half (2,893) of these were eliminated because the total debt value returned as zero[154]. The fact that so many of the firm years reported to have a debt value of zero raises a number of interesting points. Firstly is the value of debt really zero? We might expect some companies not to have any debt but it is unlikely that 22% (2,893/12,855) of the procured sample of observations have no debt. This raises issues as to how we are to treat these items. If the company has indeed zero debt, it is likely to be in a rather healthy and solvent position, or alternately the debt may be on the balance sheet of one of its subsidiaries. The only method available to check whether or not these zero debt observations are correct would be the manual inspection of each company's annual reports, something very time

[154] A subsequent examination of these 2,893 missing/zero debt values involved the re-collection of these items through Thomson Financial in an attempt to ascertain errors in the original collection process. This re-collection (some 24 months after the original collection process) yielded values for 418 observations (14.7%) which were previously unobservable. The re-collection of the Total Debt values, and the fact that there are still some 19% of desired observations within the potential sample, highlights some significant shortcomings/limitations in the Thomson Financial database.

consuming when we are talking about some 2,893 firm year observations. Nevertheless, a small sample of 20 companies was selected at random to be included in a manual collect of the total debt figure through annual reports obtained through the Perfect Information Navigator (PIN) database.

The results of this small investigation indicated that eleven out the twenty companies reported zero total debt in their annual reports, suggesting that potentially half of the missing values may be discoverable via manual collection. In practice however, this is something that would take a considerable amount of time.

In the cases where zero debt values are true, and not simply an error or missing value on the database, then this raises a further issue problematic to models marked for analysis. Many of these models contain formulations or ratios which depend upon total debt as a denominator, which is of course problematic when dividing by zero. Therefore the only option available for the treatment of these zero debt values, with respect to this empirical study at least, and not withstanding recollection from a different source, is to eliminate these observations all together.

To satisfy that the data sample used within my study provides a fair representativeness of the wider potential population of firms, the characteristics of these firms with zero debt values (2,893 observations) are compared to those firms which are used throughout this thesis as the sample of firms to which all data requirements were satisfied (6595 observations). The characteristics of the final sample of firms are also compared to the results reported by Lara et al (2006) in which several key statistics are compared across multiple databases.

Table 11.1: Comparison of included and excluded samples

	Excluded sample		**Included sample**		
Variable	**Mean**	**SD**	**Mean**	**SD**	**t-stat**
Total Assets	37.93	122.96	1423.61	8258.90	13.62*
Market value of equity	91.58	381.60	1253.40	6561.81	14.31*
Net Income	0.94	20.75	56.20	728.01	6.16*
Working Capital	10.69	30.87	14.22	574.54	0.50
Profitability (NI/TA)	-0.94	20.49	-0.08	0.54	2.23*

* Significantly different at the 5% level
Values reported in £millions

Table 11.1 compares two groups of firms; the actual (included) sample of firms on which the empirical section of this thesis is conducted, and the sample of firms which the Thomson database reports zero debt values (excluded). The five variables chosen for comparison represent the key characteristics of the two groups in summary; Total Assets and Market value of equity representing company size; Net Income and Net Income / Total assets representing profitability and; Working Capital representing liquidity. Although the liquidity variable demonstrates no difference in mean at the 5% level between the two groups, the same cannot be said for the remaining four variables. Both size variables show a clear difference between those firms which are included in the study and those which are not, the latter being considerably smaller on average. The two profitability measures also indicate clear differences between the included and excluded samples, although when controlled for size, this difference is only just apparent at 5% significance and not with any higher confidence level.

The results therefore show that there are clear differences between the firms which are included and analysed within this thesis, and those which are excluded due to missing values. A consequence therefore is that a certain amount of bias is prevalent in the conclusion and results of this study, but the true extent of this bias, by the very nature of the missing variables, is unobservable.

In a secondary analysis between these two samples I investigated the frequency of industry, to determine whether the exclusion/inclusion of observations has any basis upon its industrial sector. The results of this analysis provide some interesting points for discussion. All observations within the two samples of firms are assigned one of 142 unique LSPD industrial codes. The frequency of each industry within each of the samples can then be calculated along with the resulting proportions. For the majority of cases (138 industries) there was little difference between the proportions contained within the included and excluded samples of firms. There were however 4 industries which showed a difference in proportions, and hence show differences in the make-up of the samples with regard to industry. Table 11.2 presents the proportions of the two samples attributable to each of these 4 industrial classifications.

Table 11.2: Differences in industry proportions

Industry	Proportion of Included Firms	Proportion of Excluded Firms
Exploration & Production	3.1%	12.2%
General Mining	2.3%	9.5%
Gold Mining	1.2%	6.2%
Business Support Services	9.5%	2.6%

The first three rows within table 11.2 can be grouped under the general classification of mining companies. What is interesting is that for the group of firms in which Thomson Financial reports to have zero debt values, the proportion of mining companies within this sample is much larger than that of the "included" population. The sum of these proportions for the excluded sample is 28% indicating that almost one third of this sample are mining and exploration companies when the remaining industrial mining categories are included. This is much larger than the 6.6% of the included sample which are classified equivalently as being within the mining industry. A selection of these non-debt mining companies was chosen at random for a manual search of their annual reports to deduce whether or not these zero debt figures were correct. In each of the 20 reports selected, the companies were indeed found to have zero debt. Whereas I can find no regulatory or reporting explanation for the high percentage of mining companies excluded due to zero debt, one can reasonably explain the cash-rich balance sheets of these companies, given the lucrative nature of the industry, particularly with regards to the large increases in oil, mineral and precious metals prices over the last decade, and deduce that any liabilities have simply been repaid. The large reduction in the proportion of Business Support Services between the two samples presented in table 11.2 would suggest that these companies in contrast, have a greater tendency to deliver a total debt figure to the balance sheet.

The penultimate part of this sub-section compares key characteristics of the final (included) sample of firms used for analysis within this thesis, with those presented by Lara et al (2006). Lara et al (2006) compare summary statistics for a number of data items collected and contrasted across several datasets for a variety of countries. By comparing the final sample used within this thesis with their results, I am able to demonstrate that the sample of firms used for my empirical study is representative of the wider population of firms. Lara et al (2006) provide three data items directly attributable to UK firms across six different databases. Table 11.3 compares these three items (the median values of shareholders' equity, market capitalisation, and net income) to ascertain whether my sample differs to any significant degree to the findings of these authors.

Table 11.3: Comparing the sample to Lara et al (2006)

Item	Thesis sample	Company Analysis	Datastream	Extel	Global Vantage	Thomson Financial	Worldscope
Shareholders' equity	37.0	44.0	45.0	45.4	43.6	42.6	42.6
Market capitalisation	75.9	100.9	99.6	99.6	101.8	101.3	101.3
Net income	2.3	5.9	5.8	5.8	5.9	5.8	5.8

Median values reported in £millions

The results suggest that no significant differences in populations are observable between this study and that of Lara et al (2006). The median of shareholders' equity is slightly lower than each of the values obtained within the prior study. Market capitalisation is perhaps lower than I might expect in comparison, considering that the prior study of Lara et al was conducted upon data a decade prior to my own (1990-1999), but raises no cause for concern within this context. The net income median value is also marginally lower than any of those reported in the prior study of Lara et al (2006) but is likely to be explained by the recessive period encompassing my own dataset. Overall the key characteristics of the datasets, despite the removal of the zero total debt observations, are comparable - suggesting that the sample of firms analysed within this thesis is representative of the UK population of public firms.

A further consideration regards the limitations of the data, which are not database related, but stem from both statutory reporting regulations, and the social and political cultures embedded in society. Insolvency prediction, and any other types of statistical models, can only be as accurate as the data on which they are based. Earnings management for example would undoubtedly have some influence upon the predictive accuracy of these types of model, but is something which is extremely difficult to assess both in terms of frequency and extent. Another factor which has particular relevance to, and impact upon, insolvency prediction and credit scoring models, is the accounting standards on which accounting statements, and the values contained within, are derived in the first instance. The method of presentation and valuation of accounting numbers is intrinsically linked to the guidelines laid out within these standards, which are continuously being reviewed and amended. The numbers obtained and used within this thesis are those reported by UK public companies, and therefore they are required to report in accordance with

International Accounting Standards. The quality of these standards therefore, will have considerable impact upon the accounting numbers which are produced as a result of their implementation. IAS32 (Financial Instruments: Presentation), IAS39 (Financial Instruments: Recognition and Measurement), and IFRS7 (Financial Instruments: Disclosures), have each received significant criticism with regards to their content and guidelines. The main criticisms focus upon the complexity and lack of comparability produced by allowing for the classification of financial assets and liabilities into one of several different categories; and that concerns pertain as to the impairment requirements which delays identification and recognition of potential problems. These issues are sought to be addressed with the mandatory adoption of IFRS9 (Financial Instruments) from January 2015, although classification and measurement issues are unlikely to be fully resolved by this time. The IASB also continue to develop guidance on other issues such as hedge accounting which is seen as being overly complex and confusing. Each of these issues is likely to have considerable impact upon the credit or insolvency risk of any company which uses these financial instruments, as their valuation could potentially show a great deal of fluctuation depending on the recognition and measurement criteria of these assets.

One major change in reporting standards which is highly relevant to this thesis is the change from UK GAAP to IFRS in 2005. Real world application of insolvency prediction and credit risk models to UK companies will gradually see these models being developed and adjusted over time, as more IFRS data becomes available. The timesplit analysis of this thesis (Section 9.2) sought to replicate the real work conditions faced by lending institutions (where model training was required upon UK GAAP data and applied to IFRS data). Those models which did not require any form of training were validated upon IFRS data only, and no UK GAAP data was introduced at this point. The results in terms of model accuracy for the non-trainable models (and those which did require training), were not significantly different to those results gained through Monte Carlo simulations (Section 9.4) which combine both UK GAAP and IFRS data for training and validation purposes. It is therefore evident that the change in accounting standards in 2005 did not (at least in the case for this particular sample of firms), affect the accuracy of any of the models to any significant degree.

11.4.4 Multiple year-ahead forecasts

Many corporate insolvency prediction studies throughout the literature have investigated the predictive ability of models, many years before the failure event occurs. Within the context of the empirical study within this thesis, the primary (and only) focus has been upon predicting failure up to one year before the event. It was my intention when designing this empirical study that I would investigate the one, two, three, four and five-year ahead predictive ability of each model. It quickly became apparent however that the scale of this idealistic thesis would be a vast undertaking. The summary statistics of section 9.1, the ROC analysis of section 9.2, the distributional properties of section 9.4 and the tests of economic value within section 9.5, would all need to be done in quintuplicate with another section combining the multi-layered analysis into something understandable. This is something which I do not think would have been possible within a three-year timeframe.

It is apparent that the majority of lending (in terms of value) would be done on a long term basis, and therefore the lender would be likely to assess the long term solvency position of its potential loan customer, and not simply whether it will fail within the next 12 months. The literature has shown us however, that predicting insolvency more than 12 months before the event is particularly difficult with model accuracies dropping remarkably two, three, four and five years before the event. In this regard it would be of interest to observe the predictive abilities of the sixteen models as the event horizon is increased and something which should be investigated within a future study.

11.5 Afterword

One final thought is in relation to the concept of insolvency prediction and credit scoring itself. When assessing the efficacy of insolvency prediction models we must be mindful of two impactors (which work contrary to one another) on model efficacy which will arise when models are applied in practice, but which have rarely been mentioned in the academic literature. The first is that an efficient model should become the victim of its own success. If it is possible to formulate a prediction model of high accuracy, then efficiency arguments would suggest that the model would come into widespread use to identify failing firms; and allow targeted interventions designed to save the firms from demise. Even a partial success rate for such interventions would lead to the apparent reduced efficacy of the prediction model (and, indeed, cause a higher concentration of failures owing to or correlated with factors not considered by the model). Contrary to this impactor is another: that a model of questionable efficiency may make a self-fulfilling prophesy of firm failure: the mere prediction of failure might cause failure since the prediction leads to withdrawal of funding and trade credit, the migration of staff and a reduced set of strategic options. The practical application of such models should therefore be treated with caution.

12 Appendices

12.1 Appendix A: Member bodies of the joint insolvency committee

- The Institute of Chartered Accountants in England and Wales (ICEAW);
- The Chartered Accountants Regulatory Board;
- The Association of Chartered Certified Accountants;
- The Insolvency Practitioners Association;
- The Secretary of State for Business, Enterprise and Regulatory Reform;
- The Solicitors Regulation Authority of the Law Society.

12.2 Appendix B: Formal procedures of the five main insolvency processes

The formal definitions contained herein were obtained primarily from Beer & Young which were in turn extracted from the insolvency and enterprise acts. Beer & Young has been established since 1998, and has become a leading company delivering turnaround equity solutions to SME's. It manages the largest and most active network of private investors active in turnaround business finance and offers practical advice to entrepreneurs who face challenging circumstances. http://www.beerandyoung.com

Appendix B.I Company voluntary arrangement

Directors, Administrator or Liquidator	The Directors, Administrator or Liquidator (as appropriate) makes a 'Proposal' to the creditors. The Proposal sets out the proposed terms of the CVA
Insolvency Practitioner	The Insolvency Practitioner agrees to act on the implementation of the Proposal and then becomes known as the Nominee
The Nominee	The Nominee will convene a Creditors Meeting giving notice to all the creditors of the Proposal and providing proxy forms for voting purposes
The Creditors Meeting	The Creditors Meeting is then held and the creditors may approve, reject or propose amendments to the Proposal. The creditors may also choose another Nominee. If a majority of 75% in value of the creditor's present and voting vote in favour, the Proposal becomes binding on all unsecured creditors.
The Supervisor	Following the approval of the Voluntary Arrangement, the Nominee becomes known as the Supervisor and implements the Arrangement in accordance with the terms of the Proposal.
Secured Creditors	A CVA cannot effect the rights of a secured creditor of the company to enforce its security except with the agreement of the secured creditor
Challenge Period	A creditor may challenge the Arrangement within 28 days of its approval if it is of the opinion that it has been unfairly prejudiced by the Proposal or if of the opinion that there is a material misstatement in the Proposal
Sources of Dividend Funding	Normally, funds come from pledging future profits for a two/three year period or from a third-party who wishes to acquire the business

Appendix B.II Administration

Prior to 15 September 2003, the procedure for the appointment of an Administrator solely involved the following:

Petition	A Petition for an Administration Order is presented to the Court. The Petition could be presented by the Company, its directors or creditors
Independent	The Petition was supported by an independent persons report,

Person's Report	usually an Insolvency Practitioner
Notice	Following the Petition, notice was given to any Floating Charge holder who could then consider alternative course of action, such as appointing an Administrative Receiver
Administration Order	If successful, the Administration Order was granted
Administrators Duties	The Administrators duties were to manage the business. Make proposals as to the conduct of the Administration and call a meeting of creditors.
Creditors Meeting	A creditors meeting was held pursuant within three months of the Administration Order

The new legislation provides for a similar procedure to the 'old style' Administration to be utilised in less common circumstances. The Enterprise Act 2003 introduced a new out-of-court procedure, which has largely replaced the 'old-style' Administration. The 'old style' Administration is available where:

- Unsecured creditors wish to make an appointment
- An Administrative Receiver has been appointed
- A Provisional Liquidator has been appointed
- A Winding-Up Petition or Administration application is outstanding
- Administration is not available to companies that have applied for one within the previous 12 months.

Post-Appointment Procedure
Administrator's proposals to creditors
Under the new regime the administrator is required to send his statement of proposals to all known creditors within a period of eight weeks
Notice of the initial creditors meeting is given with the proposals and the meeting must now be held within a period of ten weeks from the date of the administration

Period of administration
Automatically ceases at the end of 12 months but can be extended
If the out-of-court procedure is used and the purpose has been fulfilled, the administration ends when the administrator files the prescribed notice at court
The court may order that the appointment ceases upon the application of a creditor

Statutory exit routes
Creditors' voluntary liquidation
Dissolution
The administrator's proposals may include a proposal for a Section 425 Scheme of arrangement or Company voluntary arrangement.

Appendix B.III Administrative receivership

The Lender	This procedure is available only to a Lender with Floating Charge security granted prior to 15 September 2003
Default	The Lender can appoint an Administrative Receiver when the borrower is in default of or in breach of the terms of the security document. The appointment of a Receiver usually follows a

	demand for repayment
Acceptance of Appointment	The Receiver must formally accept his appointment
The Receiver	The Receiver then notifies Companies House, the Company itself and its creditors. The Receiver must also advertise his appointment in The London Gazette and an appropriate newspaper
Realisation of Assets	The powers and the capacity of the Receiver depend upon the security document which should be referred to in all cases. The security document would normally enable the Receiver to carry on a company's business and realise its assets. Whilst in office, the Receiver acts as the agent of the company until liquidation
Creditors Report	Within three months of his appointment, the Receiver must send a report to the creditors and convene a Creditors Meeting to receiver the report. If, in the meantime, the company has gone into liquidation, the report need only go to the Liquidator
Conclusion of Receivership	The receivership is concluded and the Receiver ceases to act in office when he has realised the security, repaid any preferential creditors (if possible) and the appointer and Companies House are notified.

Appendix B.IV Creditors’ voluntary liquidation

Notice & Proxy Forms	Notice and Proxy Forms must be sent to shareholders and creditors seeking a resolution to place the company into liquidation
Advertising	A Creditors meeting is advertised in The London Gazette and two appropriate newspapers
Shareholders Meeting	A Shareholders Meeting is held first and an Extraordinary Resolution must be passed to wind-up the company and an Ordinary Resolution is passed to appoint a Liquidator
Creditors Meeting	A Creditors Meeting must be held. This must be within 14 days of the passing of the shareholders resolution but normally follows immediately after the Shareholders Meeting. At the Creditors Meeting a Statement of Affairs and report on the history of the business and causes of the insolvency should be presented. The shareholders choice of Liquidator can be changed by the creditors, who can either confirm the shareholders choice or vote for an alternative Liquidator. Voting is by value of debt.
The Liquidator	The duties of a Liquidator are, as with Compulsory Liquidation, to realise the assets, investigate the company's affairs, agree creditors’ claims and distribute the funds

Appendix B.V Compulsory liquidation

Issue of Petition	This is usually presented to the Court by a creditor on the grounds of the company's insolvency which is normally proven by the failure of the company to satisfy a Statutory Demand. It may however also be presented by the company itself of the shareholders. A hearing date will be given by the Court.
Service	The Petition must be served on the company
Advertisement	After the expiration of seven days after service on the company

	the Petition must be advertised, normally in The London Gazette
Winding-Up Order	Unless successfully disputed at the hearing, a Winding-Up Order is made by the Court. The date of the commencement of the liquidation is deemed to be when the Petition is presented and dispositions after that date are void. When a Winding-Up Order has been made or a Provisional Liquidator has been appointed, no action or proceeding may be presented or commenced against the company or its property except by the leave of the Court.
The Official Receiver	The Official Receiver becomes Liquidator and has a duty to investigate the company's affairs and send a report to the creditors. The Official Receiver should advertise the Order in The London Gazette and any appropriate newspapers
Conduct of the Liquidation	If the assets are likely to cover the administrative costs, the Official Receiver will call a creditors meeting to appoint a Liquidator other than himself who

12.3 Appendix C: Theoretical derivation of the MDA methodology

Following the methodology described in Dhrymes (1974), in general it can be said that a company will be characterised, and categorised, by a set of variables and attributes, specified by the vector

$$x = (x_1, x_2, \dots, x_m) \qquad \text{(C.1)}$$

and *if* the company belongs to the *i*th class group, then *x* has a well-defined probability density function $f_i(\cdot)$.

$$x \sim f_i(\cdot) \qquad i = 1,2, \dots, k \qquad \text{(C.2)}$$

where *k* is the number of class groups, (for the case of insolvency prediction this is just two, the firm is either failed or non-failed).

The density function in most cases is taken to be the normal density function, although this is not a necessity upon the theory of the MDA methodology; it does however become beneficial when constructing tests of significance. For future reference the normal density function is provided here.

$$f(x|\, \mu, \sigma^2) = \frac{1}{2\pi\sigma^2}\ exp\left(-\ \frac{(x-\mu)^2}{2\sigma^2}\right) \qquad \text{(C.3)}$$

where μ is the *mean* and σ^2 is the *variance* of the distribution.

In order to solve the classification problem, we first require some rule which classifies a company with particular vector of attributes into one of the classes. We can partition the *m*-dimensional Euclidean sample space into k disjoint sets,

$$E_i, \qquad i = 1,2, \dots, k \qquad \text{(C.4)}$$

such that if a company is characterised by a vector of attributes and is a member of set E_i, then that company will be classified into the *i*th class. During this process, however, it is possible that we may commit the error of classifying a company into the *j*th class when it is actually observed to belong to the *i*th class, and is also reasonable to attach a (positive) cost to this error. We can therefore let the cost of misclassification be denoted as c_{ij} and the cost of a correct prediction be zero.

$$c_{ii} = 0 \qquad i = 1,2,\dots,k \tag{C.5}$$

The probability that a company will be classified into class *j* when it in fact belongs to class *i* is given by

$$p_{ij} = \int_{E_j} f_i(x)dx \tag{C.6}$$

and therefore the expected cost of correctly classifying a company into class *i* is

$$Ci = \sum_{j=1}^{k} c_{ij} p_{ij} \tag{C.7}$$

To find the expected cost of the entire classification procedure, we must weigh the quantities C_i by the probabilities attaching to membership of each class; i.e. C_i is a conditional cost of misclassification given that the company belongs to the *i*th class. The unconditional cost of the procedure is

$$C = \sum_{i=1}^{k} \pi_i C_i = \sum_{j=1}^{k} \int_{E_j} S_j(x)dx \tag{C.8}$$

where

$$S_j(x) = \sum_{i=1}^{k} \pi_i C_{ij} f_i(x) \tag{C.9}$$

and π_i is the probability of membership of the *i*th population.

The problem now is to choose regions E_j so that we minimise the unconditional expected cost of the procedure. It can be shown that if the E_j are chosen so that

$$x \in E_j \qquad j = 1,2,\dots,k \tag{C.10}$$

this implies

$$S_j(x) \le S_i(x) \qquad i = 1,2,\dots,k \tag{C.11}$$

then (C.8) is minimised.

Heuristically we can prove this as follows. Choosing any index, say j_0, we can write

$$C = \int_{E^m} S_{j0}(x)dx + \sum_{j\neq j0} \int_{E_j} \left[S_j(x) - S_{j0}(x)\right] dx \tag{C.12}$$

where E^m denotes the *m*-dimensional Euclidean space. The first term in the right-hand side of equation (C.12) is independent of the choice of regions. Thus *C* will be minimised if the second term is minimised. We can do this if the E_j are defined so that they *include* points *x* for which $S_j(x) - S_{j0}(x) \le 0$ and exclude points for which $S_j(x) - S_{j0}(x) > 0$. But the definition of E_j in (C.10) and (C.11) shows that no matter what the choice of j_0 is, the term $S_j(x) - S_{j0}(x)$ is non-positive over E_j. Therefore the choice of regions minimises C.

The quantity $-S_j(x)$ is termed the *j*th discriminant score of a company with an attribute vector *x*. The classification rule implied by the procedure is as follows: *If a company has an attribute vector x, compute its discriminant scores and assign it to that population for which its discriminant score is highest*. The difficulty with this procedure is that, typically, the probability π_i will not be known and the misclassification cost C_{ij} could not be assessed with great confidence. We can illustrate this procedure for the k =2, failed *vs* non-failed firm example, assuming the densities in (C.2) are $N(\mu^{(i)}, \sum_i)$. The classification criterion can now be expressed:

Classify in E_1 if

$$\frac{S_1(x)}{S_2(x)} \le 1 \tag{C.13}$$

and classify in E_2 otherwise. But

$$\frac{S_1(x)}{S_2(x)} = \frac{\pi_2 c_{21}}{\pi_1 c_{12}} \frac{|\Sigma_1|^{\frac{1}{2}}}{|\Sigma_2|^{\frac{1}{2}}} exp - \frac{1}{2}\{(x - \mu^{(2)})' \Sigma_2^{-1} (x - \mu^{(2)}) - (x - \mu^{(1)})' \Sigma_1^{-1} (x - \mu^{(1)})\}$$

(C.14)

A criterion equivalent to that in (C.14) is

$$\frac{1}{2}\{(x - \mu^{(1)})' \Sigma_1^{-1} (x - \mu^{(1)}) - (x - \mu^{(2)})' \Sigma_2^{-1} (x - \mu^{(2)})\}$$
$$+ \frac{1}{2} ln \frac{|\Sigma_1|}{|\Sigma_2|} \leq ln\left(\frac{\pi_1 c_{12}}{\pi_2 c_{21}}\right) \quad \text{(C.15)}$$

which in the case that the two covariance matrices are equal ($\Sigma_1 = \Sigma_2$), then (C.15) may be reduced to

$$(\mu^{(1)} - \mu^{(2)})' \Sigma^{-1} x \geq \frac{1}{2} (\mu^{(1)} + \mu^{(2)})' \Sigma^{-1} (\mu^{(1)} - \mu^{(2)}) - ln\left(\frac{\pi_1 c_{12}}{\pi_2 c_{21}}\right) \quad \text{(C.16)}$$

Notice that the assumption of a common covariance matrix for the two classes has reduced the criteria from a *quadratic function* in (C.15), (e.g. Altman et al. (1977)), to a *linear function* in (C.16), (e.g. Altman (1968)).

The left hand side of the inequality in (C.16) is commonly known as the *discriminant criterion*[155]. This formulation is in terms of known parameters of the two populations (i.e. the group means and the common covariance matrix), noting that when equal costs of errors and prior probabilities of failure are assumed, the final term on the right-hand side is eliminated. Zavgren (1983) notes that "*Although these assumptions are unrealistic, they are usually made when error costs and prior probabilities are unknown*".

Dhyrmes (p.68-69) shows that for cases where the parameters are not known, and the assumptions of equal error costs and equal prior probabilities is made, the estimator can be written as follows:

$$D^*(x) = (\mu^{(1)} - \mu^{(2)})' \Sigma^{-1} x - \frac{1}{2} (\mu^{(1)} + \mu^{(2)})' \Sigma^{-1} (\mu^{(1)} - \mu^{(2)}) \quad \text{(C.17)}$$

[155] A more direct approach to this special case is provided by Fisher (1938) in which the resulting equations differ only by a constant of proportionality.

which reduced to the rule: Classify in population 1 if $D^*(x) \geq 0$, and to population 2 otherwise.

The problem with the discriminant function formulation thus far is that we typically do not know the parameters of the populations involved. Dhrymes (p.70) suggests that although the above approach has a certain intuitive appeal, it is not clear that it would possess the optimal properties that are known to characterise the procedure indicated by (C.16) when the parameters of the population are known. Therefore, the following alternative procedure for classification, based on the likelihood ratio principle, is presented. This formulation is confined the assumption of two (normal) populations with the same covariance matrix.

Suppose that we have n_1 observations from population1 characterised by the density $N(\mu^{(1)}, \sum)$; denote the ith observation by x_i, $i = 1, 2, \ldots, n_1$. We also have n_2 observations from population 2, whose density is $N(\mu^{(2)}, \sum)$; denote the jth observation by y_j, $j = 1, 2, \ldots, n_2$. Finally we have an observation z whose origin is unknown except for the fact that it belongs to either population 1 or to population 2. Of course, all observations are known to be mutually independent. The problem is how to classify z. Consider the null hypothesis

$$\begin{aligned} H_0&: (x_1, x_2, \ldots, x_{n1}, z) && \text{belong to population 1} \\ &: (y_1, y_2, \ldots, y_{n2}) && \text{belong to population 2} \end{aligned} \qquad \text{(C.18)}$$

as against the alternative

$$\begin{aligned} H_1&: (x_1, x_2, \ldots, x_{n1}) && \text{belong to population 1} \\ &: (y_1, y_2, \ldots, y_{n2}, z) && \text{belong to population 2} \end{aligned} \qquad \text{(C.19)}$$

Using the assumptions already made concerning the densities of the two populations, the maximum likelihood estimators of $\mu^{(1)}, \mu^{(2)},$ and $\sum$ under the null hypothesis are

$$\hat{\mu}^{(1)} = \frac{z+\sum_{i=1}^{n_1} x_i}{n_1+1} \qquad \hat{\mu}^{(2)} = \frac{\sum_{j=1}^{n_2} y_j}{n_2}$$

$$\widehat{\sum}_{H_0} = \frac{1}{n_1+n_2+1}\Big[\sum_{i=1}^{n_1}\big(x_i - \hat{\mu}^{(1)}\big)\big(x_i - \hat{\mu}^{(1)}\big)' + \big(z_i - \hat{\mu}^{(1)}\big)\big(z_i - \hat{\mu}^{(1)}\big)'$$
$$+\sum_{j=1}^{n_2}\big(y_j - \hat{\mu}^{(2)}\big)\big(y_j - \hat{\mu}^{(2)}\big)'\Big] \qquad \text{(C.20)}$$

It follows, therefore, that the maximum of the *likelihood function*, L(z, x_1, x_2, ..., x_{n1}, y_1, y_2, ..., y_{n2} ; $\mu^{(1)}, \mu^{(2)}, \sum$), under the null hypothesis is given by

$$\mathrm{H_0 Max\ L}\big(\mathrm{z} \ldots ;\ \mu^{(1)}, \mu^{(2)}, \Sigma\big) = (2\pi)^{-\frac{1}{2}m(n_1+n_2+1)}\left|\widehat{\sum}_{H_0}\right|^{-\frac{1}{2}(n_1+n_2+1)}$$
$$*\ \exp\left[-\tfrac{1}{2}m(n_1+n_2+1)\right]$$

(C.21)

Under the alternative hypothesis, the maximum likelihood estimators of $\mu^{(1)}, \mu^{(2)},$ and $\sum$ are given by

$$\hat{\mu}^{(1)} = \frac{\sum_{i=1}^{n_1} x_i}{n_1} \qquad \hat{\mu}^{(2)} = \frac{z+\sum_{j=1}^{n_2} y_j}{n_2+1}$$
$$\widehat{\sum}_{H_1} = \frac{1}{n_1+n_2+1}\Big[\sum_{i=1}^{n_1}\big(x_i - \hat{\mu}^{(1)}\big)\big(x_i - \hat{\mu}^{(1)}\big)'$$
$$+\sum_{j=1}^{n_2}\big(y_j - \hat{\mu}^{(2)}\big)\big(y_j - \hat{\mu}^{(2)}\big)' + \big(z_i - \hat{\mu}^{(2)}\big)\big(z_i - \hat{\mu}^{(2)}\big)'\Big] \qquad \text{(C.22)}$$

Thus

$$\mathrm{H_1 Max\ L}\big(\mathrm{z} \ldots ;\ \mu^{(1)}, \mu^{(2)}, \Sigma\big) = (2\pi)^{-\frac{1}{2}m(n_1+n_2+1)}\left|\widehat{\sum}_{H_1}\right|^{-\frac{1}{2}(n_1+n_2+1)}$$
$$*\ \exp\left[-\tfrac{1}{2}m(n_1+n_2+1)\right]$$

(C.23)

It therefore follows that that the likelihood ratio is

$$\lambda = \frac{\left|A + \frac{n_2}{n_2+1}(z-\bar{y})(z-\bar{y})'\right|}{\left|A + \frac{n_1}{n_1+1}(z-\bar{x})(z-\bar{x})'\right|}$$

$$= \frac{\left|1 + \frac{n_2}{n_2+1}(z-\bar{y})'A^{-1}(z-\bar{y})\right|}{\left|1 + \frac{n_1}{n_1+1}(z-\bar{x})'A^{-1}(z-\bar{x})\right|}$$

(C.24)

where

$$A = \sum_{i=1}^{n_1} (x_i - \bar{x})(x_i - \bar{x})' + \sum_{j=1}^{n_2} (y_j - \bar{y})(y_j - \bar{y})$$

$$\bar{x} = \frac{\sum_{i=1}^{n_1} x_i}{n_1} \,, \qquad \bar{y} = \frac{\sum_{j=1}^{n_2} y_j}{n_2} \tag{C.25}$$

12.4 Appendix D: A brief guide to gambler's ruin theory

For reference, the text within this section is extracted from several sources to summarise the gambler's ruin theory. The sources include Wilcox (1976), Scott (1981) and Altman (1983). The gambler's ruin strand of insolvency prediction is fairly uncommon today due, in part, to the lack of convincing empirical test results achieved in past studies.

The gambler in this theory begins the game with an arbitrary amount of money. He wins a dollar with probability P or loses a dollar with probability Q (1 - P). The game continues until the gambler loses all his money or until his component loses his. The theory based on this simple scenario is well developed; there are expressions for the probability of ultimate ruin, for the gambler's expected ultimate gain or loss, for the expected duration of the game, and so on. In financial applications, the firm is viewed as the gambler, and bankruptcy occurs when the firm's net worth falls to zero. Scott (1981, pp. 322-333).

The following summary of Wilcox (1976) largely follows Altman (1983, pp. 161-163). Wilcox adapted the classic gambler's ruin problem (Feller, 1968) to measuring business risk and focused upon net liquidation value (NLV) and the factors which cause it to fluctuate. NLV is simply a dollar level determined by liquidity inflow and outflow rates. For a given period, Wilcox defined the inflow rate as net income minus dividends and the outflow rate as the increase in book value of assets minus the increase in the liquidation value of those assets. Using the familiar bathtub (or following the course of this thesis, the liquid reservoir) analogy, NLV is the water level in the tub (reservoir) with both its spigot and drain open. When the inflow rate exceeds the outflow rate, NLV increases; when the opposite occurs, NLV decreases. Combining inflow and outflow variables, Wilcox referred to this net flow as the "adjusted cash-flow," the period-to-period change in NLV, excluding stock issues, accounting changes, and other non-continuing processes.

The main concern of the model was with the predicting when NLV will be negative, which often portends bankruptcy. Assuming a "stable process," Wilcox postulated that the probability of NLV being reduced to zero (insolvency) is a f unction of (1) the current NLV or current wealth, (2) the average adjusted cash-flow, and (3) the variability of the adjusted cash-flow, measured by its variance, σ^2. To summarise,

$$P_r(NLV < 0) = f(NLV, \mu, \sigma^2). \tag{D.1}$$

Other things being equal, the smaller the NLV, the smaller the adjusted cash-flow, and the larger the variation of the adjusted cash-flow, the greater the chance of failure (P_r). To determine how much NLV and average adjusted cash-flow are needed for a given degree of safety, Wilcox introduced a concept called the "size of bet" (S). He interprets S as the adjusted cash-flow at risk each year or as the simplest probabilistic process underlying the NLV cash-flows.

In terms of the gambler's ruin model, NLV = NS (where N is the number of states away from failure and S is the size of the bet) = $\mu^2 + \sigma^2$. As an approximation, Wilcox measured S^2 as the average of the squared annual changes in NLV. the firm was viewed at time t, as being in state N with wealth of NLV = NS. In period $t + 1$, the firm (the gambler) moves either to state $N - 1$ with wealth of (NLV – S) or to state $N +1$ with wealth of (NLV +S). In period $t + 2$, the firm moves from its new state with a change in wealth of either + S or – S depending upon whether it "wins" or "loses". If the probability of winning is denoted by P and the probability of losing by Q (or $1 - P$), the probability of ruin (failure) is:

$$P_r(F) = (Q/P)^n \tag{D.2}$$

If $P \leq Q$ then failure is inevitable.

A firms average winnings per period, μ, is equal to S ($P - Q$). Defining μ/S as X, the ratio Q/P is equal to $(1 - X) / (1 + X)$. Thus,

$$P_r(F) = \left(\frac{1 - X^N}{1 - X}\right) \cong 1 - 2XN. \tag{D.3}$$

While Wilcox conceded that the gambler's ruin model (in the business failure context) is an extreme over-simplification, he argued that the parameters *X* and *N* are fundamental indicators of relative financial risk, especially if they are based upon five years of adjusted cash-flows.

12.5 Appendix E: Derivation of the conditional probability models

The following derivation is adapted from the proof which is found within Gujarati (2003).

Within this thesis I focus upon two different forms of the conditional probability model (CPM), the logit and the probit. Both models are derived from the Linear Probability Model (LPM), (E.1), (also a CPM), which in the first instance resembles a traditional linear regression. For simplicity we can use a single dependant variable X_i.

$$Y_i = \beta_1 + \beta_2 X_i + \mu_i \qquad \text{(E.1)}$$

However, because the independent regressand is a dichotomous variable (taking the value 1 for failed or 0 for non-failed), the model is called a LPM. The reason for this is because the conditional expectation of Y_i given X_i, $E(Y_i|X_i)$, can be interpreted as the *conditional probability* that firm failure will occur given X_i, that is, Pr ($Y_i = 1|X_i$). If we assume that the expected errors are zero and that the model is an unbiased estimator then we obtain

$$E(Y_i|X_i) = \beta_1 + \beta_2 X_i \qquad \text{(E.2)}$$

Now, if P_i is the probability that $Y_i = 1$(that is, failure occurs), and $(1 - P_i)$ is the probability that $Y_i = 0$ (failure does not occur), then Y_i follows the Bernoulli probability distribution. Therefore, by the definition of mathematical expectation we obtain:

$$E(Y_i) = 0(1 - P_i) + 1(P_i) = P_i \qquad \text{(E.3)}$$

and therefore

$$E(Y_i|X_i) = \beta_1 + \beta_2 X_i = P_i \qquad \text{(E.4)}$$

that is, the conditional expectation of the model (E.1) can be interpreted as the conditional probability of Y_i where, in general, the expectation of a Bernoulli random variable is the probability that the random variable equals 1. Since the probability P_i must lie between 0 and 1, we have the restriction

$$0 \geq E(Y_i|X_i) \leq 1 \tag{E.5}$$

that is, that the conditional expectation (or conditional probability) must lie between 0 and 1.

Additionally, the probability model requires two key features: (1) As X_i increases, $P_i = E(Y = 1|X)$ increases but never breaks the bounds of the 0-1 interval, and (2) the relationship between P_i and X_i is non-linear where X_i approaches zero at slower and slower rates as X_i decreases, and approaches one at slower and slower rates as X_i gets very large. Geometrically the model which we desire would resemble something like we have in Figure E.1.

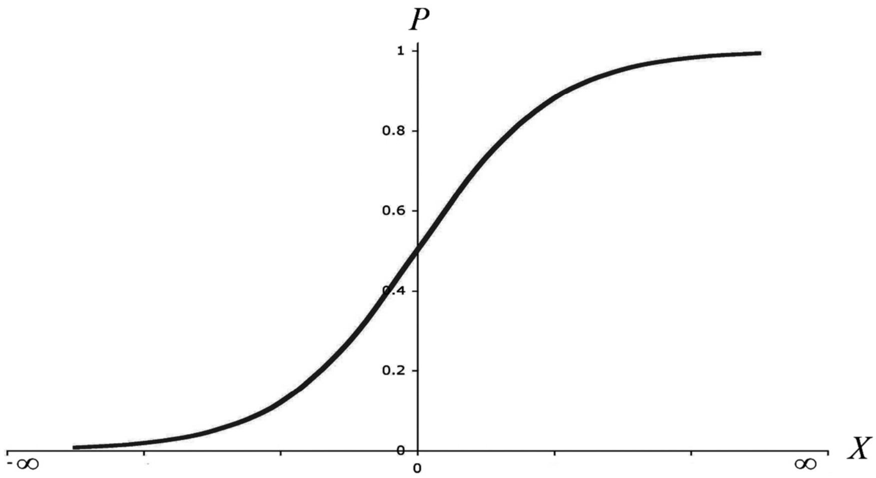


Figure E.1: A cumulative distribution function.

The S-shaped sigmoid curve very much resembles the cumulative distribution function (CDF) of a random variable, and therefore we can easily use the CDF to model regressions where the independent variable is dichotomous. For each random variable there is a unique CDF so we therefore need to establish which of these we are to use. For historical as well as practical reasons, the CDF's commonly chosen to represent 0-1 response models are the logistic and the normal, where the logistic CDF gives rise to the logit model and the normal CDF gives rise to the probit (or normit) model.

12.5.1 The logit model

The (cumulative) logistic distribution function (LDF) has the following form

$$CDF = \frac{1}{1 + e^{-z_i}} = \frac{e^z}{1 + e^z} \tag{E.6}$$

where $Z_i = \beta_1 + \beta_2 X_i$

Therefore,

$$P_i = E(Y = 1|X) = \frac{1}{1 + e^{-(\beta_1 + \beta_2 X_i)}} = \frac{1}{1 + e^{-z_i}} = \frac{e^z}{1 + e^z} \tag{E.7}$$

It is easy to verify that as Z_i ranges from -∞ to +∞, P_i ranges between 0 and 1and that P_i is non-linearly related to Z_i (i.e., X_i), thus satisfying the two model requirements. In satisfying these requirements we have inadvertently created an estimation problem because P_i is non-linear not only in X but also in the β's as can be seen clearly in (E.7). This inherent problem means that we cannot simply use Ordinary Least Squares (OLS) estimation to estimate the parameters of the model, but instead we must use the maximum likelihood (ML) estimation method. This problem is more apparent than real however, because (E.7) can be linearised as follows:

If P_i, the probability of firm failure, is given by (E.7), then $(1 - P_i)$, the probability of firm survival, is

$$1 - P_i = \frac{1}{1 + e^{z_i}} \tag{E.8}$$

Therefore we can write

$$\frac{P_i}{1-P_i} = \frac{1+e^{z_i}}{1+e^{-z_i}} = e^{z_i}$$
(E.9)

Now P_i / $(1 - P_i)$ is simply the odds ratio – the ratio of the probability that a firm will fail to the probability that a firm will survive. Thus, if P_i = 0.8, it means that odds are 4 to 1 in favour of the firm becoming insolvent. Now if we take the natural log of (E.9), we obtain a very interesting result, namely,

$$L_i = \ln\left(\frac{P_i}{1-P_i}\right) = Z_i = \beta_1 + \beta_2 X_i$$
(E.10)

that is, *L*, the log of the odds ratio, is not only linear in *X*, but also (from the estimation viewpoint) linear in the parameters. *L* is called the **logit**, and hence the name **logit model**.

To estimate (E.10), we need, apart from X_i, the values of the regressand or logit L_i. As we have data on the individual firm level, OLS estimation of (E.10) is infeasible, and is easy to demonstrate. $P_i = 1$ if the firm has failed and $P_i = 0$ if the firm survives, but if we put these values directly into the logit L_i, we obtain:

$$L_i = \ln\left(\frac{1}{0}\right) \text{ if a firm has failed}$$

$$L_i = \ln\left(\frac{0}{1}\right) \text{ if a firm survives}$$

Obviously, these expressions are meaningless and we therefore must us ML estimation thus:

Suppose we have a random sample of *n* firm observations. Letting $f_i(Y_i)$ denote the probability that Y_i = 0 or 1, and the joint probability of observing the *n* *Y* values, i.e. $f_i(Y_1, Y_2, \ldots Y_n)$ is given as:

$$f_i(Y_1,\ Y_2, \dots Y_n) = \prod_{1}^{n} f_i(Y_i) = \prod_{1}^{n} P_i^{\,Y_i}\,(1 - P_i)^{1-Y_i}$$

(E.11)

Note that we can write the joint probability density function as a product of individual density functions because each Y_i is drawn independently and each Y_i has the same (logistic) density function. The joint probability given in (E.11) is known as **the likelihood function** (LF). (E.11) is a tad awkward to manipulate, but if we take its natural logarithm, we obtain what is called the **log likelihood function** (LLF).

$$\ln f_i(Y_1,\ Y_2, \dots Y_n) = \sum_{1}^{n} [\,Y_i \ln P_i + (1 - Y_i)\ln(1 - P_i)]$$

$$= \sum_{1}^{n} [\,Y_i \ln P_i - Y_i \ln(1 - P_i) + \ln(1 - P_i)]$$

$$= \sum_{1}^{n} \left[Y_i \ln\left(\frac{P_i}{1 - P_i}\right)\right] + \sum_{1}^{n} \ln(1 - P_i)$$

(E.12)

From (E.7) it is easy to verify that

$$(1 - P_i) = \frac{1}{1 + e^{(\beta_1 + \beta_2 X_i)}}$$

(E.13)

as well as

$$\ln\left(\frac{P_i}{1 - P_i}\right) = \beta_1 + \beta_2 X_i$$

(E.14)

Using (E.13) and (E.14) we can write the LLF (E.12) as:

$$\ln f_i(Y_1,\ Y_2, \dots Y_n) = \sum_{1}^{n} Y_i(\beta_1 + \beta_2 X_i) - \sum_{1}^{n} \ln\left[1 + e^{(\beta_1 + \beta_2 X_i)}\right]$$

(E.15)

We can see from (E.15), the LLF is a function of the parameters $\beta_1$and β_2, since the X_i are known.

In ML our objective is to maximise this LLF, that is, to obtain the values of the unknown parameters in such a manner that the probability of observing the given Y's is as high (maximum) as possible. To maximise this function we must differentiate (E.15) partially with respect to each unknown, set the resulting expression to zero and solve the resulting expressions. One can then apply the second-order condition of maximisation to verify that the values of the parameters we have obtained do in fact maximise the LLF. In reality, however, in solving these expressions, the reader will quickly realise that the resulting expressions become highly non-linear in the parameters and no explicit solutions can be obtained. We can, however, solve the LLF by using one of several non-linear estimation techniques.[156]

For this particular study, the LLF estimation of some 13,000 observations is, of course, impractical to solve by hand. I adopted two methods (for verification purposes) to solve the logit models within this study. Firstly, by simply using the SAS procedure PROC_LOGISTIC, the parameters of each model were easily estimated. To verify these results it is, again, fairly simple to construct an Excel spread sheet which makes use of the SOLVER function to maximise the LLF.

[156] Further discussion of these can be found in Gujarati (2003).

12.5.2 The probit model

As previously noted, to explain the behaviour of a dichotomous dependant variable, we require the use of a suitably chosen CDF. Whereas the logit model uses the LDF (E.6), an alternative CDF which has been proven useful is the normal CDF. The estimating model which emerges from the normal CDF is the **probit**. Written explicitly in the continuation of the corporate failure example, the standard normal CDF is:

$$\mathrm{CDF} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(\beta_1+\beta_2 X_i)} e^{-z^2/2}\, \mathrm{d}z \tag{E.16}$$

In principle we can substitute the normal CDF in place of the LDF as per (E.7) and proceed in exactly the same fashion. It is however, widely recognised that the resulting ML estimation expression becomes rather more complicated in comparison to what has previously derived for the logit model. I will therefore not pursue this any further as I do not use any probit estimations within the empirical section of this thesis. An alternative and more manageable form of the probit estimation methodology can be found in McFadden (1973) and is based upon utility theory, or rational choice perspective on behaviour.

12.6 Appendix F: A brief guide to neural networks and back-propagation

Much of this discussion is adapted or re-quoted from Fletcher and Goss (1993) which is in-turn a discussion of Rumelhart et al. (1986). A full and concise derivation of the back-propagation algorithm can be found in Rojas (1996), which unfortunately is deemed to be too lengthy to reproduce here in its entirety. This Appendix therefor is designed to give the reader a brief overview of the feed-forward neural network with back-propagation training algorithm in order to aid his or her understanding of some of the processes involved within this thesis.

The feed-forward process

In a back-propagation neural network, as developed by Rumelhart et al. (1986), independent variable patterns, or values, are normalised between zero and one at the input layer to produce the signal O_i prior to presentation to the hidden node layer. Each connection between the input layer and a hidden node has an associated weight W_{ij}. The net signal I_j to an individual hidden node is expressed as the sum of all the connections between the input layer nodes and that particular hidden node plus the connection value W_{gj} from a bias node. This relationship may be expressed as:

$$\mathrm{I}_j = \sum W_{ij} O_i + W_{gj} \tag{F.1}$$

The signal from the hidden layer is then processed with a sigmoid function which again normalises the values between 0 and 1 to produce O_j prior to being sent to the output layer. The normalisation procedure is performed according to:

$$\mathrm{O}_j = \frac{1}{1 + \exp(-\mathrm{I}_j)} \tag{F.2}$$

The net signal to an output node I_k is the sum of all connections between hidden layer nodes and the respective output node, expressed as:

$$\mathrm{I}_k = \sum W_{jk} O_j + W_{gk}$$

(F.3)

where W_{gk} represents a single connection weight from a bias node with a value of 1 to the output layer. The net signal is again normalised with the sigmoid function to produce the final output value O_k, where

$$\mathrm{O}_k = \frac{1}{1 + \exp(-\mathrm{I}_k)}$$

(F.4)

In terms of corporate insolvency prediction modelling, O_k represents the probability of failure for the individual firm.

The process of error back-propagation

At the output layer the net signal, O_k, (the estimated probability of failure, or dependent variable), is compared to the actual value of the dependant variable, T_k, to produce an error signal which is propagated back through the network. Widrow and Hoff (1960) were the first to develop a learning rule based upon least mean squares, which became known as the Generalised Delta Rule, where the output layer error signals are propagated back through the network to perform appropriate weight adjustments after each pattern presentation. Rumelhart et al. (1986) describe the process of weight adjustment by:

$$\Delta W_{ij}(\mathrm{n} + 1) = \eta \delta_{pk} O_{pj} + \alpha \Delta W_{ik}(\mathrm{n})$$

(F.5)

where η is a learning coefficient and α is a “momentum” factor. The momentum factor determines the effect of past weight changes on the current direction of movement in weight space and proportions the amount of the last weight change to be added into the new weight change.

The error signal δ, back-propagated to the connection weights between the hidden and output layers is defined as the difference between the target value T_{pk} for a particular input pattern p and the neural network's feed-forward calculations of the signal from the output layer O_k as:

$$\delta_{pk} = (T_{pk} - O_{pk})O_{pk}(1 - O_{pk}) \tag{F.6}$$

Then connection weights between the input and hidden layers are changed by:

$$\delta_{pj} = O_{pj}(1 - O_{pj})\sum_{k} \delta_{pk}W_{jk} \tag{F.7}$$

The data feed-forward and error back-propagation process is continued until a "stopping" point is reached. This point is determined by comparing the errors in the training set and the test set. This methodology prevents over-learning or fitting the training set too closely with consequent large errors in the test set.

12.7 Appendix G: Derivation of the Down-and-out call option formulae

The pricing of a down-and-out call option follows and adapts the derivation found in Reisz and Purlich (2007) which in turn follows that of Merton (1973). Here I focus upon the case where the barrier *B* is presumed to be smaller than the strike price *X* and the initial (or starting) value of the firm's assets V_t. I begin by stating the subsidiary proposition Lemma A1.

Lemma A1. *Let K(s) = σW(s), where {W(s)} is a standard Brownian motion. Then for β < 0 and k ≥ β,*

$$\frac{\partial}{\partial k}\mathrm{P}(K(\mathrm{t}) < k, \min_{s\le t} K(s) < \beta) = \frac{1}{\sigma\sqrt{2\pi t}}\mathrm{e}^{-\frac{(2\beta-k)^2}{2\sigma^2 t}} \tag{G.1}$$

Proof.

$$\mathrm{P}(K(\mathrm{t}) < k, \min_{s\le t} K(s) < \beta) = \mathrm{P}(K(\mathrm{t}) < 2\beta - k, \min_{s\le t} K(s) < \beta) = \mathrm{P}(K(\mathrm{t}) < 2\beta - k) \tag{G.2}$$

where the left hand side equality results from the reflection principle and the second equality from the fact that$k \geq \beta$. Now since

$$\mathrm{P}(K(\mathrm{t}) < k, \min_{s\le t} K(s) < \beta) = \mathrm{P}(\min_{s\le t} K(s) < \beta) - \mathrm{P}(K(\mathrm{t}) > k, \min_{s\le t} K(s) < \beta),$$

$$\frac{\partial}{\partial k}\mathrm{P}(K(\mathrm{t}) < k, \min_{s\le t} K(s) < \beta) = -\frac{\partial}{\partial k}\mathrm{P}(K(\mathrm{t}) < 2\beta - k) = -\frac{\partial}{\partial k}\int_{-\infty}^{2\beta-k}\frac{e^{-\frac{u^2}{2\sigma^2 t}}}{\sigma\sqrt{2\pi t}}\mathrm{d}u$$

$$= -\frac{\partial}{\partial k}\phi\left(\frac{2\beta - k}{\sigma\sqrt{t}}\right) = \frac{1}{\sigma\sqrt{t}}\varphi\left(\frac{2\beta - k}{\sigma\sqrt{t}}\right) \tag{G.3}$$

where ϕ and φ are denoted as being the standard normal cumulative probability function and density function respectively.

Lemma A2. *Let {Y(t)} be a Brownian motion with mean αt (with α ≠ 0) and variance $\sigma^2 t$ (in the sequel: a (α, σ^2)-BM, then for β < 0 and y ≥ β,*

$$\frac{\partial}{\partial y}\mathrm{P}(\mathrm{Y(t)} < y, \min_{0\le s\le t} Y(s) < \beta) = e^{\frac{2\alpha\beta}{\sigma^2}} \frac{1}{\sigma\sqrt{2\pi t}} e^{-\frac{(y-2\beta-\alpha t)^2}{2\sigma^2 t}} \equiv \mathrm{g}_t(\mathrm{y}, \beta)$$

(G.4)

Proof. Let $\widehat{W}(t) = W(t) - (\alpha/\sigma)t$; then, by Girsanov's theorem, $\{\widehat{W}(t)\}$ is a standard Brownian motion on [0,T] under the probability measure $\hat{P}$ defined by

$$\frac{d\hat{P}}{dP} = \exp\left\{\left(\frac{\alpha}{\sigma}\right) W(T) - \frac{1}{2}\left(\frac{\alpha}{\sigma}\right)^2 \mathrm{T}\right\} \equiv \mathrm{Z(T)}$$

(G.5)

Hence $\sigma W(t) = \sigma\widehat{W}(t) + \alpha t$ is a (α,σ^2)-BM under $\hat{P}$.

Let

$$\underline{M_t} = \min_{0\le s\le t} K(s)$$

where $\{X(t)\}$ is, similarly to above, equal to $\sigma\{W(t)\}$, i.e., a $(0,\sigma^2)$-BM under the probability measure P. Then

$$\begin{aligned}\widehat{\mathrm{P}}\big(\sigma \mathrm{W(t)} \le \mathrm{y}, \underline{M_t} \le \beta\big) &= \widehat{\mathrm{E}}\big[1\{\sigma\mathrm{W(t)} \le \mathrm{y}, \underline{M_t} \le \beta\}\big] = \mathrm{E}\big[\mathrm{Z(t)}1\{\sigma\mathrm{W(t)} \le \mathrm{y}, \underline{M_t} \le \beta\}\big] \\ &= \mathrm{E}\left[\exp\left(\frac{\alpha}{\sigma^2} K(t) - \frac{1}{2}\left(\frac{\alpha}{\sigma}\right)^2 t\right) 1\{\mathrm{K(t)} \le \mathrm{y}, \underline{M_t} \le \beta\}\right] \\ &= \int_{-\infty}^{y} e^{\left(\frac{\alpha}{\sigma^2}\right)k - \frac{1}{2}\left(\frac{\alpha}{\sigma^2}\right)t} P(K(t) \in dk, \underline{M_t} \le \beta)\end{aligned}$$

(G.6)

where 1 is the indicator function, i.e., 1{A}=1 if event A ⊂ Ω happens and 0 otherwise. This in turn is equal to, by Lemma A1,

$$= \int_{-\infty}^{y} e^{\left(\frac{\alpha}{\sigma^2}\right)k - \frac{1}{2}\left(\frac{\alpha}{\sigma^2}\right)t} \frac{1}{\sigma\sqrt{2\pi t}} \mathrm{e}^{-(2\beta-k)/(2\sigma^2 t)} \mathrm{dk}$$

(G.6)

Differentiating with respects to *y* gives, after completing the square, the desired result. If we assume here that a firm's assets follow a log normal diffusion, i.e., $dV_s = (\mu - \sigma) V_s dt + \sigma_A V_s dW(s)$. We can rewrite this as $dV_s = (r - \delta) V_s d\widehat{W}(s)$, where $\widehat{W}(s) \equiv W(s) + [(\mu - r) / \sigma_A] s$ is, by Girsanov's theorem, a standard BM under the probability measure $\widehat{P}$ defined by $(d\widehat{P} / dP) = \exp\{-[(\mu - r) / \sigma_A] W(T) - [(\mu - r) / \sigma_A]^2 T/2\}$. Hence

$$V_T = V_t e^{\left(r-\delta-\frac{\sigma_A^2}{2}\right)(T-t)+\sigma_A(\widehat{W}(T)-\widehat{W}(t))} \equiv V_t e^{Y(t,T)}$$

(G.7)

where $\{Y(t, T)\}$ is a $\widehat{\mathrm{P}}$-Brownian motion with drift $\left(r - \delta - \frac{\sigma_A^2}{2}\right)(T-t) \equiv \alpha\,(T - t) \equiv \alpha\tau$ and variance $\sigma_A^2\tau$ (note that τ represents the maturity of the debt contract $T - t$). In complete markets, the down-and-out option is replicable and its time-*t* price can be written as

$$e^{-r\tau}\,\widetilde{\mathrm{E}}\left[(\mathrm{V}_T - X)^+\left\{1 - 1\left(\min_{t\le s\le T} V(s) \le B\right)\right\}|X_t\right]$$
$$= e^{-r\tau}\,\widetilde{\mathrm{E}}\,[(\mathrm{V}_T - X)^+|\,X_t] - e^{-r\tau}\,\widetilde{\mathrm{E}}\left[(\mathrm{V}_T - X)^+ 1\left(\min_{t\le s\le T} V(s) \le B\right)|X_t\right]$$

(G.8)

where $\widetilde{\mathrm{E}}\,[\cdot]$ refers to the expectation under the measure $\tilde{P}$ and $(k)^+$ denotes max(k, 0). The first term of equation (H.8) is the price of a standard (Black-Scholes) call option equation and the second term is that of a down-and-in call option. Since the only stochastic term in $V(t)$ is $Y(t,T)$ as described above, the conditioning on F_t can be dropped. Reisz and Purlich (2007) now concentrate on the second term noting that

$$\min_{t\le s\le T} V(s) \le B \;\Leftrightarrow\; \min_{t\le s\le T} Y(s) \le \beta$$

where $\beta = \ln(B/V_t) < 0$. Then, by Lemma A2 the time-*t* price of the down-and-in call option is:

$$e^{-rt}\int_{\ln\left(\frac{X}{V_t}\right)}^{\infty}(V_t e^y - X)\,\tilde{P}(Y_t \in dy, \min_{t\le s\le T} Y(s) \le \beta)$$
$$= e^{-rt}V_t\int_{\ln\left(\frac{X}{V_t}\right)}^{\infty} e^y g_\tau(y,\beta)\,dy, -e^{-rt}X\int_{\ln\left(\frac{X}{V_t}\right)}^{\infty} g_\tau(y,\beta)\,dy$$
$$\equiv e^{-rt}[V_t I_1 - X I_2]$$

Now,

$$I_2 = \exp\left(\frac{2\alpha\beta}{\sigma_A^2}\right)\int_{\ln\left(\frac{X}{V_t}\right)}^{\infty}\left(\frac{1}{\sigma_A\sqrt{2\pi\tau}}\right)exp\left(-\frac{[y-2\beta-\alpha\tau]^2}{2\sigma_A^2\tau}\right)dy$$

$$= exp\left(\frac{2\alpha\beta}{\sigma_A^2}\right)\cdot N\left(\frac{2\beta+\alpha\tau-\ln(X/V_t)}{\sigma_A\sqrt{\tau}}\right)$$

$$= \left(\frac{B}{V_t}\right)^{\frac{2(r-\delta)}{\sigma_A^2}-1}\cdot N\left(\frac{\ln\left[\frac{B^2}{V_t\,X}\right]+\left(r-\delta-\frac{\sigma_A^2}{2}\right)\tau}{\sigma_A\sqrt{\tau}}\right),$$

and

$$I_1 = \exp\left(\frac{2\alpha\beta}{\sigma_A^2}\right)\int_{\ln\left(\frac{X}{V_t}\right)}^{\infty}\left(\frac{e^y}{\sigma_A\sqrt{2\pi\tau}}\right)exp\left(-\frac{[y-2\beta-\alpha\tau]^2}{2\sigma_A^2\tau}\right)dy^{v=(y-2\underline{\beta})/\sigma_A\sqrt{\tau}}$$

$$\cdot\exp\left(\frac{2\alpha\beta}{\sigma_A^2}+2\beta\right)\int_{\left(\ln\left(\frac{X}{V_t}-2\beta\right)/\sigma_A\sqrt{\tau}\right)}^{\infty}\left(\frac{e^{v\sigma_A\sqrt{\tau}}}{2\pi}\right)exp\left(-\left[\frac{v-\alpha\sqrt{\tau/\sigma_A}}{2}\right]^2\right)dv$$

$$= e^{(r-\delta)\tau}\left(\frac{B}{V_t}\right)^{\frac{2(r-\delta)}{\sigma_A^2}+1}\cdot N\left(\frac{\ln\left[\frac{B^2}{V_t\,X}\right]+\left(r-\delta+\frac{\sigma_A^2}{2}\right)\tau}{\sigma\sqrt{\tau}}\right)$$

after completing the square. Putting the parts together we arrive at the formula which values the equity of the company as per equation (8.9) in the main text.

$$DOC = V_t\mathrm{e}^{-\delta\tau}\mathrm{N}(d_1) - X\mathrm{e}^{-r\tau}\mathrm{N}\left(d_1-\sigma\sqrt{\tau}\right)$$
$$-\left[V_t\mathrm{e}^{-\delta\tau}\left(\frac{B}{V_t}\right)^{\frac{2(r-\delta)}{\sigma^2}+1}\mathrm{N}(d_1^B) - X\mathrm{e}^{-r\tau}\left(\frac{B}{V_t}\right)^{\frac{2(r-\delta)}{\sigma^2}-1}\mathrm{N}\left(d_1^B-\sigma\sqrt{\tau}\right)\right]$$

$$d_1 = \frac{\ln(V_t/X)+\left(r-\delta+(\sigma^2/2)\right)\tau}{\sigma\sqrt{\tau}}$$
$$d_1^B = \frac{\ln(B^2/V_tX)+\left(r-\delta+(\sigma^2/2)\right)\tau}{\sigma\sqrt{T-t}}$$

Computing the probability of bankruptcy

Let us first remind ourselves of how insolvency is defined under the barrier options approach. Insolvency is deemed to occur when either; the underlying asset value crosses the barrier *B* from above at any time before *T*, or the underlying asset value stays above the barrier throughout time *T,* but is less than amount of debt owed at the date of maturity, i.e. at time *T*. Using this definition we can derive the actual probability of insolvency occurring using the actual asset drift. Let t* be the first time the asset value crosses the insolvency barrier *B*. The probability of crossing the barrier before maturity is equal to:

$$\mathrm{P}(\mathrm{t}^* \leq \mathrm{T}) = \mathrm{P}(\min_{t \leq s \leq T} Y(s) \leq \beta) = \mathrm{P}(\mathrm{Y(T)} \leq \beta) + \mathrm{P}(\min_{t \leq s \leq T} Y(s) \leq \beta, Y(T) > \beta)$$

$$= \int_{-\infty}^{\beta} f_\tau(y)dy + \int_{\beta}^{\infty} g_\tau(y,\beta)dy = N\left(\frac{\beta-\alpha\tau}{\sigma_A\sqrt{\tau}}\right) + e^{2\alpha\beta/\sigma_A^2} \cdot N\left(\frac{\beta-\alpha\tau}{\sigma_A\sqrt{\tau}}\right), \tag{G.9}$$

where the expression for $g_\tau(y,\beta)$ comes from Lemma A2. It is integrated in the same way as for the computation of I_2, replacing ln(*X/V*$_t$) with ln(*B/V*$_t$*)* as the lower limit of the integration. Since we are interested in actual, and not risk neutral probabilities of default, α is now equal to (μ - δ - $\sigma_A^2/2$), and the probability of early insolvency is equal to

$$\mathrm{P}(\mathrm{t}^* \leq \mathrm{T}) = \mathrm{N}\left(\frac{\ln\left(\frac{B}{V_t}\right) - \left(\mu - \delta - \frac{1}{2}\sigma_A^2\right)(T-\tau)}{\sigma_A\sqrt{T-\tau}}\right)$$

$$+ \left(\frac{B}{V_t}\right)^{\frac{2(\mu-\delta)}{\sigma_A^2}-1} \cdot N\left(\frac{\ln\left(\frac{B}{V_t}\right) + \left(\mu - \delta - \frac{1}{2}\sigma_A^2\right)(T-\tau)}{\sigma_A\sqrt{T-\tau}}\right) \tag{G.10}$$

As for insolvency occurring at the end of the period where the asset value is greater than the barrier but below the value of debt, it suffices to realise that

$$\mathrm{P}(\mathrm{Y(T)} \in \mathrm{dy}, \min_{t \leq s \leq T} Y(s) \geq \beta) = \mathrm{P}(\mathrm{Y(T)} \in \mathrm{dy}) - \mathrm{P}(\mathrm{Y(T)} \in \mathrm{dy}, \min_{t \leq s \leq T} Y(s) < \beta)$$
$$= f_\tau(y)dy - g_\tau(y,\beta)dy$$

to write the probability of this event as:

$$P(B \leq V_T < X, \min_{t \leq s \leq T} Y(s) \geq \beta)$$

$$P(B \leq V_t e^{Y(\tau)} < X, \min_{t \leq s \leq T} Y(s) \geq \beta) = \int_{\beta}^{\ln(\frac{X}{V_t})} f_\tau(y)dy - \int_{\beta}^{\ln(\frac{X}{V_t})} g_\tau(y,\beta)dy$$

$$= \left\{ \int_{\beta}^{\infty} f_\tau(y)dy - \int_{\beta}^{\infty} f_\tau(y)dy - \int_{\beta}^{\infty} g_\tau(y,\beta)dy + \int_{\ln(\frac{X}{V_t})}^{\infty} g_\tau(y,\beta)dy \right\}$$

These four integrals are integrated in much the same way as the integral I_2 was computed. For integrals involving $g_\tau(y,\beta)$, or as $N(d_2)$ is computed in the Black-Scholes case for integrals involving $f_\tau(y)$. We come to the result:

$$P(B \leq V_T < F, \min_{t \leq s \leq T} Y(s) \geq \beta)$$

$$= N\left(\frac{\alpha\tau - \beta}{\sigma_A\sqrt{\tau}}\right) - N\left(\frac{ln\left(\frac{V_t}{X}\right) + \alpha\tau}{\sigma_A\sqrt{\tau}}\right) - e^{\frac{2\alpha\beta}{\sigma_A^2}} \cdot \left\{ N\left(\frac{\alpha\tau + \beta}{\sigma_A\sqrt{\tau}}\right) - N\left(\frac{ln\left(\frac{B^2}{XV_t}\right) + \alpha\tau}{\sigma_A\sqrt{\tau}}\right)\right\}$$

$$= N\left(\frac{ln\left(\frac{V_t}{B}\right) + \left(\mu - \delta - \frac{1}{2}\sigma_A^2\right)(T-\tau)}{\sigma_A\sqrt{T-\tau}}\right) - N\left(\frac{ln\left(\frac{V_t}{X}\right) + \left(\mu - \delta - \frac{1}{2}\sigma_A^2\right)(T-\tau)}{\sigma_A\sqrt{T-\tau}}\right)$$
$$- \left(\frac{B}{V_t}\right)^{\frac{2(\mu-\delta)}{\sigma_A^2}-1}$$
$$\cdot \left\{ N\left(\frac{ln\left(\frac{B}{V_t}\right) + \left(\mu - \delta - \frac{1}{2}\sigma_A^2\right)(T-\tau)}{\sigma_A\sqrt{T-\tau}}\right)\right.$$
$$\left. - N\frac{ln\left(\frac{B^2}{XV_t}\right) + \left(\mu - \delta - \frac{1}{2}\sigma_A^2\right)(T-\tau)}{\sigma_A\sqrt{T-\tau}} \right\}$$

(G.11)

Finally we can sum the two insolvency conditions of equations (H.10) and (H.11) to get the actual probability of insolvency occurring:

$$
\mathrm{P(Insolvency)} = 1 - N\left(\frac{\ln\left(\frac{V_t}{X}\right) + \left(\mu - \delta - \frac{1}{2}\sigma_A^2\right)(T-\tau)}{\sigma_A\sqrt{T-\tau}}\right)
+ \left(\frac{B}{V_t}\right)^{\frac{2(\mu-\delta)}{\sigma_A^2}-1} \cdot N\left(\frac{\ln\left(\frac{B^2}{XV_t}\right) + \left(\mu - \delta - \frac{1}{2}\sigma_A^2\right)(T-\tau)}{\sigma_A\sqrt{T-\tau}}\right)
$$

12.8 Appendix H: Derivation of the European call option pricing formulae

In order to derive the price of a European call option upon a firm's assets (assuming in the first instance a firm which pays no dividends), we could solve the Black-Scholes Partial Differential Equation either by employing appropriate boundary conditions, or by calculating the expectation explicitly using techniques from probability. For this derivation I will use the latter due to its relative simplicity. To begin, we assume that the dynamic of the assets (*V*) follows a geometric Brownian motion and under a risk-free neutral measure is given by:

$$dV_t = rV_t dt + \sigma V_t dz_t \tag{H.1}$$

Where r is the risk-free rate. The solution to the above equation is given by the following which has previously been derived.

$$V_T = V_t e^{\left(r - \frac{\sigma_A^2}{2}\right)(T) + \sigma_A z_T} \tag{H.2}$$

where $Z_T \sim N(0,T)$.

The value of the option, with a strike price equal to the companies' debt *X* and maturity *T* is given by the expected value of the options' payoff, discounted back to today:

$$V_e = e^{-rT}\mathbb{E}[\max(V_T - K, 0)] \tag{H.3}$$

Equation (H.3) can be evaluated by deriving explicitly the value for V_e by first integrating against the density of a random variable with distribution N(0,1)

$$e^{-rT}\mathbb{E}[\max(V_T - K, 0)] = e^{-rT}\mathbb{E}[\max\left(V_0 e^{\left(r - \frac{\sigma_A^2}{2}\right)(T) + \sigma_A z_T} - K, 0\right)]$$

$$= e^{-rT}\mathbb{E}[\max\left(V_0 e^{\left(r-\frac{\sigma_A^2}{2}\right)(T)+\sigma_A\sqrt{T}z_T} - K, 0\right)]$$

$$e^{-rT}\int_{V_0 e^{(r-(\sigma_A^2)/2)(T)+\sigma_A\sqrt{T}z}\geq X}\left(V_0 e^{\left(r-\frac{\sigma_A^2}{2}\right)(T)+\sigma_A\sqrt{T}z} - X\right)\frac{1}{\sqrt{2\pi}}e^{-z^2/2}dz \tag{H.4}$$

Determine the bounds and separate the subtraction

$$e^{-rT}\int_{z\geq\frac{\ln(X/V_0)-(r-(\sigma_A^2)/2)T}{\sigma\sqrt{T}}}\left(V_0 e^{\left(r-\frac{\sigma_A^2}{2}\right)(T)+\sigma_A\sqrt{T}z} - X\right)\frac{1}{\sqrt{2\pi}}e^{-z^2/2}dz$$

$$e^{-rT}\int_{\frac{\ln(X/V_0)-(r-(\sigma_A^2)/2)T}{\sigma\sqrt{T}}}^{\infty} V_0 e^{\left(r-\frac{\sigma_A^2}{2}\right)(T)+\sigma_A\sqrt{T}z}\frac{1}{\sqrt{2\pi}}e^{-z^2/2}dz - Xe^{-rT}\int_{\frac{\ln(X/V_0)-(r-(\sigma_A^2)/2)T}{\sigma\sqrt{T}}}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-z^2/2}dz \tag{H.5}$$

Change the limits using the property of the normal density function, complete the square and simplify

$$V_0 e^{-\sigma_A^2/2T}\int_{-\infty}^{\frac{\ln(V_0/X)+(r-(\sigma_A^2)/2)T}{\sigma\sqrt{T}}}\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}+\sigma\sqrt{T}}dz - Xe^{-rT}\int_{-\infty}^{\frac{\ln(V_0/X)+(r-(\sigma_A^2)/2)T}{\sigma\sqrt{T}}}\frac{1}{\sqrt{2\pi}}e^{-z^2/2}dz$$

$$V_0\int_{-\infty}^{\frac{\ln(V_0/X)+(r-(\sigma_A^2)/2)T}{\sigma\sqrt{T}}}\frac{1}{\sqrt{2\pi}}e^{-(z-\sigma\sqrt{T})^2/2}dz - Xe^{-rT}\int_{-\infty}^{\frac{\ln(V_0/X)+(r-(\sigma_A^2)/2)T}{\sigma\sqrt{T}}}\frac{1}{\sqrt{2\pi}}e^{-z^2/2}dz$$

Let $x = z - \sigma\sqrt{T}$ and change the limits of the integration

$$V_0\int_{-\infty}^{\frac{\ln(V_0/X)+(r-(\sigma_A^2)/2)T}{\sigma\sqrt{T}}}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx - Xe^{-rT}\int_{-\infty}^{\frac{\ln(V_0/X)+(r-(\sigma_A^2)/2)T}{\sigma\sqrt{T}}}\frac{1}{\sqrt{2\pi}}e^{-z^2/2}dz$$

$$= V_0 N(d_1) - Xe^{-rT} N(d_1 - \sigma_A\sqrt{T})$$

(H.6)

where

$$d_1 = \frac{ln\left[\frac{V_0}{X}\right] + \left(r - \frac{1}{2}\sigma_A^2\right)T}{\sigma_A\sqrt{T}}$$

Finally we need to adjust the formula incorporate dividends, reflecting that the dividends which are paid by the firm will accrue to equity holders. Firstly we adjust V_0 to account for the reduction in asset value due to the payment of dividends prior to time *T*. Secondly, the risk free rate of return *r* is replaced by the "real" rate of return of the firm. This return is given by the return on assets μ, minus the dividend rate δ. Finally we include an additional term $\left(1 - \mathrm{e}^{-\delta T}\right)V_0$ as it is the equity holders who receive the dividends. When the dividend rate is equal to zero, this term is valueless and disappears. The resulting formula is therefore:

$$V_E = V_0 \mathrm{e}^{-\delta T} \mathrm{N}(d_1) - Xe^{-rT} \mathrm{N}\left(d_1 - \sigma_A\sqrt{T}\right) + \left(1 - \mathrm{e}^{-\delta T}\right)V_0$$

(H.7)

where

$$d_1 = \frac{\ln\left[\frac{V_0}{X}\right] + \left(\mu - \delta + \frac{1}{2}\sigma_A^2\right)T}{\sigma_A\sqrt{T}}$$

This resulting model assumes that the natural log of future assets values is normally distributed with mean $\left[\ln[V_0] + \left(r - \frac{1}{2}\sigma_A^2\right)T\right]$, and variance σ_A^2. McDonald (2002) shows therefore that the probability of insolvency occurring, i.e. $V < X$ at the end of the period *T* is:

$$\text{Prob} = \mathrm{N}\left[-\frac{\ln\left[\frac{V_0}{X}\right] + \left(\mu - \delta - \frac{1}{2}\sigma^2\right)T}{\sigma\sqrt{T}}\right]$$

(H.8))

12.9 Appendix I: Detailed results of the elimination test

Year 1

BANK	*SIZE*	*B/M*	*BV*	*AZ*	*TZ*	*OL*	*BS*	*HKCL*	*DOC*	*JW*
Value of Good Loans	23874	18587	15072	5556	9115	8668	3650	3490	6686	3759
Value of Defaulting Loans (£m)	118	711	263	64	171	97	0	30	79	8
TOTAL LOAN VALUE (£m)	23993	19298	15335	5620	9286	8765	3650	3520	6765	3767
Market Share	**24.0%**	**19.3%**	**15.3%**	**5.6%**	**9.3%**	**8.8%**	**3.6%**	**3.5%**	**6.8%**	**3.8%**
Average lending rate	1.2%	1.2%	1.3%	0.9%	1.1%	1.1%	0.7%	0.8%	1.0%	0.8%
Share of Good Loans	24.2%	18.9%	15.3%	5.6%	9.3%	8.8%	3.7%	3.5%	6.8%	3.8%
REVENUE (£m)	**286**	**228**	**201**	**52**	**104**	**99**	**27**	**27**	**68**	**28**
Share of Defaults	7.7%	46.1%	17.1%	4.2%	11.1%	6.3%	0.0%	2.0%	5.1%	0.5%
LOSS (£m)	**53**	**320**	**118**	**29**	**77**	**44**	**0**	**14**	**36**	**3**
PROFIT (£m)	**233**	**-92**	**83**	**23**	**26**	**55**	**27**	**13**	**33**	**25**
RETURN ON ASSETS	**0.97%**	**-0.48**	**0.54%**	**0.41%**	**0.29%**	**0.63%**	**0.75%**	**0.37%**	**0.48%**	**0.66%**

Year 2

BANK	*SIZE*	*BV*	*AZ*	*TZ*	*OL*	*BS*	*HKCL*	*DOC*	*JW*
Value of Good Loans	29093	18661	8627	10578	9402	4650	3939	7888	5620
Value of Defaulting Loans (£m)	265	475	87	214	155	50	51	117	128
TOTAL LOAN VALUE (£m)	29358	19136	8714	10792	9556	4700	3990	8005	5748
Market Share	**29.4%**	**19.1%**	**8.7%**	**10.8%**	**9.6%**	**4.7%**	**4.0%**	**8.0%**	**5.7%**
Average lending rate	1.4%	1.6%	1.1%	1.2%	1.3%	1.1%	0.9%	1.2%	1.2%
Share of Good Loans	29.5%	19.0%	8.8%	10.7%	9.5%	4.7%	4.0%	8.0%	5.7%
REVENUE (£m)	**402**	**296**	**98**	**132**	**120**	**50**	**35**	**91**	**70**
Share of Defaults	17.2%	30.8%	5.6%	13.9%	10.0%	3.3%	3.3%	7.6%	8.3%
LOSS (£m)	**119**	**214**	**39**	**96**	**70**	**23**	**23**	**53**	**58**
PROFIT (£m)	**283**	**82**	**59**	**36**	**50**	**27**	**12**	**39**	**13**
RETURN ON ASSETS	**0.96%**	**0.43%**	**0.68%**	**0.33%**	**0.53%**	**0.58%**	**0.31%**	**0.48%**	**0.22%**

Year 3

BANK	*SIZE*	*BV*	*AZ*	*TZ*	*OL*	*BS*	*HKCL*	*DOC*
Value of Good Loans	29384	19162	8846	10750	9708	6974	5250	8384
Value of Defaulting Loans (£m)	266	493	87	214	160	126	72	125
TOTAL LOAN VALUE (£m)	29649	19655	8933	10964	9868	7100	5322	8509
Market Share	**29.6%**	**19.7%**	**8.9%**	**11.0%**	**9.9%**	**7.1%**	**5.3%**	**8.5%**
Average lending rate	1.4%	1.6%	1.2%	1.3%	1.3%	1.3%	1.0%	1.2%
Share of Good Loans	29.8%	19.5%	9.0%	10.9%	9.9%	7.1%	5.3%	8.5%
REVENUE (£m)	**408**	**305**	**102**	**135**	**124**	**87**	**53**	**100**
Share of Defaults	17.2%	31.9%	5.6%	13.9%	10.4%	8.2%	4.7%	8.1%
LOSS (£m)	**120**	**222**	**39**	**96**	**72**	**57**	**32**	**56**
PROFIT (£m)	**289**	**84**	**63**	**39**	**52**	**31**	**20**	**44**
RETURN ON ASSETS	**0.97%**	**0.43%**	**0.71%**	**0.35%**	**0.53%**	**0.43%**	**0.38%**	**0.52%**

Year 4

BANK	*SIZE*	*BV*	*AZ*	*OL*	*BS*	*HKCL*	*DOC*
Value of Good Loans	30389	21565	12788	11777	7410	5491	9038
Value of Defaulting Loans (£m)	281	533	168	231	126	72	132
TOTAL LOAN VALUE (£m)	30670	22098	12955	12008	7536	5563	9170
Market Share	**30.7%**	**22.1%**	**13.0%**	**12.0%**	**7.5%**	**5.6%**	**9.2%**
Average lending rate	1.4%	1.6%	1.3%	1.4%	1.3%	1.0%	1.2%
Share of Good Loans	30.9%	21.9%	13.0%	12.0%	7.5%	5.6%	9.2%
REVENUE (£m)	**434**	**356**	**166**	**166**	**97**	**56**	**111**
Share of Defaults	18.2%	34.6%	10.9%	15.0%	8.2%	4.7%	8.6%
LOSS (£m)	**126**	**240**	**75**	**104**	**57**	**32**	**60**
PROFIT (£m)	**307**	**116**	**91**	**62**	**41**	**24**	**52**
RETURN ON ASSETS	**1.00%**	**0.52%**	**0.70%**	**0.52%**	**0.54%**	**0.43%**	**0.56%**

Year 5

BANK	*SIZE*	*BV*	*AZ*	*OL*	*BS*	*DOC*
Value of Good Loans	30740	21892	12933	12050	10049	10794
Value of Defaulting Loans (£m)	281	534	168	232	150	178
TOTAL LOAN VALUE (£m)	31021	22426	13101	12282	10199	10971
Market Share	**31.0%**	**22.4%**	**13.1%**	**12.3%**	**10.2%**	**11.0%**
Average lending rate	1.4%	1.6%	1.3%	1.4%	1.3%	1.3%
Share of Good Loans	31.2%	22.2%	13.1%	12.2%	10.2%	11.0%
REVENUE (£m)	**439**	**361**	**168**	**170**	**134**	**136**
Share of Defaults	18.2%	34.6%	10.9%	15.0%	9.7%	11.5%
LOSS (£m)	**126**	**240**	**75**	**104**	**67**	**80**
PROFIT (£m)	**313**	**121**	**93**	**65**	**66**	**56**
RETURN ON ASSETS	**1.01%**	**0.54%**	**0.71%**	**0.53%**	**0.65%**	**0.51%**

Year 6

BANK	*SIZE*	*BV*	*AZ*	*OL*	*BS*
Value of Good Loans	32966	23402	14877	12750	14461
Value of Defaulting Loans (£m)	289	644	197	251	163
TOTAL LOAN VALUE (£m)	33255	24046	15074	13001	14624
Market Share	**33.3%**	**24.0%**	**15.1%**	**13.0%**	**14.6%**
Average lending rate	1.5%	1.7%	1.3%	1.4%	1.4%
Share of Good Loans	33.5%	23.8%	15.1%	13.0%	14.7%
REVENUE (£m)	**487**	**399**	**195**	**184**	**209**
Share of Defaults	18.7%	41.7%	12.7%	16.3%	10.5%
LOSS (£m)	**130**	**290**	**88**	**113**	**73**
PROFIT (£m)	**357**	**109**	**107**	**71**	**136**
RETURN ON ASSETS	**1.07%**	**0.46%**	**0.71%**	**0.55%**	**0.93%**

Year 7

BANK	*SIZE*	*AZ*	*OL*	*BS*
Value of Good Loans	35097	18826	27598	16891
Value of Defaulting Loans (£m)	363	224	711	244
TOTAL LOAN VALUE (£m)	35460	19050	28309	17135
Market Share	**35.5%**	**19.1%**	**28.3%**	**17.1%**
Average lending rate	1.6%	1.5%	1.9%	1.6%
Share of Good Loans	35.7%	19.1%	28.0%	17.2%
REVENUE (£m)	**558**	**282**	**517**	**267**
Share of Defaults	23.5%	14.5%	46.1%	15.8%
LOSS (£m)	**163**	**101**	**320**	**110**
PROFIT (£m)	**395**	**181**	**197**	**157**
RETURN ON ASSETS	**1.11%**	**0.95%**	**0.70%**	**0.92%**

Year 8

BANK	*SIZE*	*AZ*	*BS*
Value of Good Loans	41503	33815	22656
Value of Defaulting Loans (£m)	620	507	325
TOTAL LOAN VALUE (£m)	42123	34321	22982
Market Share	**42.4%**	**34.5%**	**23.1%**
Average lending rate	1.8%	1.8%	1.7%
Share of Good Loans	42.4%	34.5%	23.1%
REVENUE (£m)	**747**	**622**	**393**
Share of Defaults	42.7%	34.9%	22.4%
LOSS (£m)	**279**	**228**	**146**
PROFIT (£m)	**468**	**394**	**247**
RETURN ON ASSETS	**1.11%**	**1.15%**	**1.07%**

Year 9

BANK	*SIZE*	*AZ*
Value of Good Loans	49652	47475
Value of Defaulting Loans (£m)	809	612
TOTAL LOAN VALUE (£m)	50461	48087
Market Share	**51.2%**	**48.8%**
Average lending rate	1.9%	1.9%
Share of Good Loans	51.1%	48.9%
REVENUE (£m)	**964**	**885**
Share of Defaults	56.9%	43.1%
LOSS (£m)	**364**	**276**
PROFIT (£m)	**600**	**609**
RETURN ON ASSETS	**1.19%**	**1.27%**

12.10 Appendix J: Genetic algorithms

This brief explanation of Genetic Algorithms is reproduced from Back et al. (1996). For a more thorough description of genetic algorithms refer to Goldberg (1989).

A genetic algorithm simulates Darwinian evolution. It maintains *a population* of chromosomes, where a *chromosome* is a candidate-solution to the problem we want to solve. Chromosomes are often called *strings* in a genetic algorithm context. A string in its turn, consists of a number of *genes*, which may take some number of values, called *alleles.* The genetic algorithm terms for genes and alleles are *features* and *values*. Associated with each string is a *fitness value*, which determines how 'good' a string is. The fitness value is determined by a *fitness function*, which we can think of as some measure of profit or goodness that we want to maximise. Basically, there are three operators that lead to good results in a genetic algorithm, namely reproduction, crossover, and mutation.

Reproduction This is a process in which strings are copied onto the next generation. Strings with a higher fitness value have more chance of making it to the next generation. Different schemes can be used to determine which strings survive into the next generation. A frequently used method is *roulette wheel selection*, where a roulette wheel is divided in a number of slots, one for each string. The slots are sized according to the fitness of the strings. Hence, when we spin the wheel, the best strings are the most likely to be selected. Another well-known method is *ranking*. Here, the strings are sorted by their fitness value, and each string is assigned an offspring count that is determined solely by its rank.

Crossover A part of one string is combined with a part of another string. This way, we hope to combine the good parts of one string with the good parts of another string, yielding an even better string after the operation. This operation takes two strings, the parents, and produces two new ones, the offspring. Many kinds of crossover can be thought of. Two kinds of crossover are illustrated in Figure 16.10a and 16.10b.

Mutation A randomly selected gene in a string takes a new value. The aim of this operator is to introduce new genetic material in the population, or at least prevent the loss of it. Under mutation, a gene can get a value that did not occur in the population before, or that has been lost due to reproduction. Mutation is illustrated in Figure J.

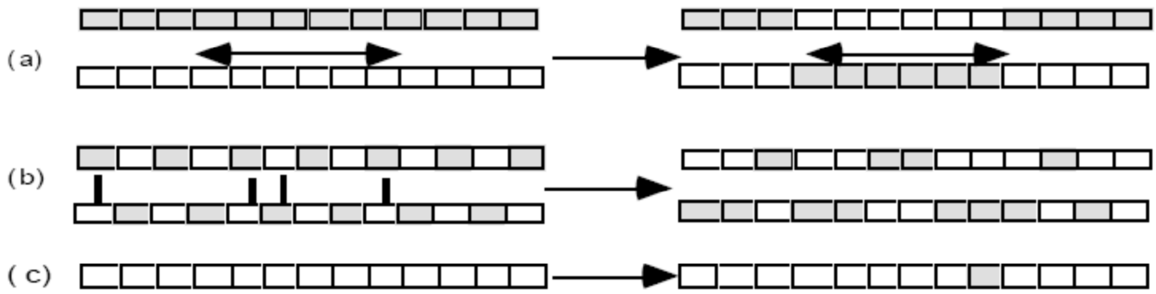


Figure J: Parts (a) and (b) illustrate two different kinds of crossover operations. Part (c) illustrates a mutation operation.

13 Bibliography

Abdel-Khalik, A.R. (1973). The Effect of Aggregating Accounting Reports on the Quality of the Lending Decision: An Empirical Investigation. *Journal of Accounting Research,* 11, 104-38.

Abrahams, C. R., & Zhang, M. (2009). A Comprehensive Credit Assessment Framework. *SAS white paper series*. Available at http://www.sas.com/resources/whitepaper/wp_4351.pdf

Agarwal, V., & Taffler, R. J. (2003). Do statistical failure prediction models work ex ante or only ex post?, *Deloitte & Touche Lecture Series on credit risk*, University of Antwerp Belgium.

Agarwal, V., & Taffler, R. J. (2007). Twenty-five years of the Taffler z-score model: does it really have predictive ability? *Accounting and Business Research*, 37(4), 285-300.

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

Alici, Y. (1995). Neural Networks in Corporate Failure Prediction: The UK Experience. *Processing Third International Conference of Neural Networks in the Capital Markets, London, October:* 393-406.

Alexander, S. S. (1949). The Effect of Size of Manufacturing Corporation on the Distribution of the Rate of Return. *Review of Economics and Statistics*, 31(3), 229-35.

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

Altman, E. I. (1973). Predicting Railroad Bankruptcies in America. *Bell Journal of Economics and Management Science*, 4(1), 184-211.

Altman, E. I. (1982). Accounting Implication of Failure Prediction Models. Journal of Accounting, *Auditing and Finance*, 6(1), 4-19.

Altman, E. I. (1983). *Corporate Financial Distress. A Complete Guide to Predicting, Avoiding, and Dealing with Bankruptcy*. Wiley-Interscience, New York.

Altman, E. I. (1984). The success of business failure prediction models : An international survey. Journal of Banking and Finance, 8(2), 171-198.

Altman, E. I. (1993). *Corporate Financial Distress and Bankruptcy. A Complete Guide to Predicting and avoiding Distress, and Profiting from Bankruptcy*. 2nd Ed. Wiley-Interscience, New York.

Altman, E. I., Baidya, T., & Riberio-Dias, L. M. (1979). Assessing Potential Financial Problems of Firms in Brazil. *Journal of International Business Studies*, Fall 1979.

Altman, E. I. & Eisenbeis, R. A. (1978). Financial Applications of Discriminant Analysis: A Clarification. *Journal of Financial and Quantitative Analysis*, 13(1), 185-195.

Altman, E. I., Haldeman, R. G., & Narayanan, P. (1977). ZETA analysis: A new model to identify bankruptcy risk of corporations. *Journal of Banking and Finance*, 1(1), 25-54.

Altman, E. I., Eom, Y. H., & Kim, D. W. (1995). Failure prediction: evidence from Korea. *Journal of International Financial Management and Accounting,* 6 (3), 230–249.

Altman, E.I., & Izan, (1982). Identifying corporate distress in Australia: An Industry Relative Analysis. Australian Graduate School of Management, Sydney.

Altman, E.I., & Lavallee, M. (1981). Business Failure Classification in Canada*. Journal of Business Administration*, 12(1), 147-164.

Altman, E. I., & McGough, T. P. (1974). Evaluation of a company as a going concern. *Journal of Accountancy*, 138(6), 50-57.

Altman, E., I., Marco, G., & Varetto, F. (1994). Corporate Distress Diagnosis: Comparisons UsingLinear Discriminant Analysis and Neural Networks. *Journal of Banking and Finance,* (18)3: 505–529.

Alves, P., Beekes, W., & Young, S. (2007). Which database should I use? A comparison of UK firms' Financial Statement Data From Six Sources. *Lancaster University Management School, Working paper.*

Anderson, T.W. (1962). An Introduction to Multivariate Statistical Analysis. New York, Wiley.

Ash, T. (1989). Dynamic node creation. *Connection Science*. 1(4), 365–375.

Atiya, A. F. (2001). Bankruptcy Prediction for Credit Risk Using Neural Networks: A Survey and New Results. *IEEE Transactions on Neural Networks,* (12)4: 929–935.

Aziz, M., & Dar, H. (2006). Predicting corporate bankruptcy: where we stand? *Corporate Governance*, 6(1), 18-33.

Back, B., Laitinen, T., Sere, K., & van Wezel, M. (1996). *Choosing Bankruptcy Predictors Using Discriminant Analysis, Logit Analysis, and Genetic Algorithms.* Technical Report No. 40. Turku Centre for Computer Science, Turku.

Balcaen, S., & Ooghe, H. (2004). Alternative methodologies in studies on business failure: do they produce better results than the classical statistical methods? Working Paper Series, Faculty of Economics and Business Administration, Ghent University (Belgium), Paper number, 04/249.

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.

Bardos, M., & Zhu, W. (1997). Comparision of discriminant analysis and neural networks: Application for the detection of company failures. *Biometric approaches in management science*, 25.

Barnes, P. (1982). Methodological implications of non-normality distributed financial ratios. *Journal of Business Finance and Accounting*, 9 (1), 51–62.

Bates, T. (1973). An econometric analysis of lending to black businessmen. *Review of Economics and Statistics*, 55(3), 272-283.

Beaver, W. (1966). Financial Ratios as Predictors of Failures. *Journal of Accounting Research (supplement)*, 4(1), 71-102.

Beaver, W. (1968). Market prices, financial ratios and prediction of failure. *Journal of Accounting Research*, 8, 179–192.

Becerra, V., Galvao, R., & Abou-Seada, M. (2005). Neural andWavelet Network Models for Financial Distress Classification. *Data Mining and Knowledge Discovery*, 11, 35-55.

Beermann, K. (1976). *Possible Ways to Predict Capital Losses with Annual Financial Statements*. Dusseldorf: 1976.

Begley, J., Ming, J. & Walls, S. (1996). Bankruptcy classification errors in the 1980's: an empirical analysis of Altman's and Ohlson's models. *Review of Accounting Studies*, 1(4), 267-284.

Berry, M.J.A., & Linoff, G. (1997). *Data Mining Techniques.* NY: John Wiley & Sons.

Betts, J., & Belhoul, D. (1982). The identification of companies in danger of failing using discriminant analysis. *7th Advances in Reliability Technology Symposium proceedings*, April (University of Bradford, UK).

Betts, J., & Belhoul, D. (1983). Applications of reliability analysis in the management of the economy. *4th Eurodata conference*, March (Venice).

Bharath, S., & Shumway, T. (2004). Forecasting default with the KMV-Merton Model. *Working Paper Series*, University of Michigan.

Bharath, S., & Shumway, T. (2008). Forecasting Default with the Merton Distance to Default Model. *The Review of Financial Studies*, 21(3), 1339-1369.

Bilderbeek, J. (1979). An Empirical Study of the Predictive Ability of Financial Ratios in the Netherlands. *Zeitschrift fur Betriebswirtschaft*, No.5 (May).

Black, F. & Cox, J. (1976). Valuing Corporate Securities: Some Effects of Bond Indenture Provisions. *Journal of Finance* 31(2), 351–367.

Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy,* 81(3), 637–654.

Bliss, J. H. (1923). *Financial and Operating Ratios in Management*. New York: The Ronald Press, 34-38.

Blum, A. (1992). *Neural Networks in C++*. NY: Wiley.

Blum, M. (1974). Failing Company Discriminant Analysis. *Journal of Accounting Research*, 12(1), 1-25.

Blums, M. (2003). D-Score: Bankruptcy Prediction Model for Middle Market Public Firms. *Working Paper, Macalester College*.

Board, J., Delargy, P.J.R., & Tonks, I. (1990). Firm Financing and R&D Intensity: A Post-SSAP 13 US-UK Comparison. *Mimeo*.

Boger, Z., & Guterman, H. (1997). Knowledge extraction from artificial neural network models. *IEEE Systems, Man, and Cybernetics Conference*, Orlando, FL.

Bolitho, N. (1972). How to spot insolvency before it happens. *Investors Chronicle*, 23rd March, 1146-1147.

Boritz, J. E., Kennedy, D. B., & Albuquerque, A. (1995). Predicting corporate failure using a neural network approach. *Intelligent Systems in Accounting, Finance and Management*, 4(2), 95-111.

Box, G., & Cox, D. (1964). An analysis of transformations. *Journal of the Royal Statistical Society*, *Series B*, 26(2), 211-52.

Cadbury, A. (1992). Financial Aspects of Corporate Governance. *FRC publications*.

Cahill, E. (1981). Irish Listed Company Failure Ratios, Accounts and Auditors' Opinions. *Journal of Irish Business and Administrative Research*. 3(1). 23-47.

Campbell, J. Y., Hilscher, J., & Szilagyi, J. (2006). In search of distress risk. *Harvard University*, Working paper.

Casey, C. (1980). Variation in Accounting Information Load: The Effect on Loan Officers' Predictions of Bankruptcy. *The Accounting Review*. 55(1), 36-49.

Cassidy, D. (2003). Maximizing shareholder value: the risks to employees, customers and the community. *Corporate Governance.* 3(2), 32-37.

Castagna, A. D, & Matolcsy, Z. P. (1981). The Prediction of Corporate Failure: Testing the Australian Experience. *Australian Journal of Management*. 6(1), 23-28.

Cawkwell, A. (2008). Don't fear the insolvency adviser. *Building Sustainable Design*. September.

Chalos, P. (1985). Financial distress: a comparative study of individual, model and committee assessments. *Journal of Accounting Research*, 23 (2), 527–543.

Chang, P. C., & Afifi, A. A. (1974). Classification based on Dichotomous and Continuous Variables. *Journal of the American Statistical Association*, 69(346).

Charalambous, C. (1992). A conjugate gradient algorithm for the efficient training of artificial neural networks. *IEE Proceedings*, Part G (139), 301–310.

Charitou, A., & Lambertides, N. (2003). Earnings management prior to bankruptcy, *Working Paper, University of Cyprus*, 1–46.

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

Citron, D. & Taffler, R. J. (1992). The audit report under going concern uncertainties: an empirical analysis. *Accounting & Business Research*, 22(8), 337-45.

Coats, P. K. & Fant, L. F. (1993). Recognizing financial distress patterns using a neural network tool. *Financial Management*, 22(3), 142-55.

Collins, R., & Green, R. (1982). Statistical Methods for Bankruptcy Prediction. *Journal of Economics and Business*, 34, 349-354.

Collongues, Y. (1977). Ratios Financiers et Prevision des Faillites des Petites et Moyennes Enterprises. *Review Banque*, No. 365.

Cooper, D. (2010). Company Insolvencies Rise Dramatically in 2009 With Little Hope of Improvement For 2010. *CooperMatthews 2010 Debt Articles*. Available at http://coopermatthews.com/articles.html

Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1), 21-27.

Cox, D.R., & Miller, H.D. (1965). *The Theory of Stochastic Processes*. John Wiley, New York.

Crosbie, P. J., & Bohn, J. R. (2003). Modelling Default Risk. *Moody's KMV*.

Cybinski, P. (2001). Description, Explanation, Prediction: the Evolution of Bankruptcy Studies? *Managerial Finance,* 27 (4), 29–44.

Dambolena, I., & Khoury, S. (1980). Ratio stability and corporate failure. *Journal of Finance*, 33 (4), 1017–1026.

Datastream plc., (1980). The datastream Z-score service: A technical review.

Davidson, R., and MacKinnon, J. (1984). Convenient specification tests for logit and probit models. *Journal of Econometrics* 25(3), 241–262.

Davies, G. (2002). *A history of money from ancient times to the present day*, *3rd ed*. Cardiff: University of Wales Press.

Daubie, M., & Meskens, N. (2002). Business failure prediction: a review and analysis of the literature, *New Trends in Banking Management, Contributions to Management Science*, 2, 71-78.

Deakin, E. (1972). A Discriminant Analysis of Predictors of Business Failure. *Journal of Accounting Research*, 10(1), 167-169

Deakin, E. (1977). *Business Failure Prediction: An Empirical Analysis*. Financial crises: Institutions and Markets in a Fragile Environment, New York: Wiley.

Dhrymes, P. (1974). *Econometrics*. New York: Springer-Verlag, 1974; corrected reprint of 1970 ed. Published by Harper and Row, New York.

Diamond, H. (1976). Pattern recognition and the detection of corporate failure. Ph.D dissertation, New York University.

Duffie, D., & Wang, K. (2004). Multi-Period corporate failure prediction with stochastic covariates. *Stanford University Working paper*.

Earl, M., & Marais, D. (1979). The prediction of corporate bankruptcy in the UK using discriminant analysis. *Oxford Centre of Management Studies*, working paper 79/5.

Edmister, R. (1972). An empirical test of financial ratio analysis for small business failure prediction. *Journal of Financial and Quantitative Analysis*, 7(2), 1477-93.

Efron, B. (1975). The efficiency of logistical regression compared to normal discriminant analysis. *Journal of the American Statistical Association,* 70, 313-323.

Eidleman, G. (1995). Z-scores - a guide to failure prediction. *The CPA Journal*, 12(9), 52-53.

Eisenbeis, R. A. (1977). Pitfalls in the Application of Discriminant Analysis in Business, Finance, and Economics. *Journal of Finance*, 32(3), 875-900.

El Hennaway R., & Morris, R. (1983). The Significance of Base Year in Developing Factor Prediction Models. *Journal of Business Finance and Accounting*, 10, 209-223.

Engelmann, B., Hayden, E., & Tasche, D. (2003). Testing rating accuracy. *Risk,* 16, 82–86.

Falkenstein, E., Boral, A., & Carty, L. (2000). RiskCalc for private companies: Moody's default model. *Moody's KMV*.

Feller, W. (1968). *An Introduction to Probability Theory and its Applications*. Third Edition. Vol.1, New York: Wiley 1968.

Feltham, G. A., & Demski, J. S. (1970). The Use of Models in Information Evaluation. *The Accounting Review,* 45, 623-40.

Financial Reporting Council (2006). The Combined Code on Corporate Governance. *FRC publications*. June 2006.

Financial Services Authority - Internal Audit Division (2008). The supervision of Northern Rock: a lessons learned review. March 2008.

Finney, D. (2009). A Brief History of Credit Rating Agencies. *Investopedia articles.* Source: http://www.investopedia.com/articles/bonds/09/history-credit-rating-agencies.asp

Fisher, R. A. (1938). The Statistical Utilization of Multiple Measurements. *Annals of Engenics*, 8, 376-386.

Fitzpatrick, P. J. (1932). A Comparison of Ratios of Successful Industrial Enterprises with Those of Failed Firms. *The CPA Journal*, 12(3), 598-605.

Fletcher, D., & Goss, E. (1993). Forecasting with neural networks. An application using bankruptcy data. *Information and Management*, 24(3), 159-67.

Foreman, R., D. (2003). A logistic analysis of bankruptcy within the US local telecommunications industry. *Journal of Economics and Business*. 55, 135-66.

Frank, R. E., Massy, W. R., & Morrison, D. G. (1965). Bias in Multiple Discriminant Analysis. *Journal of Marketing Research II*, August 1965.

Frank, W. G., & Weygandt, J. J. (1971). A prediction model for convertible debentures. *Journal of Accounting Research*, 9(1), 116-26.

Gaganis, C., Pasiouras, F., Spathis, C., and Zopounidis, C. (2007). A comparison of nearest neighbours, discriminant and logit models for auditing decisions. *International Journal of Intelligent Systems in Accounting, Finance & Management,* 15, 23-40.

Gentry, J.A., Newbold, P., & Whitford, D.T., (1985). Classifying Bankrupt firms with Funds Flow Components. *Journal of Accounting Research* 23, 146–160.

Gentry, J.A., Newbold, P., & Whitford, D.T., (1987). Funds flow components, financial ratios and bankruptcy. *Journal of Business Finance & Accounting,* 14 (4), 595–606.

Geroski, P. A., & Gregg, G. (1997). Coping with Recession: Company Performance in Adversity, *Oxford University Press.*

Geske, R. (1977). The Valuation of Corporate Liabilities as Compound Options. *Journal of Financial and Quantitative Analysis,* 12(4), 541–552

Gilbert, E. (1968). On discrimination using qualitative variables. *Journal of the American Statistical Association,* 63, 1399-1412

Goldberg, D. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addsion-Wesley, Reading, MA,

Gombola, M. J., Haskins, M. E., Ketz, E. J., & Williams, D. D. (1987). Cash Flow in Financial Prediction. *Financial Management,* 16(1), 55–65.

Gönen, M. (2006). Receiver operating characteristic curves. *SAS Users Group International* (31), San Francisco, CA. March 2006.

Gujarati, D. (2003). *Basic Econometrics, Fourth Ed*. McGraw-Hill, London.

Hamer, 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., (2004). Marginal classifier improvement and reality, *Paper presented at the Symposium on Data Mining*, Ghent University, Belgium, May 2004.

Hanley, J., & McNeil, B. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143(April), 29-36.

Hassibi, B., Stork, D., & Wolff, G, (1993). Optimal brain surgeon and general network pruning. *IEEE International Conference on Neural Networks*, 293–299.

Haug, E., & Taleb, N. (2009). Why We Have Never Used the Black-Scholes-Merton Option Pricing Formula. Working Paper, University of Michigan.

Hebb, D. (1949). *Organization of Behavior*. New York: John Wiley.

Heine, M. L. (2000). Predicting Financial Distress of Companies: Revisiting the Z-score and ZETA® Models. New York University, July 2000.

Hickman, W. B. (1958). Corporate Bond Quality and Investor Experience. Princeton University Press, 390-421.

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.

Hopfield, J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Science*, 79, 2554-8.

Horrigan, J. (1968). A Short History of Financial Ratio Analysis. *The Accounting Review*, 43(2), 284-94.

Ingersoll, J. (1987). *Theory of Financial Decision Making*. Rowan and Littlefield, Savage, MD.

Jamieson, B. (2008). "Why Northern Rock annual report is a collectors' item". *The Scotsman*, 19 February 2008.

Joy, O. M., & Tollefson, J. O. (1975). On the financial applications of discriminant analysis. *Journal of Financial and Quantitative Analysis*, 10(5), 723-739.

Joy, O. M., & Tollefson, J. O. (1978). Financial Applications of Discriminant Analysis: A Clarification. *Journal of Financial and Quantitative Analysis*, 13(1), 185-195.

Kealhofer, S. (2003) Quantifying credit risk I: Default prediction. *Financial Analysts Journal,* 59, 30-44.

Kealhofer, S. (2003) Quantifying credit risk II: Debt Valuation. *Financial Analysts Journal,* 59, 78-92.

Kealhofer, S., & Kurbat, M. (2001). The default prediction power of the Merton approach, relative to debt ratings and accounting variables. *KMV research paper series*.

Keasey, K., & McGuinness, P. (1990). The failure of UK industrial firms for the period 1976-1984, logistical analysis and entropy measures. *Journal of Business Finance and Accounting,* 17(1), 119-135.

Keasey, K., & Watson, R. (1987). The Prediction of Small Company Failure: Some Behavioural Evidence for the UK. *Accounting and Business Research*, 18(65), 49-58.

Keasey, K., & Watson, R. (1991). Financial distress models: a review of their usefulness. *British Journal of Management,* 2 (2), 89–102.

Kerling, M., & Poddig, T. (1994). Klassifikation von Unternehmen mittels KNN. *Neuronale Netze in der Ökonomie*. München: Vahlen Verlag, 424–490.

Kida, T. (1980). An Investigation into Auditors' Continuity and Related Qualification Judgments. *Journal of Accounting Research*, 18(2), 506-23.

King, G., & Zeng, L. (2001). Logistic Regression in Rare Events Data. Society for Political Methodology, Washington University.

Kiviluoto, K. (1998), Predicting Bankruptcies with the Self-organizing Map. N*eurocomputing,* 21(1–3), 191–201.

Ko, C. J. (1982). *A Delineation of Corporate Appraisal Models and Classification of Bankrupt Firms in Japan*. New York University.

Koh, H. C. (1991). Model predictions and auditor assessments of going concern status. *Accounting and Business Research*, 21(84), 331-338.

Koziol, J. A. (1982). A class of invariant procedures for assessing multivariate normality. *Biometrika*, 69, 423-27.

Kun, Y., Denker, J., and Solla, S. 1990. Optimal Brain Damage. *Advances in Neural Information Processing Systems*, Touretzky D. (Ed.), San Mateo, Calif.: Morgan Kaufmann, 598–605.

Lachenbruch, P. A. (1967). An almost unbiased method of obtaining confidence intervals for the probability of misclassification in discriminant analysis. *Biometrics*, 23(4), 639-645.

Lachenbruch, P. A. (1975). *Discriminant Analysis*. Hafner Press: New York.

Lachenbruch, P. A., Sneeringer, C., & Revo, L. T. (1973). Robustness of the Linear and Quadratic Discriminant Function to Certain Types of Non-Normality. *Communications in Statistics*, 1(1), 39-56.

Laitinen, E. K. (1994). Traditional versus operating cash flow in failure prediction. *Journal of Business Finance and Accounting* 21 (2), 195–217

Laitinen, T., & Kankaanpää, M. (1999). Comparative analysis of failure prediction methods: the Finnish case. *The European Accounting Review*, 8(1), 67-92.

Lara, J. M., Osma, B., & Noguer, B. G. (2006). Effects of database choice on empirical accounting research. *Abacus,* 42(3).

Law, S. (2010). Media Statement: R3 comments on the insolvency statistics for Q1. *R3 Press Release*, 7th May.

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.

Lennox, C., & Pittman, J. (2011). Voluntary Audits verses Mandatory Audits. *The Accounting Review*, 86(5), 1655-1678.

Leshno, M., & Spector, Y. (1996). Neural Network Prediction Analysis: The Bankruptcy Case. *Neurocomputing,* 10(1), 125–147.

Lev, B. (1971). Financial Failure and Informational Decomposition Measures. *Accounting in Perspective: Contributions to Accounting Thoughts by Other Disciplines,* 102-111, Southwest Publishing Co. Cincinnati.

Lev, J. (1969). *Elementary Statistical Methods*. Holt, Rinehart and Winston (New York)

Libby, R. (1975). Accounting Ratios and the Prediction of Failure: Some Behavioural Evidence. *Journal of Accounting Research*, 13(1), 150-161.

Lincoln, E. E. (1925). *Applied Business Finance, 3rd revised edition*. A. W. Shaw Company, 339-58.

Littleton, A. C. (1926). The 2 to 1 Ratio Analyzed. *Certified Public Accountant*, August, 244-46

Lo, A., (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 International Journal of Management Science,* 19 (6), 673–678.

Maddala, G. (1983). *Limited-dependent and qualitative variables in econometrics*. Econometric Society monographs. Cambridge University Press.

Mader, F. (1975). Les Ratios et l'Analyse du Risque. *Analyse Financiere*, 2nd Quarter, 1975.

Mader, F. (1979). Un Enchantillon d'Entreprises en Difficulte. *Journee des Centrales de Bilan*.

Malkovich, J. F., & Afifi, A. A. (1973). On Tests for Multivariate Normality. *Journal of the American Statistics Association*, 68(341).

Mason, R. J., & Harris, F. C. (1979). Predicting Company Failure in the Construction Industry. *Proc. Inst. Civil Engnrs*, 66(2), 301-307.

McCulloch, W., & Pitts, W. (1943). A logical calculus of the ideas imminent in nervous activity. *Bulletin of Mathematical Biophysics.* 5, 119-33.

McFadden, D. (1973). Conditional Logit Analysis of Qualitative Choice Behaviour. *Frontiers in Econometrics*, New York Academic Press.

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

McGurr, P., & DeVaney, S. (1998). Predicting Business Failure of Retail Firms: An Analysis Using Mixed Industry Models. *Journal of Business Research*, 43, 169-176.

McKee, T. E., & Greenstein, M. (2000). Predicting Bankruptcy Using Recursive Partitioning and a Realistically Proportioned Data Set. *Journal of Forecasting,* 19(3), 219–230.

Melumad, N., & Thoman, L. (1990). On auditors and the courts in an adverse selection setting. *Journal of Accounting Research,* 28(1), 77–120.

Merton, R. C. (1973). Theory of rational option pricing. *The Bell Journal of Economics and Management Science*, 4(1), 141-183.

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

Merwin, C. (1942). Financing Small Corporations in Five Manufacturing Industries, 1926-36. *New York: National Bureau of Economic Research*.

Messier, W. F., & Hansen, J. (1988). Inducing rules for expert system development: An example using default and bankruptcy data. *Management Science*, 34(12), 403-415.

Moizer, P. (1985). *Independence*. Current Issues in Auditing, Harper & Row (Chapter), pp. 34-44.

Morris, R. (1997). *Early warning indicators of corporate failure.* Ashgate Publishing

Moses, O. D. (1990). On analysts' earnings forecasts for failing firms. *Journal of Business Finance & and Accounting*, 7(1), 108-118.

Moyer, R. C. (1977). Forecasting financial failure: a re-examination. *Financial Management*, 6(1), 11-17.

Neophytou, E., Charitou, A., & Charalambous, C. (2000). *Predicting Corporate Failure: Empirical Evidence for the UK.* Southampton: Department of Accounting and Management Science, University of Southampton, 1-173.

Nguyen, H. G. (2005). Using Neutral Network in Predicting Corporate Failure. *Journal of Social Sciences*, 1(4), 199-202.

Odom, M., & Sharda, R. (1990). A Neural Network Model for Bankruptcy Prediction. *Proceeding of the International Joint Conference on Neural Networks, San Diego, 17–21 June 1990,* Volume II. IEEE Neural Networks Council, 163–171.

Ohlson, J. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting*, 18(1), 109-131.

Olmeda, I., & Fernandez, E. (1997). Hybrid Classifiers for Financial Multicriteria Decision Making: The Case of Bankruptcy Prediction. *Computational Economics*, 10(4), 317–352.

Ooghe, H., & Balcaen, S. (2002). Are failure prediction models transferable from one country to another? An empirical study using Belgian financial statements, *Proceedings of the 9th Annual Conference of the Multinational Finance Society*, Cyprus.

Ooghe, H., & Joos, P. (1990). Failure prediction, explanation of misclassifications and incorporation of other relevant variables: result of empirical research in Belgium. Working paper, Department of Corporate Finance, Ghent University (Belgium).

Ooghe, H., Joos, P., De Vos, D., & De Bourdeaudhuij, C. (1994). Towards an improved method of evaluation of financial distress models and presentation of their results. Working Paper, Department of Corporate Finance, Ghent University (Belgium),

Ooghe, H., & Verbaere, E. (1985). Predicting business failure on the basis of accounting data: The Belgian experience. *The International Journal of Accounting*, 9(2), 19-44.

Paquette, L. R., & Skender, C. J. (1996). Using a bankruptcy model in the auditing course: The evaluation of a company as a going concern. *Journal of Accounting Education*, 14(3), 319-329.

Peel, M. J. (1987). Timeliness of Private Company Accounts and Predicting Corporate Failure. *Investment Analyst*. 83, 23-27.

Peel, M. J., & Peel, D. A. (1987). Some further empirical evidence on predicting private company failure. *Accounting and Business Research*. 18(69), 57-66.

Pinches, G. (1980). Factors influencing classification results from multiple discriminant analysis. *Journal of Business Research*, 8(4), 429-456.

Pinches, G., Mingo K, & Caruthers J. (1973). The Stability of Financial Patterns in Industrial Organisations. *Journal of Finance*, 28(2), 389-396.

Piramuthu, S., & Raghavan, H. – Shaw, M. (1998), Using Feature Construction to Improve the Performance of Neural Networks. *Management Science,* 44(2), 416–430.

Platt, H. D. & Platt, M. B. (1990). Development of a Class of Stable Predictive Variants: The Case of Bankruptcy Prediction. *Journal of Business Finance & Accounting,* 17(1). 31–44.

Platt, H. D. & Platt, M. B. (1994). Bankruptcy Prediction with Real Variables. *Journal of Business Finance & Acccounting*. 21(4), 491-510.

Pompe, P. P. M., & Feelders, A. J. (1997). Using machine learning, neural networks, and statistics to predict corporate bankruptcy. *Microcomputers in Civil Engineering*, 12, 267-76.

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

R3 (2010). The “insolvency lag”: risks for 2010: Evidence from previous recessions shows early recovery is a dangerous time for struggling businesses and individuals. *R3 Briefing Paper.* Available at https://www.r3.org.uk/

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

Rich, D. (1994). The mathematical foundations of barrier options. *Advances in Futures and Options Research* 7, 267–311.

Rojas, R. (1996). *Neural Networks*. Springer-Verlag, Berlin.

Rosendale, W. M. (1908). Credit Department Methods. *Bankers' Magazine*, 183-184.

Rosenblat, F. (1957). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*. 65, 386-408.

Rosner, R. L. (2003). Earnings manipulation in failing firms. *Contemporary Accounting Research,* 20 (2), 361–408.

Rumelhart, D., Hinton, G., & Williams, R. (1986). Learning Internal Representations by Error *Propagation. Parallel Distributed Processing Exploration in the microstructure of Cognition, Vol. 1*. MIT Press Cambridge.

Salchenberger, L. – Cinar, E. – Lash, N. (1992). Neural Networks: A New Tool for Predicting Thrift Failures. *Decision Sciences,* 23(4), 899–916.

Santomero, A. M., & Vinso, J. D. (1977), Estimating the probability of failure for commercial banks and the banking system, *Journal of Banking & Finance*, 1(2/3), 185-205.

Saulnier, R. J., Halcrow, H. G., & Jacoby, N. H. (1958). *Federal Lending and Loan Insurance*. Princeton University Press, 456-81.

Scott, E. (1978). On the Financial Applications of Discriminant Analysis: Comment. *Journal of Financial and Quantitative Analysis*, 13(1), 201-205.

Scott, J. (1981). The Probability of Bankruptcy: A Comparison of Empirical Predictions and Theoretical Models. *Journal of Banking and Finance*, 5(3), 317-344.

Setiono, R. & Hui, L.C.K. (1995). Use of a quasi-newton method in a feedforward neural network construction algorithm. *IEEE Trans. Neural Networks*, 6(1), 273–277.

Shannon, C.E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, 27.

Shapiro, S.S., & Francia, R.S. (1972). Approximate Analysis of Variance Test for Normality. *Journal of the American Statistical Association*, 67(337), 215-216.

Shumway, T. (2001). Forecasting bankruptcy more accurately. A simple hazard model. *Journal of Business*, 74(1), 101-24.

Sinkey, J. (1975). A Multivariate Statistical Analysis of the Characteristics of Problem Banks. *Journal of Finance*, 30(1), 21-36.

Smith, R. F. & Winakor, A. H. (1935). Changes in Financial Structure of Unsuccessful Corporations. *Bureau of Business Research, University of Illinois*.

Sobehat, J., & Keenan, S. (1999). An introduction to market-based credit analysis. *Moody's KMV.*

Sobehart, J., Stein, R., Mikityanskaya, V., & Li, L. (2000). Moody's Public Firm Risk Model: A Hybrid Approach To Modeling Short Term Default Risk. *Moody's KMV.*

Stein, R. (2002). Benchmarking Default Prediction Models: Pitfalls and Remedies in Model Validation. *Moody's KMV.*

Swingler, K. (1996). *Applying Neural Networks: A Practical Guide*. London: Academic Press.

Taffler, R. J. (1976). Finding Those Firms in Danger. Accountancy Age, 16.

Taffler, R. J. (1982). Forecasting Company Failure in the UK using Discriminant and Financial Ratio Data. *Journal of the Royal Statistical Society*, Series A (General), Part 3, 342-358.

Taffler, R. J. (1983). The Assessment of Company Solvency and Performance Using a Statistical Model. *Accounting and Business Research*, 15(52), 295-308.

Taffler, R. J. (1984). Empirical Models for the Monitoring of UK Corporations. *Journal of Banking and Finance*, 8(2), 199-227.

Taffler, R. J. (1995). The Use of the Z-score Approach in Practice. *Centre for Empirical Research in Finance and Accounting*, City University Business School Working Paper.

Taffler, R. J., & Tisshaw, H. (1977). Going, Going, Gone - Four Factors Which Predict. Accountancy, 88(1003), 50-54.

Taffler, R. J., & Tseung, M. (1984). The audit going concern qualification in practice – conspicuous by its absence? *British Accounting Association*. University of East Anglia.

Tetko, I.V., Livingstone, D.J., & Luik, A.I. (1995). Neural Network Studies. 1. Comparison of Overfitting and Overtraining. *Journal of Chemical Information and Computer Sciences*, 35, 826-833.

Theil, H. (1967). Economics and Information Theory. Center Math. Stud. Business Econ., Univ. Chicago. Chicago, IL

Takahashi, K., Kurokawa, K, & Watese, K. (1979). Predicting Corporate Bankruptcy through Financial Statements. *Society of Management Science of Keio University*, November 1979.

Tam, K. (1991). Neural network models and the prediction of bank bankruptcy. *Omega*, 19(5), 429-45.

Tam, K., & Kiang, M. (1992). Managerial Applications of the Neural Networks: The Case of Bank Failure Predictions. *Management Science,* 38(7), 416–430.

Tamari, M. (1966). Financial ratios as a means of forecasting bankruptcy. *Management International Review*, 4, 15-21.

Tan, C. N. W. (1999), *An Artificial Neural Networks Primer with Financial Applications Examples in Financial Distress Predictions and Foreign Exchange Hybrid Trading System.* Australia: School of Information Technology, Bond University.

Tang, T. C., & Chi, L. C. (2005). Neural networks analysis in business failure prediction of Chinese importers: A between-countries approach. *Expert Systems with Applications*, 29(2), 244-55.

Trigueiros, D. andTaffler,R. 1996. Neural networks and empirical research in accounting. *Accounting and Business Research*, 26, 347–355.

Tudela, M., & Young, G. (2003). Predicting default among UK companies: a Merton approach. *Financial Stability Review*, June 2003.

Udo, G. (1993). Neural network performance on the bankruptcy classification problem. *Computers & Industrial Engineering*, 25, 377-380.

Van Frederikslust, R. (1978) Predictability of Corporate Failure. Leiden: *Martinus Nijhoff Social Science Division*.

Van De Mieroop, M. (2004). King Hammurabi of Babylon: A Biography (Blackwell Ancient Lives). Wiley-Blackwell.

Vasicek, O. (1984). *Credit Valuation*, KMV Corporation.

Vassalou, M., & Xing, Y. (2004). Default Risk in Equity Returns. *Journal of Finance*, 59(2), 831-868.

Virag, M. & Kristof, T. (2005). Neural Networks in Bankruptcy Prediction – A Comparative Study on the Basis of the First Hungarian Bankruptcy Model. *Acta Oeconomica.* 55 (4), 403–425.

Von Stein, J. H. (1981). *Identifying Endangered Firms*. Hohenheim University.
Wall, A. (1919). Study of Credit Barometrics. *Federal Reserve Bulletin*, March, 229-43.

Wall, A. & Duning, R. W. (1928). *Ratio Analysis of Financial Statements*. Harper and Brothers, 152-79.

Walter, J. E. (1957). Determination of Technical Solvency. *The Journal of Business*, January, 30-43.

Ward, T. J., & Foster, B. P. (1997). A note on selecting a response measure for financial distress. *Journal of Business Finance & Accounting*, 24(6), 869–879.

Watson, C. (1990). Multivariate Distributional Properties, Outliers, and Transformation of Financial Ratios. *The Accounting Review*, 65(3), 682-695.

West, R. G. (1985). A factor-analytic approach to bank condition. *Journal of Banking and Finance*, 9(2), 253-266.

Weston, F. J., & Brigham, E. F. (1966). *Managerial Finance. 2nd ed.* New York: Holt, Rinehart and Winston, 1966.

White, R. W., & Turnbull, M. (1975a). The probability of bankruptcy: American railroads. Working paper, Institute of Finance and Accounting, London University Graduate School of Business, February.

White, R. W., & Turnbull, M. (1975b). The probability of bankruptcy for American industrial firms. Working paper, Institute of Finance and Accounting, London University Graduate School of Business, July.

Widrow, B., & Hoff, M. (1960). Adaptive switching circuits. *IRE-WESCON Convention Record*, 4, 96-104

Wilcox, J.W. (1976). The Gambler's Ruin Approach to Business Risk. *Sloan Management Review*. March 1976.

Wilson, R., & and Sharda, R. (1994). Bankruptcy prediction using neural networks. *Decision Support Systems*, 11(5), 545-57.

Wood, D., & Piesse, J. (1987). The information value of MDA based financial indicators. *Journal of Business Finance & Accounting* 14(1), 27–38.

Yang, Z. (2001). *A New Method for Company Failure Prediction Using Probabilistic Neural Networks.* Exeter: Department of Computer Science, Exeter University.

Yang, Z. R., Platt, M. B. & Platt, H. D. (1999). Probabilistic neural networks in bankruptcy prediction. *Journal of Business Research*, 44(2), 67-74.

Yao, X. (1999). Evolving artificial neural networks. *Proceedings of the IEEE*, 87(9), 1423–1447.

Yim, J., & Mitchell, H. (2005). A Comparison of Corporate Distress Prediction Models in Brazil: Hybrid Neural Networks, Logit Models and Discriminant Analysis. *Nova Economia BeloHorizonte,* 15(1), 73–93.

Zavgren, C. (1983). The prediction of corporate failure: the state of the art. *Journal of Accounting Literature*, 2, 1-38.

Zavgren, C. (1985). Assessing the vulnerability to failure of American industrial firms: a logistic analysis. *Journal of Business Finance & Accounting*, 21(1), 19-45.

Zhang, Q. & Benveniste, A. (1992). Wavelet Networks. *IEEE Trans. Neural Networks*, 3(6), 889–898.

Zhang, G. – Hu, M. – Patuwo, B. (1999). Artificial Neural Networks in Bankruptcy Prediction: General Framework and Cross-validation Analysis. *European Journal of Operational Research,* 116, 16–32.

Zmijewski, M. (1984). Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research*, 22(suppl.), 59-86.