PREDICTING CORPORATE FAILURE THROUGH A COMBINATION OF INTELLIGENT TECHNIQUES

BY

SVERRE GUNNERSEN
B.BUS.SYS. (HONS)

A THESIS SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DOCTOR OF PHILOSOPHY

FACULTY OF INFORMATION TECHNOLOGY
MONASH UNIVERSITY

SUBMITTED DECEMBER 2013

Notice 2

I certify that I have made all reasonable efforts to secure copyright permissions for third-party content included in this thesis and have not knowingly added copyright content to my work without the owner's permission.

Declaration of Originality

I, Sverre Gunnersen, declare that this thesis is my own work and has not been submitted in any form for another degree or diploma at any university or other institute of tertiary education. Information derived from the published and unpublished work of others has been acknowledged in the text and a list of references is given in the bibliography.

Signature: [redacted] **Date:** 24/4/2014

Acknowledgements

This research would like to acknowledge funding provided by the Australian Research Council (ARC) Linkage Projects (LP), Australian Postgraduate Award Industry (APAI), grant LPO453884.

Sincere appreciation to Lincoln Indicators Pty Ltd who initiated the APAI grant that funded this research, supplied much of the Australian data used, and provided assistance in post-processing that data.

Special thanks go to Professor Kate Smith-Miles and Associate Professor Vincent CS Lee, the supervisors of this thesis, as without them this work would not have been possible.

Abstract

Corporate failure is one of the most popular prediction problems because early identification of at-risk companies presents such a clear economic benefit to creditors, investors and society as a whole. Throughout the years statistically based classification systems, intelligent systems such as Neural Networks with its many variants, and newer techniques such as Genetic Programming have been applied to this problem. Indeed when a new variation or technique is proposed, the prediction of corporate failure is often one of the first test domains for the new methodology. Likewise, the cause of corporate failure is a topic that has received much academic and literary attention, including case studies investigating the trajectories that failing companies take or post hoc qualitative analysis as to whether certain fundamental causes such as one-man-rule can be attributed to the subsequent collapse of a company. However, throughout the history of this topic a number of challenges emerge that remain unaddressed within the literature.

The first challenge is that while many papers outlining new classification techniques compare results with another popular classification system as a baseline, little research exists that comprehensively compares many classification techniques across multiple datasets. This thesis finds that intelligent techniques such as Neural Networks, Genetic Programming and Support Vector Machines outperform statistical techniques such as Discriminant Analysis and Logistic Regression.

The second challenge is that the desire of researchers to compare results has resulted in the use of the same cross-section of factors, with little analysis as to whether or not the factors being used are impacting on the classification accuracy of the method. This thesis finds that an objective factor selection methodology leads to performance gains.

The third is that far less research exists that considers whether share market or macroeconomic data can have a positive impact on classification accuracy. While this research did find some performance gains when including share market information, the difficulty of linking financial information with share market information leads to data loss that outweighs the small performance improvement.

The fourth is that while most classificatory research on this problem focuses on the accuracy of the technique, less attention is given to whether the subjective clustering methods used (e.g. by “industry”) are effective, and this research finds that an objective clustering technique improves classification accuracy. Furthermore, this research builds on the existing cluster visualisation methods by developing a new and more effective cluster visualisation algorithm.

Finally this research attempts to contribute to the theoretical understanding of corporate failure by analysing the classificatory surface of the resulting predictive models and performing a case study analysis of failed companies. In doing so, the model’s strengths and limitations are discussed and some of the causes of failure from the literature are identified.

In summary, this research makes the following contributions to the field of bankruptcy prediction: a literature review of notable bankruptcy prediction research, a comparison of popular classification techniques, the development and testing of a new objective factor selection methodology, an examination of the effect of share market and macroeconomic data on classification accuracy, the development and testing of a new cluster visualisation method that overcomes limitations in existing methods, an examination of the effect of objective clustering on classification accuracy utilising the new visualisation method, and a case-study analysis on selected failed companies that relates the reasons for failure outlined in secondary sources to the company’s failure prediction trajectory.

Table of Contents

List of Tables

List of Figures

1. Introduction

Between 1999 and 2001, the Australian telephone carrier One.Tel consumed over $1 billion in cash before entering voluntary administration. More recently, American Airlines' parent company filed for Chapter 11 bankruptcy protection, and the bankruptcies of WorldCom and Enron preceded the Global Financial Crisis. Predicting financial distress is a well-researched problem, though one in which there is no consensus on the methodology to use. Research into ratio analysis dates back as far as 1935 (Smith & Winakor, 1935), and building predictive models dates back to 1966 (Beaver, 1966). With a trend away from linear, statistical analysis techniques towards non-linear intelligent systems, developments that seek to improve predictive accuracy continue today. Not only do these new techniques demonstrate greater accuracy, but some allow opportunities to build a greater understanding of corporate failure, its causes and its symptoms.

Many of these developments focus on the modelling system itself, using financial ratios from key previous research, such as Altman (1968) with small, hand-picked samples known to be similar. Very little of the literature, with the notable exception of Edmister (1972), advocates alternative sets of inputs to the classification system. Furthermore most research considers only one dataset, and questions are therefore raised as to whether or not the methodology proposed in such research is extendable to alternative datasets.

Due to this focus on the classificatory model itself, little research exists on the effect of clustering bankruptcy data. Specifically Deboeck & Kohonen (1998) utilise a Self-Organising Feature Map, which was then subsequently shown to be an excellent clustering algorithm, but employ the methodology only as an unsupervised learning algorithm, leaving an opportunity for the hybridisation of this technique and supervised learning algorithms to measure the effect of objective clustering.

Finally, while Altman's Z-Score (1968) can be used to better understand the key performance indicators of corporate failure, many of today's proposed models of corporate failure prediction are highly complex non-linear algorithms, from which it is difficult or impossible to gain an understanding into the nature of corporate failure and therefore it can be difficult to trust the resulting model.

1.1 Goals of this Research

The goals of this thesis can be roughly divided into a number of sections. Firstly, this thesis seeks to examine the current state of corporate failure prediction and corporate failure theory to identify opportunities to make contributions to both fields. To this end, specific questions will be developed that shape the direction of this thesis.

Secondly, this thesis seeks to prepare the data, evaluate the available modelling techniques, and develop an objective factor-selection methodology that improves classification accuracy. Then through the use of multiple datasets support or refute the proposition that a particular set of factors can be utilised in different scenarios without needing to re-perform the objective factor search.

Thirdly, this thesis aims to consider whether additional external information, such as share market data and macroeconomic data, provides additional information above and beyond that located in a company's annual financial statements, in regards to the company's failure classification.

Fourthly, it considers whether clustering, specifically building on the works of Deboeck & Kohonen (1998), can be used to improve classification accuracy, in particular when hybridised with supervised learning algorithms. This thesis is distinctive in that it specifically compares objective clustering versus the effect of using the industry standard classification systems typically employed within corporate failure classification research. To do this, this thesis examines common cluster visualisation methods and considers whether improvements can be made in this area.

Finally this thesis considers whether the results of the objective factor selection, clustering, and the use of classification systems that result in successful models can be used to build a greater understanding of the key indicators of corporate failure and how they relate to the theoretical causes of failure as outlined by authors such as Argenti (1976).

1.2 Summary of Findings

This thesis demonstrates that the objective factor selection methodology proposed significantly increases out-of-sample accuracy on both a U.S.-centric and an Australian-centric dataset. It supports the findings of Edmister (1972) that the best selection of factors is unique to that particular dataset, that therefore it is necessary to re-perform factor selection when the dataset changes.

It is found that neither the inclusion of the chosen share market data nor macroeconomic data reliably increase the accuracy of the classification models and in the case of share market data in fact severely limits the availability of data.

While investigating the objective clustering of companies, this thesis identifies theoretical and practical weaknesses in a common cluster visualisation technique, then goes on to develop, test and utilise a new cluster visualisation methodology. It then uses that methodology to perform an objective clustering of companies using a multi-level Self-Organising Feature-Map, hybridised with supervised learning algorithms, finding that it significantly increases the classification accuracy beyond un-clustered data. It further finds that the grouping of companies by industry, a technique often used for research into corporate failure, can in fact decrease the net accuracy of the model.

This research finds that while new and useful key performance indicators of corporate failure can be identified, they are often utilised by the model in a non-linear sense which helps explain the generally poorer performance of linear classification models. It finds evidence for some of the commonly believed causes of failure such as overtrading, but that income or sales related data is often the best predictor of corporate failure.

1.3 Thesis Structure

This research proceeds with chapter 2 which is an analysis of literature in the area of predicting corporate failure, identifying three major styles of modelling systems used from 1966 through to today. Chapter 3 goes beyond the modelling of failure to identify the theorised causes and symptoms of failure to identify opportunities for contributions to the literature.

Chapter 4, Methodology, includes an assessment of the available data and the preparation of that data, followed by an evaluation of available modelling techniques, including Discriminant Analysis, Neural Networks, Support Vector Machines, Genetic Programming and Logistic Regression. It then outlines the methodology that will be used for objective factor selection in the following chapters, and compares the classification accuracy of that methodology to one in which objective factor selection is not performed.

The data analysis is divided into two halves. The first half, found in chapter 5, examines the effect of adding additional data such as share market or macroeconomic data, while the second half, in chapter 6, builds a new cluster visualisation technique and then uses that technique to test the effect of objective clustering on bankruptcy data.

Having determined the effective techniques in the previous chapters, chapter 7 uses these techniques to predict corporate failure and selects specific cases within those datasets for additional analysis.

Finally chapter 8 summarises the findings of this thesis and presents opportunities for additional research.

This structure can be visualised in Figure 1-1.

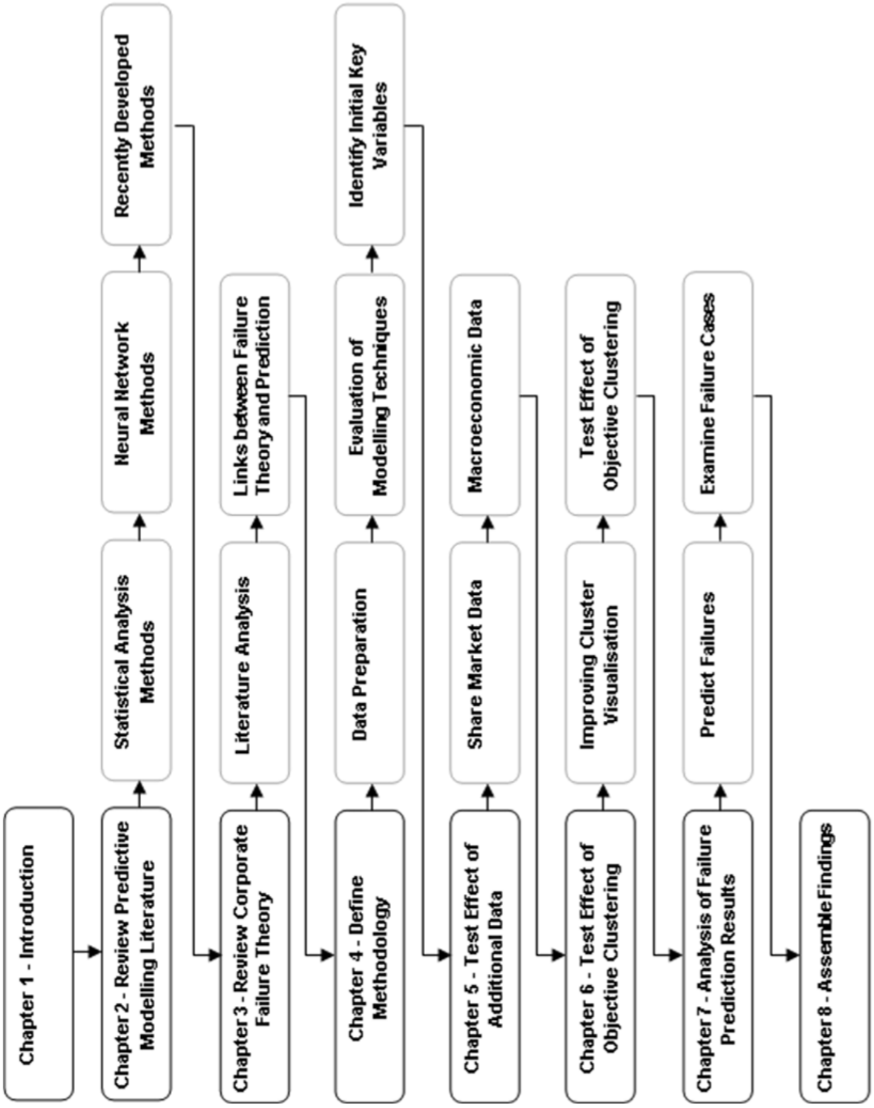


Figure 1-1 - Thesis Stucture

2. Review of Predictive Modelling Literature

The “problem” of corporate failure prediction covers two fields of study: firstly the building of predictive computer models sits comfortably in the field of information technology, or more specifically, intelligent computing, while the theoretical model of corporate failure sits more comfortably in the field of finance. While this chapter will address the development of corporate failure predictive models, chapter 3 looks at the finance aspect.

The arena of quantitative prediction of corporate failure spans back to Beaver (1966, p. 91), “Financial Ratios as Predictors of Failure” which found that “ratio analysis can be useful in the prediction of failure for at least five years before failure”. This finding spurred two key pieces of research, Altman (1968) and Beaver (1968), which in turn has led to a collection of papers referenced in almost any publication to do with the prediction of corporate failure (Edmister, 1972; Deakin, 1972; Wilcox, 1973; Blum, 1974; Libby, 1975; Altman, et al., 1977; Ohlson, 1980; Odom & Sharda, 1990; Coats & Fant, 1993; Altman, et al., 1994; Wilson & Sharda, 1994).

The evolution of corporate failure prediction initially began with statistically based Univariate Analysis, progressed to Multivariate and Multiple Discriminant Analysis, Probabilistic Theory, Logit Regression and Rough Sets Theory. Since then, the techniques typically used have begun to fall into Mitchell’s now widely-cited definition of machine learning (Mitchell, 1997, p. 2) by using techniques such as Artificial Neural Networks and Genetic Algorithms. This classification of techniques as “statistical” or “intelligent” is often used in the modern literature, such as Chen et al. (2011) who state “these models have progressed from statistical methods to the artificial intelligence (AI) approaches” and so this research’s structure reflects those commonly accepted classifications of techniques.

While this thesis has roughly divided popular techniques into “statistical” or “intelligent”, it must be acknowledged that iterative “statistical” techniques could be argued to meet the definition of machine learning, and that many “intelligent” techniques such as Support Vector Machines utilise many statistical methods. Nevertheless, the following sections are broken up into each of the major areas with the goal of maintaining a degree of structure while introducing research papers in a sequential manner.

Within each section, the major dimensions of the research are discussed, including research type, data collection technique, sample, theory, findings, strengths and weaknesses. Each section will attempt to analyse, compare and critique the pieces of research within.

2.1 Selection of Literature to Review

Since it is not reasonable to review every piece of research that has been written about corporate failure, it is necessary to clearly identify the scope of the review, and outline the method for selecting which of the available sources should be examined.

Papers on corporate failure prediction were selected on the following criteria:

- Must be predicting corporate failure using a computational methodology
- Must be published in an academic journal
- Must be referenced by other peer-reviewed papers

To build an initial population of papers on corporate failure prediction, searches were performed on the following online databases for articles:

- ProQuest – http://www.proquest.com
- EBSCOhost – http://ejournals.ebsco.com
- ScienceDirect – http://www.sciencedirect.com
- Wiley InterScience – http://www.interscience.wiley.com

These databases were chosen as they contained a higher number of related journals than other sources.

From these searches, recent papers relevant to predicting corporate failure were selected and their references noted. From the list of references in each of those publications, other relevant articles were accessed, added to the review and their own references similarly followed. From this population of research, the “reverse citations” feature of the above databases were utilised. Research was selected for review based on its relevance to this thesis and perceived contribution to the body of knowledge.

In an attempt to compare each paper, a number of criteria common across corporate failure prediction research have been specified and investigated. Those criteria are:

- The definition of corporate failure used
- The sample selection used
- The dimension reduction method used
- The variables ultimately chosen for the model
- The research methodology
- The key findings

It should be noted the dimension reduction is typically two separate stages: feature selection plus feature extraction. However in the literature included in this study, these two stages are often executed using the one algorithm (such as Principal Component Analysis), and as such it is convenient to consider it one “stage” in the following literature review.

This process resulted in the 34 highly relevant articles found in each of the below sections within this chapter. Of those articles, a pattern of the evolution of corporate failure prediction starts to emerge. The selected papers can be roughly divided into those from 1966 to 1989 that focus on the use of statistical analysis, those that are generally from 1990 to 1999 that focus on Neural Networks and those that utilise more recent developments in the field of corporate failure prediction from 2000 onwards. This structure has been represented in the design of this chapter and each section.

2.2 Prediction by Statistical Analysis

Statistical analysis includes the use of univariate and multivariate statistical techniques, including Discriminant Analysis, as a means of predicting corporate failure. While Discriminant Analysis is by far the most popular predictive statistical technique in the field of statistical analysis, there are some exceptions that will also be analysed in this section.

Discriminant Analysis is a statistical methodology that forms a (generally quadratic) equation that can be used to describe the separation of groups of data. In this case, the quadratic equation describes the separation between failed and non-failed corporations, and uses an array of input variables, such as financial ratios, to do so. Discriminant Analysis can be traced back to the early 1900's, though it was Beaver's paper (1966) that first applied it to the field of corporate failure prediction. Since then it has been used as the benchmark for all other techniques, such as in Altman et al.'s paper (1994), "Corporate distress diagnostic: Comparisons using linear Discriminant Analysis and Neural Networks (the Italian experience)".

Though Discriminant Analysis is a somewhat older methodology for calculating the probability of group membership, it has a number of benefits such as a simple discriminant function and extremely fast processing time. Compare this to the discriminant function that results from a fully trained multi-layered Neural Network (discussed in more detail later on), in which each neuron represents a generally sigmoidal function, the cross-connectivity of neurons creates a highly complex algorithmic result and it becomes clear that a simpler discriminant function is far more "explainable" when discussing group assignments. Furthermore, Discriminant Analysis can be highly accurate. Altman's exploratory paper (1968), "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy" identified a discriminant function that was 95% accurate in the dataset in the year prior to failure.

Given the strengths and impact of statistical analysis, especially that of Discriminant Analysis, statistical analysis in general is a principal methodology in the field of corporate failure prediction. The following subsections examine each of the elements of the various papers based on statistical analysis, with the intention of critically analysing and comparing them, and also providing an overall insight into the evolution of the field in general.

2.2.1 Definition of Corporate Failure

Prior to any statistical analysis taking place, it is necessary for the research to establish which companies within the sample selection (see section 2.2.2 below) have failed, and which have not. In turn this can be used for pairing the samples and deducing an algorithm to predict failure. To establish which companies have failed and which have not, a definition of corporate failure needs to be established. The different definitions used within the relevant literature have been outlined in Table 2-1 below.

Author	**Title**	**Year**	**Definition of Corporate Failure**
Beaver, W.	Financial Ratios as Predictors of Failure	1966	"'Failure' is defined as the inability of a firm to pay its financial obligations as they mature. Operationally, a firm is said to have failed when any of the following events have occurred: bankruptcy, bond default, an overdrawn bank account, or non-payment of a preferred stock dividend."
Altman, E.	Financial Ratios. Discriminant Analysis and the Prediction of Corporate Bankruptcy	1968	"The bankrupt group (1) are manufacturers that filed a bankruptcy petition under Chapter X of the National Bankruptcy Act".
Beaver, W.	Market Prices, Financial Ratios, and the Prediction of Failure	1968	Same definition as Beaver (1966) used.
Edmister, R.	An Empirical Test of Financial Ratio Analysis for Small Business Failure Prediction	1972	Businesses reporting a "loss loan" case to the Small Business Administration.
Deakin, E.	A Discriminant Analysis of Predictors of Business Failure	1972	Firms that "experienced bankruptcy, insolvency, or were otherwise liquidated for the benefit of creditors."

Author	Title	Year	Definition of Corporate Failure
Wilcox, J.	A Prediction of Business Failure Using Accounting Data	1973	Firms that had made a “court filing for Chapter X or XI bankruptcy”.
Blum, M.	Failing Company Discriminant Analysis	1974	“The operational definition of failure is based on the criteria of *International Shoe*, that is, events signifying an inability to pay debts as they come due, entrance into a bankruptcy proceeding, or an explicit agreement with creditors to reduce debts.”
Elam, R.	The Effect of Lease Data on the Predictive Ability of Financial Ratios	1975	“[F]irms are defined as bankrupt when they have undertaken at least one of the following actions: (1) filed for reorganization under Chapter XI of the Federal Bankruptcy Act (Chandler Act); (2) filed for reorganisation under Chapter X of the Chandler Act; (3) voted in a stockholders' meeting to file either under Chapter X or Chapter XI; (4) reached agreement with creditors to reduce firm's liabilities at a loss to the creditors.”
Altman, et al.	A new model to identify bankrupt risk of corporations	1977	Defined as a company that has filed a “bankruptcy petition”.
Ohlson, J.	Financial Ratios and the Probabilistic Prediction of Bankruptcy	1980	“The failed firms must have filed for bankruptcy in the sense of Chapter X, Chapter XI, or some other notification indicating bankruptcy proceedings.”
Chen, K. & Shimerda, T.	An Empirical Analysis of Useful Financial Ratios	1981	Failure not explicitly defined.
Barniv, R. & Raveh, A.	Identifying financial distress: a new nonparametric approach	1989	Definition of failure adopted from the study that is being used as the benchmark.

Table 2-1 - Comparison of Definition of Corporate Failure for Statistical Analysis Methodologies

The majority of this research has opted for a bankruptcy court filing definition, though including firms that are unable to meet loans that are due is often used. Note that the small sample sizes used (see section 2.2.2), allow the researchers to manually identify failed companies according to whatever criteria they see fit.

While the differences in chosen definitions make comparing the research more difficult, it is certainly not surprising as different stakeholders will have different perspectives on when the company that they have an interest in should be considered “failed”.

Regardless, assuming the resulting model is able to provide some kind of index describing the mathematical “closeness” to failure, even a model using a conflicting definition of failure should be able to provide great insight into the overall risk of the firm for creditors or investors with different perspectives.

2.2.2 Sample Selection

In any quantitative research, sample selection is a key foundational aspect that impacts on the entire research process. A poor sample selection method can lead to results that are not indicative of the population.

The sample selection techniques used by the statistical analysis research this literature review is critiquing is outlined in Table 2-2, below.

Author	**Title**	**Year**	**Sample Selection Method**
Beaver, W.	Financial Ratios as Predictors of Failure	1966	Moody’s Industrial Manual identified 79 failed firms, which were paired with 79 non-failed firms with the “same industry and asset size”.
Altman, E.	Financial Ratios. Discriminant Analysis and the Prediction of Corporate Bankruptcy	1968	33 manufacturing firms that filed a bankruptcy petition, paired with 33 non-failed firms selected on a “stratified random basis”.
Beaver, W.	Market Prices, Financial Ratios, and the Prediction of Failure	1968	Same sample as Beaver (1966) used.
Edmister, R.	An Empirical Test of Financial Ratio Analysis for Small Business Failure Prediction	1972	42 small businesses selected from data provided by Small Business Administration and Robert Morris Associates using restrictive criteria, and 562 small businesses selected based on less restrictive criteria, ensuring an “equal number of loss and non-loss cases”.
Deakin, E.	A Discriminant Analysis of Predictors of Business Failure	1972	32 failed firms selected and matched against 32 non-failed firms from the same “industry classification, year of the financial information provided and asset size”.
Wilcox, J.	A Prediction of Business Failure Using Accounting Data	1973	Moody’s Industrial Manual identified 52 failed firms, which were paired with 52 non-failed firms selected alphabetically within the same industry, asset size, and met data availability criteria.
Blum, M.	Failing Company Discriminant Analysis	1974	“Paired sample of 115 failed and 115 nonfailed firms” of unknown origin.

Elam, R.	The Effect of Lease Data on the Predictive Ability of Financial Ratios	1975	48 failed firms that were listed as bankrupt by The Wall Street Journal Index and a Dun and Bradstreet supplied list of >$1M failures, met criteria such as having financial information available in Moody's Industrial Manual as well as having leave data available, were matched with 48 non-failed firms selected from the Compustat Annual Industrial Tape and met criteria such as being in the same industry as well as having leave data available.
Altman, et al.	A new model to identify bankrupt risk of corporations	1977	"Two samples of firms consist of 53 bankrupt firms and a matched sample of 58 non-bankrupt entities. The latter are matched to the failed group by industry and year of the data" of unknown origin.
Ohlson, J.	Financial Ratios and the Probabilistic Prediction of Bankruptcy	1980	"The data were obtained … from 10-K financial statements as reported at the time", with the "final sample [being] made up of 105 bankrupt firms", compared with "a sample of nonbankrupt firms obtained from the *Compustat* tape."
Chen, K. & Shimerda, T.	An Empirical Analysis of Useful Financial Ratios	1981	All 1053 firms from the COMPUSTAT tape during the specified time period.
Barniv, R. & Raveh, A.	Identifying financial distress: a new nonparametric approach	1989	"The first sample includes 200 industrial firms and was employed by FAK (1985). They randomly selected 142 non-bankrupt manufacturing firms and retailing companies from the COMPUSTAT files as their control group, while the analysis subsample composed 58 bankrupt companies". "The second sample includes 69 non-life insurance companies which failed during 1975-1983. The solvent group consists of 69 non-life insurers randomly selected from A. M. Best files."

Table 2-2 - Comparison of Sample Selection Methods for Statistical Analysis Methodologies

In all cases examined here, extreme case and stratified sampling has been used (Patton, 2002) to select as many failed firms as possible and select an approximately equal number of non-failed firms. In most cases, paired sampling has been used, such as in Beaver (1966), but in a few cases independent samples have been used such as in Altman et al. (1977).

Given the deviant nature of corporate failure, extreme case sampling is to be expected, and to ensure a lack of model bias towards non-failed firms stratified sampling is the obvious choice for selecting non-failed firms. The weakness in using this approach is that the frequency of non-failed firms to failed firms is no longer representative of the population, disallowing many types of statistical analysis. This does not necessarily translate to a weakness in the research itself, due to the comparative nature between failed and non-failed firms, but does make it difficult to apply such models to the greater non-paired population.

2.2.3 Variable Selection & Dimension Reduction

The process of dimension reduction in most statistical analysis techniques is highly resource intensive. Unless the researcher intends to perform their choice of predictive technique on every available combination of variables, it becomes necessary to reduce down the hundreds of variables, ratios and indicators to a small number of key parameters that can be used to formulate a classificatory model.

There is a vast array of variable selection and dimension reduction methods shown in Table 2-3.

Author	**Title**	**Year**	**Variable Selection Method**
Beaver, W.	Financial Ratios as Predictors of Failure	1966	30 financial ratios selected based on popularity in the literature, performance in previous studies, or "that the ratio be defined in terms of a 'cash flow' concept".
Altman, E.	Financial Ratios. Discriminant Analysis and the Prediction of Corporate Bankruptcy	1968	An array of 22 ratios selected by popularity, potential relevancy, "and a few 'new' ratios". From this, 5 ratios were selected based on predictive performance.
Beaver, W.	Market Prices, Financial Ratios, and the Prediction of Failure	1968	Annual rates of return computed for all firms included in the study.
Edmister, R.	An Empirical Test of Financial Ratio Analysis for Small Business Failure Prediction	1972	Ratios "advocated by theorists" or found to be "significant predictors of business failure in previous empirical research".
Deakin, E.	A Discriminant Analysis of Predictors of Business Failure	1972	Replicated the Beaver **(1966)** study.
Wilcox, J.	A Prediction of Business Failure Using Accounting Data	1973	Used net income, dividends, stock issued, cash, current assets, total assets, total liabilities, adjusted cash flow and adjusted cash position.
Blum, M.	Failing Company Discriminant Analysis	1974	Selected ratios using a "cash-flow framework", grouped into "liquidity, profitability, and variability".
Elam, R.	The Effect of Lease Data on the Predictive Ability of Financial Ratios	1975	Used ratios that are "commonly discussed in financial literature and textbooks".
Altman, et al.	A new model to identify bankrupt risk of corporations	1977	Variables "found in other studies to be helpful", as well as "several 'new' measures that were thought to be potentially helpful as well".
Ohlson, J.	Financial Ratios and the Probabilistic Prediction of Bankruptcy	1980	"The criterion for choosing among different predictors was simplicity."
Chen, K. & Shimerda, T.	An Empirical Analysis of Useful Financial Ratios	1981	From "thirty-four financial ratios [that had] been found by researchers to be significant variables in the prediction of firm failure in recent studies", principal component analysis was used to identify key variables.
Barniv, R. & Raveh, A.	Identifying financial distress: a new nonparametric approach	1989	Variable selection method adopted from the study that is being used as the benchmark.

Table 2-3 - Comparison of Variable Selection Methods for Statistical Analysis Methodologies

In all cases, the initial ratio selection method was highly subjective, ranging from the researcher's perception of popularity to ratios found in a proposed theoretical cash flow model.

The importance of ratio selection is somewhat de-emphasised when developing a "proof of concept", such as Beaver's (1966; 1968) or Altman's (1968) papers. In these instances the goal is simply to demonstrate the potential of the technique for corporate failure prediction. Fine tuning the model to ensure maximum predictive ability is not of great concern. However, by the

1970's papers such as Elam's paper (1975) were being published, which aimed to test the effect of lease data on the predictive model. Arguably there would be value in ensuring a good initial ratio selection technique has been used. By Barniv & Raveh's (1989) paper, the standard of adopting variables that are identical to that used in a benchmark study had been set, and as will be discussed in the following chapters, in general, the benchmark study is often Beaver's (1966) or Altman's (1968) paper.

Of particular interest is the paper of Chen & Shimerda (1981), which examined all the ratios used by various key pieces of research in the past, and empirically identifies the importance of these variables, making the broader statement that any single variable from each factor (such as liquidity, profitability etc.) tends to cover the majority of the available information.

Some variables are arguably more popular than others in these studies. Table 2-4 below outlines the variables that were selected by each of the major research papers for inclusion in their statistical analysis.

Author	**Title**	**Year**	**Variables**
Beaver, W.	Financial Ratios as Predictors of Failure	1966	Cash flow to sales, cash flow to total assets, cash flow to net worth, cash flow to total debt, net income to sales, net income to total assets, net income to net worth, net income to total debt, current liabilities to total assets, long-term liabilities to total assets, current plus long-term liabilities to total assets, current plus long-term plus preferred stock to total assets, cash to total assets, quick assets to total assets, current assets to total assets, working capital to total assets, cash to current liabilities, quick assets to current liabilities, current assets to current liabilities, cash to sales, accounts receivable to sales, inventory to sales, quick assets to sales, current assets to sales, working capital to sales, net worth to sales, total assets to sales, cash to fund expenditures for operations, defensive assets to fund expenditures for operations, defensive assets minus current liabilities to fund expenditures for operations.

Author	Title	Year	Variables
Altman, E.	Financial Ratios. Discriminant Analysis and the Prediction of Corporate Bankruptcy	1968	Working capital to total assets, retained earnings to total assets, earnings before interest and taxes to total assets, market value equity to book value of total debt, sales to total assets.
Beaver, W.	Market Prices, Financial Ratios, and the Prediction of Failure	1968	Cash dividend paid on security during previous period plus current price for security minus adjusted price for security during previous period to adjusted price for security during previous period, residual rate of return during previous period,
Edmister, R.	An Empirical Test of Financial Ratio Analysis for Small Business Failure Prediction	1972	Current assets minus inventory to current liabilities, current assets to current liabilities, inventory to net working capital, net working capital to total assets, current assets to total debt, total debt to equity, fixed assets to equity, cash flow to current liabilities, current liabilities to equity, equity and long-term debt to fixed assets, inventory to sales, fixed assets to sales, total assets to sales, net working capital to sales, equity to sales, earnings before taxes to sales, earnings before taxes to total assets, earnings before taxes to equity, earnings before taxes plus depreciation to total debt.
Deakin, E.	A Discriminant Analysis of Predictors of Business Failure	1972	Cash flow to total debt, net income to total assets, total debt to total assets, current assets to total assets, quick assets to total assets, working capital to total assets, cash to total assets, current assets to current liabilities, quick assets to current liabilities, cash to current liabilities, current assets to sales, quick assets to sales, working capital to sales, cash to sales.
Wilcox, J.	A Prediction of Business Failure Using Accounting Data	1973	Net income, dividends, non-cash current assets, long-term assets, stock issued, total liabilities.
Blum, M.	Failing Company Discriminant Analysis	1974	Cash plus notes receivable plus market securities plus sales to cost of goods sold minus depreciation plus selling and administrative expense plus interest, net quick assets to inventory, cash flow to total liabilities, market value to total liabilities, book value to total liabilities, rate of return to stockholders, standard deviation of net income, trend breaks for net income, slope for net income, standard deviation of net quick assets to inventory, trend breaks for net quick assets to inventory, slope for net quick assets to inventory.

Author	Title	Year	Variables
Elam, R.	The Effect of Lease Data on the Predictive Ability of Financial Ratios	1975	Cash to current liabilities, current assets to current liabilities, current assets minus inventories to current liabilities, cash flow to sales, cash flow to total assets, cash flow to net worth, cash flow to current liabilities, net worth to total liabilities, net worth to long-term liabilities, net worth to fixed assets, net operating profit to interest, sales to inventory, sales to accounts receivable, sales to working capital, sales to current assets minus inventories, sales to cash, net operating profit to sales, net profits to sales, sales to fixed assets, sales to total assets, sales to net worth, net income to net worth, net operating profit to total assets, net operating profit to total debt, current liabilities to total assets, long-term liabilities to current assets, current plus long-term liabilities to total assets, current plus long-term liabilities plus preferred stock to total assets.
Altman, et al.	A new model to identify bankrupt risk of corporations	1977	Earnings before interest and taxes to total assets, net available for total capital to total capital, sales to total assets, sales to total capital, earnings before interest and tax to sales, net available for total capital to sales, log tangible assets, interest coverage, log interest coverage and working capital to long-term debt, fixed charge coverage, earnings to debt, earnings to 5 year mats, cash flow to fixed charges, cash flow to total debt, working capital to long-term debt, current assets to current liabilities, working capital to total assets, working capital to cash expenses, retained earnings to total assets, book equity to total capital, market value equity to total capital, 5 year market value equity to total capital, standard error of estimate of earnings before interest and taxes to total assets (norm), earnings before interest and taxes drop, margin drop, capital lease to total assets, sales to fixed assets.
Ohlson, J.	Financial Ratios and the Probabilistic Prediction of Bankruptcy	1980	Total assets to GNP price-level index, total liabilities to total assets, working capital to total assets, current liabilities to current assets, boolean test for total liabilities being greater than total assets, net income to total assets, funds provided by operations to total liabilities.
Chen, K. & Shimerda, T.	An Empirical Analysis of Useful Financial Ratios	1981	Any single ratio "to represent a factor can account for most of the information provided by all the ratios of that factor." "Still, the question of which ratio should represent a factor has yet to be resolved."

Author	Title	Year	Variables
Barniv, R. & Raveh, A.	Identifying financial distress: a new nonparametric approach	1989	For industrial companies: Net income to total assets, current assets to total assets, log of total assets, market value of equity to total capitalization, current assets to total assets, cash flow to total debt, quick assets to total assets, quick assets to total liabilities, earnings before interest and taxes to total assets, log of interest coverage. For insurance companies: Stability of the ratio of earnings to revenues, liability size decomposition measures, absolute value of the decomposition measures on the liability size.

Table 2-4 - Comparison of Selected Variables for Statistical Analysis Methodologies

There are many financial ratios in common between these research papers, however it is limiting that this consensus is generally built through subjective ratio choices rather than objective variable selection.

2.2.4 Research Methodology

While many of the papers found in this analysis have used Multiple Discriminant Analysis, it is apparent from Table 2-5 that many different research methodologies were used.

Author	Title	Year	Research Methodology
Beaver, W.	Financial Ratios as Predictors of Failure	1966	Beaver uses a dichotomous classification test as follows, "the data are arrayed (i.e., each ratio is arranged in ascending order). The array of a given ratio is visually inspected to find an optimal cutoff point–a point that will minimize the percent of incorrect predictions. If a firm's ratio is below (or above, as in the case of the total debt to total-assets ratio) the cutoff point, the firm is classified as failed. If the firm's ratio is above (or below, for the total debt to total-assets ratio) the critical value, the firm is classified as nonfailed." "Trial and error" was used to find the optimal cutoff points.
Altman, E.	Financial Ratios. Discriminant Analysis and the Prediction of Corporate Bankruptcy	1968	Profiles the available variables using basic statistical tests, before "observation of the statistical significance of various alternative functions", "evaluation of inter-correlations", "observation of the predictive accuracy" and "judgement of the analyst" is used to develop a variable profile, and develop a discriminant function.

Author	Title	Year	Research Methodology
Beaver, W.	Market Prices, Financial Ratios, and the Prediction of Failure	1968	Beaver compares a cross-sectional analysis on the market prices with a cross-sectional analysis of ratios. A time-series analysis is then used, and the results from both are used to compare the predictive ability of market prices and financial ratios.
Edmister, R.	An Empirical Test of Financial Ratio Analysis for Small Business Failure Prediction	1972	"Multiple discriminant analysis (MDA) is used to form a linear model which classifies individual cases based upon historic financial ratios." "In this research a limitation is placed on variables entering a discriminant function through the normal step-wise procedure in order to limit multicollinearity while systematically selecting variables."
Deakin, E.	A Discriminant Analysis of Predictors of Business Failure	1972	Deakin expands on Beaver's (1966) research, but modifies the definition of failure somewhat. The research finds "linear combinations of the 14 ratios used by beaver" using MDA, and generates a "decision rule" that is tested on the selected firms.
Wilcox, J.	A Prediction of Business Failure Using Accounting Data	1973	Expanding on earlier work, Wilcox outlines a theoretical framework for calculating the probability of a business failing. The results are used to generate scatter plots, on which regression analysis or Discriminant Analysis could be performed. The results are used to refine the theoretical framework and allow Wilcox to explore the theory of business failure.
Blum, M.	Failing Company Discriminant Analysis	1974	The research uses a cash flow framework to identify financial ratios that will be used in the analysis. Blum uses Discriminant Analysis to develop the Failing Company Model, before comparing its accuracy with a non-ratio model, market anticipation, Altman's multivariate analysis, and Beaver's univariate model.
Elam, R.	The Effect of Lease Data on the Predictive Ability of Financial Ratios	1975	The paper outlines the methods used to calculate the "amounts to be paid for leases during each year", before performing some univariate statistical analysis on the data. Elam uses single-ratio, multi-ratio predictive models and Multiple Discriminant Analysis to test the effects of adding lease data.
Altman et al.	A new model to identify bankrupt risk of corporations	1977	"Bankruptcy classification is attempted via the use of a multivariate statistical technique known as discriminant analysis. In this study, the results using both linear and quadratic structures are analyzed." Altman et al. used "forward stepwise discriminant analysis", "backward stepwise discriminant analysis", "scaled vector test", "separation of means test", "conditional deletion test", and the "univariate F-statistic" to "reduce [the] variable set to an acceptable number.
Ohlson, J.	Financial Ratios and the Probabilistic Prediction of Bankruptcy	1980	Ohlson develops a probabilistic model of bankruptcy and then performs some analysis on the initially selected ratios. "Three sets of estimates were computed for the logit model", the models are compared and the results are used to draw conclusions on the performance of each model.

Author	Title	Year	Research Methodology
Chen, K. & Shimerda, T.	An Empirical Analysis of Useful Financial Ratios	1981	A number of "factors" are identified (e.g. liquidity, profitability), and each of the factors are discussed. Principal Component Analysis is used on the ratios within each of the factors, and conclusions regarding which ratios should be used are drawn.
Barniv, R. & Raveh, A.	Identifying financial distress: a new nonparametric approach	1989	The "classical approach to discriminant analysis is based on choosing linear combinations of the variables that will maximize the ratio of the between-groups (1 and 2) to within-group variances." "The method proposed in this paper is based on a different 'separation' rule, namely a different quantity to be maximised. We suggest using a linear combination of the observations, and choosing the coefficients so that the scores Z_1, given to group 1 will be *greater than* (or less than) the scores Z_2, of group 2. The method searches for an optimal linear combination which yields minimum overlapping between the two groups of scores."

Table 2-5 - Comparison of Research Methods for Statistical Analysis Methodologies

While the various research papers tend to seek the answers to varying questions – such as Elam (1975) examining the effect of lease data, compared with earlier papers such as Altman (1968) which tend to develop a proof of concept of corporate failure prediction – there are some aspects of the research that are more directly comparable between papers. For example, while Discriminant Analysis is by far the most popular statistical technique used in this field, there are exceptions such as Beaver (1966), but even these exceptions follow the basic premise of MDA with slightly different statistical processes. For example, Barniv & Ravea (1989) use a technique that is based on MDA, but they modify the separation rule to minimise overlap between distressed and non-distressed firms.

Of particular note is the techniques used for eliminating collinear variables. The stepwise procedure appears to be the most popular technique, though backwards stepwise and the F-statistic were also used, especially in Altman et al. (1977). These techniques, being an objective means for variable elimination, represented an opportunity to strengthen the research as a whole. Unfortunately, these techniques were only used on the limited list of subjectively selected variables, rather than on a broader selection of available factors.

2.2.5 Key Findings

Each of the papers examined concluded their research by outlining the findings that the researchers felt is of most importance. It is these key findings that are used to develop the academic knowledge in the field, and so they are outlined below in Table 2-6.

Author	**Title**	**Year**	**Key Findings**
Beaver, W.	Financial Ratios as Predictors of Failure	1966	• "That the ratio can convey useful information in determining solvency for at least five years before failure." • "It is slightly more risky for a firm to have a very high cash-flow to total-debt ratio than to have a lower one in a range where the bulk of the nonfailed firms appear." • "The profile analysis indicated that the mean current ratio of the failed firms was above the magic '2:1' standard in all five years." "The evidence hints that failed firms may appear to window dress." • "Attempts to window dress may tend to improve the predictive power of ratios rather than impair it, as is often suggested." • "Not all ratios predict equally well. The cash-flow to total-debt ratio has excellent discriminatory power throughout the five-year period. However, the predictive power of the liquid asset ratios is much weaker." • "In the first year before failure the error is only 13 per cent, while in the fifth the error percentage is 22"
Altman, E.	Financial Ratios. Discriminant Analysis and the Prediction of Corporate Bankruptcy	1968	• "Based on the above results it is suggested that the bankruptcy prediction model is an accurate forecaster of failure up to two years prior to bankruptcy and that the accuracy diminishes substantially as the lead time increases." • "All of the observed ratios show a deteriorating trend as bankruptcy approached", "that the most serious change in the majority of these ratios occurred between the third and second years prior to bankruptcy." • "The discriminant-ratio model proved to be extremely accurate in predicting bankruptcy correctly in 94 per cent of the initial sample with 95 per cent of all firms in the bankrupt and non-bankrupt groups assigned to their actual group classification."

Author	Title	Year	Key Findings
Beaver, W.	Market Prices, Financial Ratios, and the Prediction of Failure	1968	• "In both analyses, the conclusion was the same as that of the initial test–the financial ratio had superior discriminatory power." • "The lack of perfect association between the forecasts indicates that investors either respond to nonratio sources of information, they did not use the ratios as they are used here, or both." • "The findings of the cross-sectional and time series analysis are uniformly consistent with respect to the two major contentions: (1) Investors recognize and adjust to the new solvency positions of failing firms. (2) That price changes of the common stocks act as if investors rely upon ratios as a basis for their assessments, and impound the ratio information into the market prices." • "In every instance, the ratio has a lower error [min 13%] than either of the return variables [min 22%]".
Edmister, R.	An Empirical Test of Financial Ratio Analysis for Small Business Failure Prediction	1972	• "The research yields results that generally affirm the advocates' belief in the value of ratio analysis and that lend some support for numerical credit scoring." • "The seven-variable function correctly discriminates in the 39 out of 42 cases (93 percent) when the decision rule is to predict failure if $z < .520$". • "Dividing a ratio by its respective industry average is show to be a desirable technique." • "Classifying ratios by quartile is a particularly valuable tool". • "The predictive power of ratios is cumulative. No single ratio predicts nearly as well as a small group, and some ratios that are not significant predictors by themselves serve to improve discriminant ability when added to a function." • "Reliable functions are most likely formed with a set of independent predictors." • "The small business function fails to discriminate when only one statement is available."
Deakin, E.	A Discriminant Analysis of Predictors of Business Failure	1972	• "Error rates of 22%, 6%, 12%, 23%, and 15% were observed for each of the five years prior to failure." • "The deterioration [of predictive accuracy] of the first year is rather severe and cannot be explained by the presence of any unusual events peculiar to the sample used." • "The application of statistical techniques, particularly discriminant analysis, can be used to predict business failure from accounting data as far as three years in advance with a fairly high accuracy."
Wilcox, J.	A Prediction of Business Failure Using Accounting Data	1973	• "The statistics x, N, and (1-x/1+x)N, derived from an explicit stochastic cash flow model of the firm's financial progress, yielded an improvement in risk-ranking power over the various financial ratios tested by Beaver."

Author	Title	Year	Key Findings
Blum, M.	Failing Company Discriminant Analysis	1974	• "Predictive accuracy of the Failing Company Model is 93-95 percent at the first year before failure, 80 percent at the second year, and 70 percent at the third, fourth and fifth years before failure." • "The cash flow to total debt ratio, found to be the best predictor by Beaver's research, received generally high rankings." • "In comparison with other studies of business failure, the Failing Company Model was demonstrated to be more reliable than a reported multivariate model. However, its accuracy was only approximately that of the leading univariate study to date." • "Inventory declined rapidly for failing companies, which shows that, in general, firms do not seem to fail for reasons of excess accumulation of inventory, at least as shown by annual financial reports." • "Total liabilities of nonfailed firms increased more steadily than those of failed firms, indicating that debt was a usual way for nonfailing firms to finance growth."
Elam, R.	The Effect of Lease Data on the Predictive Ability of Financial Ratios	1975	• "The conclusion indicated by the single-ratio research is that the inclusion of lease data does not appear to improve the predictive power of individual ratios for any of the first five years before bankruptcy." • "Only one attempt of the fourteen indicated that lease data significant improved the model's predictive power. Based on this evidence it appears impossible to conclude that lease data improve the predictive power of multiple discriminant models."
Altman, et al.	A new model to identify bankrupt risk of corporations	1977	• "The ZETA model for assessing bankruptcy risk of corporations developed in this paper demonstrates significantly improved accuracy over an existing failure classification model and, perhaps more importantly, is based on data most relevant to current conditions." • "The model's bankruptcy classification accuracy ranges from over 96 (93% holdout) percent one period prior to bankruptcy to 70% five annual reporting periods prior."
Ohlson, J.	Financial Ratios and the Probabilistic Prediction of Bankruptcy	1980	• "Both sets of variables [financial state and performance] contribute significantly and independently of each other" • "Size appears as an important predictor" • "Misclassified bankrupt firms seem to lack any 'warning signals' of impeding bankruptcy" • "Significant improvement probably requires additional predictors." • The statistic "Percent Correctly Predicted"' equals 96.12 percent; it is tabulated on the basis of a cutoff point of .5."

Author	Title	Year	Key Findings
Chen, K. & Shimerda, T.	An Empirical Analysis of Useful Financial Ratios	1981	• "Thirty-four financial ratios have been found by researchers to be significant variables in the prediction of firm failure in recent studies." • "[T]he financial ratios investigated can be classified by a substantially reduced number of factors". • "[T]he selection of one ratio to represent a factor can account for most of the information provided by all the ratios of that factor."
Barniv, R. & Raveh, A.	Identifying financial distress: a new nonparametric approach	1989	• "It is demonstrated that the [nonparametric] models slightly outperform [8% error rate] the three other models for almost all costs and risks three years prior to the event" • "The [nonparametric] models substantially outperform the [Discriminant Analysis] and logit/probit models in terms of validation results."

Table 2-6 - Comparison of Key Findings for Statistical Analysis Methodologies

Perhaps the most interesting comparison is the startlingly different success rates experienced by each of the predictive models. While each has used different datasets, different ratios, different pre-processing and different statistical methodologies, it highlights the fact that this classification technique, as with any other, is highly dependent on the data provided to make accurate classifications.

It is worthwhile to note that Deakin (1972) has used the smallest sample size. Furthermore, Deakin's research has used organisations across multiple industries while other models have restricted their larger samples to a single industry. Deakin (1972) has mostly replicated the methodology of the Beaver (1966) study, and as a result has not modified the choice of variables.

The research analysed is consistent in many of the conclusions drawn out. For example:

- Financial ratios are generally an effective predictive indicator.
- Financial analysis, and Multiple Discriminant analysis especially, is a generally effective predictive methodology.

- A larger set of financial variables increases the predictive power of the model, except where the variables are highly correlated.
- Ratios that describe cash and liabilities tend to be of greater predictive power.

2.2.6 Conclusion

While the research in this area is generally high quality and rigorous, the subjective selection of financial data, subjective choice of firms to analyse and small sample sizes indicate that there are still a number of opportunities to further develop models using a statistical methodology. For example, Altman et al. (1977) developed the “ZETA” model, which is still widely used in both research and practice even to this day, yet the ZETA model suffers from some limitations such as subjectively choosing financial variables to consider in the model.

2.3 Prediction by Artificial Neural Networks

By the late 1980’s, Artificial Neural Networks were the popular choice for bankruptcy prediction. The Backpropagation Learning Algorithm (LeCun, 1985; Parker, 1985) quickly became the de-facto methodology, and a number of Neural Network journals had been founded. By the early 1990’s, all commonly referenced papers in the field of corporate failure prediction were using Neural Networks of one form or another, and they continue to be used today.

The principal behind Artificial Neural Networks is the modelling of a highly simplified version of the organic brain. Neurons connected by synapses that activate when sufficient input is received from other neurons, are used to model the brain’s decision-making processes. A “learning algorithm” that modifies the weights of the connecting synapses is used to model the brain’s natural learning processes which strengthen and weaken synapses as they are used.

While Neural Networks are one of many types of intelligent techniques, the technique represents one of the most diverse and popular areas within intelligent computing. There are many different network structures available to a researcher, in combination with many different learning algorithms. Further still, Neural Networks can be combined with many other statistical and intelligent techniques to address weaknesses and capitalise on other techniques' strengths.

Neural Networks have a number of distinct advantages over the statistical methodologies used in the papers analysed earlier in this thesis, such as group memberships not needing to be linearly separable, the ease of incorporating a larger number of input variables, and the ease of building a highly accurate model on both in-sample and out-of-sample data. The popularity of Neural Networks in corporate failure prediction is certainly not surprising.

Neural Networks, however, are limited in their "explainability". Even in the very smallest of networks the resulting output function from a Neural Network is highly complex due to the many cross-connections between neurons. This limits the network's usefulness in being able to justify its prediction to a human, no matter how accurate that prediction might be. Furthermore Neural Networks that are too simple can be unable to sufficiently model the data, while networks that are too complex can become so specialised in the prediction of the dataset on which they were trained that they become unable to provide reasonable accuracy on a set of data that they have not previously been exposed to, a condition known as "over-fitting" (Fine, 1999, p. 231), though there are techniques that can be used to address this. The Neural Networks algorithm is explained in more detail in section 4.2.

Even with these limitations, Neural Networks represent a major step forward in the field of corporate failure prediction. The following subsections will look at each of the elements of notable papers that utilise Neural Networks to predict business failure, in order to compare and contrast them.

2.3.1 Definition of Corporate Failure

Predictive Neural Networks are used in a “supervised” manner, that is, the network is trained on data where the correct answers have been provided to give the network an opportunity to learn relationships between inputs and outputs. In order to do so, the researcher needs to establish which firms within the training data should be classified as failed or non-failed. Like statistical analysis, a definition of failure is required. Table 2-7 below outlines the different definitions used within the literature.

Author	Title	Year	Definition of Corporate Failure
Odom, M. & Sharda, R.	A neural network model for bankruptcy prediction	1990	When “the firms declared bankruptcy”
Raghupathi, W., Schleade, L. & Raju, B.	A neural network approach to bankruptcy prediction	1991	Adopted from the Wall Street Journal Index, and companies that are deleted from Moody’s Industrial Manual.
Coats, P. & Fant, L.	Recognising financial distress patterns using a neural network tool	1993	Opted to identify “financially troubled firms”, using “auditors’ reports rather than the traditional bankruptcy filings”.
Altman, E., Marco, G. & Varetto, F.	Corporate distress diagnostic: Comparisons using linear discriminant analysis and neural networks (the Italian experience)	1994	“(1) some form of bankruptcy proceeding, (2) were wound up in temporary receivership or (3) had stated that they were in dire straits with regard to their payments to the banks.”
Wilson, R. & Sharda, R.	Bankruptcy prediction using neural networks	1994	Wilson & Sharda used “bankruptcy”, but did not specifically define it. They did use Moody’s Industrial Manual as a data source, which may indicate the identification of deleted companies within it, as with other similar studies.
Fanning, K & Cogger, K.	A comparative analysis of artificial neural networks using financial distress prediction	1994	“[F]iling for Chapter X or XI bankruptcy”.
Boritz, J. & Kennedy, D.	Effectiveness of neural network types for prediction of business failure	1995	“The sample consisted of 171 companies which filed for bankruptcy between 1971 and 1984 inclusive.”
Lee, K., Han, I. & Kwon, Y.	Hybrid neural network models for bankruptcy predictions	1996	“We define the state of bankruptcy as follows: 1. The firms which applied for, have started, or are under the process of corporate clearance. 2. The firms which quit or closed business. 3. The firms which have had losses for the consecutive three years and are currently under legal control. 4. The firms which reported the withdrawal of listing or terminated to be listed by the Korea Stock Exchange.”
Yang, Z., Platt, M. & Platt, H.	Probabilistic Neural Networks in Bankruptcy Prediction	1999	Does not explicitly define bankruptcy, but it is implied that the definition is the same as in “Platt, Platt, and Pedersen’s (1994)” study, as the data from this study is used.

Author	Title	Year	Definition of Corporate Failure
Lee, K., Booth, D. & Alam, P.	A comparison of supervised and unsupervised neural networks in predicting bankruptcy of Korean firms	2005	"Korean firms that filed for bankruptcy during 1995–1998"

Table 2-7 - Comparison of Definition of Corporate Failure for Neural Network Methodologies

Much like in section 2.2, the definitions between papers vary somewhat. Again the majority of papers have adopted the legal definition of bankruptcy, while some have chosen to classify distress rather than collapse. One paper of particular interest is that of Coats & Fant (1993) which relied on auditor's reports, therefore their intelligent system was not predicting financial distress so much as predicting what auditors classify as financial distress – correct or otherwise.

This review now focuses its attention on the sample selection used within the Neural Network methodology.

2.3.2 Sample Selection

The following table outlines the different sampling selection techniques used by papers that focused on Neural Networks as a model for predicting corporate failure.

Author	Title	Year	Sample Selection Method
Odom, M. & Sharda, R.	A neural network model for bankruptcy prediction	1990	"The sample, obtained from Moody's Industrial Manuals, consisted of a total of 129 firms, 65 of which went bankrupt during the period and 64 nonbankrupt firms matched on industry and year."
Raghupathi, W., Schleade, L. & Raju, B.	A neural network approach to bankruptcy prediction	1991	51 bankrupt firms "chosen from listings in the Wall Street Journal Index for the years 1980-1988 and from a list of deleted companies in the Moody's Industrial Manual", paired with 51 non-bankrupt firms from "the same industry and approximately the same asset size", selected from "the same sources".
Coats, P. & Fant, L.	Recognising financial distress patterns using a neural network tool	1993	94 "distressed" firms "collected from the Standard and Poor's COMPUSTAT financial database covering the period 1970-1989", and listed in the "Industrial Research File". 188 "viable" firms selected "from the Full Coverage File". "These firms were matched 2-to-1".

Altman, E., Marco, G. & Varetto, F.	Corporate distress diagnostic: Comparisons using linear discriminant analysis and neural networks (the Italian experience)	1994	"404 each of healthy, unsound and vulnerable firms. A second independent sample of 453 companies was used, 151 of each type, with data limited to the last year prior to bankruptcy. A final sample, independent of the other two, was analysed comprising 900 healthy and 900 vulnerable companies for three years of historical series."
Wilson, R. & Sharda, R.	Bankruptcy prediction using neural networks	1994	"The sample of firms for which these ratios were obtained consisted of firms that either were in operation or went bankrupt between 1975 and 1982. The sample, obtained from *Moody's Industrial Manuals*, consisted of a total of 129 firms, 65 of which went bankrupt during the period and 65 non bankrupt firms matched on industry and year." "We created three proportions (or base rates) for each of the training and testing set compositions. The first factor level (or base rate) was a 50/50 proportion of bankrupt to non bankrupt cases, the second level was a 80/20 proportion (80% non-bankrupt, 20% bankrupt), and the third factor level, an approximate 90/10 proportion".
Fanning, K & Cogger, K.	A comparative analysis of artificial neural networks using financial distress prediction	1994	"[T]he author tested this theory on a sample of matched pairs of 52 failed firms with 52 non-failed firms from one to five years prior to failure. Firms were matched in terms of size and industry characteristics."
Boritz, J. & Kennedy, D.	Effectiveness of neural network types for prediction of business failure	1995	"The sample of bankrupt companies used for the present study was obtained from Boritz et al. 1995 which was based on the dataset developed by Kennedy and Shaw (1991). The sample consisted of 171 companies which filed for bankruptcy between 1971 and 1984 inclusive." "The non-bankrupt companies used in the present study also were obtained from Boritz et al., 1995. The sample was collected from the *Compustat II Database* and consisted of 6,153 non-bankrupt companies selected from the same time period as the bankrupt companies."
Lee, K., Han, I. & Kwon, Y.	Hybrid neural network models for bankruptcy predictions	1996	"Bankruptcy cases reported in Korea from 1979 through 1992. We collected a sample of 83 bankrupt firms", "listed in the Korea Stock Exchange". "A failed firm was matched with a nonfailed firm in terms of (1) asset size, (2) capital size, (3) number of employees, and (4) age."
Yang, Z., Platt, M. & Platt, H.	Probabilistic Neural Networks in Bankruptcy Prediction	1999	"Platt, Platt, and Pederson (1994) built an early warning bankruptcy model for the U.S. oil and gas industry. Their data for 122 companies for the period 1984 to 1989 are used in this study."

Lee, K., Booth, D. & Alam, P.	A comparison of supervised and unsupervised neural networks in predicting bankruptcy of Korean firms	2005	"Each failed firm is matched with a non-failed firm in terms of (1) asset size and (2) a two-digit Standard Industrial Classification (SIC) code as control measures. The asset size of a non-failed firm is matched with that of a failed firm using the 3-year period prior to bankruptcy filings. As a result, we have a matched sample of 168 firms, 84 failed firms and 84 non-failed firms."

Table 2-8 - Comparison of Sample Selection Methods for Neural Network Methodologies

It can be seen that the sample selection techniques have remained fairly constant since the first few corporate failure prediction papers that utilised statistical analysis. The most interesting difference is the number of papers that used different sample sizes for failed and non-failed firms, presumably due to Neural Networks ability to deal with unpaired data, unlike many statistical techniques.

The size of the samples is also generally larger in comparison to previous research, possibly due to the increasing availability of data on failed firms between 1990 and 2000, though still surprisingly small especially for more recent research. For example, Lee et al. (2005) used just 168 firms.

The following section will examine the dimension reduction techniques used.

2.3.3 Variable Selection & Dimension Reduction

Author	Title	Year	Variable Selection Method
Odom, M. & Sharda, R.	A neural network model for bankruptcy prediction	1990	"We have chosen to use the same financial ratios that Altman used in his 1968 study."
Raghupathi, W., Schleade, L. & Raju, B.	A neural network approach to bankruptcy prediction	1991	"Thirteen financial ratios used", "selected from ratios proven popular (and useful) in earlier research [Harris (1989)] on bankruptcy prediction".
Coats, P. & Fant, L.	Recognising financial distress patterns using a neural network tool	1993	"The financial information we chose to describe each firm is the set of five ratios from Altman's Z score model".

Author	Title	Year	Variable Selection Method
Altman, E., Marco, G. & Varetto, F.	Corporate distress diagnostic: Comparisons using linear discriminant analysis and neural networks (the Italian experience)	1994	The best network in the first series of tests contained "ten financial ratios: four relative to the firms' financial structure and indebtedness, two to liquidity, and four representative of company profitability and internal-financing". The second series of networks used "fifteen business ratios; these are a broader set than the one in the ten-ratio network described in the previous section." The third series of networks used the "ones used in discriminant functions". The final series of networks used "ratios that are representative of [each respective] characteristic".
Wilson, R. & Sharda, R.	Bankruptcy prediction using neural networks	1994	"We used the same financial ratios as Altman [1]."
Fanning, K & Cogger, K.	A comparative analysis of artificial neural networks using financial distress prediction	1994	"To fairly compare all methods, only the three variables, X, N, T, can be used" (note these are defined in the following subsection).
Boritz, J. & Kennedy, D.	Effectiveness of neural network types for prediction of business failure	1995	Variables used by Altman **(1968)** and Ohlson **(1980)** were used both separately and combined to compare the results from these studies.
Lee, K., Han, I. & Kwon, Y.	Hybrid neural network models for bankruptcy predictions	1996	"A MDA-assisted neural network is a neural network model operating with input variables selected by MDA method. Similarly, an ID3-assisted neural network indicates a neural network model operating with input variables selected by the ID3 method. A SOFM-assisted neural network is a neural network model combining a supervised neural network model with an unsupervised neural network model. We use a SOFM model as preprocessing mechanism."
Yang, Z., Platt, M. & Platt, H.	Probabilistic Neural Networks in Bankruptcy Prediction	1999	Used the ratios used in "Platt, Platt, and Pederson (1994)".
Lee, K., Booth, D. & Alam, P.	A comparison of supervised and unsupervised neural networks in predicting bankruptcy of Korean firms	2005	"Each firm is described by Altman's five variables since the prediction capabilities of these ratios are well documented in the previous literature"

Table 2-9 - Comparison of Variable Selection Methods for Neural Network Methodologies

Even though the field of corporate failure is maturing by this stage, the variable selection techniques have still remained highly subjective, with Altman (1968) becoming the de-facto choice across multiple datasets. In many cases, the same variables have been used as in prior (statistical) research, as a means of comparing the two techniques.

The most notable paper in this respect is Lee et al. (1996), which used Multiple Discriminant Analysis, ID3 (a decision-tree algorithm), and Self Organising Feature Map assisted techniques to select variables. Moreover, Lee et al. (1996) found that more objective variable selection techniques may improve the performance of other researcher's models.

The following table notes the variables used within each piece of analysed research.

Author	Title	Year	Variables
Odom, M. & Sharda, R.	A neural network model for bankruptcy prediction	1990	Working Capital/Total Assets, Retained Earnings/Total Assets, Earnings before Interest and Taxes/Total Assets, Market Value of Equity/Total Debt, Sales/Total Assets
Raghupathi, W., Schleade, L. & Raju, B.	A neural network approach to bankruptcy prediction	1991	Current assets to current liabilities, cash plus short term investments plus net receivables to current liabilities, income from operations plus depreciation plus depletion plus amortization to current liabilities plus long-term debt, current liabilities plus long-term debt to total assets, current assets less current liabilities to total assets, income from operations to total assets, income from operations plus taxes plus interest expense to total assets, net sales to total assets, retained earnings to total assets, current assets to net sales, current assets less current liabilities to total sales, current assets to total assets, cash plus short-term investments to total assets.
Coats, P. & Fant, L.	Recognising financial distress patterns using a neural network tool	1993	Working capital to total assets, retained earnings to total assets, earnings before interest and taxes to total assets, market value of equity to book value of total debt, sales to total assets.
Altman, E., Marco, G. & Varetto, F.	Corporate distress diagnostic: Comparisons using linear discriminant analysis and neural networks (the Italian experience)	1994	Variables used not explicitly specified.
Wilson, R. & Sharda, R.	Bankruptcy prediction using neural networks	1994	Working capital to total assets, retained earnings to total assets, earnings before interest and taxes to total assets, market value of equity to total debt, sales to total assets.
Fanning, K & Cogger, K.	A comparative analysis of artificial neural networks using financial distress prediction	1994	The mean adjusted cash flow divided by its standard deviation, the firm's adjusted cash position divided by its standard deviation.

Boritz, J. & Kennedy, D.	Effectiveness of neural network types for prediction of business failure	1995	Working capital to total assets, retained earnings to total assets, earnings before interest and taxes to total assets, market value of equity to total debt, sales to total assets, total assets to GNP price-level index, total liabilities to total assets, working capital to total assets, current liabilities to current assets, 1 if total liabilities > total assets, net income to total assets, funds provided by operations to total liabilities, 1 if net income negative for last two years, change in net income, 1 if listed on exchange.
Lee, K., Han, I. & Kwon, Y.	Hybrid neural network models for bankruptcy predictions	1996	"For Group 1, 10 financial variables were selected as important input variables for predicting bankruptcy. Similarly, 18 and 17 financial variables were chosen for Group II and II, respectively." "The number of financial variables chosen by ID3 is 7 for Group I", "7 and 9 financial variables were selected by using ID3 for Group II and III, respectively."
Yang, Z., Platt, M. & Platt, H.	Probabilistic Neural Networks in Bankruptcy Prediction	1999	Net cash flow to total assets, total debt to total assets, exploration expenses to total reserves, current liabilities to total debt, and the trend in total reserves.
Lee, K., Booth, D. & Alam, P.	A comparison of supervised and unsupervised neural networks in predicting bankruptcy of Korean firms	2005	Working capital to total assets, retained earnings to total assets, earnings before interest and taxes to total assets, market value of equity to book value of total debt, sales to total assets.

Table 2-10 - Comparison of Selected Variables for Neural Network Methodologies

This thesis will now briefly discuss the research methodologies used by papers using the Neural Network approach.

2.3.4 Research Methodology

Author	Title	Year	Research Methodology
Odom, M. & Sharda, R.	A neural network model for bankruptcy prediction	1990	Compared discriminate analysis (SAS DISCRIM) with a 5-node hidden layer using the back propagation learning algorithm.
Raghupathi, W., Schleade, L. & Raju, B.	A neural network approach to bankruptcy prediction	1991	Normalized the input variables, applied the "backpropagation algorithm" to half of the sample selection, using trial and error to determine the optimum combination of hidden layers and hidden nodes."

Author	Title	Year	Research Methodology
Coats, P. & Fant, L.	Recognising financial distress patterns using a neural network tool	1993	Applied “a learning paradigm called ‘Cascade-Correlation’ (or ‘Cascor’)”. Used “four models [to] represent four different lead times, i.e., the year for which distressful conditions in a firm are reported by auditors, and the one, two, and three years prior.”
Altman, E., Marco, G. & Varetto, F.	Corporate distress diagnostic: Comparisons using linear discriminant analysis and neural networks (the Italian experience)	1994	“The method considered here is the well-known Error Back Propagation Algorithm by Rumelhart et. al (1986).” Altman et al. also attempted to develop a model “sensitive to the passing of time, and the changes of the companies’ business patterns.” Different styles of models were built, starting with networks designed to separate healthy from unsound companies, progressing to multi-layer networks, progressing to a comparison of Neural Networks to discriminant function using the same ratios, and finally to breaking the model up into “simpler networks connected to each other.”
Wilson, R. & Sharda, R.	Bankruptcy prediction using neural networks	1994	Used the “backpropagation training algorithm”, with “5 input neurons (one for each financial ratio), 10 hidden neurons and 2 output neurons (one indicating bankrupt firm, the other indicating non-bankrupt firm) was used. Such a network structure was chosen on the basis of previously espoused heuristic guidelines”.
Fanning, K & Cogger, K.	A comparative analysis of artificial neural networks using financial distress prediction	1994	Trained a Neural Network on the first 40% of cases using a NN with six nodes in the first hidden layer and seven in the second hidden layer using sigmoidal activation functions and a backpropagation learning algorithm, and a Generalised Adaptive Neural Network using quadratic functions.
Boritz, J. & Kennedy, D.	Effectiveness of neural network types for prediction of business failure	1995	Uses “Back-Propagation and Optimal Estimation Theory”. “Within the back-propagation training method, four different models (Back-Propagation, Functional Link Back-Propagation With Sines, Pruned Back-Propagation, and Cumulative Predictive Back-Propagation) are tested.”
Lee, K., Han, I. & Kwon, Y.	Hybrid neural network models for bankruptcy predictions	1996	Divided the sample data into three groups separated by date. Used different variable selection techniques before constructing a “3-layer network”, with “the number of hidden units are set to the same as the number of input units”. Results were obtained for MDA-only techniques, ID3-only techniques, MDA-assisted NN, ID3-assisted NN, SOFM(MDA)-assisted NN and SOFM(ID3)-assisted NN.

Author	Title	Year	Research Methodology
Yang, Z., Platt, M. & Platt, H.	Probabilistic Neural Networks in Bankruptcy Prediction	1999	Uses a probabilistic NN which "employ Bayesian decision-making theory based on an estimate of the probability density in data space." Used 33 non-bankrupt and 11 bankrupt companies in the training set, 26 non-bankrupt companies and 14 bankrupt companies in the validation set, and 30 non-bankrupt and 8 bankrupt companies in the test set.
Lee, K., Booth, D. & Alam, P.	A comparison of supervised and unsupervised neural networks in predicting bankruptcy of Korean firms	2005	A three-layered (one hidden layer) back-propagation Neural Network was used with a logistic transfer function, testing one to ten hidden nodes with a Levenberg-Marquardt algorithm. This is compared with a two-dimensional Kohonen Self-Organising Feature Map (SOM) using 200 output nodes with 4 clusters as well as compared with Quadratic Discriminant Analysis and Logistic Regression.

Table 2-11 - Comparison of Research Methods for Neural Network Methodologies

In almost all of the instances of Neural Networks being used, the standard feed-forward model was selected using a back-propagation learning algorithm (this methodology is discussed in section 4.2. Coats & Fant (1993) break from this mould by using the Cascade-Correlation learning paradigm which dynamically adds hidden neurons as algorithms. Boritz & Kennedy (1995) also deviate by employing a number of different learning techniques.

Key findings from the various analysed papers will now be outlined.

2.3.5 Key Findings

Author	Title	Year	Research Methodology
Odom, M. & Sharda, R.	A neural network model for bankruptcy prediction	1990	• "The neural network appears to be more robust, performing better than the discriminant analysis method in each of the three situations. The neural network also appears to be more consistent than the discriminant analysis method." • "When the training sample was reduced to the 90/10 proportion, the discriminant analysis had a correct prediction rate of 59.26% and the neural network had a correct prediction rate of 77.78% for the holdout subsample."

Author	Title	Year	Research Methodology
Raghupathi, W., Schleade, L. & Raju, B.	A neural network approach to bankruptcy prediction	1991	• "The configuration with 15 nodes in the first hidden layer and 2 in the second seems to have the best percentage of correct classifications (86%). • "Neural networks might provide suitable models for the bankruptcy prediction process."
Coats, P. & Fant, L.	Recognising financial distress patterns using a neural network tool	1993	• "Some networks required up to 1,400 training cycles and installed as many as eight hidden nodes" to achieve 100% accuracy on the training data. • "Test results suggest that the NN approach is more effective than MDA for pattern classification."
Altman, E., Marco, G. & Varetto, F.	Corporate distress diagnostic: Comparisons using linear discriminant analysis and neural networks (the Italian experience)	1994	• "The best results were obtained with a three-layer network: one initial hidden layer of ten neurons, a second layer with four neurons and an output layer consisting of a single neuron" in the first series of tests. • "The most satisfactory results were obtained with a three-layer network, comprising fifteen neurons in the first hidden layer, six neurons in the second hidden layer and one neuron in the output layer" in the second series of tests. • "At the end of training, the [best second series] network was able to recognize correctly 97.7% of healthy and 97% of unsound companies." • "The greatest problem concerns the existence of non-acceptable types of behaviour in the network, combining a large number of variables several times over in a complex fashion. These behaviour patterns are characteristic of networks of any complexity that have at least two inputs."
Wilson, R. & Sharda, R.	Bankruptcy prediction using neural networks	1994	• Achieved 95%+ accuracies with models trained from a 50/50 training set. • Accuracy dropped to low 70%s as the training set composition moved to 90/10. • "In every instance, neural networks outperformed discriminant analysis in classification accuracy, especially in the prediction of bankrupt firms".
Fanning, K & Cogger, K.	A comparative analysis of artificial neural networks using financial distress prediction	1994	• Artificial Neural Networks and Generalised Adaptive Neural Networks rivalled predictive accuracy of model-based approaches. • No single approach was uniformly superior • Generalised Adaptive Neural Network seems preferable to Artificial Neural Networks due to not imposing a priori network architecture

Author	Title	Year	Research Methodology
Boritz, J. & Kennedy, D.	Effectiveness of neural network types for prediction of business failure	1995	• "The Optimal Estimation Theory neural networks had the lowest rate of Type I error but the highest rate of Type II error." • "The Back-Propagation neural networks had high Type I error but lower rates of Type II error." • "While the neural network models' performance is in line with that of the more conventional techniques such as discriminant analysis and logit/probit, it is noteworthy that their performance is not a dramatic improvement over those conventional techniques." • "Comparing results across the three sets of variables shows that the relative performance of neural networks and traditional statistical techniques is affected by the choice of variables in the learning sample. We demonstrate that the performance of the neural networks tested is sensitive to the choice of variables selected and that the networks cannot be relied upon to "sift through" variables and focus on the most important variables."
Lee, K., Han, I. & Kwon, Y.	Hybrid neural network models for bankruptcy predictions	1996	• "On the average, the SOFM(MDA)-assisted neural network model performs the best. This model shows an outstanding prediction accuracy (84%) for Group I." • "Hybrid neural network models perform better than MDA and ID3."
Yang, Z., Platt, M. & Platt, H.	Probabilistic Neural Networks in Bankruptcy Prediction	1999	• Best overall results obtained using Fisher MDA and deflated data, achieving 87% accuracy. • "Deflation improves the discriminant ability of bankruptcy prediction models."
Lee, K., Booth, D. & Alam, P.	A comparison of supervised and unsupervised neural networks in predicting bankruptcy of Korean firms	2005	• "The [back-propagation] network consistently outperforms logistic regression as well as other classification techniques" and "the prediction accuracy of the Kohonen self-organizing feature map, as expected, is lower than the other supervised classification techniques".

Table 2-12 - Comparison of Key Findings for Neural Network Methodologies

Presenting the key findings of each of these research papers in the above format demonstrates some interesting similarities and differences.

Almost all of the research analysed concluded that Neural Networks present themselves as an equal or superior methodology in their predictive ability. Most of the research found that a 3-layered network was most effective, and this is a particularly interesting finding since a trained two-layer network has an output function that uses the same structure to that of Discriminant Analysis. Most of the results found huge disparity in the optimum number of hidden neurons, which is expected considering the differences of datasets and their relative complexities, though this can present a difficult problem for networks that need to adapt to an environment where the complexity of the data is changing.

Having analysed the papers focusing on the use of Neural Networks, some broader conclusions can be drawn.

2.3.6 Conclusion

Given the broad nature of Neural Networks and the different combinations of model structure and learning algorithm, it is surprising just how few papers analysed here deviate from the feed-forward backpropagation method, especially given the findings of Coats & Fant (1993) and Boritz & Kennedy (1995) which demonstrate the predictive ability of the alternatives. While more specialised papers, such as Chauhan, et al.'s (2009) paper "Differential evolution trained wavelet neural networks: Application to bankruptcy prediction in banks", such methodology's remain rare.

Furthermore, it is surprising that of all the papers considered here, only one author (Lee, et al., 1996) and (Lee, et al., 2005) utilised an unsupervised Neural Network to assist the model.

Given the number of Neural Network techniques, datasets and factor selection methodologies that remain untested in this field, and the lack of consensus on the best approach to use and expected outcomes, it appears there are still avenues of research yet to be undertaken using Neural Networks.

2.4 Recent Developments in Predicting Corporate Failure

While Discriminant Analysis and Neural Networks represent the two most popular methodologies for predicting corporate failure, they are not the only methodologies available.

Rough Set Theory (RST), Genetic Algorithms (GA), Genetic Programming (GP) and Support Vector Machines (SVM) are some of the alternatives that have recently received much attention. This section will examine relevant recent papers to finalise the analytical aspect of this literature review, and identify weaknesses in the current academic state of corporate failure prediction. In doing so, the direction that the rest of this thesis will take can be defined.

Around 1998 there was a fairly sudden reduction in the number of published papers using Neural Networks, and a corresponding rise in the number of published papers using alternative methods. Therefore papers published after 1998 that use statistical analysis or Neural Networks but focus on a non-methodological aspect of the research – such as Kane et al. (1998), which focuses on the pre-processing of data – have also been included in this section.

Firstly, some background into some of the more popular alternative techniques will be provided. This is not an exhaustive list of available methodologies or an in-depth analysis of each method, simply sufficient background that the following literature review has enough context to be useful.

One of the more popular alternatives to Neural Networks is Genetic Algorithms (GA), discovered by Holland (1975) and based on the idea of natural selection. The assumption behind Genetic Algorithms is that there is a sequence of data that represents the "best" solution to a given problem. The best sequence of data could be the optimum stock levels to minimise inventory costs while maximising product availability, or the most appropriate parameters for a Multiple Discriminant Analysis function, or the weights for the synapses joining the neurons in a Neural Network. Each datum is considered to be a "gene" and a process of reproduction, natural selection and random mutation is used to move towards the result with the least error. In this respect, Genetic Algorithms is a learning algorithm, with the structure of the solution left to the researcher to decide. One of the strengths of Genetic Algorithms is that in comparison with the learning algorithms typically used with Neural Networks, their random mutations tend to perform a more complete search of the error surface theoretically resulting in a better solution over a longer period of time. On the other hand Genetic Algorithms do not attempt to identify individual genes for their role in the resulting error and therefore do not follow the path of steepest decent down the error surface. Therefore it could be argued that the process is less “intelligent” than that of Neural Networks, yet more “exhaustive”.

Further utilising the concept of natural selection is that of Genetic Programming. Like Genetic Algorithms, Genetic Programming is based on reproduction, natural selection and mutation – but unlike Genetic Algorithms, Genetic Programming uses a mathematical operation as the gene, as opposed to discrete or continuous values. In doing so, Genetic Programming is designed to discover a mathematical function that generates the minimum error. This overcomes a weakness in both Neural Networks and Genetic Algorithms which both require a predefined equation structure, but potentially at the cost of a large increase in the search space because each gene can now take on non-numerical forms such as mathematical operators or functions.

Another popular tool used in recent corporate failure prediction research is that of Rough Sets. Rough Sets Theory (RST) is credited to Pawlak (1982) and is a tool that is used to describe incomplete, uncertain or inconsistent data by way of building decision rules that can describe relationships between attributes and outcomes in large volumes of information. The premise of rough set theory is that for a given array of data with input variables and decision variables, classes can be approximated using upper bound and lower bound sets. In doing so, rule induction can be performed on the incomplete and inconsistent data. While Rough Sets Theory does not explicitly define a learning algorithm that should be used, ROSE2 is a popular software application that applies rough set theory to classify data, which uses “a modified version of the LEM2 algorithm” (Prosoft, 1999). The LEM2 learning algorithm serves a similar purpose as the backpropagation learning algorithm does to Neural Networks, which is to efficiently train the model through an iterative error-based approach.

Finally, Support Vector Machines (SVM) are most easily understood when explained as two groups of two-dimensional data-points, in which an SVM linear algorithm seeks to find the line with the greatest margin that separates two groups. In dimensions greater than two, the line becomes a hyperplane but in fact still linearly separates the groups. Mathematically, the optimal hyperplane can be calculated by way of optimising a quadratic function. To deal with non-linearly separated data, a "kernel function" is used to map the data-points into a higher dimensional space in which they become linearly separable.

Having provided some background to some of the more popular recent methodologies, the following subsections will examine the various recent papers used in the field of corporate failure prediction. Due to the more recent papers not yet having had sufficient time to become highly cited (one of the conditions for being included in this literature review), this section contains an additional sub-section that briefly discusses notable research to provide the necessary context to the work contained within this thesis.

2.4.1 Definition of Corporate Failure

Like the papers published that utilise statistical analysis or Neural Networks, a definition of failure is required to establish which firms have failed and which firms have not failed. Table 2-13 below outlines the different definitions used within the literature.

Author	Title	Year	Definition of Corporate Failure

Kane, G., Richardson, F. & Meade, N.	Rank transformations and the prediction of corporate failure	1998	"The failure event is defined as the date of occurrence of the Chapter 7 or Chapter 11 bankruptcy petition filing, or the date of initiation of an involuntary liquidation proceeding, as provided by the *Wall Street Journal Index*. Because the initiation dates of involuntary liquidation are not usually available, the COMPUSTAT delisting date was typically used as an approximation."
Dimitras, A., Slowinski, R., Susmaga, R. & Zopounidis, C.	Business failure prediction using rough sets	1999	While not explicitly defined, "healthy" firms were defined as "firms that did not [file] for bankruptcy".
Varetto, F.	Genetic algorithms applications in the analysis of insolvency risk	1999	"The definition of unsound companies includes not only those actually declared bankrupt but also those that have been considered insolvent by the member banks of the Centrale dei Bilanci (bank credits defined as "overdue").
McKee, T.	Developing a bankruptcy prediction model via rough sets theory	2000	"Non-bankruptcy companies were defined as those companies that had positive cash flow from operations for the most recent five-year period. Bankrupt companies were defined as companies that have either filed for bankruptcy or had a significant subsidiary file for bankruptcy."
Anandarajan, M., Lee, P., & Anandarajan, A.	Bankruptcy prediction of financially stressed firms: an examination of the predictive accuracy of artificial neural networks	2001	While bankruptcy was not explicitly defined, the phrase "filed for bankruptcy" is used, implying the definition used.
McKee, T. & Lensberg, T.	Genetic programming and rough sets: A hybrid approach to bankruptcy classification	2002	Bankruptcy not explicitly defined, but assumed to be the same as the research that it being built upon, McKee **(2000)**, that is "companies that have either filed for bankruptcy or had a significant subsidiary file for bankruptcy."
Wang, Z.	Financial Ratio Selection for Default-Rating Modelling: A Model-Free Approach and Its Empirical Performance	2004	"[The research considers] the firms whose long-term domestic issuers are rated as default (D) or selected default (SD) by Standard & Poor's as firms under distress. In S&P's definition, a firm is rated default when interest payments or principal payments are not made on the due date even if the applicable grace period has not expired, unless S&P believes that such payments will be made during such grace periods. S&P also assigns default rating to a firm upon the filling of bankruptcy petition if debt service payments are jeopardized."
Wu, C.	Using non-financial information to predict bankruptcy: a study of public companies in Taiwan	2004	Failed – "judicially declared a special arrangement company by authorities when the company has operational difficulties" Non-Failed – "no special stock arrangement, which are listed on the TSE market. Their stocks are allowed to trade publicly."
Shin, K., Lee, T. & Kim, H.	An application of support vector machines in bankruptcy prediction model	2005	Korean manufacturing firms which "filed for bankruptcy" "from 1996 to 1999".

Lensberg, T., Eilifsen, A., & McKee, T.	Bankruptcy theory development and classification via genetic programming	2006	While the definition of failure is not explicitly defined, it appears to adopt the definition used in previous research **(McKee & Lensberg, 2002)**.
Etemadi, H., Rostamy, A. & Dehkordi, H.	A genetic programming model for bankruptcy prediction: Empirical evidence from Iran	2009	“Paragraph 141 of Iran Trade Law”
Chen, H., Yang, B., Wang, G., Liu, J., Xu, X., Wang, S., Liu, D.	A novel bankruptcy prediction model based on an adaptive fuzzy k-nearest neighbor method	2011	The Wieslaw dataset defines failure as bankruptcy, while the Australian credit dataset is not related to corporate failure and is not relevant to this research.

Table 2-13 - Comparison of Definition of Corporate Failure for Recent Research

2.4.2 Variable Reduction & Sample Selection

Author	Title	Year	Sample Selection Method
Kane, G., Richardson, F. & Meade, N.	Rank transformations and the prediction of corporate failure	1998	“The sample consists of firms from the New York Stock Exchange (NYSE), the American Stock Exchange (AMEX), and the National Association of Securities Dealers Automated Quotation System (NASDEQ).” “’Non-failed’ firms are those contained in a randomly selected sample of 2,000 companies.”
Dimitras, A., Slowinski, R., Susmaga, R. & Zopounidis, C.	Business failure prediction using rough sets	1999	“A large number of firms which failed in Greece in the years 1986-1990 were collected. From this large set, 40 firms from 13 industries meeting the criteria of (a) having been in business for more than five years and (b) data availability were selected.” “The 40 failed firms were matched one by one to 40 “healthy” firms”. “The healthy firms were chosen among those of the same industry and having also similar total assets and number of employees for the year -1 to the corresponding failed firm.”
Varetto, F.	Genetic algorithms applications in the analysis of insolvency risk	1999	“The GAs for the identification of linear functions were applied to the sample of 3840 firms, with tests on an independent sample of 898 companies."
McKee, T.	Developing a bankruptcy prediction model via rough sets theory	2000	100 bankrupt and 100 non-bankrupt firms were “randomly selected for the fiscal years 1986 to 1988 from *Compact Disclosure* (Disclosure Inc, 1990).”
Anandarajan, M., Lee, P., & Anandarajan, A.	Bankruptcy prediction of financially stressed firms: an examination of the predictive accuracy of artificial neural networks	2001	“[The] final sample resulted in 265 distress firmed with dividend omission or reduction, 319 distress firms with technical defaults or default on loan payments, 91 distress firms restructuring their debt and 104 firms that filed for bankruptcy.”

Author	Title	Year	Sample Selection Method
McKee, T. & Lensberg, T.	Genetic programming and rough sets: A hybrid approach to bankruptcy classification	2002	“Used data from Disclosure Incorporated (1997) to identify 146 bankruptcy companies”, “matched to 146 non-bankruptcy companies first by industry and then by assets size”. “One non-bankrupt company was subsequently dropped due to missing data, resulting in a total sample size of 291 companies.”
Wang, Z.	Financial Ratio Selection for Default-Rating Modelling: A Model-Free Approach and Its Empirical Performance	2004	“The data used in [the] analysis is from Compustat of Standard & Poor’s (S&P).” Firms without complete data entries for financial ratio calculation were deleted, as were firms with negative equities (“because they usually generate negative ratios”). 1992 firms were present in the sample, of which 46 were rated as default or selected default.
Wu, C.	Using non-financial information to predict bankruptcy: a study of public companies in Taiwan	2004	“There are 31 failed companies and 31 non-failed companies that qualify according to the above definition by TSE, during 1995 to 2000 in the study.” “The failed and non-failed companies are matched up, whereby the samples are matched by some characters, such as they belong to the same industry, their sizes are similar, and/or they sell a similar product.”
Shin, K., Lee, T. & Kim, H.	An application of support vector machines in bankruptcy prediction model	2005	“externally non-audited 2320 medium-size manufacturing firms”
Lensberg, T., Eilifsen, A., & McKee, T.	Bankruptcy theory development and classification via genetic programming	2006	Used the “Norwegian Register of Bankruptcies” to identify 1953 entities, before finding the corresponding entry in a financial accounting database developed by Dun and Bradstreet. Eliminating the private companies and the companies where the required 6 variables were not available left 568 bankrupt companies which were matched with non-bankrupt companies by 5-digit industry code.
Etemadi, H., Rostamy, A. & Dehkordi, H.	A genetic programming model for bankruptcy prediction: Empirical evidence from Iran	2009	“The dataset used for this research consists of 144 Iranian companies. All of them were or still are listed on the Tehran Stock Exchange (TSE). 72 companies went bankrupt under paragraph 141 of Iran Trade Law from 1998 through 2005. The other 72 companies are ‘matched’ companies, from the same period of listing on the TSE.”
Chen, H., Yang, B., Wang, G., Liu, J., Xu, X., Wang, S., Liu, D.	A novel bankruptcy prediction model based on an adaptive fuzzy k-nearest neighbor method	2011	The Wieslaw dataset “which contains 30 financial ratios and 240 cases in total”, and the Australian credit dataset, “307 instances of creditworthy applicants and 383 instances where credit is not creditworthy”.

Table 2-14 - Comparison of Sample Selection Methods for Recent Research

The samples used for this research are much more varied. Kane et al. (1998), Varetto (1999), Wang (2004) and Lensberg (2006) all used large sample sizes, often in the thousands of cases.

By comparison, Dimitras et al. (1999), used a sample size of 40. Typically, data that was matched was matched by industry and size, though a large amount of research was not matched at all.

Having discussed the samples used, the following section will discuss the variable selection techniques adopted by the analysed papers.

2.4.3 Variable Selection & Dimension Reduction

Like the papers that focus on statistical analysis or Neural Networks, the need to identify a reasonably sized set of input variables is recognised.

Author	Title	Year	Variable Selection Method
Kane, G., Richardson, F. & Meade, N.	Rank transformations and the prediction of corporate failure	1998	"[W]e constructed two logistic regression models: one using the variables recommended by Hopwood, et al. 1994 and a second (for confirmation of results) using the variables from the classic Altman 1968 linear discriminant model with a variable added to control for size differences."
Dimitras, A., Slowinski, R., Susmaga, R. & Zopounidis, C.	Business failure prediction using rough sets	1999	"The credit manager of a large Greek bank was employed to act has a decision maker (DM)", the DM "played an important role in:" "the choice of the attributes (financial ratios) entering the information table."
Varetto, F.	Genetic algorithms applications in the analysis of insolvency risk	1999	While the GA was responsible for the selection of ratios, a "financial analyst establishes, on the basis of economic reasoning", the "list of ratios which belong to each family".
McKee, T.	Developing a bankruptcy prediction model via rough sets theory	2000	"Prior research (McKee, 1995a, b) used both theoretical work concerning bankruptcy theory and prior bankruptcy research to identify the following eight potential predictive variables:" "The first six ratios were selected from the Hopwood et al. (1989) study to build on the stream of prior research. The last two ratios were selected from a publication (McKee, 1989) which suggested they might be significant in analyzing the financial liquidity of a company." "In the prior research (McKee, 1995b) couple recursive partitioning with continuity theory to reduce this set of eight potential predictive variables to two variables."

Author	Title	Year	Variable Selection Method
Anandarajan, M., Lee, P., & Anandarajan, A.	Bankruptcy prediction of financially stressed firms: an examination of the predictive accuracy of artificial neural networks	2001	Quantitative: "Rather than use a multiplicity of financial ratios as in previous studies we used a model that incorporates ratios measuring profitability, solvency, and liquidity. The model selected is the Zmijewski score. Zmijewski (1984) developed a weighted probit bankruptcy prediction model." Qualitative: Selected qualitative variables, "According to evidence provided by prior research (e.g. Gajpal et al., 1994; Gilson et al, 1990; Giroux and Wiggins, 1984; Turetsky, 1997)"
McKee, T. & Lensberg, T.	Genetic programming and rough sets: A hybrid approach to bankruptcy classification	2002	"Factors were selected [from the literature] if they were highly significant predictors in multiple studies and had significant theoretical support from continuity theory." "This process led to the identification of eleven predictive factors". "Rough sets theory was then used to develop a bankruptcy prediction model from the 11 variables representing the 11 factors." From the model, four key variables were identified.
Wang, Z.	Financial Ratio Selection for Default-Rating Modelling: A Model-Free Approach and Its Empirical Performance	2004	"[The research used] the financial ratios used in Frydman, Altman, and Kao (1985), were a classification tree is constructed by [recursive partition tree], as a candidate to start my ratio selection."
Wu, C.	Using non-financial information to predict bankruptcy: a study of public companies in Taiwan	2004	"Independent variables consist of two categories, one is the financial ratio related group, and the other is the non-financial group. The financial-related group consists of 18 financial ratios from the database of Taiwan Economic Journal. After using factor analysis, the study selects some variables which have the highest loadings". The non-financial-related group consists of three variables that describe the board holding ratio, the use of an external auditor and the stock price trend, but the research does not discuss why these variables were chosen.
Shin, K., Lee, T. & Kim, H.	An application of support vector machines in bankruptcy prediction model	2005	"We apply two stages of the input variable selection process. At the first stage, we select 52 variables among more than 250 financial ratios by independent-samples t-test between each financial ratio as an input variable and bankrupt or non-bankrupt as an output variable. In the second stage, we select 10 variables using a MDA stepwise method to reduce dimensionality."

Author	Title	Year	Variable Selection Method
Lensberg, T., Eilifsen, A., & McKee, T.	Bankruptcy theory development and classification via genetic programming	2006	"We judgmentally selected 15 basic ratios which were strong bankruptcy indicators in multiple prior bankruptcy prediction models studies", "we decided to include the prior audit opinion", "we decided to include 10 possible fraud indicators", and "we used a dummy variable in this study to reflect start-up status." "We performed an initial variable analysis on the sample" (using a Genetic Programming tournament), which "reduced the number of variables from 28 to 6".
Etemadi, H., Rostamy, A. & Dehkordi, H.	A genetic programming model for bankruptcy prediction: Empirical evidence from Iran	2009	"At the first stage, bankruptcy prediction literature was reviewed and 65 variables among more than 250 financial ratios were selected as predictive variables. These financial ratios were chosen based on popularity in literature. In the second stage, 43 variables were selected based on availability of the necessary data. Table 1 shows the selected variables. In the third stage, using stepwise discriminant analysis (SDA) was used to select final variables."
Chen, H., Yang, B., Wang, G., Liu, J., Xu, X., Wang, S., Liu, D.	A novel bankruptcy prediction model based on an adaptive fuzzy k-nearest neighbor method	2011	"We fill focus on exploring the PSO-based [Particle Swarm Optimization] parameter optimization and feature selection approach. The continuous PSO algorithm will be employed to evolve an adaptive [Fuzzy K-means Neural Network], where the neighbourhood size k and the fuzzy strength parameter m are adaptively specified. On the other hand, the binary PSO will be sued as a feature selection vehicle to identify the most informative features as well".

Table 2-15 - Comparison of Variable Selection Methods for Recent Research

Perhaps the most interesting aspect of this subsection is that only two papers (Shin, et al., 2005; Chen, et al., 2011) began the variable selection process with more than 18 factors. While many other reviewed papers can be commended for their usage of objective ratio selection techniques, papers such as Kane et al. (1998), Varetto (1999), McKee (2000), Anandarajan et al. (2001), McKee & Lensberg (2002), Wang (2004), Wu, (2004) and Lensberg et al. (2006) all use subjectively selected ratios or simply use the ratios used in prior research. This situation represents a limitation in the available literature on corporate failure prediction.

That being said, some objective ratio selection techniques were used on the limited set of initial factors, from Logistic Regression (Kane, et al., 1998), Genetic Algorithms (Varetto, 1999), Rough Sets Theory (McKee & Lensberg, 2002), Recursive Partitioning Trees (Wang, 2004), Factor Analysis (Wu, 2004), Multiple Discriminant Analysis (Shin, et al., 2005), Genetic Programming (Lensberg, et al., 2006) and Particle Swarm Optimization (Chen, et al., 2011).

The following table outlines the different variables that were used.

Author	**Title**	**Year**	**Variables**
Kane, G., Richardson, F. & Meade, N.	Rank transformations and the prediction of corporate failure	1998	"The Hopwood, McKeown, and Mutchler (1994) model", net income to total assets, current assets to total assets, current assets to current liabilities, cash to total assets, current assets to sales, long-term debt to total assets, natural log of firm sales. "The Altman (1968) model", working capital to total assets, retained earnings to total assets, operating income to total assets, sales to total assets, market value equity to book value debt, natural log of firm sales ("added to allow for possible size effects").
Dimitras, A., Slowinski, R., Susmaga, R. & Zopounidis, C.	Business failure prediction using rough sets	1999	Net income to gross profit, gross profit to total assets, net income to total assets, net income to net worth, current assets to current liabilities, quick assets to current liabilities, long term debt plus current liabilities to total assets, net worth to net worth plus long term debt, net worth to net fixed assets, inventories to working capital, current liabilities to total assets, working capital to net worth.
Varetto, F.	Genetic algorithms applications in the analysis of insolvency risk	1999	List of financial ratios made available to the GA not published.
McKee, T.	Developing a bankruptcy prediction model via rough sets theory	2000	Current assets to current liabilities, net income to total assets.
Anandarajan, M., Lee, P., & Anandarajan, A.	Bankruptcy prediction of financially stressed firms: an examination of the predictive accuracy of artificial neural networks	2001	Quantitative: Net income to total assets, total debt to assets, current assets to current liabilities. Qualitative: Negative cash flows, reduction or omission of dividends, debt default, troubled debt restructuring (1 for true, 0 for false)
McKee, T. & Lensberg, T.	Genetic programming and rough sets: A hybrid approach to bankruptcy classification	2002	Working capital to net worth, net income to total assets, cash to current liabilities, investment cash flow to net income.

Author	Title	Year	Variables
Wang, Z.	Financial Ratio Selection for Default-Rating Modelling: A Model-Free Approach and Its Empirical Performance	2004	Cash to total assets, cash to total sales, cash flow to total debt, current assets to current liabilities, current assets to total assets, current assets to total sales, earnings before interest and taxes to total assets, log of interest coverage plus 15, log of total assets, market value of equity to total capitalization, net income to total assets, quick assets to current liabilities, quick assets to total assets, quick assets to total sales, retained earnings to total assets, total sales to total assets, total debt to total assets, working capital to total assets, working capital to total sales, standard deviation of earnings before interest and taxes to total assets.
Wu, C.	Using non-financial information to predict bankruptcy: a study of public companies in Taiwan	2004	Long-term capital ratio to fixed assets, current ratio, quick ratio, times interest earned, inventory turnover, total assets turnover, return on assets, return on total equity, net profit before taxes to capital issued, cash reinvestment ratio, board holding ratio to capital issued, does the sample firm change its external auditor?, stock price trend.
Shin, K., Lee, T. & Kim, H.	An application of support vector machines in bankruptcy prediction model	2005	Total asset growth, contribution margin, operating income to total asset, fixed asset to sales, owner's equity to total asset, net asset to total asset, net loan dependence rate, operating asset constitute ratio, working capital turnover period, net operating asset turnover period.
Lensberg, T., Eilifsen, A., & McKee, T.	Bankruptcy theory development and classification via genetic programming	2006	Audit opinion (coded 1, 2 or 3), log of total assets, cash plus short-term investment to current liabilities, log of year founded, share capital to total assets, operating income plus interest expenses to interest expense.
Etemadi, H., Rostamy, A. & Dehkordi, H.	A genetic programming model for bankruptcy prediction: Empirical evidence from Iran	2009	"(1) Operational income to sales ratio (X36), (2) Total liability to total assets (X9), (3) Sales to current assets ratio (X43), (4) Interest expense to gross profit (X25), and (5) Quick assets to total assets (X20)."
Chen, H., Yang, B., Wang, G., Liu, J., Xu, X., Wang, S., Liu, D.	A novel bankruptcy prediction model based on an adaptive fuzzy k-nearest neighbor method	2011	Cash to current liabilities, cash to total assets, current assets to current liabilities, current assets to total assets, working capital to total assets, working capital to sales, sales to inventory, sales to receivables, net profit to total assets, net profit to current assets, net profit to sales, gross profit to sales, net profit to liabilities, net profit to equity, net profit to equity plus long term liabilities, sales to receivables, sales to total assets, sales to current assets, 365 receivables to sales, sales to total assets, liabilities to total income, current liabilities to total income, receivables to liabilities, net profit to sales, liabilities to total assets, liabilities to equity, long term liabilities to equity, current liabilities to equity, EBIT to total assets, current assets to sales.

Table 2-16 - Comparison of Selected Variables for Recent Research

While the sets of variables used are generally similar to those used in the research studied in the previous sections, one exception is that Anandarajan et al. (2001), Wu (2004) and Lensberg et al. (2006) used some qualitative variables such as boolean variables for things such as “troubled debt restructuring” or “audit opinion”.

2.4.4 Research Methodology

Unlike in the previous research methodology sections above, this section deals with a number of fundamentally different techniques used to predict corporate failure. Therefore, unlike the previous research methodology sections, this section will discuss not only the methodology used within the chosen technique, but also the technique itself. In doing so, conclusions can be drawn from comparisons between techniques across some of the more recent research.

Author	Title	Year	Research Methodology
Kane, G., Richardson, F. & Meade, N.	Rank transformations and the prediction of corporate failure	1998	Used the statistical methodology logistical regression to formulate two models with the two sets of variables as specified in Table 2-16. The two models were tested both with the raw data and with data that was ranked in comparison to other values for the same variable.
Dimitras, A., Slowinski, R., Susmaga, R. & Zopounidis, C.	Business failure prediction using rough sets	1999	Used “the credit manager of a large Greek bank” as a decision maker to identify the “choice of attributes”, the “discretization of the continuous attributes”, the “selection of a satisfactory reduct of attributes from among all reducts calculated for the learning sample”, and “the test of decision rules on the testing sample”. Rough set analysis was performed on the coded data using RoughDAS and ProFIT. The analysis yielded a number of “reducts” (predictive models), which were presented to the decision maker to select the best one. From this, three sets of decision rules were generated, and the results were then compared with the results from discriminant analysis and logit analysis.

Author	Title	Year	Research Methodology
Varetto, F.	Genetic algorithms applications in the analysis of insolvency risk	1999	Genetic algorithms were used to create both a linear function that could be used to model corporate health, and a credit score that could be used to identify an organisations risk. In cases were continuous values were identified, the continuous values were broken down into n discrete values, where n is the number of intervals chosen between the high and low bounds of the continuous variable. All values now being discrete, they were coded within the model using binary representation.
McKee, T.	Developing a bankruptcy prediction model via rough sets theory	2000	Continuous variables were recoded into discrete variables, before the RoughDAS software was used to implement Rough Sets Theory and create atoms for the decision model. The RoughDAS software was then used to generate a 27-rule decision model. “The 27-rule decision model was then run against both the 100-companies sample from which it was developed and a separate holdout sample of 100 companies.”
Anandarajan, M., Lee, P., & Anandarajan, A.	Bankruptcy prediction of financially stressed firms: an examination of the predictive accuracy of artificial neural networks	2001	Compares the effectiveness of backpropagation Neural Networks (using sigmoid transfer function) in comparison with a Genetic Algorithm Neural Network and Multiple Discriminant Analysis, including “qualitative ‘bad news’ variables” (negative cash flows from operations, dividend reductions or omissions, debt default or troubled debt restructuring) into the model, while using “only financially distressed firms [in the] control sample”.
McKee, T. & Lensberg, T.	Genetic programming and rough sets: A hybrid approach to bankruptcy classification	2002	“In this paper, we suggest dealing with the lack of a causal basis for bankruptcy prediction by means of a two-stage hybrid model: Stage 1 uses a *rough sets* model (Pawlak, 1982) to identify subsets of potentially important explanatory variables, and Stage 2 a *genetic programming* algorithm (Koza, 1992) to develop a structural model of bankruptcy based on those variables. The aim is to let the data speak for itself as far as possible, by minimizing the amount of a priori structure imposed by functional forms and statistical selected procedures.”

Author	Title	Year	Research Methodology
Wang, Z.	Financial Ratio Selection for Default-Rating Modelling: A Model-Free Approach and Its Empirical Performance	2004	The research aims to identify the key financial ratios when faced with the problem of predicting corporate failure, but acknowledges the weakness of Principal Component Analysis, "[T]he components identified by PCA will not be sensitive to corporate failure indicators and will not be able to uncover which ratios distinguish these two distributions." Therefore the research uses sliced average variance estimation (SAVE) to "construct a set of factors and to identify the financial ratios that are informative and will be used to model firm failure." To test the SAVE methodology, "multivariate discriminant analysis", "generalized smoothing spline" models, and "recursive partition tree methods" were used as a predictive models.
Wu, C.	Using non-financial information to predict bankruptcy: a study of public companies in Taiwan	2004	The research uses factor analysis to reduce the initial number of financial ratios, before using Logistic Regression on models both with and without non-financial information to predict corporate failure.
Shin, K., Lee, T. & Kim, H.	An application of support vector machines in bankruptcy prediction model	2005	Used Support Vector Machines with a radial basis function as the kernel function, varying the upper bound and kernel parameters to find optimal prediction performance. For comparison a back-propagation Neural Network was designed using the standard 3-layer model with 10 hidden nodes and a sigmoidal transfer function.
Lensberg, T., Eilifsen, A., & McKee, T.	Bankruptcy theory development and classification via genetic programming	2006	Used a Genetic Programming tournament firstly to identify key variables, and secondly to develop a predictive model using an expanded sample (due to there being more companies with data available for the key variables). 2,020,000 tournaments were used, some random noise was introduced to avoid over training, and a penalty was applied for complexity. Two logit models were also developed to "benchmark" the Genetic Programming methodology using the same 6 variables.
Etemadi, H., Rostamy, A. & Dehkordi, H.	A genetic programming model for bankruptcy prediction: Empirical evidence from Iran	2009	Compared a Genetic Programming model that was using number-of-hits as a fitness function to Multiple Discriminant Analysis model. For both models, a randomised division of training and holdout sample was used.
Chen, H., Yang, B., Wang, G., Liu, J., Xu, X., Wang, S., Liu, D.	A novel bankruptcy prediction model based on an adaptive fuzzy k-nearest neighbor method	2011	"Based on an adaptive fuzzy k-nearest neighbor (FKNN) method, where the neighborhood size k and the fuzzy strength parameter m are adaptively specified by the continuous particle swarm optimization (PSO) approach."

Table 2-17 - Comparison of Research Methods for Recent Research

As discussed in the introduction for this section, the four major techniques used include Genetic Algorithms, Rough Sets Theory, Support Vector Machines and Genetic Programming.

Generally speaking, the existing research has not made many comparisons between major research techniques, though Anandarajan et al. (2001) compared Neural Networks, Genetic Algorithms and Discriminant Analysis, while Shin et al. (2005) compared Support Vector Machines with Neural Networks and Chen et al. (2011) compared various forms of Neural Networks and Support Vector Machines with their proposed algorithm.

2.4.5 Key Findings

Author	Title	Year	Key Findings
Kane, G., Richardson, F. & Meade, N.	Rank transformations and the prediction of corporate failure	1998	• "Comparison of the results reveals that the use of ranked predictor variables improves the explanatory and prediction capabilities of both models."
Dimitras, A., Slowinski, R., Susmaga, R. & Zopounidis, C.	Business failure prediction using rough sets	1999	• "The attribute with the highest frequency of occurrence in reducts is a_{11} [managerial performance ratio] • "The reduct selected was the #16, which includes: a_4 (profitability ratio), a_5 [current assets to total liabilities], a_7 [long term debt plus current liabilities to total assets], a_9 (solvency ratios) and a_{11} (managerial performance ratio)." • Achieved 76.3% test set accuracy in the year prior to failure (correctly classified 94.7% of bankrupt firms and 57.9% of healthy firms) using the "'strong', partly discriminating rules".
Varetto, F.	Genetic algorithms applications in the analysis of insolvency risk	1999	• Achieved 96.94% accuracy on sound companies and 92.97% accuracy on unsound companies using Genetic Algorithms to generate a linear function. • Achieved 89.96% accuracy on sound companies and 94.98% accuracy on unsound companies using Genetic Algorithms to generate a credit score. • "Our overall results indicate that the discriminant analysis technique proved to be slightly better than those obtained with the GAs."

McKee, T.	Developing a bankruptcy prediction model via rough sets theory	2000	• Achieved 88% accuracy when using valued closeness relation matching strategy when neither Exact Match or non-deterministic matching strategy match is possible. • “The rough sets model from the current research, therefore, appears to be more robust than the recursive partitioning model since it had a 23% higher prediction accuracy on the 100-company validation sample.”
Anandarajan, M., Lee, P., & Anandarajan, A.	Bankruptcy prediction of financially stressed firms: an examination of the predictive accuracy of artificial neural networks	2001	• “The genetic algorithm neural network (ANN-GA) had the highest accuracy [95% for bankrupt firms, 94% for non-bankrupt]” for the holdout sample.
McKee, T. & Lensberg, T.	Genetic programming and rough sets: A hybrid approach to bankruptcy classification	2002	• “The preferred genetic program utilized only three of the four variables selected by the rough sets algorithm as the variable investment cash flow to net income was not picked up by the genetic model. This presumably occurred because this variable did not contain additional information beyond that in the other three variables.” • Achieved 80% accuracy on the validation sample using Genetic Programming in comparison to the 67% achieved through rough sets theory. • “[N]ot only negative profits, but also abnormally high ones, are a signal of high bankruptcy risk, except in very small companies.” “[T]he risk of bankruptcy decreases with companies size only if profits are positive” “[A] currently unprofitable company may still be considered a good risk if it is small and liquidity [is] good.”
Wang, Z.	Financial Ratio Selection for Default-Rating Modelling: A Model-Free Approach and Its Empirical Performance	2004	• “Based on [the sliced average variance estimation] results, return on investment ([Earnings before interest and taxes to total assets], [standard deviation of earnings before interest and taxes to total assets]), capital turnover ([working capital to total sales], [current assets to total assets]), short-term ([quick assets to current liabilities]), cash position ([cash to total sales], [cash to total assets]), inventory turning ([current assets to total sales]), and receivables turnover ([quick assets to total assets], [quick assets to total sales]) contribute to the two factors.” • “For both the [principal component analysis] and [sliced average variance estimation] ratio selection methods, [generalized smoothing spline] and [recursive partition tree] predict default-rated firms more accurately compared to [multiple discriminant analysis].

Wu, C.	Using non-financial information to predict bankruptcy: a study of public companies in Taiwan	2004	• "The correct classification results for the prediction model, which is based upon both financial and non-financial information, are superior to the prediction model, which is based upon financial information." • "The model is able to correctly predict some 87.1% of the companies in the sample."
Shin, K., Lee, T. & Kim, H.	An application of support vector machines in bankruptcy prediction model	2005	• "[Support Vector Machines] approach outperforms [Backpropagation Neural Networks]". • "The choice of the kernel function and the determination of optimal values of the parameters have a critical importance on the performance of the resulting system".
Lensberg, T., Eilifsen, A., & McKee, T.	Bankruptcy theory development and classification via genetic programming	2006	• Achieved 81% accuracy on the validation sample. • Genetic programming performed better than the logic models • "An unfavourable audit report has a more negative bankruptcy status impact for a large company than a small one." Interest paying ability has a more positive bankruptcy status impact for large firms than small ones.
Etemadi, H., Rostamy, A. & Dehkordi, H.	A genetic programming model for bankruptcy prediction: Empirical evidence from Iran	2009	• Found the GP model had a 90% accuracy in the holdout sample while MDA achieved 73% in the holdout sample.
Chen, H., Yang, B., Wang, G., Liu, J., Xu, X., Wang, S., Liu, D.	A novel bankruptcy prediction model based on an adaptive fuzzy k-nearest neighbor method	2011	• "PTVPSO-FKNN [the new method] outperforms all other methods with the AUC of 81.69%, except the type II error which slightly higher than that of PNN." • The proposed approach are implemented in a parallel environment which reduces computational time.

Table 2-18 - Comparison of Key Findings for Recent Research

Unlike the research studied in the previous sections of this thesis, the key findings in this research are of particular interest because of the new research directions that their findings create.

Aside from the (generally high) accuracies shown by the research, of particular note are findings that discuss the relationships between variables and bankruptcy risk as calculated by the model. McKee & Lensberg (2002) comment, "[N]ot only negative profits, but also abnormally high ones, are a signal of high bankruptcy risk, except in very small companies", "the risk of bankruptcy decreases with companies size only if profits are positive", and "a currently unprofitable company may still be considered a good risk if it is small and liquidity [is] good." Conversely Lensberg et al. (2006) comment, "An unfavourable audit report has a more negative bankruptcy status impact for a large company than a small one", and that the firm's ability to pay interest has a more positive bankruptcy status impact for large firms than small ones.

Perhaps unsurprisingly, each piece of research – with the exception of Veretto (1999) – found that more modern techniques such as Genetic Programming or Support Vector Machines outperform less modern techniques such as Neural Networks or Multiple Discriminant Analysis.

2.4.6 Research Limitations

Unlike sections 2.2 and 2.3 above, this section focuses on the more recent developments in the field of corporate failure prediction. Therefore, the limitations of individual papers are of greater importance, as those limitations are less likely to be addressed in newer papers and thus represent opportunities for future research. Consequently, the following section will outline important limitations that are worth drawing attention to.

Author	Title	Year	Research Limitations
Kane, G., Richardson, F. & Meade, N.	Rank transformations and the prediction of corporate failure	1998	• Only compared the effect of rank transformations on a "recent model (Hopwood, et al. 1994) and a classic model (Altman 1968)". Since rank transformation primarily addresses the limitations inherent in using non-normalised data, this research is assuming that the data in these two papers lacked normality. • Limited choice of initial variables.

Author	Title	Year	Research Limitations
Dimitras, A., Slowinski, R., Susmaga, R. & Zopounidis, C.	Business failure prediction using rough sets	1999	• While the acknowledgement of many other methodologies is made, the paper focuses on a comparison between rough sets and "discriminant analysis and logit analysis". This is unusual since it could easily be argued that Neural Networks were the defacto standard for corporate failure prediction by 1999. • This application of rough set theory had a large dependence on the subjective opinions of an expert. • Subjective choice of initial variables.
Varetto, F.	Genetic algorithms applications in the analysis of insolvency risk	1999	• Did not publish the original or reduced selection of variables, nor discuss which variables the GA selected. • Did not publish the resulting linear function that was generated or the resulting ruleset for the credit scoring system. • Subjective choice of initial variables.
McKee, T.	Developing a bankruptcy prediction model via rough sets theory	2000	• While the research was generally sound, it did not compare the results with those that would have been achieved by using Neural Networks or Discriminant Analysis. This severely limits its comparative ability. • Limited choice of initial variables.
Anandarajan, M., Lee, P., & Anandarajan, A.	Bankruptcy prediction of financially stressed firms: an examination of the predictive accuracy of artificial neural networks	2001	• Used only distressed firms in the sample, which limits the models ability to evaluate the financial status of a firm that does not meet the researchers criteria of distressed. • Did not discuss the structure of the Neural Network used. • Limited choice of initial variables.
McKee, T. & Lensberg, T.	Genetic programming and rough sets: A hybrid approach to bankruptcy classification	2002	• The use of rough sets theory to reduce the variables given to the Genetic Programming technique is an objective technique, but there would have been benefit to giving all the ratios to the technique. • Subjective choice of initial variables.
Wang, Z.	Financial Ratio Selection for Default-Rating Modelling: A Model-Free Approach and Its Empirical Performance	2004	• Sliced Average Variance Estimation is a relatively esoteric means of financial ratio selection. • Neural networks was a popular predictive technique available at the time that was not tested with the chosen variables. • Limited choice of initial variables.
Wu, C.	Using non-financial information to predict bankruptcy: a study of public companies in Taiwan	2004	• Neural networks was a popular predictive technique available at the time that was not tested with the chosen variables. • Limited choice of initial variables.

Author	Title	Year	Research Limitations
Shin, K., Lee, T. & Kim, H.	An application of support vector machines in bankruptcy prediction model	2005	• Optimised input variable selection using something other than the accuracy of the model in question, potentially selecting the non-optimal factors. • On the complete training dataset, the SVM parameters that yielded the best accuracy resulted in a very poor out-of-sample accuracy.
Lensberg, T., Eilifsen, A., & McKee, T.	Bankruptcy theory development and classification via genetic programming	2006	• Neural networks was a popular predictive technique available at the time that was not tested with the chosen variables. • Limited choice of initial variables.
Etemadi, H., Rostamy, A. & Dehkordi, H.	A genetic programming model for bankruptcy prediction: Empirical evidence from Iran	2009	• Limited sample size. • No comparison with other intelligent techniques such as Neural Networks. • Optimised factors based on Stepwise Discriminant Analysis, with no evidence that a factor set optimised to Genetic Programming would not have performed better.
Chen, H., Yang, B., Wang, G., Liu, J., Xu, X., Wang, S., Liu, D.	A novel bankruptcy prediction model based on an adaptive fuzzy k-nearest neighbor method	2011	• Limited sample size when using bankruptcy data, and acknowledgement that the proposed method may not yield increased accuracy on large data sets. • Limited initial choice of variables.

Table 2-19 - Comparison of Limitations for Recent Research

One common research limitation is the lack of comparison between major techniques such as Neural Networks. While this is understandable since most research tends to focus on a particular methodology, a greater comparison of techniques would help bring some unity to the area of corporate failure prediction.

Another common limitation, as discussed previously in 2.4.3, is the subjective choice of original variables, as only Shin et al. (2005) used a wide array of factors and applied an objective factor selection methodology. There is definite insight to be gained by performing some kind of objective dimension reduction on a larger array of available variables.

2.4.7 Additional Recent Noteworthy Research

Beyond the research identified previously in this section are a number of recent papers that should be highlighted as they provide context to the rest of this thesis.

In “Bankruptcy prediction models based on multinorm anlaysis: An alternative to accounting ratios” (Andrés, et al., 2012), the authors find that by computing industry norms using non-parametric quantile regression, and then detecting deviations from those industry norms, some improvements using existing classifiers can be obtained on their dataset of Spanish firms. “Enhanced default risk models with SVM+” (Ribeiro, et al., 2012) examines the effect of non-financial information from additional sources on a selection of French companies and finds that prediction performance using SVM can be increased using from the baseline SVM that did not include such additional data. In “Simple instance selection for bankruptcy prediction” (Tsai & Cheng, 2012), a study is undertaken in which outliers in the data are identified and filtered out across four datasets using Neural Networks, Decision Trees, Logistic Regression and Support Vector Machines, finding that removing some outliers can improve performance but that removing outliers can in some scenarios decrease predictive accuracy.

“Performance of corporate bankruptcy prediction models on imbalanced dataset: The effect of sampling methods” (Zhou, 2013) investigates the effect of various forms of oversampling (re-using the minority class to achieve distribution balance), various forms of under-sampling (using only some of the majority class to achieve distribution balance), and in general finds that under-sampling performs better than over-sampling. “The application of brute force logistic regression to corporate credit scoring models: Evidence from Serbian financial statements” (Nikolic, et al., 2013) uses clustering to identify highly correlated variables across 350 financial ratios to generate a shortlist of 24 ratios, and then tests all possible variable combinations that contain between 5 and 14 variables to find an optimum combination of those 24 ratios in a Logistic Regression model, finding the highest scoring model to contain 8 ratios.

2.4.8 Conclusion

McKee & Lensberg (2002) comment, “A common approach to bankruptcy prediction is to review the literature to identify a large set of potential predictive financial and/or non-financial variables and then develop a reduced set of variables, through some combination of judgmental and mathematical analysis, that will predict bankruptcy. However, a problem exists in that the various models developed normally use both different variables and different forms to specify the relationships between these variables. Thus, after 30 years of research on this topic, there is no generally accepted model for bankruptcy prediction that has its basis in a causal specification of underlying economic determinants. Clearly, research convergence will be necessary for this situation to improve.” This situation remains largely unresolved over 10 years later. Many papers such as those included in the previous sections use arbitrarily selected ratios (though some papers do use an objective variable selection method), use small sample sets (though some do use large sample sizes), compare just one or two newer methodologies (though some compare more), use just one data set (though some use two), and very few perform any kind of post-prediction analysis that investigate how or why their models have classified the cases in the datasets correctly or incorrectly.

3. Review of Corporate Failure Theory

So far, the literature review has drawn attention to academic work focused on the prediction and classification of corporate failure. While some of those papers utilise their findings to develop overall theory on corporate failure, such as Lensberg et al. (2006), with "Bankruptcy theory development and classification via genetic programming", the theoretical framework behind bankruptcy is generally not examined in the scope of corporate failure prediction. It is worthwhile to review the theory of corporate failure from an finance viewpoint, as doing so highlight opportunities to forge links between corporate failure prediction and corporate failure theory.

Unlike the prediction of corporate failure however, the theory of bankruptcy is a much more matured area. While the *Journal of Banking & Finance*, for instance, yields many more journal articles focusing on the prediction of failure than discussing the underlying theory of failure, the number of published books in relation to general corporate failure far outnumbers the number of published books on bankruptcy prediction. The motivations for reviewing the literature are also different. Instead of critically analysing the research to identify potential weaknesses and develop a new direction, the analysis of corporate failure theory instead aims to identify a commonly accepted theoretical framework that can be used to develop questions that relate to corporate failure prediction.

As a result of this review, there is an opportunity to put a magnifying glass to some of the recurring themes discussed and identify questions that can potentially be answered by this thesis.

3.1 The Literature

3.1.1 Selection Methodology

There were not a sufficient number of books on the topic to develop specific criteria such as those used for corporate failure prediction journal articles. Instead, texts were selected by a combination of subjective and objective criteria including the number of sources referencing the material, how recently the text was published, its contribution to the overall body of knowledge and its availability. For example Ross & Kami (1973) has been included in the literature review because that research appears to have heavily influenced the findings of Argenti (1983), which in turn has become the basis of a generally accepted theory of corporate failure.

3.1.2 Ross & Kami (1973)

Ross & Kami's "Corporate Management in Crisis: Why the Mighty Fall" is heavily referenced in Argeni (1976), Kharbanda & Stallworthy (1985), and Clarke et al. (1997), and was one of the first texts to build a theoretical framework of corporate failure. Ross & Kami propose that while there are a number of circumstances surrounding a company failing, the ultimate cause is bad management.

Ross & Kami identify the "Ten Commandments of Management" (p. 21), as follows:

1. Develop and communicate a *strategy …a unified sense of direction* to which all members of the organisation can relate.
2. If you want to achieve plans, programs, and policies, then *overall controls* and *cost controls* must be established.
3. Exercise care in the selection of a *Board of Directors* and require that they actively *participate in management*.
4. Avoid *one-man rule.*
5. Provide *management depth*.
6. Keep informed of change and *react to change*.
7. Don't overlook the customer and the *customer's new power*.

8. Use but don't misuse computers.
9. Do not engage in *accounting manipulations*.
10. Provide for an *organizational structure* that meets the *need of people*.

The book goes on to study numerous cases of corporate crisis, finding that in each situation most of the Ten Commandments have been violated.

In each case, Ross & Kami discuss the "lessons to be learned". Of those lessons, some are worth noting since they go beyond the above Ten Commandments and come up so regularly in future research.

- "If you take the calculated risk of losing money–you usually do"
- "The acquisition chain letter can't go forever"
- "Change is accelerating and so must the company's reaction to changing times"

3.1.3 Argenti (1976)

This book, "Corporate Collapse: the causes and symptoms" is particularly useful because it is one of the first theory-focussed texts on corporate failure, and has come to be a very popular secondary source for future corporate failure research.

Through an extensive literature review and interviews with a number of notable experts, Argenti found that there was agreement between experts and the literature on a number of causes and symptoms of business failure, such as problems with management, lack of accounting information, overtrading, lack of adaptation to change, and creative accounting. On the other hand there were possible causes of failure that were found in one source but not the other; for example the experts discussed high gearing while the literature discussed economic cycles.

Argenti also documented the findings of his analysis on two significant corporate failures of the time, Rolls-Royce and Penn Central. In doing so, Argenti found evidence to support management problems, accounting information, change, creative accounting, a big project, inflation and gearing as possible causes and symptoms at Rolls-Royce. Meanwhile the research at Penn Central uncovered evidence to support management problems, accounting information, change, constraints, economic cycle, a big project, creative accounting, and gearing.

Argenti went on to claim, “no one seems ever before to have tried to coordinate all the knowledge about failure that lies scattered through the literature and in the minds of innumerable experts all over the world”, but that is not what makes Argenti’s research so unique– the text goes beyond a simplistic list of causes and symptoms, instead to the “dynamics of failure, the sequencing of events” (p. 121). After discussing a combined list of causes and symptoms of failure, he identified three main “trajectories” of corporate failure as follows in Figure 3-1.

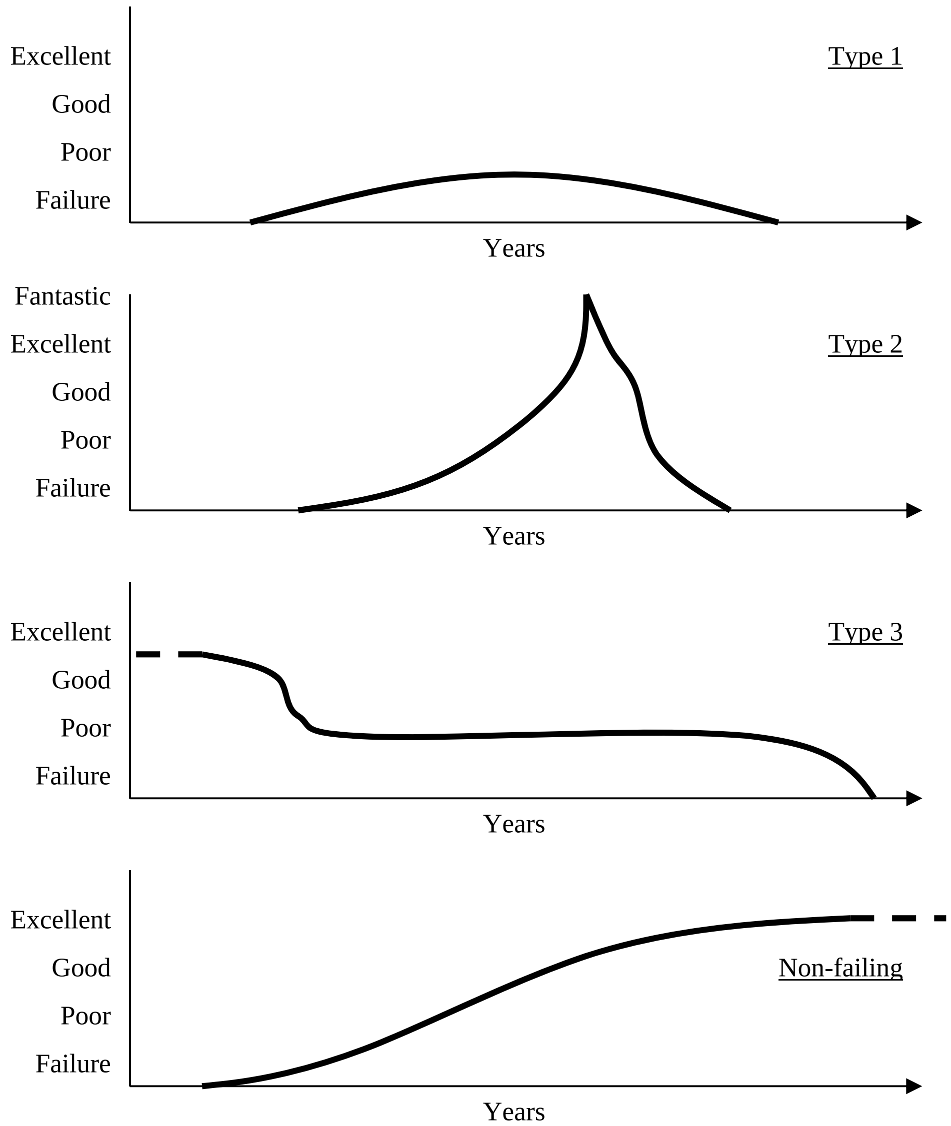


Figure 3-1 - The three types of Failure Trajectory (Argenti, 1976, p. 150)

Argenti wrote, “Type 1 failures occur only to companies newly formed and, almost invariably therefore, affect only small ones. Argenti, however, describes a company which could almost certainly have been classified as Type 1, launched with a capital outlay of $17m” (p. 153). It could be argued that even companies with a capital outlay of $17m are vulnerable to Type 1 failure.

Argenti goes on to say that the launch of a company often brings with it a number of key defects such as one-man rule, lack of management depth, an unbalanced top team, no budget, no cash flow plan, no costing system, loan interest and depreciation are not considered by management, there has been no allowance for losses, no knowledge of marginal cost or contribution, high gearing, and of course the big project. The cash flows and profits are negative, resulting in poor financial ratios. Creative accounting may begin, overtrading may begin in a vain attempt to control the crises, a further loan may increase gearing, but finally capital is exhausted and the company fails (pp. 154-156).

“Type 2, on the other hand, shoots upwards to ‘fantastic’ heights before crashing down again as did IOS, Atlantic Acceptance, Stirling Homex, and others” (p. 151). Similarly to Type 1 companies, the same management defects are present, “but there is one very prominent and identifiable difference, namely that while the proprietor of a Type 1 company is not notable for his outstanding personality, the proprietor of a Type 2 is.” “While the Type 1 proprietors are engineers, technicians, marketing men, hairdressers, welders, builders and other mortals, Type 2 proprietors are super-salesmen; they are leaders of men, flamboyant, loquacious, restless and bubbling with ideas. The scale of their ambition is almost pathological. They never accept advice, they ‘know it all’” (p. 158). The Type 2 trajectory differs from Type 1 due to the personality traits of the proprietor and a great product, resulting in a monumental increase in sales. As sales grow, capital resources are required. Argenti notes that so far, the trajectory is not so different from the non-failing trajectory. But instead of levelling off, the sales, profits and capital all continue to expand. The press gets involved and soon the company becomes so large that formal management is required. But one-man-rule continues, as do a number of other key defects. It is not long before turnover is continuing to increase but profits are not, immediately resulting in creative accounting. “In a frantic attempt to keep turnover and profits rising… [the proprietor] now reaches into the absurd” (p. 159). Argenti gives the example of Atlantic Acceptance beginning to lend money to those who could not afford to give it back, having lent money to everyone that could, an example that is reminiscent in the 2008 Global Financial Crisis, the effects of which continue today. The result in this case was overtrading, and all it takes is a normal business hazard and failure is not far away. The press draw attention to the problems, the banks and stock market punish the company, and so the capital runs dry.

Finally Type 3 failure occurs in companies that have been successfully trading for a long time. Argenti notes some key defects very early in the trajectory such as “one-man rule or chairman-chief executive or unbalanced top team or non-participating board or lack of management depth or weak finance function” (p. 161). Further to this is a lack of attention to budgetary control or lack of a timely cash flow forecast. At some point a change in the market occurs, but management do not respond to the change. As a result of these defects, perhaps years later, a big project will be launched or overtrading will occur. Argenti argues that the first crash takes place as a result of two things going wrong at the same time, such as a failed project and a business hazard, resulting in a profit fall that is reflected in financial ratios. Morale begins to falter and profits do not recover. Creative accounting starts, a loan is obtained and gearing rises. Once the competitive edge is lost and the gearing has risen, Argenti calls the company waterlogged (p. 162). The high gearing results in high interest repayments and therefore low profits, and the lack of competitive edge means that it won’t take much to cause an unrecoverable spiral toward failure. A big project is undertaken to try and slingshot the company back into the black, but the project is too resource intensive for the company to handle. On the other hand, “a small project would be safe but, of course, it would be too small to solve the problem” (p. 164). The company is in a fix.

Argenti states “while refusing to admit that more than three trajectories are needed to explain the vast majority of failures that occur, I do feel bound to admit that a company following one trajectory could switch to another” (p. 166). That could, if the right changes in the company were made, include switching to a non-failure trajectory.

Two particularly recurring themes in this text are creative accounting, and a focus on whether the company is currently profitable, however hidden by creative accounting.

3.1.4 Kharbanda & Stallworthy (1985)

While Kharbanda & Stallworthy's book, "Corporate Failure: Prediction, Panacea and Prevention" has not been as heavily referenced as Argenti (1976), it is useful to review because it acknowledges the Argenti trajectories and causes of failure, but focuses entirely on the effect of management.

The chapter "Management is the Crux" breaks the success of a company down into a number of important qualities including "'back to basics' management'", "consisting of the assimilation of news, information and comment, leading to knowledge", which in turn is associated with "commitment and discipline"; interpersonal relations; effective communication; the company culture; company excellence, which includes "a bias for action", being "close to the customer", "autonomy and entrepreneurship", "productivity through people", "hand-on, value-driven", staying "close to the business they know", an "elegant structure", and a "subtle combination of centralized and decentralized control"; acknowledgement of management lessons taken from eastern cultures; teamwork; and finally, a good leader.

One interesting comment made in Kharbanda & Stallworthy is that "one-man rule *can* be a success, but it depends so much on the man". Kharbanda & Stallworthy have appeared to support one-man rule in some circumstances, while previous authors have discredited it. Argenti acknowledges the instances of one-man rule that have not failed, but instead of qualifying the personality of the autocrat, attributes these cases to an insufficient number of other management defects to cause major problems (Argenti, 1976, p. 124).

Kharbanda & Stallworthy have used a number of cases to test their hypotheses. The first being Penn Central, the finding was made that while a “hostile environment” and “low profits” were contributing factors, faulty management “is the only item on which action could have been taken”. The next case is Rolls-Royce, where the findings of Argenti are discussed. However, rather than structuring the blame back to “the big project” or “lack of accounting information”, these causes are instead attributed to management problems.

3.1.5 McRobert & Hoffman (1997)

“Corporate Collapse: An Early Warning System for Lenders, Investors and Suppliers” was published as a direct follow-on to Argenti (1976). The forward, written by Argenti, acknowledges some of the weaknesses that are present in his own theoretical framework, mainly in Asian environments, and praises McRobert & Hoffman for having “updated my studies into the modern world”, but also having “shifted the centre of gravity from the Anglo-American economic areas” (McRobert & Hoffman, 1997, p. v).

It is not necessary to rehash the elementary similarities between Argenti (1976) and McRobert & Hoffman (1997). Being based on the same theoretical framework it is not surprising that the fundamental model is the same. However, McRobert & Hoffman identify “inadequate strategic understanding”, under the management heading, as one of the primary causes of failure.

While similar to overtrading (which also features), McRobert & Hoffman identify diversification as a cause of failure. The argument is made that “those who succeed on the first, steep segment of the development curve often have difficulty with the second”, as a result the organisation diversifies into areas of business that allow it to focus on that same development curve. The result is that the new operations lead to overtrading and loss of control. It is possible that the addition of this cause of failure to the theoretical framework is in response to many recent company’s failures through over-diversification. The example is given in the book of the

Westpac Banking Corporation, which between 1987 and 1993 very nearly reached insolvency for this reason.

McRobert & Hoffman highlight the need for internal controls in an organisation. They comment, “the controls may be absent because an autocratic chief executive has over-ridden them, or because the industry or company culture chooses to disregard them, or because a rapid growth in activity has outpaced the usefulness of the existing system”. While both the first and third reasons for controls being absent are in themselves causes of failure, it is a valid point that the presence of internal controls can help identify, avoid and rectify problems caused by one of the many other causes of failure.

Overtrading features heavily in both Argenti (1976), and in McRobert & Hoffman (1997). However McRobert & Hoffman break overtrading down into the exhaustion of one of three types of resources: physical, human and financial. Comment is also made on the possible exhaustion of management structure resources. McRobert & Hoffman argue that while these resources can be stretched, the situation becomes “very unstable”. “A critical piece of machinery will fail suddenly and expensively through lack of maintenance. A key executive, unable to withstand the pressure, will collapse or resign. A major creditor will lose faith and/or patience and will present the company with an ultimatum. Each of these instances can lead to and has led to failure and collapse.”

Unlike Argenti, McRobert & Hoffman separate the causes of failure and their symptoms. Instead of repeating their findings, Figure 3-2 will summarise the relationships they theorise exist between causes and symptoms.

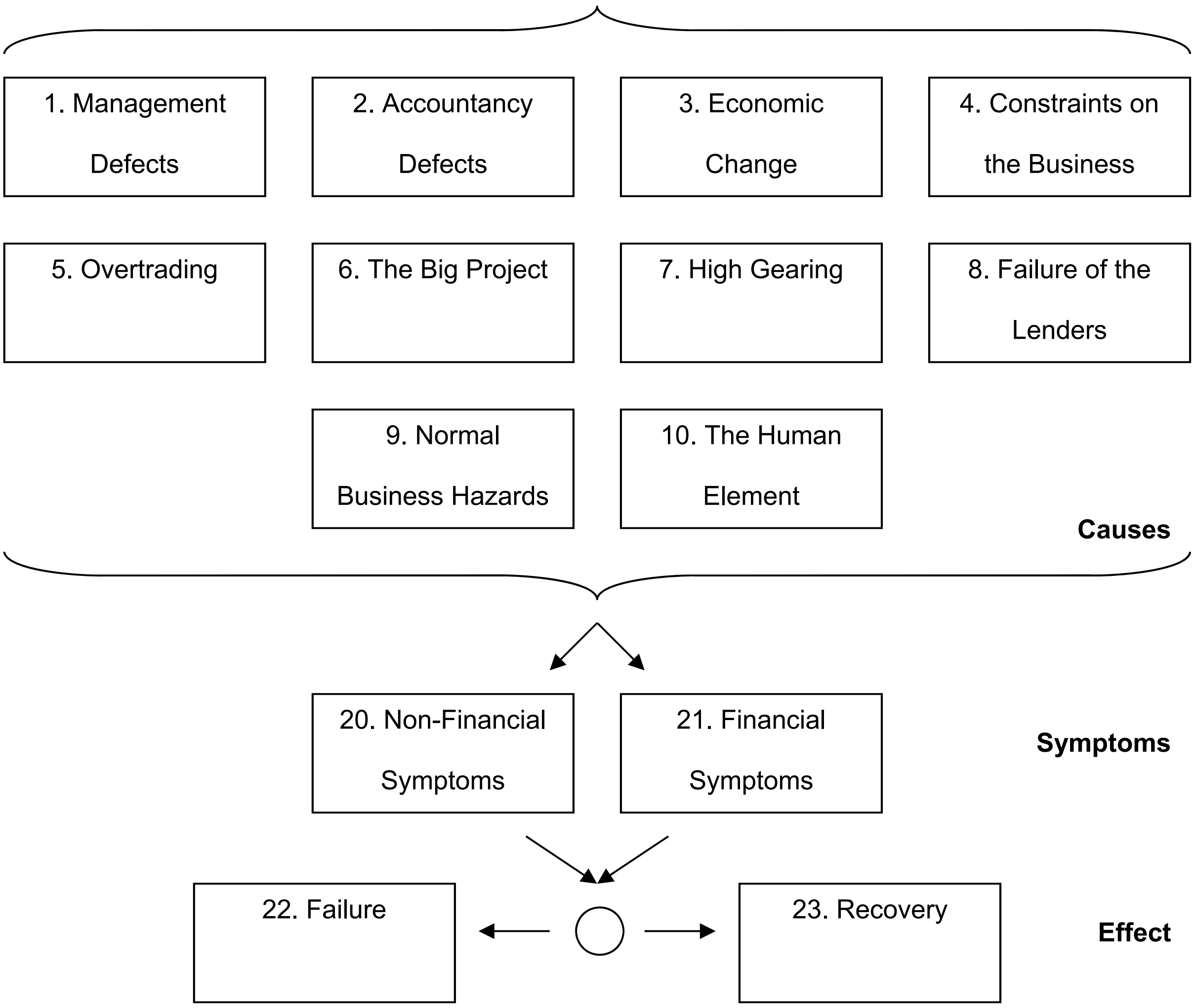


Figure 3-2 - Causes & Symptoms of Failure Part 1 (McRobert & Hoffman, 1997)

3.1.6 Clarke, Dean & Oliver (1997)

The revised edition of "Corporate Collapse: Accounting, Regulatory and Ethical Failure" is an interesting text because it focuses on how corporations can fail unexpectedly, immediately after publishing audited financial statements giving a clean bill of health.

Attention is drawn to the example of Reid Murray (pp. 55-65), in which Clarke, Dean & Oliver identify the acquisition of a number of companies in a short period, all the while artificially inflating profits on the financial statements. Furthermore there is evidence given of high gearing. Clarke, Dean & Oliver also comment, “At the management level, the RMH Board was dominated by its founding chief executive” – yet another Argenti cause of failure. Finally, the comment is made “The pattern of the rise and fall of Reid Murray is familiar”; it “shows what John Argenti described as a trajectory pattern of ‘remarkable ascent and rapid demise’”. Interestingly, the comments of B.L. Murray QC and B.J. Shaw QC, “reporting to the Victorian Parliament in 1963 on the causes of Reid Murray’s collapse”, are highlighted. “They concluded that the defects in RMH’s accounts were partly responsible for the collapse”.

Perhaps the most relevant finding to this thesis is the effectiveness with which the current accounting system allows for insolvent companies to appear perfectly healthy – even after a perfectly legitimate audit. Clarke, Dean & Oliver take the position that the prediction of failure cannot be accurate when it is based on financial ratios since the ratios themselves are so open to manipulation. This opinion seems to be at odds with the findings of Beaver (1966), Altman (1968) and the journal articles on corporate failure prediction that followed.

This disparity can be resolved somewhat by acknowledging two key points. First that not all companies fail like those in Clarke, Dean & Oliver (1997). Altman’s predictive mechanism may be highly successful at predicting failure, but may not be so accurate when faced with a company that has gone so far out of its way to act fraudulently. Second, that even though a high profit, and good quick ratio may fool some people into thinking a company is healthy, statistics can be harder to trick, especially if it is designed to find not just poor performance, but also abnormal performance.

3.1.7 Probst & Raisch (2005)

This journal article, published in the *Academy of Management Executive* and titled, “Organizational crisis: The logic of failure”, is one of the few more recent journal articles that discusses the causes of organisational failure.

In acknowledgement of six major bankruptcies from the year 2000, Probst & Raisch (2005) undertake a multiple case study of “the 100 largest organizational crises of the last five years” with the aim of achieving a “more complete theoretical explanation of the failure of successful firms”.

What makes this publication so unique is that regardless of the considerable amounts of bankruptcy research and the maturity of the area, Probst & Raisch have undertaken exploratory research, and this is reflected by their choice of the multiple case study as a research technique. Yet their findings are quite different to the findings of other exploratory research.

The choice of exploratory research appears to be in response to a surprisingly large number of big company bankruptcies. The article begins, “Reports of crises in once highly regarded companies dominated the business news during the first three years of the new millennium. WorldCom, Enron, Conseco, Global Crossing, United Airlines, Kmart… each month brought the sound of another titan crashing to earth.” The logic follows, if corporate failure and how to avoid it is well understood, why does it unexpectedly happen with companies that are thought to be financially sound?

Through the case study, Probst & Raisch identify two primary causes for unexpected crash or bankruptcy, “The Burnout Syndrome”, and the “Premature Aging Syndrome”. These syndromes are identified by a number of characteristics, as demonstrated in Figure 3-3.

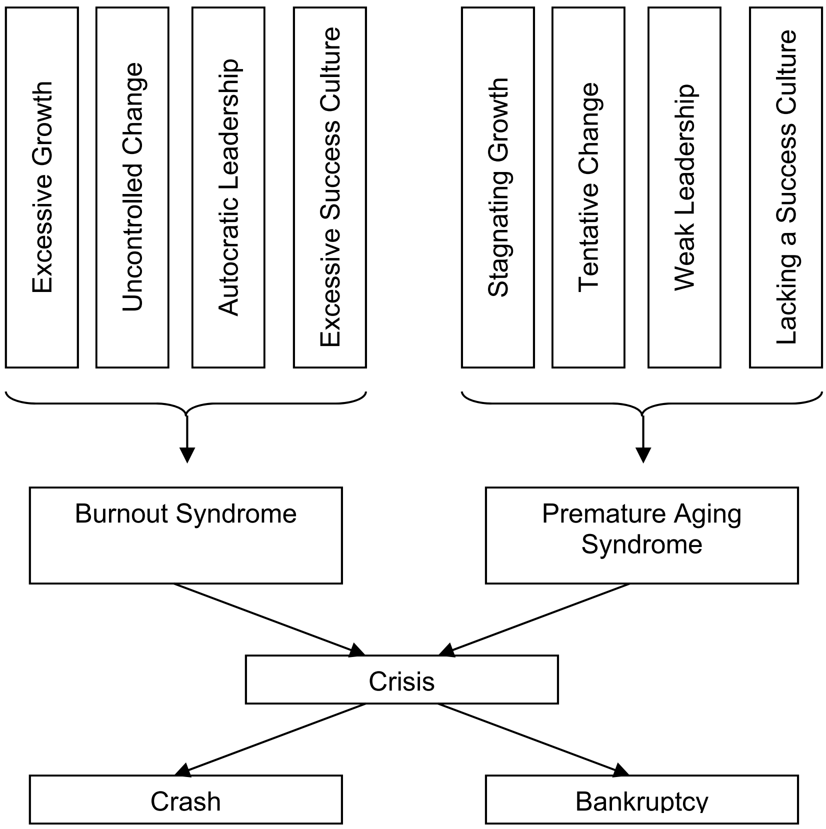


Figure 3-3 - Theoretical Framework for Probst & Raisch (2005)

3.2 The Links between Corporate Failure Theory and Corporate Failure Prediction

Having now analysed the primary existing literature on corporate failure theory, there are some recurring themes that in turn raise questions that can potentially be addressed through the use of research in corporate failure prediction. Each of those themes will be addressed.

3.2.1 Lack of Accounting Information

One of the common causes of failure identified in the previous section is a lack of accurate accounting information available to management, or the availability of bad information. In particular it is argued that other measurements of a company's financial health will be more useful in classifying failure.

Fusaro & Miller (2002, p. 146) note that on news that Enron's quality of earnings might be lower than anticipated their share price fell from $42 to $40 per share, suggesting that the market is able to assess a firm's financial health better than is reflected in the company's annual financial statements themselves. Therefore the question is asked "Does the inclusion of share market information increase the classification accuracy?"

3.2.2 Accounting Manipulation

Very much related to a lack of accounting information is the deliberate reporting of misleading accounting information, designed to make the company look healthier than it actually is. It is argued therefore that, like section 3.2.1 above, the inclusion of non-financial information such as share market information will increase accuracy of a predictive model. It is further argued that companies in which accounting manipulation is known to have occurred are likely to be misclassified as non-failure when only accounting information is available to the model.

3.2.3 Overtrading

Overtrading is generally accepted in the literature to be a major cause of failure for Argenti Type 2 failures. While it is difficult to measure the exhaustion of physical or human resources from outside the company, insight into a company's financial resources – in particular its cash position – can be gained from annual financial statements. The question is therefore raised, "What impact does cash have a model's ability to predict failure?" However to simply identify cash factors as a useful predictor does not prove that overtrading was present, or whether or not it was a cause of failure. If a model is developed that appears to detect overtrading, it is necessary to identify a case for which overtrading occurred and validate the results of the model against it.

3.2.4 High Gearing

Another common cause of failure identified in the literature is the level of debt a company faces relative to its earning ability. As gearing increases, a firm's survivability is expected to decrease. Loans have interest repayments, and often clauses that make the entire loan due at the most difficult of times. Even Enron's use of pre-pays, while classified as "trading liabilities", had their own sequence of interest-like expenses (McLean & Elkind, 2003, p. 159).

As gearing increases, so does the company's cost of interest, until eventually the loan related expenses exceed the company's gross profit. Alternatively, high gearing can leave a company profitable but cause cash flow problems as the various forms of credit become due. A recent Australian example is that of the Nine Network, which in 2012 faced external administration with "forecast earnings before interest, tax, depreciation and amortisation (EBITDA) of $253 million" (Gluyas, 2012). The problem, however, was the "$3.2 billion debt" (Whalley, 2012), which was resulting in interest repayments of "$379 million in 2010-11" (Shoebridge, 2011).

Therefore this thesis asks the question "Is gearing an important factor in the successful classification of company failure?", and furthermore, "Is there evidence of high gearing in cases that were correctly classified?"

3.2.5 Industry Classifications

The previous chapters touched on the lack of industry-specific discussion in the existing research. While the fundamental causes of corporate failure may be the same in all companies, some industries (such as manufacturing which has high initial capital costs) may be more prone to overtrading, while other industries (such as the highly regulated airline industry) may be more susceptible to external constraints. In turn, these industries are expected to exhibit different failure symptoms.

Furthermore, there are market-wide forces that influence all industries simultaneously, such as a general downturn in the economy. There are also forces that simultaneously affect a cross-section of industries, such as industries that deal internationally being heavily affected by changes in the value of the dollar. There are also external forces that simultaneously affect a number of specific businesses, such as businesses with a small number of employees being similarly affected by changes to unfair dismissal laws, even though they may come from different industries. Finally there are forces that influence an individual organisation. Thus, forces have a "scope", as follows in Figure 3-4.

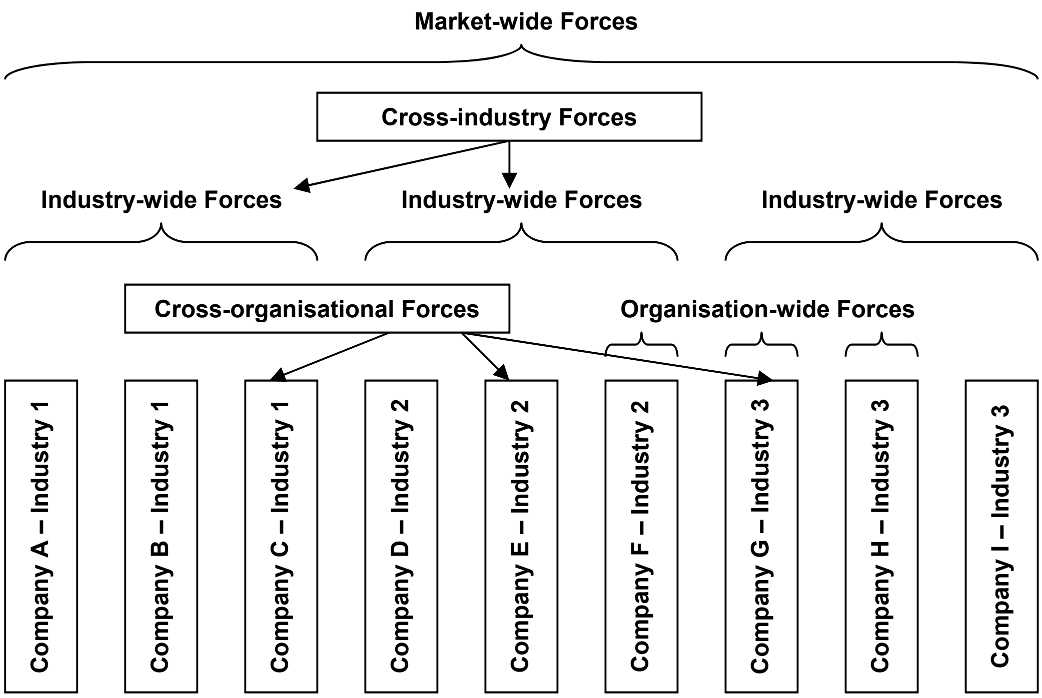


Figure 3-4 - Scope of Forces on Organisations

As can be seen from this figure, a given company may be under the influence of market-wide forces, cross-industry forces, industry-wide forces, cross-organisational forces and organisation-wide forces. Meanwhile a different company is under the influence of the same market-wide forces, some of the same cross-industry forces, the same industry-wide forces, some of the same cross-organisational forces, with entirely different organisation-wide forces.

While two companies may have very different causes (and therefore symptoms) of failure, a predictive model can only be successful on the assumption that companies exhibit similar symptoms prior to failure or non-failure. The question then becomes, how can a predictive model accurately classify companies, when the symptoms of failure between companies may vary so greatly?

A possible solution to this problem is to use different models on groups of companies. However, as the number of models increases, the number of cases within each group that the model can learn from is reduced. In the extreme case of developing a different model for every company, the models would not be exposed to even a single bankruptcy condition until the company in question had already failed, and therefore the models would not be predicting anything at all.

It is tempting to manually classify companies into groups, perhaps based on industry and size as much predictive research in the literature review with small sample sizes have done, but this method is arbitrary and subjective, and as shown in the above figure companies in the same industry may be under the influences of different forces.

This thesis therefore asks the question, “Can objective clustering be used to improve classification accuracy?”, “Is such a clustering method more effective than grouping by industry?”, and “Does the inclusion of market-wide macroeconomic factors increase classification accuracy?”

4. Methodology

With a review of the existing research complete, it becomes necessary to outline the methodology that will be used by this thesis to address the questions raised in section 3.2.

4.1 Data Preparation

4.1.1 Data Sources

For this research, two data sources have been considered. The first data source is companies with the country code "USA" from the Compustat "Legacy" Global Industrial/Commercial dataset provided by Wharton Research Data Services (WRDS). WRDS provides many datasets, and at first glance it may seem unusual to select the "Legacy" dataset which only contains from 1989 to 2008 inclusive, when there are many similar alternatives including the Compustat North American Annual database which includes data up until the current year. However, when compiling data, it became apparent that in some datasets key fields required for the following chapter had low data availability. For example, many fields had 70% data availability in the North American Annual dataset, and 95% data availability in the Legacy Global Industrial/Commercial dataset which allowed for a more comprehensive testing of factors.

The Legacy Global Industrial/Commercial data source is provided as a comma separated file (CSV), with each row representing a given company at a given date, and uses field names such as "data75" as the label for the data "Current Assets – Total". This dataset includes approximately 93,220 company-years represented as rows, with each containing some or all of 189 financial variables represented as columns.

Within the dataset, companies are uniquely identified by their "GVKEY", a proprietary number assigned to each corporate entity that does not change even when the company name or the stock ticker symbol are changed over the company's life. Companies in the Compustat database also contain a “CUSIP” code which is a 9 character code that is assigned by a third party, in which the first 6 characters identify the company (stored separately as the field “CNUM”), the next two characters identify the asset issued by that company, and the last character serves as a check digit. This CUSIP code becomes important when joining this dataset with others, as will be done when investigating the effect of stock market information in chapter 5.

Date information within this Compustat data source is stored through a number of fields, including “FYR” which indicates the month that ends that financial statement’s fiscal year, and the “YEAR” field which indicates either the year starting the fiscal period (if the FYR field is between 1 and 5 inclusive) or the year ending the fiscal period (if the FYR field is between 6 and 12). The Compustat Legacy data also includes the field “SCALE” which indicates whether the data is stored in millions (scale 3), billions (scale 6) or trillions (scale 9).

The second data source has been provided by Lincoln Indicators, who have provided data files from Aspect Financial Pty Ltd (referred to as “Aspect”) from 1987 through 2006 inclusive:

- AspectFullReplace.mdb – The original file of companies from Aspect
- FailedData.mdb – An addendum file of failed companies from Aspect
- AspectFailed2006-07-13.mdb – An addendum to FailedData.mdb

The data, in Microsoft Access Format, contains a number of related tables – notably:

- tblFinancialValueAnnualRaw – “Raw” data supplied by the company
- tblFinancialValueAnnualTotal – Data modified by Aspect to create consistency
- tblFinancialItemRaw – a data definition of ItemIDs
- tblFinancialItemTotal – a data definition of TotalItemIDs

Once the data had been pre-processed (see section 4.1.2) this dataset contains 11,239 company-years with 670 financial variables.

4.1.2 Data Pre-processing

The Compustat data source requires some modification prior to using it within the research. Firstly, some USA companies report in currencies other than the United States dollar, so any companies reporting in currencies other than USD were excluded. Secondly, the Compustat data source includes the variable "scale" which indicates the multiplier of the inputs contained within and needed to be applied to each row.

Within each financial variable, some items contained non-numeric information. For example, data19, "Interest and Related Income" sometimes contains the value "C". These non-numeric data fields generally indicate some kind of missing data, and are able to be substituted with *NULL* for simplicity.

Unlike Compustat, the Aspect data source arrived in multiple files. In order for the data to be analysed by various software, the disparate data sources and tables needed to be combined. As such, a large amount of pre-processing was needed to create a file with a horizontal format with company-years appearing as each row, and each financial variable appearing as each column.

The method used for pre-processing the Aspect dataset is shown in Figure 4-1, resulting in 706 variables that become columns, and 12,193 company-years that become rows:

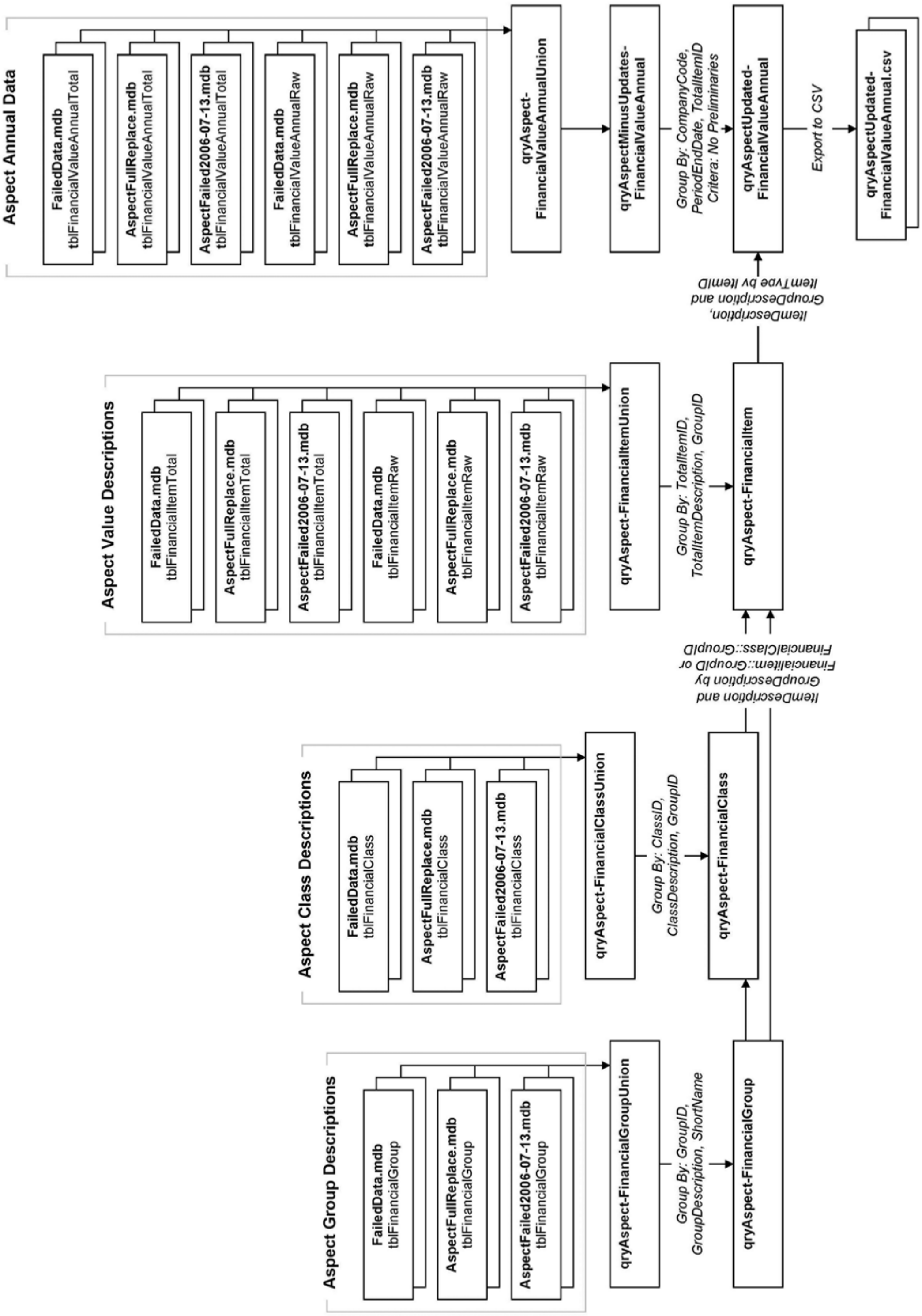


Figure 4-1 – Initial Pre-processing of Aspect Data

4.1.3 Definition of Failure

As this thesis will utilise a number of supervised learning techniques such as Neural Networks, it is necessary to include a single output column that indicates “failure”, and the calculation of that column requires an operationalised definition of failure. Beaver (1966) defined failure as bankruptcy, bond default, overdrawn bank account or non-payment of a preferred stock dividend, while Altman (1968) used “Chapter X of the National Bankruptcy Act”. While the definitions of failure vary, particularly within Australia where only individuals can become bankrupt, what holds true across previous research is that a firm that fails outside of a specific period of time is considered “non-failed” though there is little discussion about what that period of time should be. The research identified in sections 2.2.1, 2.3.1 and 2.4.1 sometimes use a “failed in 12 months” definition, sometimes classifies every company-year as failed if the firm ever fails, or sometimes uses a different failure horizon.

Argenti (1976) outlines how failure manifests itself in the years leading up the actual bankruptcy event, and that research highlights how symptoms of bankruptcy will begin to appear well before the actual bankruptcy event, something that this research intends to test. The definition of failure that is selected here will change the nature of supervised learning algorithms in that the system will optimise towards successful predictions of the chosen definition, so it is important not to select a failure horizon that is too short and will therefore be unable to perceive longer term symptoms. Likewise selecting a failure horizon that is too long will cause a learning algorithm to perceive a company that was healthy at that time to be treated as a company-year that exemplifies a company that will fail, which would reduce accuracy of the predictive model. This hypothesis is supported in many papers which examine the effect of modifying the failure horizon, for example in Deakin (1972). Because this thesis is, among other things, interested in examining the links between corporate failure theory and bankruptcy prediction research, it was decided to utilise a medium-term horizon of failure (a “failure event” within a 4-year period), this research can therefore avoid the unnecessary exclusion of long-term failure symptoms in the factor selection process. There is existing evidence that failure can be successfully identified within this time period (Beaver, 1966).

As with the research reviewed in section 2, it is also necessary to clearly define what constitutes a "failure event" within the dataset. Within a United States based dataset, bankruptcy filing is a commonly used definition. Within an Australian dataset, the close equivalent "external administration" can be used. While these are perfectly adequate definitions of failure, it must be acknowledged that this decision was primarily driven by the available data as will be examined in the following subsections. For example, the Compustat data source includes information for each company-year on whether that unique company is still active, and if not the reason and date for its inactive status. This provides an excellent source of failure data that does not require the combining of disparate data sources.

Note that the definition of failure within each dataset causes some inconsistencies between the United States and Australian cases. In the case of Australian data, external administration typically occurs when a company is no longer able to meet its debts as they fall due, and to trade a company while "insolvent" is a criminal act under section s588G(3) of the Corporations Act. Conversely, Chapter 11 bankruptcy protection in the United States allows management to continue trading while protecting the company from its creditors during restructuring, and is therefore typically used earlier in the failure cycle.

The Compustat financial data source includes failure data by way of the field "INCO", which contains a number as to why that particular company is no longer listed, and "INCOD" that contains the date that the company was delisted. Only some INCO codes are relevant to failure, however, so a brief analysis of the failure codes is presented here. Within the data source of 93,220 rows, 46,620 contain an inactive code representing just over 50% of the dataset. However, of those inactive codes, 36,560 company-years were delisted due to "Acquisition or merger", while only 1,880 were delisted due to "Bankruptcy", 1,140 due to "Liquidation", and 7,040 due to other reasons such as becoming a private company or otherwise not specified. While it is certain that some financially troubled companies are included in the "Acquisition or merger" set, there are certainly companies in this set that were not in financial difficulty and therefore this INCO code cannot be used in either failure or non-failure data. The reduced set therefore contains 3,020 company-years of failed companies, and 46,600 company-years of non-failed companies, which means that failed company-years constitute approximately 6.5% of the dataset.

By comparison, the Aspect dataset does not include any indication regarding whether a given company has been delisted or gone into administration, and it therefore becomes necessary to source such data elsewhere. To this end, the Australian Securities and Investments Commission (ASIC) maintain accurate records of both private and public companies entering administration, the date on which it occurred, and the Australian Company Number (ACN) used as the unique identifier. This dataset was purchased by Monash University and funded by the Australian Research Council's Australian Postgraduate Award Industry (APAI), project ID 0453884.

The ASIC data then needed to be cross-referenced to the resulting Aspect dataset. Both datasets contain the Australian Company Number (ACN), which can be cross referenced against the list of companies that have entered administration. This allowed the status code, the date the company entered this status, and the failure date to be included in the Aspect dataset as additional columns, however there were some companies in the Aspect dataset which did not include the ACN and had to be excluded from the dataset entirely. As a result of this process, the Aspect dataset contained 1,322 of failed company-years and 9,917 of non-failed representing approximately 11.8% of the sample as failed.

Once the failed and non-failed company-years had been identified within each dataset, the failed and non-failed sets were randomly divided into equal sets of in-sample training data, in-sample validation data, and out-of-sample validation data ("applied data"). While the in-sample training data is used to directly train the models, the in-sample validation data is used by many techniques (such as Genetic Programming) to measure the performance of a solution on data that the technique has not been directly exposed to, and thus determine whether that solution has successfully identified underlying relationships in the data. However, as solutions that perform well on the in-sample training data but poorly on the in-sample validation data are automatically discarded, the model is also being trained on the in-sample validation data (though indirectly). It therefore becomes necessary to use an out-of-sample validation set to independently measure the methodology's accuracy. Some techniques, such as the Neural Network method used in this chapter, use techniques that do not require an in-sample validation set, and in these cases the in-sample training and in-sample validation sets are combined, but the out-of-sample set is still reserved to independently measure accuracy.

There are two things worth highlighting regarding the division of these sets. The first is that the division of data into equal portions was done to ensure sufficient failure cases were present in each set to provide an accurate percent-hit and percent-miss result. The second is that when using time series data such as this it is generally accepted to use the most recent in the out-of-sample set (with either random or sequential attribution used only for training and testing sets). The implicit assumption in this practice is that the out-of-sample set is used to ensure that the predictive system is able to predict *future* cases from *past* data (in spite of the environment changing) – that is, that the resulting predictive accuracy is indicative of how the predictive system would perform on real out-of-sample data in a changing environment. While this will be tested in chapter 7 of this thesis, for now the changing environment (such as the lead up to the Global Financial Crisis) could impact on the results, so hence random sampling across time-series data is used.

4.2 Evaluation of Modelling Techniques

It is important to perform an analysis and selection of available modelling techniques that are appropriate to the datasets being used. As it is impossible to evaluate every possible modelling technique, this thesis will focus on modelling techniques that have been popularly or recently used in the literature review in section 2, which can be summarised as follows:

- Discriminant Analysis
- Linear Regression
- Neural Networks
- Genetic Algorithms
- Genetic Programming
- Support Vector Machines

Unlike the other techniques noted here, Genetic Algorithms are generally used as a component in a larger classifier algorithm, rather than acting as a standalone classification technique. Once exception is Varetto (1999), who predefined a Genetic Linear Score (GLS) as:

$$\mathrm{GLS} = a_0 + a_1 R_{h1} + a_2 R_{j2} + \cdots + a_n R_m$$

"in which a_0 indicates the constant, a_j the *i*th coefficient, R_{ki} the *k*th variable (ratio) of the *i*th family of variables", therefore building in the assumption that corporate failure can be effectively modelled by this arbitrarily selected linear classification algorithm. More commonly however are examples such as Anandarajan et al. (2001) who utilise Genetic Algorithms as the learning algorithm within a Neural Network solution, or Min et al. (2006) who predominantly use the Genetic Algorithm component to optimise the parameters of Support Vector Machines. As it is not commonly used as a standalone classification algorithm, but as a component in other classification algorithms, it has been excluded from this comparison.

To ensure a fair comparison, it was necessary to define ahead of time exactly how the data was to be presented to each technique. Firstly, a selection of factors is required, but until the modelling techniques have been chosen, the experiments which identify key factors that are outlined in section 4.3 cannot be performed, creating somewhat of a “catch 22”. Therefore this chapter will use the financial ratios from Altman (1968) for which data is consistently available as follows:

- Working capital to total assets
- Retained earnings to current assets
- Earnings before interest and taxes to current assets
- Sales to total assets

Another question that needs to be answered is whether the provided financial ratios should be normalised prior to being presented to each modelling technique. To assist in this decision, normalised with default parameters, normalised with optimised parameters and unnormalised data from the datasets was provided to all available techniques.

4.2.1 Data Normalisation

Some of the techniques in question, particularly statistically based methods such as Discriminant Analysis, are highly sensitive to data where the distribution is non-normal. For this reason, it is common to apply a number of normalisation techniques in which the goal is to transform the data to a normal distribution while minimising the loss of information. For example, linearly scaling data to values between 0 and 1 has the effect of preventing outliers from dominating the statistical functions used, but if most of the data is centred on the mean then the ability to differentiate between cases can be reduced. By comparison, logistic normalisation has the effect of maintaining a mostly linear relationship around the mean, but becoming more and more non-linear as the extremes of the data are reached. This has the effect of allowing a modelling system to differentiate between the majority of cases, but potentially hiding the size of outliers in the dataset.

For the purposes of an example, the ratio working capital to total assets from the Compustat database is presented in Figure 4-2 as a histogram, showing that almost all cases have values between zero and one, but outliers cause the histogram to require buckets that can span between -250 and 50, so for convenience a histogram with buckets showing only between -1 and 1 is presented as well.

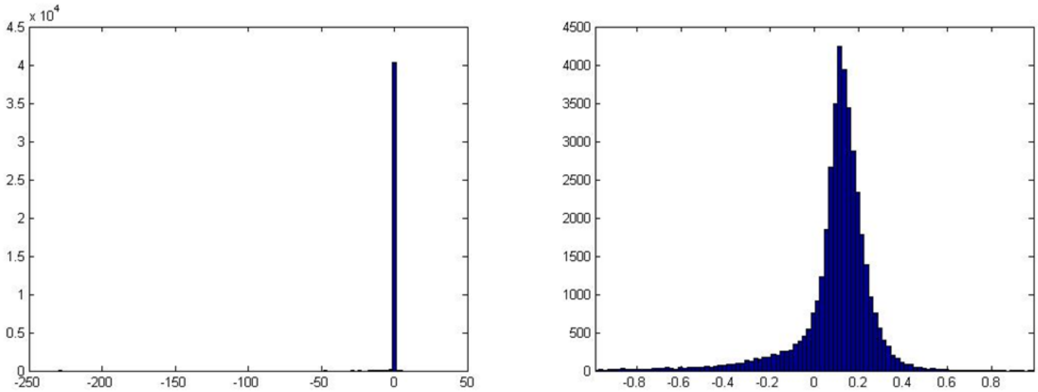


Figure 4-2 - Histograms of Working Capital to Total Assets

It can be seen that a normal curve is apparent, though outliers could potentially cause issues for modelling systems. The question then becomes one of deciding which normalisations to apply. The most commonly used normalisations with their respective algorithms are shown below, and these normalisations are supported by most data modelling implementations:

- Variance $\left(x_{norm} = \frac{x-\bar{x}}{\sigma}\right)$
- Range $\left(x_{norm} = \frac{x-x_{min}}{x_{max}-x_{min}}\right)$
- Logarithmic $\left(x_{norm} = ln(grad(x - offset) + 1)\right)$
- Logistic $\left(x_{norm} = \frac{1}{1+e^{-grad(x-offset)}}\right)$
- Sigmoidal $\left(x_{norm} = \tanh\left(grad(x - offset)\right)\right)$

The logarithmic function applies a mostly linear mapping around x_{min}, becoming less linear as values of x increase, which makes it an excellent choice for exponentially distributed values. The sigmoidal and logistic functions are mathematically similar, due to the fact that $\tanh x = \frac{e^x - e^{-x}}{e^x + e^{-x}}$, though when applied with a default gradient of 1 and an offset of 0 the logistic function maps non-linearly between 0 and 1 and has a higher resolution around outliers, while the sigmoidal function maps non-linearly between -1 and 1 and has a higher resolution around the mean.

In terms of which normalisations to apply, different implementations use different rules, and have different algorithms for determining the gradient and offset. For example the SOM Toolbox software exclusively uses default parameters (e.g. a gradient of 1 and an offset of 0), then leaves it to the user to decide which normalisation to apply to each input. By comparison the SOMine Viscovery software uses a gradient of $\frac{2}{s}$ for sigmoidal normalisations, where s is the standard deviation, and an offset of $\bar{x}$, where $\bar{x}$ is the mean (Viscovery Software GmbH, 2007, p. 65). In any case, the goal of the functions are to make the distribution normalised, and this can be measured by comparing the maximum distance between the Cumulative Distribution Function (CDF) of the normalised variable and the CDF of the standard normal distribution. In Figure 4-3, the variable working capital to total assets has been normalized using a sigmoidal function with a gradient of 1 and an offset of 0, and the resulting histogram and CDF (with a maximum distance of 0.16629) are shown with an overlay of the normal distribution (shown in red) for comparison purposes.

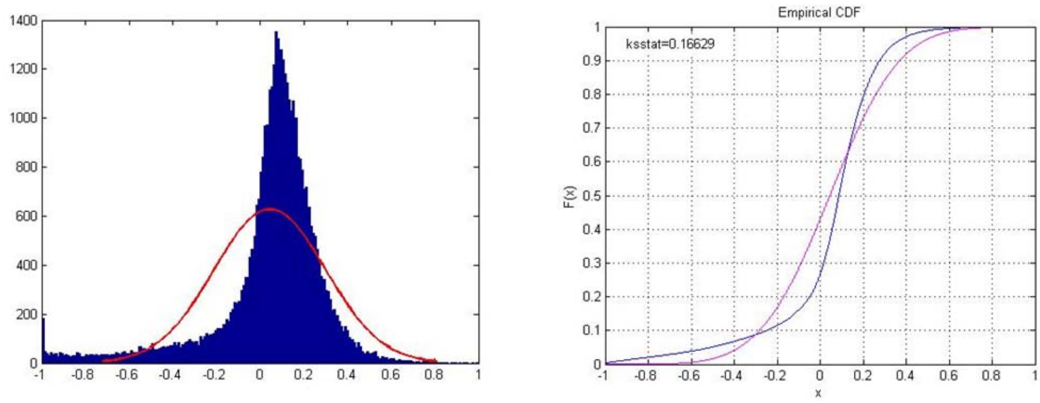


Figure 4-3 - Histogram and CDF of Sigmoidal Normalisation on Compustat Working Capital to Total Assets

Differing normalisation functions, offsets and gradients will have different maximum distances between the cumulative distribution function and the standard normal distribution, so this measure can be used to test functions and function parameters with the goal of minimising that distance. This process was performed on all variables used for this chapter, and the resulting normalisation function and parameters for each variable can be found in Appendix A. By way of example, the variable working capital to sales is used in Figure 4-4, this time using a logistic function with a gradient of 11.5279 and an offset of 0.0685, reducing the maximum distance between the CDF and the CDF of the normal distribution to 0.10443.

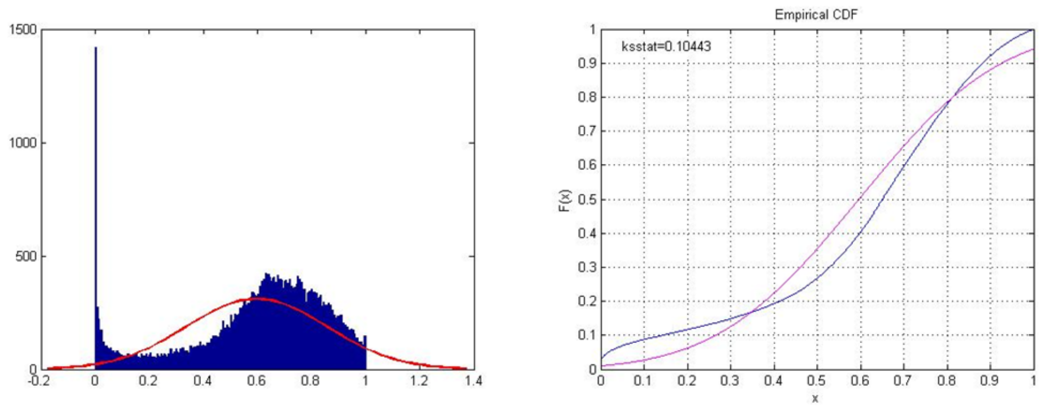


Figure 4-4 - Histogram and CDF of Logistic Normalisation on Compustat Working Capital to Total Assets

To test the effectiveness of the normalisation process, the variables are presented to each technique as a set of unnormalised, a set of normalised with default parameters, and a set of normalised with optimised parameters.

4.2.2 Discriminant Analysis

Discriminant Analysis is a methodology for determining a discriminant function for each class of training objects that can be used to classify additional objects that were not part of the original training set. Discriminant Analysis assumes that the observations are part of a multivariate normal distribution, which allows the formula $P(x|i) = \left(\frac{1}{(2\pi)^{\frac{n}{2}}|C_i|^{\frac{1}{2}}}\right)\exp\left(-\frac{1}{2}(x-\mu_i)^T C_i^{-1}(x-\mu_i)\right)$ for the probability of an observation given a class membership to be calculated based only on the mean and covariance matrices of each of the classes, where μ_i is the mean and C_i is the covariance of group i. In turn Bayes Theorem, $P(i|x) = \frac{P(x|i).P(i)}{\sum_j P(x|j).P(j)}$, can be used to calculate the inverse probability – that is the probability of class membership given an observation. The application of Bayes Theorem to the formula for a multivariate normal distribution results in the "quadratic discriminant function", and if it can be further assumed that the covariance matrices are equal, as is the case if the discrimination between classes is to be linear, then many of the terms in the quadratic discriminant function cancel out resulting in the "linear discriminant function". Whether the quadratic or linear discriminant function is used the function for each class is calculated on each unknown observation, with the largest class membership being used to identify which class the observation should belong to.

The resulting memberships for each observation in each class are often referred to as the parameters of the discriminant functions, and can be considered data points in an n-dimensional space for which each discriminant function is an axis. In this respect Discriminant Analysis is often said to be similar to Principal Component Analysis (PCA) as both methods transform the observations into a multi-dimensional space based on the differences between the observations. A detailed explanation of the implementation of Discriminant Analysis can be found in Fisher (2011).

Discriminant Analysis is a very fast and effective supervised algorithm for the classification of data points. Like any method, a number of well documented limitations exist, including the assumption that the observations are normally distributed. That being said, Discriminant Analysis has been used with great success even when the assumptions of normality are violated (Altman, 1968). On a more practical level, Discriminant Analysis has repeatedly been shown to perform more poorly than other techniques, most notably Neural Networks (Odom & Sharda, 1990; Coats & Fant, 1993; Wilson & Sharda, 1994; Lee, et al., 1996) though exceptions do exist (Yang, et al., 1999).

4.2.3 Logistic Regression

Logistic Regression is a classification method that uses "Maximum Likelihood Estimation" to determine the best set of parameters that linearly separates classes of observations. Logistic Regression assumes that a set of independent variables and their resulting dependant variable can be modelled probabilistically as a logistic function, that is: $P(y|x) = \sigma(w_0 + w_1x_1 + w_2x_2 + \ldots + w_nx_n)$ where $\sigma(\alpha) = \frac{1}{1+e^{-\alpha}}$. Maximum Likelihood Estimation is a method for determining the best set of parameters w given the set of observations x and known classifications y. More specifically, the probability that the data comes from any choice of w can be calculated, so maximum likelihood estimation seeks to find the maximum probability for the given data x, y for each possible w: $\arg\max_w P(x, y|w)$. However, in terms of maximising this probability, the observations x and their classifications y are fixed, so in fact Logistic Regression is seeking to maximise the likelihood function $\arg\max_w L(w|x, y)$. The likelihood function itself can be defined as the product of the function for each observation, but applying the natural log allows the product to be converted into a summation. Given that the natural log is a monotone transformation, it is therefore mathematically simpler to maximise the “log-likelihood” rather than the likelihood.

Unlike Linear Regression in which it is possible to solve analytically for the parameters that give the maximum likelihood, the use of the non-linear logistic function means that it must be solved algorithmically, “[Ordinary Least Squares] estimation is in this sense a special case of maximum likelihood estimation, one in which it is possible to calculate a solution directly without iteration” (Menard, 2002). In this respect, Logistic Regression is less computationally efficient than Linear Regression, but the Linear Regression Algorithm builds in the assumption that that the dependent variable is continuous. In many problems (including bankruptcy prediction) the dependent variable is nominal (failure versus non-failure). Furthermore Linear Regression includes issues such as predicted outcomes that are greater than one or less than zero (which in categorical data is inadmissible), an assumption that the variance is constant, and the assumption that prediction errors will be normally distribution. So while Logistic Regression is a very common methodology for building classificatory or predictive models, similarly to Discriminant Analysis is restricted to modelling linearly separable data which means that in complicated datasets, its performance is often poorer than other techniques (Lee, et al., 2005).

4.2.4 Artificial Neural Networks

Neural Networks attempt to model the structure of the biological brain by interconnecting artificial neurons (with links referred to as synapses) and modifying the strength of respective synapses to allow the model to “learn”.

In the biological brain, a neuron “fires” according to the type of neuron, it’s respective discharge pattern, and the strengths of the signals coming from connected neurons – so this is modelled computationally by applying a mathematical function, called an activation function, to the sum of input values. The mathematical function is the choice of the user, though the most common choice is a sigmoidal activation function because it ensures that the output of the neuron is between 0 and 1 and will never result in an undefined output as might be the case if a logarithmic function was used.

The interconnections between the neurons are weighted, and it is these weights that are modified to find a model that can effectively differentiate between observations in the training data. The methodology that is used to update the weightings of the synapses is referred to as the learning algorithm, and the learning algorithms can be roughly classified into two types: supervised and unsupervised. In supervised learning, the learning algorithm allows a naïve Neural Network to process an observation, and then identifies the error between the networks output and the actual output from some training data. The learning algorithm is then tasked with identifying the weights that most contributed to the error and adjusting them by a small amount, such that if that if the same observation was provided to the network the output error would be slightly smaller. In unsupervised learning, model's weights are adapted to reflect the training data distribution, and each neuron's output is calculated as the degree of similarity of its synapse structure to an input pattern. As observations are provided iteratively, the model learns to activate the same output neurons for similar inputs. The learning algorithms available will be outlined once the structure of the Neural Network model has been discussed.

The structure of the Neural Network changes the networks ability to model any underlying relationships between observations: If the structure of the network is inappropriate or too simple, the network will be unable to achieve good classification or good clustering (supervised and unsupervised respectively) on the observations provided to it. If the network is too complex, then the learning algorithm can "overfit" to the training data in which case the network has essentially memorised each individual case rather than learn any underlying relationships. This can be detected by testing the model on data that it has not been exposed to during training. Many structures of network are available to the user, including the most common supervised "feed-forward" structure in which there is an input neuron for each dimension of the observations, which feed into an optional layer of "hidden" neurons (the number of hidden neurons is the choice of the user), and these feed into the output layer which can be a single neuron for binary classifications or multiple neurons for multiclass classifications. Extensions and alternatives to the feed-forward structure abound, including structures in which neurons outputs feedback into earlier neurons, neurons whose synapses skip layers, structures that do not use layers at all, and so on. In many cases the structure also implies an activation function for how the inputs to a neuron are converted to outputs, for example the Radial Basis Function (RBF) Neural Network uses an input layer, a hidden layer and an output layer, but the hidden layer uses the nonlinear RBF.

An example of a typical feed-forward Neural Network is shown diagrammatically in Figure 4-5.

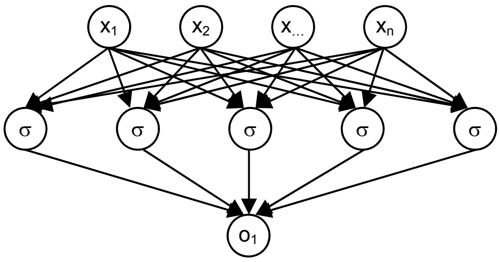


Figure 4-5 - Example of a Single Output Feedforward Neural Network

The choice of learning algorithm is another aspect of Neural Networks in which choices abound, with the most ubiquitous choice for a supervised network being the "backpropagation" algorithm (LeCun, 1985). The backpropagation algorithm works using a method called "steepest gradient descent" in which the gradient of the error versus the weight of a synapse can be calculated, and the weight of each synapse can be adjusted in the direction that most reduces the error. This algorithm of course has limitations, one of which is that the global error minimum cannot be found without a complete search of the error surface (network error versus the weight of each synapse), so the steepest gradient descent method can tend to find local minima, rather than global minimum, in the error surface. To address this, the backpropagation algorithm allows parameters and methods to be chosen to try and avoid the network converging on small local minima.

In an unsupervised network, a common choice of learning algorithm is that used by Self-Organising Maps (SOM), which is a kind of "Winner Takes Most" learning in which the winning output neuron's (and its neighbours) most similar synapses to the input pattern are rewarded by strengthening their weights to make them more connected to the winning neuron and its neighbours, while the least connected synapses are updated to make them less connected to the winning neuron and its neighbours. This can result in neurons that become entirely disconnected from the network if they do not assist in classifying any of the observations.

A somewhat hybridised supervised technique, used in Coats & Fant (1993), and the one used in this thesis when utilising Neural Networks, is the "Cascade-Correlation" algorithm (Fahlman & Lebiere, 1990) which starts with no neurons in the hidden layer and incrementally add neurons to find the optimum number of hidden neurons in an attempt to incorporate an element of "self-organising" into an otherwise supervised network.

In comparison to Logistic Regression, Neural Networks are far more flexible because they can model data that is not linearly separable, thanks to the use of the hidden layer of neurons. This increased flexibility comes with a decrease in computational efficiency, as will be seen in the methodology comparison in the conclusion of this chapter. Similar to Logistic Regression however is the predefined mathematical structure that is enforced on the eventual solution, for example the use of a sigmoidal activation function which may or may not be ideal for the problem at hand, leading to alternatives such as Gaussian activation functions.

4.2.5 Genetic Programming

Like Neural Networks, Genetic Programming (and Genetic Algorithms) model a biological process to achieve machine learning. Genetic Programming and Genetic Algorithms use the principles of Darwin's natural selection to pass genetic information from the "fittest" parents to children, who go on to be evaluated for fitness and in turn may become parents. Genetic Algorithms consider the "genes" to be a representation of a possible solution such as combinations of values for key financial ratios that cause bankruptcy. However Genetic Algorithms require the fitness algorithm to be pre-defined – information which is often not readily available. Genetic Programming on the other hand considers that the "genes" can be inputs as well as mathematical operations, constants or functions, and the fitness of the model is simply defined by how close the output of the model is to the known solution in the training data. In this respect, Genetic Programming does not require the fitness function to be known a priori, and also does not enforce a mathematical structure on the solution like Neural Networks, Logistic Regression or Discriminant Analysis.

Genetic Programming begins by randomly generating an initial population of solutions, which can be represented as trees of inputs, operations, constants or functions, then evaluating the error for each solution against each observation. From this population, the best performing programs are selected for reproduction. Parents reproduce to generate child solutions that have undergone "crossover", in which tree branches of the parents are combined or replaced in the children, plus some degree of random mutation. Each occurrence of this process is a "generation", and programs are free to evolve for as many generations as is necessary until the stopping criteria for a "run" is met. An example stopping criteria might be a certain number of generations, or a number of Generations Without Improvement (GWI). Runs are often performed a number of times from different random initial populations, after which the stopping criteria can be re-evaluated. Similarly to Neural Networks, the parameters of the Genetic Program model, such as the GWI stopping criteria, directly influence whether the model is sufficiently complex to find an adequate solution to the problem, or conversely allow the model to become overly complex and begin memorising cases.

An example of two parents resulting in two children is given in Figure 4-6, where the underlying training data is the values of a, b, and dependent variable c, from Pythagoras' Theorem, is used. In this example, one parent (top left) is a representation of the algorithm $\sqrt{\frac{a}{\frac{1}{b}}}$, another parent (top right) is the representation of the algorithm $b + a^2b^2$, and these two parents result in one child (bottom left) with $\frac{a}{b} + b$, and a second child (bottom right) with a small mutation that has resulted in $\sqrt{a^2 + b^2}$, therefore successfully discovering Pythagoras' algorithm.

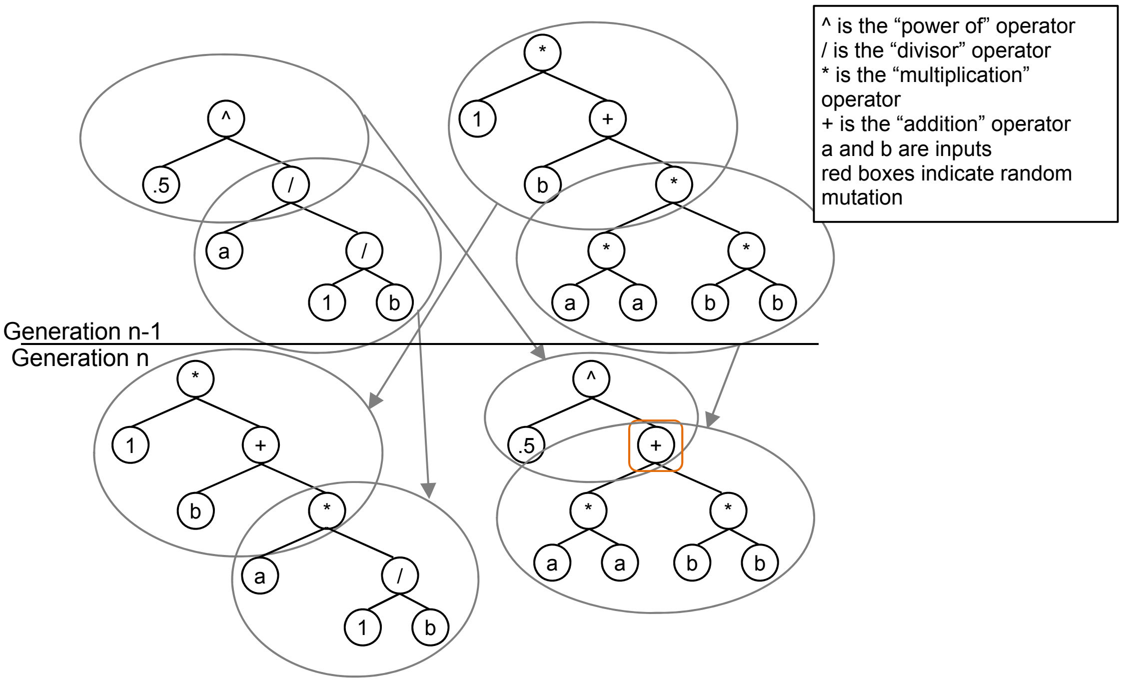


Figure 4-6 - Example of Genetic Programming Evolution to find Pythagoras' Theorem

Due to the fact that Genetic Programming is solving for the mathematical equation that gives the best fitness, the search space for Genetic Programming is exponentially larger meaning that it can take a much longer time for the stopping criteria and an adequate solution to be found. Since all mathematical solutions are within the search space, all of the solutions from other techniques such as Neural Networks, Logistic Regression or Discriminant Analysis could in theory be tested by the Genetic Program, however in practicality the search space is so large that the probability of testing any randomly selected mathematical structure approaches zero.

4.2.6 Support Vector Machines

Support Vector Machines (SVM), like Genetic Programming, is a comparatively new technique that can address a classification or regression problem. Support Vector Machines work by mapping non-linear classification problems into a higher linearly separable dimensionality using a “kernel function”, and then attempting to find the linear division between the data points that maximises the margins of that division (the maximum-margin hyperplane). In the simplest 2-dimensional linear problem with no higher dimensional mapping, SVM is simply seeking the widest line possible that accurately separates the data, as seen in Figure 4-7. The points which bound the margin of the division are referred to as the support vectors.

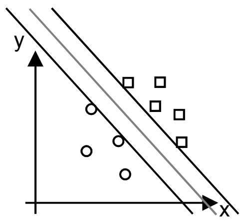


Figure 4-7 - Support Vector Machines Maximum-Margin Hyperplane

Finding the optimal hyperplane in an n-dimensional space, such as the above diagram, can be treated as a quadratic optimisation problem for which there are a number of well-known algorithms including the Interior Point Method (Mehrotra, 1992). However to classify points that are non-linearly separable it becomes necessary to map the n-dimensional space into a higher dimensionality using a “kernel function” (Aizerman, et al., 1964). There are many kernels available to users of Support Vector Machines including “Quadratic”, “Polynomial”, “Gaussian Radial Basis Function” and so on, with each having strengths and drawbacks, so ultimately it becomes a matter of choosing the most appropriate kernel for the task at hand (Scholkopf, et al., 1998), though work exists to automatically select the best kernel based on the features of the data (Ali & Smith-Miles, 2006). Support Vector Machines also require the user to specify parameters for “upper bound C and the bandwidth of the kernel function” (Shin, et al., 2005). Similarly to Neural Networks, the resulting classification solution from Support Vector Machines can be very difficult to interpret, in particular because the kernel function can be quite mathematically complex, however Support Vector Machines have been shown to provide good classificatory performance in bankruptcy prediction (Shin, et al., 2005).

4.2.7 Comparison of Modelling Techniques on Available Data

Perhaps the most telling aspect of the literature reviewed in chapter 2 is that there is no consensus on the best technique to use for a corporate failure classification problem. While various models have various strengths and benefits, such as the interpretability of the resulting classificatory equation or the time required to train the model, it is best to consider these in light of the performance of these techniques on the available data. As noted earlier in the chapter, the analysis cannot cover every possible classificatory system available, but the methodologies chosen in this comparison are popular in bankruptcy prediction literature. To ensure fairness, all techniques are given sufficient time to “converge”, that is the algorithm has either completed its computations (such as in Discriminant Analysis), or its continuation is no longer producing better results (such as in Genetic Programming). When comparing the results, the time required for the model to reach convergence can be compared. One thing noted in the literature review is that some techniques, for example Support Vector Machines, are highly sensitive to the parameters selected prior to training the model, so it is necessary to briefly discuss and justify the choice of parameters for each technique before comparing results.

As discussed previously, Discriminant Analysis builds in a number of assumptions about the data, including that the measurements for each class are distributed normally. In Linear Discriminant Analysis there is a further assumption that the covariance between classes is the same, but Quadratic Discriminant Analysis does not require this assumption to be met so it has been selected for this experiment. Quadratic Discriminant Analysis is also a very popular Discriminant Analysis technique for bankruptcy prediction, probably due to the relaxed assumptions being made on the data. Discriminant Analysis requires the covariance matrices to be estimated, and in situations where there are many dimensions available the covariance matrix may not be positive definite. For this reason it is often necessary to use a “stepwise” algorithm which is similar to the best-first forwards accuracy based search that will be carried out using the chosen modelling techniques in section 4.3, however the limited number of factors being considered in this section meant that such an algorithm was unnecessary. Similarly, the Logistic Regression algorithm is often found not to converge when exposed to many factors, but again the reduced dimensionality of the data in this section meant that a stepwise algorithm was not required and allowed for a fair comparison of techniques.

As discussed earlier in this chapter, the Neural Network model that will be used is the “Cascade-Correlation” algorithm because it eliminates the need to specify the number of hidden neurons a-priori. Instead, the Cascade-Correlation algorithm starts with a Neural Network model that contains no hidden neurons, and sequentially determines the optimum number of hidden neurons for the training dataset (Fine, 1999). Furthermore for this technique this thesis will utilise the “Jack-Knife” method for cross-validation, in which the observation the network has trained on is left out when calculating the accuracy of the network for that observation (Rojas, 1996), thus the combined in-sample training and in-sample validation sets are combined as discussed in section 4.1.3. The Neural Network algorithm was configured to use a maximum of 80 hidden neurons, as it was experimentally found that convergence of the model was reached within that limit.

For the Genetic Programming model used in this chapter it was opted to configure the algorithm to perform runs with a “Generations Without Improvement” (GWI) stopping criteria of 80. Once ten runs had been performed, the GWI was doubled, and this process was repeated until the model reached convergence. The GP fitness during training was calculated using squared error, using an instruction set that included addition, arithmetic, comparison, condition, transfer, division, multiplication, subtraction and trigonometric operators, using randomised constants and a population size within each run of 500.

For Support Vector Machines, it was decided to use the Radial Basis Function (RBF) kernel due to its proven effectiveness at classifying bankruptcy cases, such as in Shin (2005), using the implementation box constraint parameter default of “N/(2*N1) for the data points of group one and by N/(2*N2) for the data points of group two, where N1 is the number of elements in group one, N2 is the number of elements in group two, and N = N1 + N2” with a scaling parameter of 1 (The MathWorks, Inc, 2011).

A dataset of the sizes outlined in the previous section is unlikely to suffer from systematic bias as a result the random division of the dataset into in-sample training, validation, and out-of-sample. As noted in Zhu & Davidson (2007, p. 249), “given a succinctly large dataset, a simple train-and-test technique is a perfectly acceptable method for estimating the true error rate”. However, to ensure no bias is introduced at this early stage, this research performs 5-fold cross validation (Stone, 1974), providing the raw results in Appendix B and noting the mean results in this chapter.

Once the choice of methodologies and parameters to use had been established, it was possible to perform a classification using these methods on the dataset outlined earlier in this chapter. It was found that on the CompuStat dataset, the best mean in-sample validation accuracy of 77.2% was achieved using Neural Networks (normalised with optimised parameters), followed by Genetic Programming with 75.9% (normalised with optimised parameters). Of these results, Neural Networks appeared to experience over-fitting when using normalised with optimised parameters with accuracy dropping to 72.4%, however when using unnormalised data Neural

Networks achieved just 73.6% on the in-sample data, but increased out-of-sample accuracy to 72.6%. Furthermore, the normalised data improved out-of-sample accuracy by just 0.7%.

On the Aspect dataset, the best in-sample accuracy was also Neural Networks (normalised with optimised parameters) with 66.3%, then Genetic Programming (normalised with optimised parameters) at 66.2%. Again, the normalising of the data has had very little impact on the out-of-sample results, accuracy increased by just 0.2% and 1.8% using Genetic Programming and Neural Networks respectively.

To test significance, the McNemar's test was used on the first fold (Everitt, 1977) as this is the most appropriate statistical test when comparing classifiers in a large dataset (Salzberg, 1997). Comparing pair-wise accuracy between techniques, all results were found to be statically significant to $p<0.01$ with the exception of SVM (normalised with optimised factors) to NN (normalised with optimised factors) in the CompuStat dataset which was found to be significant to $p<0.05$.

One interesting development was the time required for the models to converge. Discriminant Analysis and Logistic Regression returned results in under one second, while the Neural Network model took approximately 20 seconds to complete as it tested each number of hidden neurons. While the Genetic Programing parameters selected gave it no predetermined stopping criteria, it was found on completion that the best performing models were discovered on run 5 in both datasets. On the CompuStat dataset 5 runs took less than 8 minutes while on the Aspect dataset 5 runs took less than 1 minute to complete. SVM took a similar amount of time to complete training for any given kernel and parameter set.

Therefore it was decided that due to the consistently good accuracy of both Neural Networks and Genetic Programming, the adequate speed of convergence of both techniques, the popularity of Neural Networks and the simplicity of the resulting predictive function in Genetic Programming, that the experiments carried out in this thesis over the following chapters would use these two techniques. That is not to say that the other techniques are inferior, just that it is critical when examining the predictive capabilities of a dataset to choose an appropriate tool not only based on the accuracy of the model but also given the usefulness of the final output and the processing time required to achieve that output. Furthermore by selecting these two techniques, both a deterministic and a non-deterministic modelling technique have been selected, and the non-deterministic behaviour of Genetic Programming makes it an ideal choice for identifying low-contributory factors with the goal of classification accuracy (as addressed in section 4.3.5). When looking exclusively at these two techniques, it was found that while the best in-sample accuracy is achieved when using normalisation, in particular normalization with optimised parameters, out-of-sample results were generally comparable. For example on the CompuStat dataset when using Neural Networks, out-of-sample accuracy decreased from 73.3% to 72.6%, while in the Genetic Programming environment on the Aspect Dataset, out-of-sample accuracy increased from 62.8% to 63.0%. Other techniques, such as Discriminant Analysis, benefited greatly from normalisation, with in-sample accuracy increasing from 53.0%, only marginally higher than a weighted naïve model, to 66.2%, with out-of-sample accuracy increasing from 53.8% to 63.4%. However, unnormalised data has the benefit of being more readily interpretable and this will make analysis in section 7.2 notably easier. Thus, it was decided that Genetic Programming and Neural Network models would be training and validated using unnormalised data in the following sections and chapters.

Having now evaluated a number of available modelling techniques, this chapter can conclude by using the selected techniques to perform factor-selection, before the following chapters can perform an in-depth analysis on the effects of share market data, macroeconomic data, and the effect of objective clustering on accuracy using these techniques.

4.3 Identifying Initial Key Variables

For the purpose of demonstrating a new application of multiple discriminate analysis (MDA), Altman et al. (1968) used an array of 22 ratios selected by popularity, potential relevancy, "and a few 'new' ratios" – from this, 5 ratios were selected based on predictive performance for their dataset (working capital to total assets, retained earnings to total assets, earnings before interest and taxes to total assets, market value equity to book value of total debt, sales to total assets). Meanwhile Beaver (1968) used 30 factors and required "that the ratio be defined in terms of a 'cash flow' concept". By comparison, Edmister (1972) examined 10 ratios "found to be significant predictors of business failure in previous empirical research", finding that many ratios adopted by one study are not necessarily adopted in others.

The motivation for this chapter is therefore focussed on addressing the factor selection process and identifying whether an optimal set of factors is specific to a dataset, and whether an objective factor selection methodology can be used to increase classification accuracy.

4.3.1 The Cost of Information

While all financial variables contain information, the inclusion of additional information in a classification system has the effect of exponentially increasing the size of the error surface – in turn this reduces how much of the error surface can be adequately searched in a given timeframe, dramatically increases the timeframe required to search it or decreases the probability that the outcome will be near the global error minimum. In many cases, a dataset with too many dimensions will simply take too long to find a reasonable solution or cannot be solved for using statistical methods such as Discriminant Analysis. Thus, the inclusion of information comes at a cost. Furthermore, due to the relationships between financial ratios, there is some degree of overlap in the information provided by any two ratios. Therefore there can be diminishing informational returns with the addition of each subsequent financial variable to the classification system. It therefore becomes useful to objectively minimise the number of factors while maximising the informational gain with each input.

4.3.2 Background on Factor Selection Strategies

Liu & Motoda (1998) discusses three principal dimensions of feature selection. The first being an evaluation measure, the second being a search strategy and the third being a generation scheme. Each dimension has a number of alternatives within it, depicted in Figure 4-8:

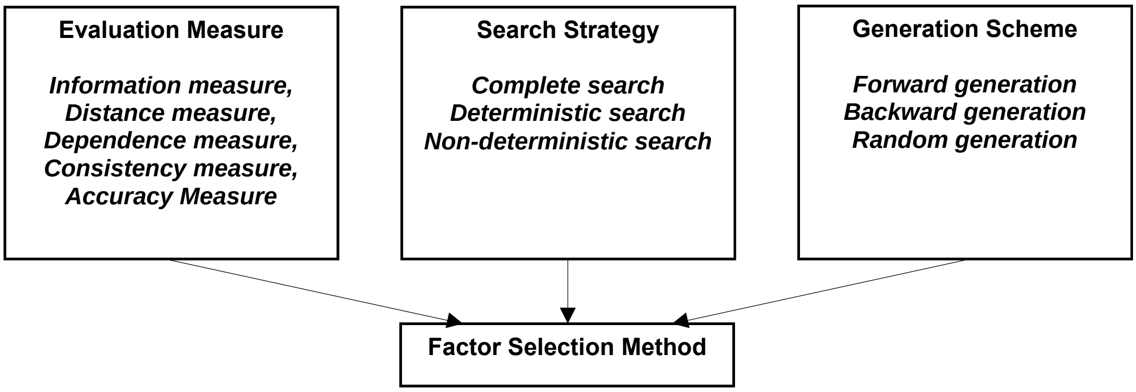


Figure 4-8 - Dimensions of Factor Selection

Information measures, the first evaluation measure, seek to calculate how much information is being added to the situation with the introduction of the factor in question. Liu outlines an information measure by calculating the difference between the prior uncertainty and the expected posterior uncertainty to find the information gain (IG) of factor X, $IG(X) = \sum_i U\big(P(c_i)\big) - E\big[\sum_i U\big(P(c_i|X)\big)\big]$, where U is an uncertainty function and $P(c_i)$ is the prior class probability for $i = 1,2,\ldots,d$. Liu also outlines a commonly used uncertainty function $-\sum_i P(c^i)\log_2 P(c_i)$.

The second evaluation measure, distance measures, are “derived from distances between the class-conditional density functions” (Liu & Motoda, 1998). That is, a feature is of higher value if it separates two classes more easily.

The third kind, dependence measures, examine associations or correlations – which while similar to an information measure, focus instead on the ability to predict one measure given another over a posterior class probability.

Fourthly, consistency measures seek to minimise the number of factors while maintaining consistent class separation as when a full set of features is being used, that is $P(C|\text{FullSet}) = P(C|\text{Subset})$.

Finally, accuracy measures rely on selecting a subset whereby the most accurate class membership can be obtained on sample testing, and therefore can only be used in a supervised manner.

Interested readers are directed to Liu & Motoda (1998) for a more detailed explanation.

Having outlined the available evaluation measure, it now becomes necessary to examine search strategies. The first possible search strategy is a complete search of all possibilities, and this strategy can be performed depth-first or breadth-first. In a case with just two factors, a forward depth-first search would try feature *a*, then features *a* with *b* together, before looking at feature *b* alone; whereas breadth-first examines *a* first, then *b*, before moving on to *a* with *b* together.

The second possible search strategy is a deterministic heuristic search, an example of which is a best-first (or “greedy”) strategy. The best-first strategy evaluates each feature in a breadth-first manner, and the highest-ranking feature is selected to have additional features added to it. Each combination at that level is evaluated, following the path of best sample accuracy. The limitation of this strategy is that the optimum combination of features may not be tested if it is not on the same path as that which is taken by the search strategy, which will take the most gain at each step without looking ahead. Liu & Motoda (1998) outline other deterministic heuristic search strategies that are far too numerous to comprehensively evaluate within the scope of this thesis, but they include popular techniques such as a Beam Search. One of the largest limitations of these search strategies is that interdependence between features can be missed, as each feature individually is worth little but the inclusion of both may be worth much (Liu & Motoda, 1998). On the other hand, a deterministic heuristic search enforces an a priori structure that can massively decrease processing cost while maintaining a high-probability that useful combinations are tested.

The third search strategy is a non-deterministic search that is designed to overcome the deterministic search interdependence limitation by introducing an element of randomness into the search function. For example, Genetic Algorithms could be used, where chromosomes with possible mutations represent the inclusion or exclusion of a given feature. Genetic Programming is another example of a non-deterministic search strategy, whereby well performing predictive algorithms become the parents of new algorithms with mutation built directly. A more detailed explanation of Genetic Programming can be found in section 4.2.5.

Whatever search strategy is selected, the underlying problem is finding an optimal feature subset in a large, hyper-dimensional search space that is defined by the choice of evaluation measure.

The last dimension is the factor selection methodology is the generation scheme. Factor sets can be produced forward, backwards, or randomly. Where a forward generation scheme starts with an empty set and adds features for evaluation, a backwards generation scheme starts with a complete set and removes features. In the case of a non-deterministic search strategy, the generation scheme must be random to avoid becoming trapped in local mimima within the search space (Liu & Motoda, 1998).

Of course it is not necessary to limit a search methodology to any single one of the strategies outlined above, and Guyon, et al. (2006) provide a more comprehensive review of the available strategies. As will be demonstrated in the following section, there is benefit to building a hybridised search strategy that combines elements of non-deterministic search strategies with a heuristic search of factors that are shown to perform well.

4.3.3 Initial Population of Factors

The first stage of the variable selection process is to identify an initial population of candidate variables. Research that uses statistical or algorithmic methodologies to analyse data often deals with massive datasets, and in many of these cases it is not practical to approach the dataset holistically. Too many factors cause some techniques to become prone to over-fitting, exponentially increase the processing time or reduce accuracy, and such models also reduce the explainability of the resulting model. “By choosing a minimal subset of features, irrelevant and redundant features are removed according to the criterion. When [the data dimensionality] is reduced, the data space shrinks and in a sense, the data set is now a better representative of the whole data population” (Liu & Motoda, 1998).

The question therefore becomes, “Which factors should be selected?” As this research focuses on the ability to classify failed companies, the simplest answer to this question is “the factors that give the best results” – and indeed this is perhaps the most important consideration, but the testing of every possible combination of inputs is simply not possible. For example, in the case of corporate failure, it is not unreasonable to have access to some 200 pieces of raw data per company per year, for example the Compustat data source used for this research (see section 4.1.1) contained 189 raw inputs. One of those raw pieces of data, net profit for example, can be used as part of a financial ratio: net profit over total expenses. In turn, the resulting ratio may be of more use than either piece of raw data alone, as the ratio now describes a relative difference that can be compared between two companies of vastly different net profit. To ensure a complete set of data, a researcher could take each of the 200 or so available variables and use them as the numerator, then for each numerator use the 199 remaining variables as the denominator. It is therefore possible to calculate every 2-value ratio available given the dataset. Even excluding reciprocals (net profit over total expenses excludes the reciprocal total expenses over net profit), this results in 20,100 ratios, and excludes potentially useful 3-value ratios (such as net profit less tax expense over total expenses).

Furthermore, the use of classification accuracy in the selection of factors would require each of those ratios to be tested by including them in a classification model and calculating the resulting accuracy. To ensure the optimal combinatorial selection of those 20,100 ratios, a classification system would need to be exposed to 5.045×10^{6050} combinations of those 20,100 ratios, calculated through the binomial coefficient. Given the enormous number of possibilities, it is likely that factors are found which are highly predictive of the in-sample set by chance, and do not reflect any underlying relationships in the data resulting in poor out-of-sample accuracy. Therefore it is worthwhile including only variables for which there is at least some theoretical relationship to corporate failure. This of course comes at a cost. It is entirely possible that there is a sound theoretical basis for including one of the many 2 or 3-variable ratios in the final model that is excluded due to the subjective process of ratio selection. This cost increases as the number of ratios considered decreases, highlighting the need to examine as many factors as is reasonably possible.

In an attempt to overcome this weakness, a wide array of ratios were considered. Rather than selecting a single piece of research and utilising just those variables, such as the ever popular Beaver (1966) or Altman (1968), this research instead considered all variables that were used across many notable publications. An initial selection of corporate failure prediction research was selected by following the path of citations from recent publications back to the initial work of Beaver (1966) and Altman (1968). From this selection, each paper was considered from oldest to newest and each paper that utilised a new factor was added to the population. As factors were added to the sample in chronological order, each new paper contributed fewer new factors to the population. While papers after Wang (2004) were examined, they did not promote factors that had not already been added to the sample. The following papers contributed new factors to the population:

- Beaver, 1968
- Altman et al., 1968
- Edmister, 1972
- Deakin, 1972
- Wilcox, 1973
- Blum, 1974
- Elam, 1975
- Altman et al., 1977
- Ohlson, 1980
- Chen, S, 1981
- Barniv & Raveh, 1989
- Raghupathi et al. 1991
- Coats, & Fant. 1993
- Altman, et al., 1994
- Wilson & Sharda, 1994
- Lee, et al., 1996
- Yang, 1999
- Kane, et al., 1998
- Dimitras, et al. 1999
- Varetto, 1999
- McKee, 2000
- Anandarajan, et al., 2001
- McKee & Lensberg, 2002
- Wang, 2004

From the resulting selection of variables outlined in these papers, factors for which no data was available were removed. Within the Compustat dataset, this resulted in 77 inputs as a combination of raw inputs, 2-variable and 3-variable ratios while in the Aspect dataset this resulted in 79 factors to consider. These factors are listed in Appendix C and Appendix D.

4.3.4 Removal of Missing Data and Dataset Creation

Missing values then had to be removed from the dataset. This could of course be achieved by removing all factors with missing data, but there were very few factors with 100% data availability. Alternatively this could be achieved by removing all cases with missing data, but again there were very few individual cases with no missing values. Rather, most cases and most factors had a high but not entirely complete set of data.

While this research has opted to use deletion as a method of dealing with missing data, there are many alternatives that could have been used that would have resulted in higher data availability. From the simplest mean substitution methods through to highly mathematical methods such as multiple imputation (Rubin, 1987), an analysis of the different methods of dealing with missing data and their resulting predictive accuracy on the modelled data would be an interesting complement to this research. Such an analysis, however, is beyond the goals of this thesis.

It could be argued that mean substitution might be the most appropriate method for the problem of corporate failure classification, and this may be true, however the later chapters of this thesis analyse in depth the relationships between variables, the resulting failure classification, and the eventual actual failure or non-failure of the company. Mean substitution, in these cases, would likely create scenarios in which a company is correctly (or incorrectly) classified, and the justification for this decision comes down to a variable that did not exist in the original dataset. It is much more difficult to make a case for the explainability of a model if the justification includes data that was non-existent in the original data. Therefore, deletion has been used to ensure that such methods do not compromise this eventual analysis.

It is therefore a matter of achieving the best balance of factors versus available data. This was achieved by plotting the number of factors versus the number of remaining company-years that met the definition of failure outlined in 4.1.3, and then choosing the number of factors that would maintain at least 85% of the failure data. The plots for the Compustat and Aspect dataset are shown in Figure 4-9 and Figure 4-10.

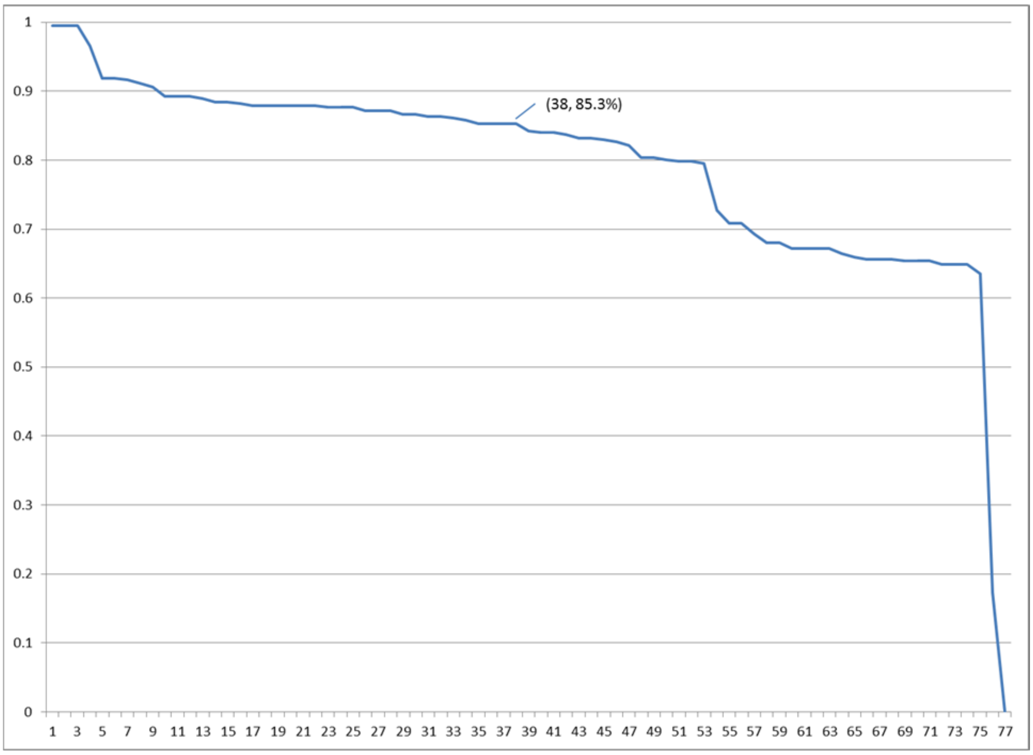


Figure 4-9 - Factors versus Availability of Compustat Failed Data

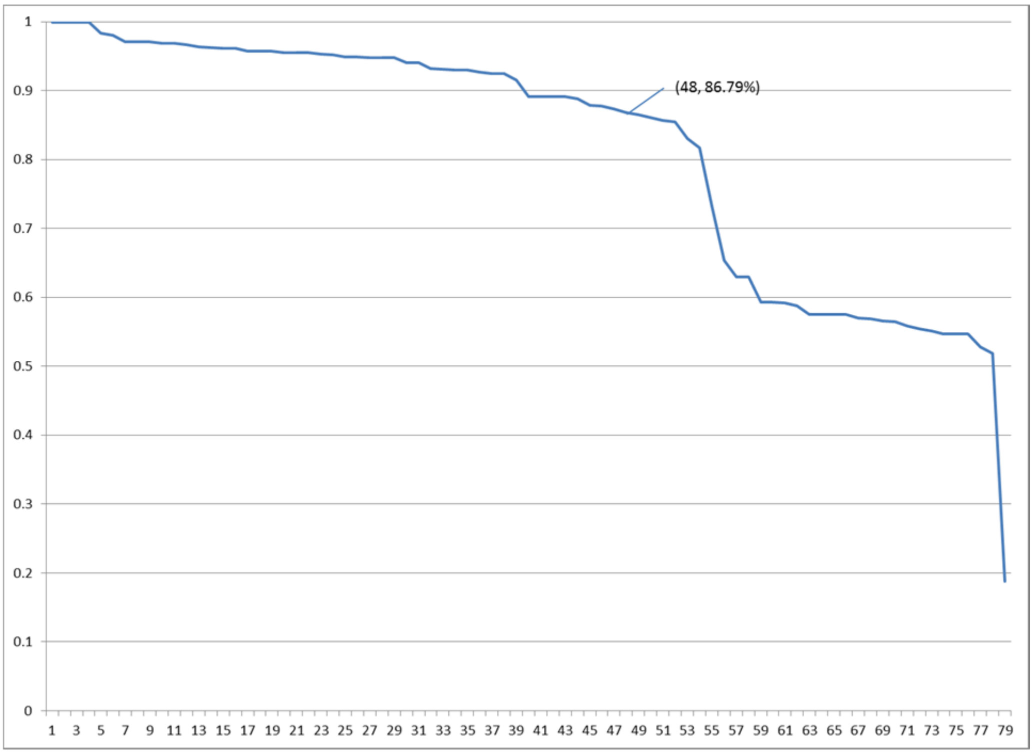


Figure 4-10 - Factors versus Availability of Aspect Failed Data

Within the Compustat dataset, this process resulted in 38 remaining factors, with 293 failed company-years and 46,160 non-failed company years. Within the Aspect dataset, this process resulted in 48 remaining factors with 351 failed company-years and 4,884 non-failed company-years, which were then divided into an in-sample training set, an in-sample validation set and an out-of-sample validation set, as per section 4.1.3.

4.3.5 Removal of Low-Contributory Factors

By selecting factors from a wide range of previous research into corporate failure prediction, and further limiting the choice of factors as a means of dealing with missing data, the number of factors to consider has shrunk from the hypothetical 20,100 possible factors to just 38 and 48 on the Compustat and Aspect datasets respectively. Even still a best-first heuristic search, as discussed in the following section, on a 38 factor set requires 741 combinations to be tested – and on a 48 factor set the number of combinations grows to 1,176. While not necessarily prohibitive, depending on the time required for a classification model to converge, there was an opportunity to use a non-deterministic search strategy to estimate which factors contribute the “most” to successful bankruptcy prediction to further minimise the combinations to test, while minimising the likelihood of removing factors that contain useful information.

Given that the ultimate goal for this chapter is the identification of factors for bankruptcy prediction, the best choice of evaluation measure is one of classification accuracy. In turn this raises the question of how accuracy should be determined. In line with the work of Liu & Motoda (1998), a non-deterministic embedded based method is an ideal choice, and such properties are found in one of the modelling techniques selected in chapter 4.2, Genetic Programming (GP).

The strength of the Genetic Programming modelling algorithm for this problem lies in its random generation of an initial population for each run, and the use of randomly selected variables within each program. While it is almost certain that factors which contain useful information will be randomly excluded due to poor combinations of operators, constants and other factors, each factor can be trialled in hundreds of thousands of potential combinations. It was therefore necessary to ensure that the Genetic Programming algorithm was configured with sufficiently conservative stopping criteria such that it is unlikely well-performing factors are excluded.

Similarly to the parameters used in section 4.2.7, this section used a Genetic Programming algorithm for each dataset performing 10 runs with stopping criteria of 80 Generations Without Improvement (GWI). Once these 10 runs had been completed, the GWI was doubled and another 10 runs performed. This process was repeated until the model had converged, such that doubling the GWI was no longer having any effect on the accuracy of the in-sample validation training set. This particular method for determining convergence is very resource inefficient, for example the runs after the GWI has been doubled 5 times take 32 times as long, but it was important to minimise the likelihood of excluding a factor due to a poor choice of stopping criteria.

Once the model reached convergence it was necessary to determine which factors had contributed to the winning programs and which ones had been excluded. To that end, the “average input impact” was calculated across the winning programs, which attributes a score to factors that relates to how well they have contributed towards the model. From this, any inputs with an average input impact of 0.00 were excluded as they have essentially been made “extinct” by the evolutionary process. The full results are available in Appendix E and Appendix F. The Compustat dataset converged after 125 minutes and 516,687,603 programs had been evaluated, with the algorithm stopping after 10,240 generations without improvement. Of the winning programs, 77% were found within 80 GWI, and 90% were found within 160 GWI. The factors that survived the evolutionary process are compared with the factors found in (Beaver, 1966) and (Altman, 1968) in Table 4-1. The time to reach convergence appears high and in turn raises questions about whether the low-efficiency nature of a Genetic Program is the best choice. However, as demonstrated by the fact that the vast majority of winning programs were identified within the first tens or hundreds of generations (as opposed to the tens of thousands that were run), the low efficiency of Genetic Programming in this case is a product of the criteria used for “convergence”. Hindsight shows that a much more aggressive convergence criteria could have been used without adversely affecting the results.

Factor	Survived	Beaver (1966)	Altman (1968)
current plus long-term liabilities to total assets	✓	✓	
total liabilities	✓		
log of total assets			
cash to total assets		✓	
net income to total assets	✓	✓	
Net income	✓		
net income to net worth	✓	✓	
cash to fund expenditures for operations	✓	✓	
book value to total liabilities			
cash flow to total assets	✓	✓	
cash flow to total liabilities			
log tangible assets			
cash flow to net worth		✓	
current liabilities to total assets	✓	✓	
long-term liabilities to total assets		✓	
current liabilities to equity			
net income to sales	✓	✓	
total assets to sales	✓	✓	
earnings before interest and taxes to total assets			✓
sales to total assets	✓		✓
earnings before taxes to sales	✓		
net operating profit to sales	✓		
net worth to sales		✓	
earnings before taxes to equity			
sales to net worth	✓		
cash to current liabilities		✓	
cash to sales	✓	✓	
sales to cash			
current assets to total assets		✓	
long-term assets	✓		
cash flow to sales		✓	
fixed assets to equity			
non-cash current assets			
retained earnings to total assets			✓
working capital to total assets		✓	✓
current assets to current liabilities		✓	
long-term liabilities to current assets			
current liabilities to current assets	✓		

Table 4-1 - Survivors of Evolutionary Process in CompuStat Dataset

By comparison, the Aspect dataset converged after 153 minutes and 1,499,777,476 programs had been evaluated, with the algorithm stopping after 40,960 generations without improvement. 10% of programs had reached convergence within 80 GWI and 23% of programs had reached convergence within 160 GWI, with the results shown in Table 4-2.

Factor	Survived	Beaver (1966)	Altman (1968)
cash flow to sales		✓	
cash flow to total assets		✓	
cash flow to net worth		✓	
net income to sales		✓	
net income to total assets	✓	✓	
net income to net worth		✓	
current liabilities to total assets	✓	✓	
long-term liabilities to total assets		✓	
current plus long-term liabilities to total assets	✓	✓	
cash to total assets	✓	✓	
current assets to total assets		✓	
working capital to total assets		✓	✓
cash to current liabilities	✓	✓	
current assets to current liabilities	✓	✓	
accounts receivable to sales		✓	
net worth to sales		✓	
total assets to sales		✓	
cash to fund expenditures for operations	✓	✓	
defensive assets to fund expenditures for operations		✓	
defensive assets minus current liabilities to fund expenditures for operations		✓	
retained earnings to total assets	✓		✓
earnings before interest and taxes to total assets	✓		✓
sales to total assets.			✓
fixed assets to equity			
cash flow to current liabilities			
current liabilities to equity			
fixed assets to sales			
equity to sales	✓		
earnings before taxes to sales			
earnings before taxes to equity			
cash to total assets	✓	✓	
Net income			
non-cash current assets			
long-term assets			
total liabilities.			
cash flow to total liabilities	✓		
book value to total liabilities			
sales to working capital			
sales to cash	✓		
net operating profit to sales	✓		
sales to net worth	✓		
net operating profit to total assets	✓		
long-term liabilities to current assets			
log tangible assets			
current liabilities to current assets			
funds provided by operations to total liabilities.	✓		
log of total assets			
income from operations to total assets			

Table 4-2 - Survivors of Evolutionary Process in Aspect Dataset

At this point some preliminary comparisons can be drawn, as the factors net income to total assets, current liabilities to total assets, current plus long-term liabilities to total assets, cash to fund expenditures for operations, net operating profit to sales, and sales to net worth survived in both the Compustat and the Aspect evolutionary models. Interestingly 2 of those 6 ratios, net operating profit to sales and sales to net worth, are not found in the usual Beaver (1966) or Altman (1968) factors, which demonstrates that there was useful information that was being excluded in chapter 4.2 which utilised only factors from Beaver (1966) and Altman (1968).

4.3.6 Best-First Forwards Search

Having reduced the search space, alternatives to the non-deterministic search strategy may be employed. The 17 Compustat factors that survived the evolutionary process yield 131,071 possible combinations of those factors as inputs to a classification model, so even with a very limited subset of factors some kind of heuristic search was required. There are many search algorithms that could be used to address this problem, such as a Beam Search (Liu & Motoda, 1998), which is an optimisation of a best-first forwards search, but to maintain consistency with forward stepwise Discriminant Analysis and forward stepwise Logistic Regression used in 4.2.7 this chapter will use an accuracy based, breadth-first heuristic search strategy, commonly referred to as a “greedy search” which resulted in 153 combinations of factors to test for the 17 factor scenario.

Where a forwards approach adds the best performing factor at each step, a backwards worst-first approach could trial the removal of each factor from the complete set. However Genetic Programming systems often find the best performing offspring do not necessarily use all inputs made available to it, meaning that removing one of the many unused variables from a complete set will not change the resulting classification accuracy. As this makes it difficult to choose between two factor choices, it was best in this case to use a forward generation scheme.

Unlike the previous section which required the non-deterministic qualities of Genetic Programming, this section has no such requirement so each dataset can be examined using both the Genetic Programming and Neural Network algorithms outlined in section 4.2.7

For the Genetic Programming-based forward search, each combination in the best-first strategy of the selected variables from the previous section will be run through a Genetic Programming environment with consistent parameters – this time for 15 runs of 150 generations without improvement. By hard limiting the number of runs each combination of variables can undergo, combinations with a large number of variables are penalised due to the inability to spend enormous amounts of time experimenting with any one given combination, thus driving the system towards finding the smallest number of useful inputs. The parameters of 15 runs for 150 GWI were selected because the results from section 4.3.5 indicated that GP models with much higher dimensionality were able to reach convergence within as little as 10 runs of 80 GWI and 10 runs of 160 GWI, and it was important to ensure the model was not prematurely stopped due to overly constrained stopping criteria. Similarly to the Neural Network algorithm used in 4.2.7 this chapter will use a Cascade Correlation Neural Network with up to 80 hidden neurons. On the Compustat dataset the Genetic Programming algorithm using these stopping criteria took approximately 9 minutes per combination to complete, while the Neural Network algorithm took approximately 2 minutes per combination.

For the Genetic Programming environment, each factor was chosen based on the resulting best program's weighted hit-rate on in-sample validation dataset, whereas for the Neural Network environment the Jack-Knife method meant that the in-sample training and in-sample validation sets had been combined, so the best factor was selected based on the accuracy of the combined in-sample training and validation datasets. In either case, the out-of-sample data was not exposed to any of the models during the factor search.

4.3.7 Diagram of Factor Selection Model

The methodology outlined in the previous sections are summarised in Figure 4-11:

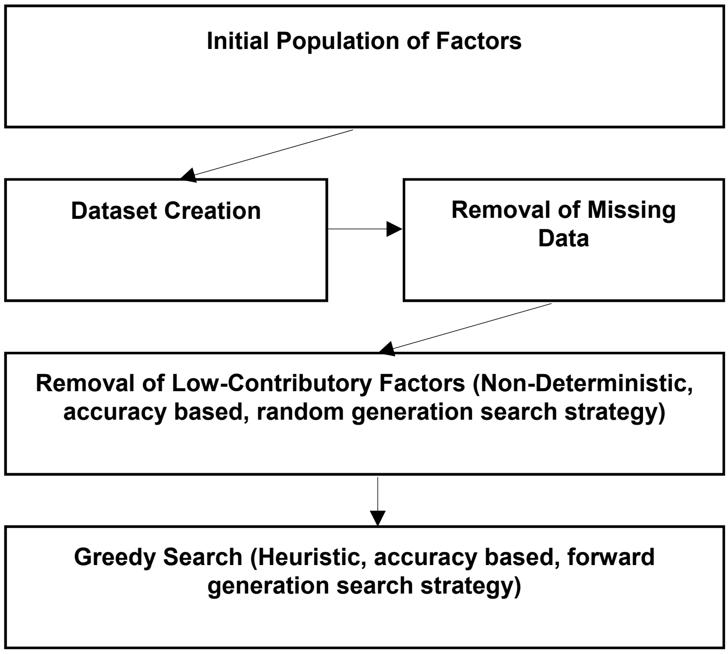


Figure 4-11 - Diagram of Factor Selection Model

4.3.8 Results & Discussion

For the Compustat dataset, the Genetic Programming algorithm found the best results on the in-sample validation dataset once the 5th factor had been added to the best-first search, yielding a validation accuracy of 76.4%. When tested on the out-of-sample dataset, the accuracy was reduced to 73.8%, indicating that some over fitting has occurred. By comparison, when all surviving factors from section 4.3.5 are included, validation accuracy drops to 75.1% and out-of-sample accuracy drops to 71.4% ($p<0.01$). When using only the factors from Beaver (1966) or only the factors from Altman (1968), accuracy drops to 75.7% and 72.6% on the in-sample set with 73.8% and 72.1% on the out-of-sample ($p<0.01$). When all available factors are used the model achieved 74.5% on the in-sample and 71.3% on the out-of-sample ($p<0.01$).

Using Neural Networks on the other hand found the best in-sample accuracy once the 13th factor had been added, with a combined training and validation accuracy of 75.1% and an out-of-sample accuracy of 70.4%. By comparison when all surviving factors from section 4.3.5 are included the accuracy of the in-sample set drops slightly to 75.0% and the out-of-sample accuracy drops just slightly to 70.1% ($p<0.01$). When using only the factors from Beaver (1966) or only the factors from Altman (1968), accuracy drops to 72.2% and 73.1% on the in-sample set with 69.8% and 67.1% on the out-of-sample ($p<0.01$) and when using all available factors, accuracy is approximately equal to that of from Beaver (1966) factors or Altman (1968) factors, achieving 74.0% with 69.8% on the out-of-sample ($p<0.01$).

A table of the full results can be found in Appendix G.

A plot of validation accuracy versus number of factors for each method is shown in Figure 4-12, and can be found numerically in Appendix H and Appendix I.

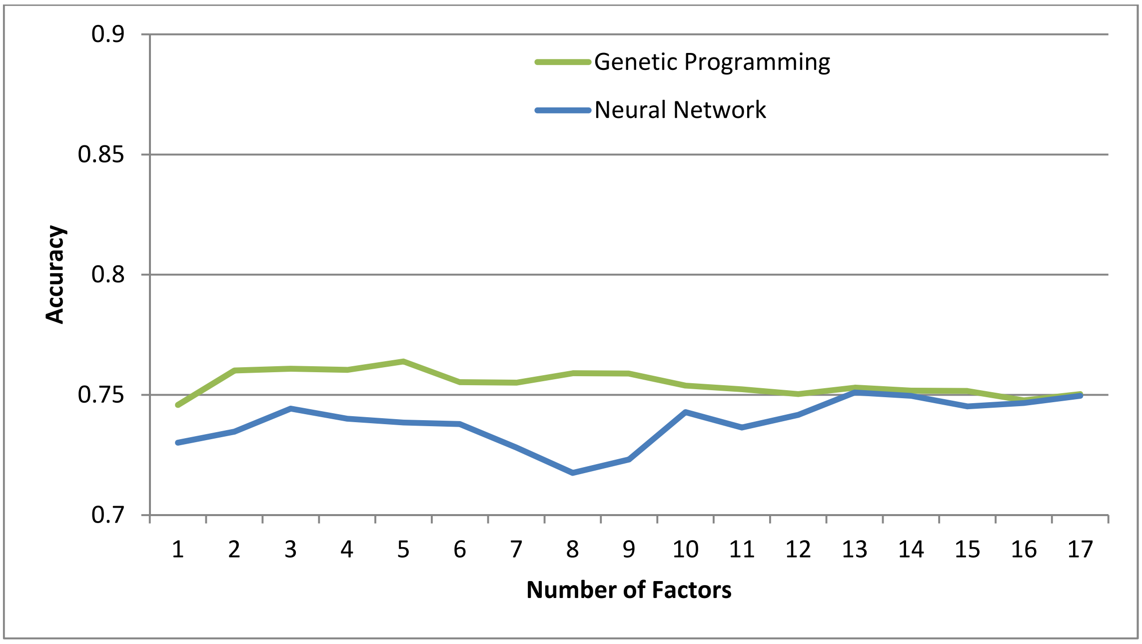


Figure 4-12 - Number of Factors versus In-Sample Accuracy for Compustat Dataset

For the Aspect dataset, the Genetic Programming algorithm found the best results on the in-sample validation dataset once the 9th factor had been added to the best-first search, yielding a validation accuracy of 68.4% with 65.3% on the out-of-sample. When all surviving factors are used, accuracy drops to 66.1% with 60.0% on the out-of-sample ($p<0.01$), indicating that over-fitting is occurring more when there are more factors for the model to consider. Furthermore, while most models noted so far had approximately equal accuracy for failure and non-failure, the out-of-sample accuracy was very unbalanced only on this result, detecting just 38.5% of failure cases but detecting 83.0% of non-failure cases. When using only the factors from Beaver (1966) or only the factors from Altman (1968), accuracy drops to 66.9% with 63.4% ($p<0.01$) on the out-of-sample, and when using all available factors accuracy drops to 66.0% with 63.2% on the out-of-sample ($p<0.01$).

When using Neural Networks, the 12^{th} factor resulted in the best in-sample accuracy of 66.0% with 64.0% on the out-of-sample, but when all surviving factors were used the model performance decreased to 59.7% with 52.3% on the out-of-sample ($p<0.01$). When using Altman (1968) factors, the network accuracy was decreased to 61.2% ($p<0.01$), though this also was very unbalanced detecting 85.4% of failure cases but just 37.0% of non-failure cases, with 60.4% on the out-of-sample in more balanced result of 57.0% and 70.3% on failure and non-failure cases respectively. When using Beaver (1966) factors the network accuracy was decreased to 58.8% with 57.5% on the out-of-sample ($p<0.01$), and when using all available factors the accuracy was decreased to 65.8% with 62.5% on the out-of-sample ($p<0.01$).

A table of these results can be found in Appendix G.

A plot of accuracy versus the resulting in-sample accuracy is shown Figure 4-13 for both the Genetic Programming and Neural Network models and the numerical results can be found in Appendix J and Appendix K.

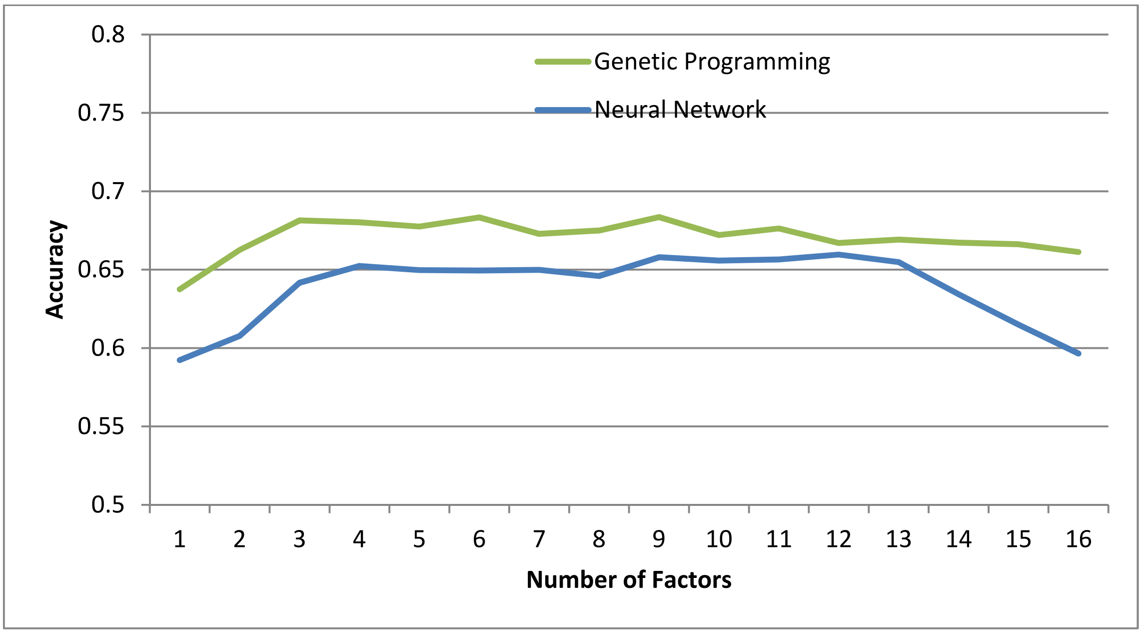


Figure 4-13 - Number of Factors versus Accuracy for Aspect Dataset

4.3.9 Conclusion

This chapter has found that a forward generation heuristic search using classification accuracy as an evaluation measure is an effective tool for finding a better subset of factors. Not only does this result in a simpler model, but this research finds it yielded a small (but statistically significant) increase in accuracy. In one case the Beaver (1966) factors resulted in an approximately equal out-of-sample accuracy, and in one case using all available factors resulted in an approximately equal out-of-sample accuracy but in no cases did the best-first search reduce accuracy in either in-sample or out-of-sample datasets.

More importantly, this research has supported the findings of Edmister (1972), that classification models (and the factors used within them) are tied very strongly to the particular dataset used. That is, the consistent utilisation of factors from prior research into new research without first testing the effectiveness of that factor set can easily become a limitation on research undertaken today.

5. Analysis Part I (Data)

The methodology outlined in the previous chapter is used throughout this and the following chapter. This chapter, Analysis Part I, takes the opportunity to consider the inclusion of additional data, while the following chapter, Analysis Part II, considers the clustering of data.

5.1 The Effect of Share Market Data

This research has shown that factors which stem from a company's financial information can be used to classify failure in a holdout sample with a higher accuracy than a naïve model. However, a company's financial information is not necessarily the only information that is available. As far back as Beaver (1968), researchers have examined whether share market information can be used instead or in addition to financial data. Agarwal & Taffler (2008) identified that because accounting statements address previous performance not future expectations accounting values may not necessarily be a true representation of the financial situation of the company, or manipulation may be present, and that financial statements are prepared on the expectation that the company is not facing imminent bankruptcy and so goes on to include information in a contingent-claims valuation approach. Another example is Atiya (2001) which states, "A problem faced by a firm will typically be reflected in the stock price well before it shows up in its balance sheet and income statement". These statements are consistent with the theoretical framework of failure identified in chapter 3.

It appears that some of the main motivations for using share market information is based on the belief that it is a distillation of all available knowledge including recent news, auditor opinion and of course accounting statements. If bankruptcy is perceived by the market, then the share price should change appropriately. Agarwal & Taffler (2008) make the argument, "in efficient markets, stock prices will reflect all the information contained in accounting statements and will also contain information not in the accounting statements".

If this is true, then key market information should assist a classification or prediction model, particularly one that is identifying failing companies. This of course comes at the cost of increased dimensionality, whereby the increase in factors in the dataset drastically increases the size of the error surface. In turn this reduces the probability that the optimal solution will be found and so potentially decreases the accuracy of the model. The issue of over-dimensionality has already been addressed in chapter 4.3, and so the same methodology that has been shown to address this issue can be used here.

The addition of market information has been investigated by others. As noted earlier, the works by Agarwal & Taffler (2008) and that of Atiya (2001) showed mixed results, so this chapter seeks to extend the usefulness of a market-inclusive versus market-exclusive comparison by performing an objective factor-selection methodology on both a market-inclusive pool of factors and a market-exclusive pool of factors, and then comparing the effectiveness of the factor-optimised models. In doing so, this chapter aims to identify which, if any, market factors are most useful in the model.

Like the previous chapter, this chapter has opted to utilise Genetic Programming and Cascade-Correlation Neural Networks, due to their performance and qualities outlined in section 4.2.7, and it is not necessary to further justify their choice in this chapter. However, the dataset that was used previously does not include share market information, so the next section of this chapter will discuss the combination of available financial information with available share market information.

5.1.1 Share Market Data

Being US-centric data, the Compustat dataset was combined with the annual US stock data available from the Centre for Research in Security Prices (CRSP), though there is some difficulty in linking the two datasets. The US share market regularly reuses “ticker” symbols as companies come into and leave the market. Compustat provides the variables “CNUM” and “CIC” which in turn match against the variable “CUSIP” from CRSP, but that can only be used to link the non-unique “PERMCO” variable, not the unique “PERMNO” variable that is needed to identify a single stock for a single company in a single year. Using such a non-unique variable results in matches of 81% of Compustat cases and 82% of CRSP cases (Palacios, 2007). However CRSP provide a “CCM linking file” which contains the Compustat primary identifier (“GVKEY”), the CRSP primary identifier (“PERMNO”), and the start date and the end date of the link. Using this file results in an 85% and 87% match rate against Compustat and CRSP respectively (Palacios, 2007).

This is an important note because much research in this area, for example the earlier cited Agarwal & Taffler (2008) as well as Atiya (2001), do not discuss that increasing the number of factors and the number of data sources often creates challenges in the form of missing or poorly linked data. Oftentimes, as discussed in chapter 2, previous research carefully selects companies or cases with full data availability which masks the real-world challenges of dealing with large datasets.

Similarly to the US-centric Compustat dataset, the Australian-centric Aspect dataset does not include share market information. To this end, this chapter turned to the “Stock Doctor” database which is a proprietary product produced by Lincoln Indicators Pty Ltd, Melbourne. The Stock Doctor database includes information such as the opening price, the daily high/low and the volume of shares traded on that day for each security.

Similarly to the Compustat/CRSP dataset, some manipulation was required to make the data ready to use. Specifically, because the available Stock Doctor data was reported daily it was necessary to determine the “closest” trading day to the date that the company’s annual financial statement was released to determine things such as the closing price, and it was also necessary to calculate the nearest trading day a year prior to determine annual returns. Also similarly to the Compustat/CRSP dataset, a portion of company-years (26.7%) in the Aspect dataset could not be matched to share market information in the Stock Doctor database and was removed from the dataset.

5.1.2 Addition of Share Market Factors

Unlike financial information, share price data represents a comparatively small set of potential additional factors. Similarly to the identification of potentially useful financial data factors in section 4.3.3, this chapter examined other research in this area including Queen & Roll (1987), Agarwal & Taffler (2008) and Atiya (2001), and in combination with looking at the available data in the CRSP and Stock Doctor datasets, the share market factors that were selected were end of year share price (adjusted), volume (adjusted), market capitalization, annual return, variance (the expected value of the squared deviation from the mean) and beta (the correlated volatility of the share with an index).

As noted in 5.1.1, there was not 100% data availability when linking the existing annual financial data to the available share market data, and so it was felt that the factor choice from chapter 4.3 may no longer be optimised. Therefore the removal of low-contributory factors (section 4.3.5) and the accuracy-based best-first forwards search (section 4.3.6) was repeated on both the Compustat and Aspect datasets. Differences between the datasets used for this chapter and the datasets used in chapter 4.3 also meant that it was inappropriate to compare the accuracy of this model directly with the results from chapter 4.3, so it was also necessary to re-perform the factor search on the same dataset but with a set of factors that did not include the additional share market information.

5.1.3 Results & Discussion

The evolutionary factor selection process on the reduced Compustat dataset, without share market data, resulted in the following factors, and is documented in Appendix L:

- current plus long-term liabilities to total assets
- net income to total assets
- net income
- net income to net worth
- cash to fund expenditures for operations
- current liabilities to total assets
- net income to sales
- sales to total assets.
- earnings before taxes to sales
- net operating profit to sales
- net worth to sales
- retained earnings to total assets
- current liabilities to current assets

While many of the factors remain the same as those found in section 4.3.5, such as net income to sales, some factors such as cash flow to total assets are no longer found to be contributory. Likewise some factors, such as retained earnings to total assets, are now found to be useful where previously they were not. This is further evidence of Edmister's (1972) statement that the optimal choice of factors is extremely sensitive to the dataset being used.

Within the reduced Aspect dataset, the following factors survived the evolutionary process:

- long-term liabilities to total assets
- current plus long-term liabilities to total assets
- cash to total assets
- current assets to total assets
- working capital to total assets
- cash to current liabilities
- cash to fund expenditures for operations
- defensive assets to fund expenditures for operations
- defensive assets minus current liabilities to fund expenditures for operations.
- retained earnings to total assets

- earnings before interest and taxes to total assets
- cash flow to current liabilities
- sales to cash
- net operating profit to sales
- net operating profit to total assets

Again some new factors are found to survive on this dataset, such as long-term liabilities to total assets, while some factors from chapter 4.3 did not survive, such as net income to total assets, and again some are found in both cases, such as cash to current liabilities.

With the surviving factors identified, the best-first forward search could be carried out. For the Compustat dataset with share market data, the Genetic Programming best-first model had the highest in-sample performance when the 6^{th} factor had been added, with accuracy on the in-sample validation set achieving 77.2% and an out-of-sample accuracy of 79.0%. However when the share market data was not available to this best-first model the in-sample validation set accuracy and the out-of-sample accuracy actually increased to 77.4% and 80.9% respectively ($p<0.01$).

When using Neural Networks, the inclusion of share market data resulted in an in-sample validation accuracy of 80.3% with 81.2% on the out-of-sample, quite a bit higher than that of Genetic Programming, but this time decreasing to 78.2% with 69.5% on the out-of-sample when share market data was excluded ($p<0.01$). Interestingly while Neural Networks performed better than Genetic Programming when share market data was available, it appears that the Neural Network model was over-fitting when share market data was not available which resulted in a very low out-of-sample accuracy.

The results on the in-sample data are shown in Figure 5-1 and Figure 5-2, and the numerical results are available in Appendix N, Appendix O, Appendix P and Appendix Q.

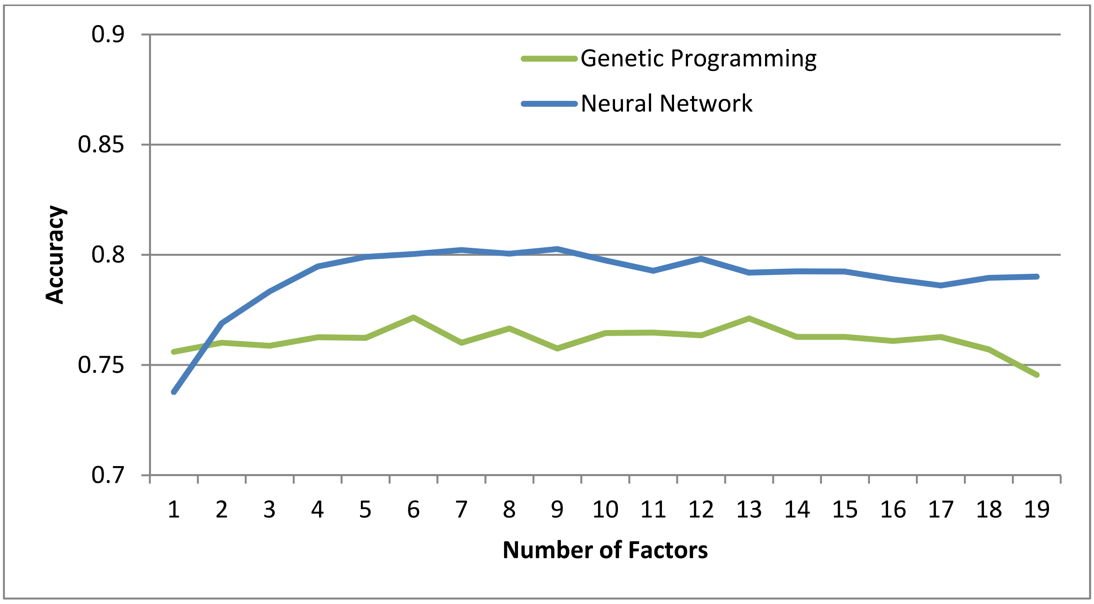


Figure 5-1 - Number of Factors versus In-Sample Accuracy for Compustat Dataset with Share Market Data

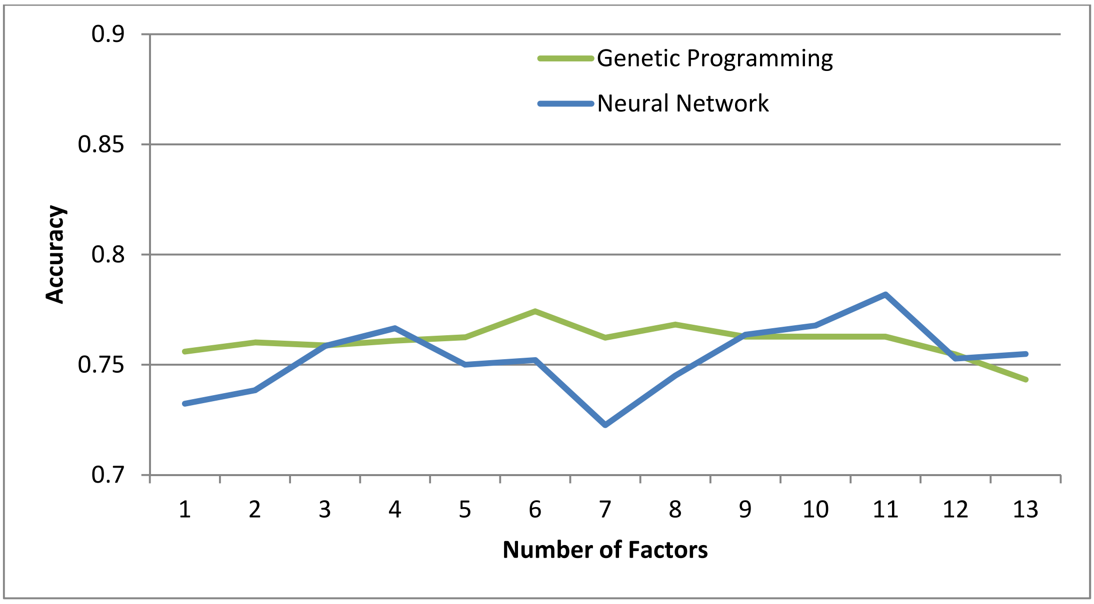


Figure 5-2 - Number of Factors versus In-Sample Accuracy for Compustat Dataset without Share Market Data

For the Aspect dataset, using Genetic Programming with share market data available the model achieved the best in-sample validation accuracy once the 4th factor had been added, yielding an accuracy of 68.9% with 62.6% on the out-of-sample, but since none of the best-first 4 factors included any share market information, there was no benefit (or drawback) to the share market data being added.

When using Neural Networks the model achieved the best accuracy once the 8th factor had been added with an in-sample accuracy of 70.7% with 63.6% on the out of sample. When share market information was not available to the model, the maximum accuracy decreased to 69.6% with 59.4% ($p<0.01$). Again the Neural Network model appears to have undergone over-fitting, evidenced by the large disparity between the in-sample and out-of-sample accuracy.

The results on the in-sample data are shown in the Figure 5-3 and Figure 5-4, and the numerical results are available in Appendix R, Appendix S, Appendix T, Appendix U.

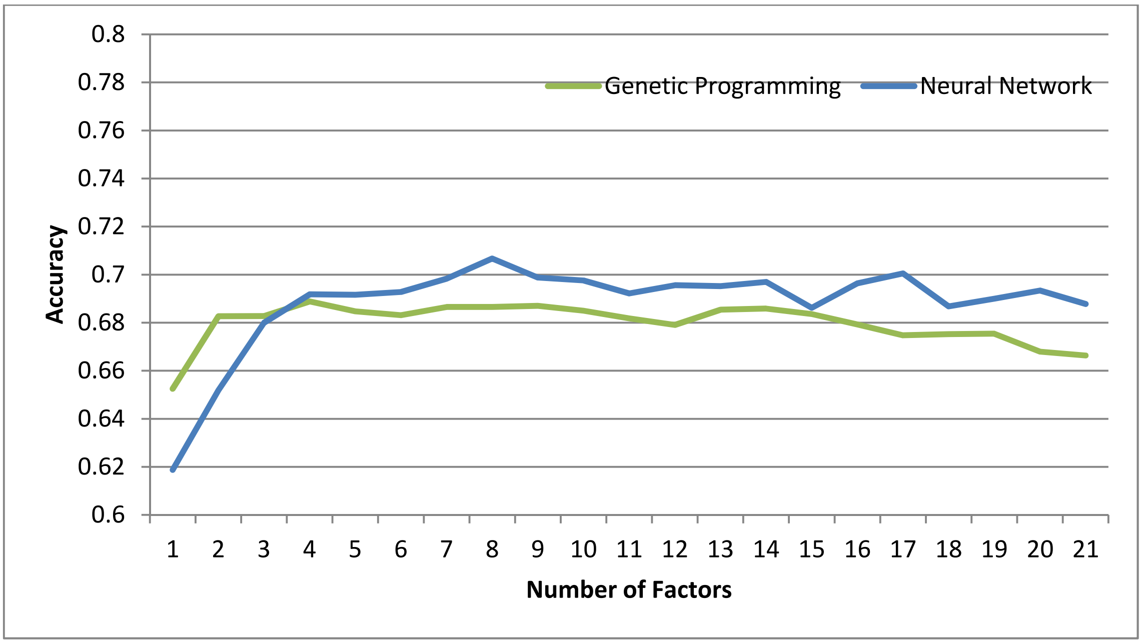


Figure 5-3 - Number of Factors versus Accuracy for Aspect Dataset with Share Market Data

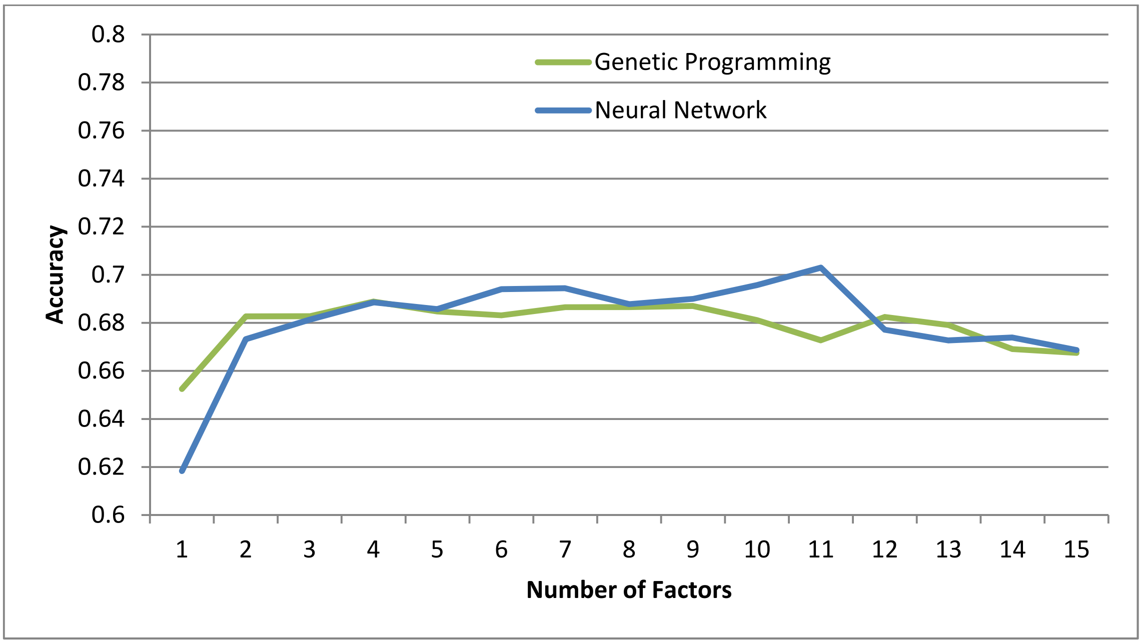


Figure 5-4 - Number of Factors versus Accuracy for Aspect Dataset without Share Market Data

Beyond a comparison of classification accuracy when share market data is added, it is worthwhile to consider the order of the share market data. In three out of four experiments in which share market data was available, the variable "share price" was included as a useful factor by the objective best-first heuristic factor search. In the one experiment in which it was not, the Compustat dataset using Neural Networks, no share market data was found to be useful at all. Neither volume nor variance, on the other hand, was found to be useful by the objective factor search in either of the datasets using either of the classificatory models.

5.1.4 Conclusion

This chapter has found that while share market data such as the end of financial year share price can contribute to the classification accuracy of bankruptcy models, the increases of up to 2.1% on in-sample accuracy come at a cost of up to 26.7% data loss due to difficulties in matching end of fiscal year financial data to share market data. Furthermore, the addition of share market data was shown to sometimes cause decreases in both in-sample and out-of-sample accuracy, or in one case to have no impact on the classification accuracy of the model at all. Given the theoretical framework behind the inclusion of share market data, these results are somewhat surprising. Due to the mixed results using the Compustat and Aspect datasets with Genetic Programming and Neural Network classification models, this thesis will revert to using datasets that do not include share market information.

5.2 The Effect of Macroeconomic Data

While certainly the factors examined in chapter 4.3, and the inclusion of share market data in section 5.1 are often considered in corporate failure prediction research, the inclusion of macroeconomic data is overlooked in key research such as Lensburg, et al. (2006), one of the first papers to utilise Genetic Programming on this problem domain. This is surprising because a body of theory exists which links macroeconomic factors with corporate failure prediction, Liu (2004) references some seven papers that “sought to determine bankruptcies and insolvencies of UK companies at the aggregated level”. That is not to say that no research on this topic exists, for example Rose, et al. (1982) used stepwise regression, applied correlation analysis and lead-lag relationships to conclude that macroeconomics conditions are highly correlated with business failure, while Rösch & Scheule (2005) found that “the inclusion of macroeconomic factors renders the systematic unobservable factors less important and diminishes the impact of correlations”. It is therefore reasonable to hypothesise that the inclusion of macroeconomic data in the best-first forward search model, now shown to increase both in-sample and out-of-sample accuracy, would benefit from the inclusion of macroeconomic factors.

5.2.1 Macroeconomic Data

This research examines macroeconomic factors that have been found to be useful in prior research, specifically the GDP Growth Rate (Rösch & Scheule, 2005; Nam, et al., 2008), the Retail Price Index/Consumer Price Index (Liu, 2004), the Interest Rate (Rösch & Scheule, 2005; Liu, 2004), as well as two factors that were used in the generation of aggregate indexes including the Exchange Rate and the Dow Jones Composite/Australian All Ordinaries. While this is far from an exhaustive list of macroeconomic factors, these factors were easily obtainable from the Development Data Group (World Bank Group, 2012), Morningstar (Morningstar Australasia Pty Limited, 2012) and Standard & Poor's (CME Group Index Services, 2011).

Similarly to the share market data used in the previous section, the macroeconomic data was taken at the time of the financial statement creating essentially distinct values for each macroeconomic factor that are repeated across all company-years from the same year of data. Since some factors (such as the Interest Rate) are typically larger than other factors (such as GDP), this research performed the last step of the best-first search on both range normalised and unnormalised data, finding that normalising the data decreased accuracy on the in-sample dataset, and therefore the results from the unnormalised data are used in the following sections.

5.2.2 Results & Discussion

Unlike the inclusion of share market data, which resulted in a portion of company-years being excluded due to a lack of share market data, the inclusion of macroeconomic data did not change the existing data in any way, and it is therefore unnecessary to carry out the removal of low-contributory factors. Instead, the factors identified from the methodology in section 4.3.5 can be used for performing the best-first forward search as outlined in section 4.3.6.

Notably, the Genetic Programming methodology on the Compustat dataset discovered the same first 3 factors in the same order with a very similar in-sample accuracy of 75.9% (earnings before taxes to sales, net income to total assets, cash to fund expenditure from operations). However in this experiment, the fourth factor was discovered to be net income to net worth, rather than sales to total assets. This is not entirely surprising because Genetic Programming is a non-deterministic methodology, and the information gain from the addition of the fourth factor is small or possibly negative (section 4.3.8 shows an accuracy decrease of 0.05% on the in-sample data). Moreover, cash flow to total assets, a factor discovered to be useful in the best-first search from section 4.3.6 is also found to be useful in this experiment.

A similar situation is found with the Aspect dataset also using Genetic Programming, the first three factors, net income to net assets, cash to current liabilities and cash flow to total liabilities, are discovered in the same order, but as the information gain diminishes (a 0.12% accuracy decrease on the in-sample data) the non-deterministic nature of Genetic Programming yields a different fourth factor. While this finding is somewhat of a tangent from the original purpose of this chapter, it is an interesting validation of the results from section 4.3 that demonstrates experimentally that appropriate stopping criteria were used for the algorithms.

Within the Compustat dataset using Genetic Programming, the peak in-sample accuracy was discovered after the inclusion of earnings before taxes to sales, net income to total assets, cash to fund expenditures for operations, net income to net worth, and cash flow to total assets, with an in-sample accuracy of 76.9% and an out-of-sample accuracy of 73.6%. While the in-sample result is marginally better than when the experiment was run without macroeconomic data of 76.4%, the out-of-sample accuracy has decreased marginally from 73.8%. These differences are entirely due to the non-deterministic nature of Genetic Programming because at this stage the datasets are identical. What is most illuminating in this instance is that the peak accuracy in the best-first model includes no macroeconomic factors at all. By comparison when using Neural Networks, peak in-sample accuracy was achieved with the inclusion of the 17^{th} factor which did in fact include some macroeconomic factors (exchange rate and the Dow Jones industrial index) as factors, but while the in-sample accuracy increased from 75.1% to 75.6%, out-of-sample accuracy fell from 70.4% to 68.2%, indicating that the higher dimensionality has resulted in more over-fitting. The results of the best-first search are shown in Figure 5-5 with the numerical results available in Appendix V and Appendix W. Note that this is the only result of this section that included macroeconomic factors in the best-first forward search (resulting in two methods to compare), and is therefore the only result in which McNemar's test for statistical significance can be applied, finding $p<0.01$.

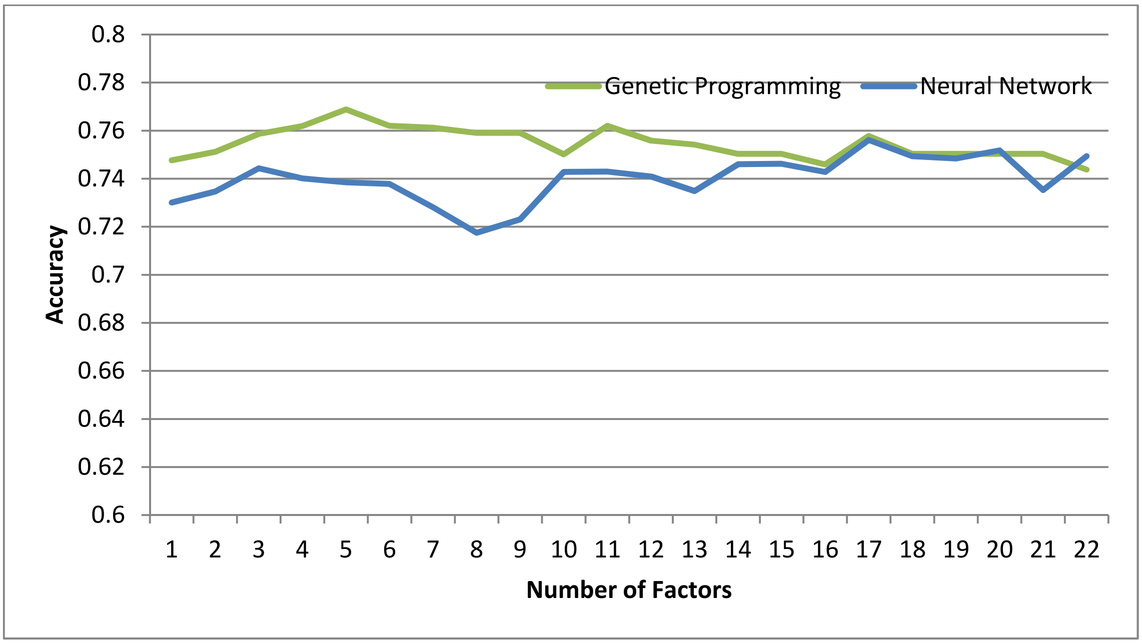


Figure 5-5 - Number of Factors versus Accuracy for Compustat Dataset with Macroeconomic Data

The Aspect dataset yields similar results. When using Genetic Programming the peak in-sample accuracy diminished from 68.4% to 68.1% and the out-of-sample accuracy reduced from 65.2% to 62.3%. When using Neural Networks, peak accuracy reduced accuracy on the in-sample data from 66.0% to 65.7%, and from 64.0% to 63.5% on the out-of-sample dataset, and also found that the peak results were obtained when macroeconomic factors were not incorporated into the network. These results are shown in Figure 5-6 and can be found numerically in Appendix X and Appendix Y.

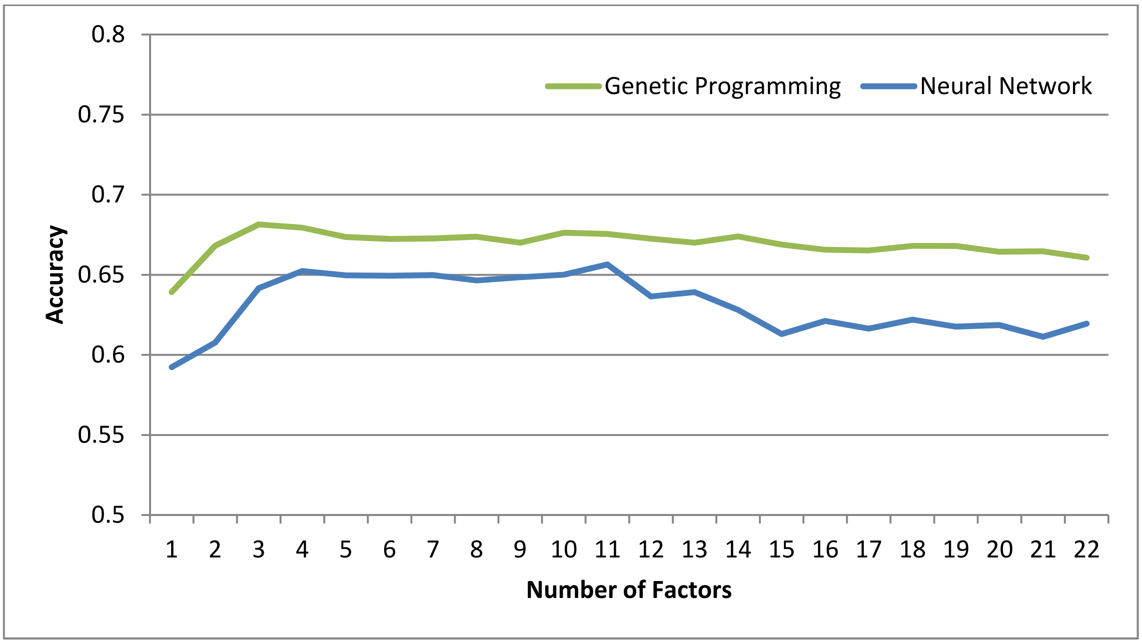


Figure 5-6 - Number of Factors versus Accuracy for Aspect Dataset with Macroeconomic Data

5.2.3 Conclusion

This section has examined the effect of adding macroeconomic data to the factors identified in chapter 4.3 to validate the findings of research such as Rösch & Scheule (2005). However, this chapter has found that the best-first forwards search approach included macroeconomic factors in only one of the four experiments, and in that case it appears that the increased accuracy on the in-sample dataset was due to over-fitting as evidenced by the decreased out-of-sample accuracy. While this result is surprising it does not indicate that companies fail independently of the circumstances of the macroeconomy, it indicates only that in the datasets used for this research the macroeconomic factors used contributed less information than the cost of increasing the dimensionality. This could be because the factors chosen were not a good representation of the macroeconomy, or more likely because the effect of the macroeconomy on the company is represented within the company-specific financial information also available to the model.

Furthermore, this chapter has validated the results of the methodology outlined in section 4.3.6, finding that re-performing the non-deterministic experiments yields highly similar results, showing that the stopping criteria selected did not prevent the model from finding near-optimal results on the in-sample data.

Given that previous research has found macroeconomic factors to be a useful lead indictor for corporate failure, there is certainly value in expanding the range of factors available to the Genetic Programming and Neural Network models used in this research, as well as considering other classification and prediction systems such as Support Vector Machines (SVM) before concluding that macroeconomic indicators provide no additional information to the datasets used in this research. However, having considered a set of reasonably popular macroeconomics factors across two classification systems and two datasets, the findings from this section are sufficiently robust to continue onto part II of the analysis without the inclusion of the macroeconomic factors used here.

6. Analysis Part II (Methodologies)

Having considered the effect of including additional data in chapter 5, this chapter will concentrate on the methodologies that can be applied to the data to increase classification accuracy.

6.1 Improving the Visualisation of Clusters

A potential improvement to the classificatory accuracy of bankruptcy data comes in the form of clustering. Certainly other research, including both Beaver (1966) as well as Altman (1968) indicated that company "similarity" should be considered, as both papers paired companies by industry. However, if the effect of clustering or even simply grouping companies by industry is to be considered in chapter 6.2 as planned, some measure of clustering effectiveness will need to be used. However, this chapter goes on to identify a number of limitations in the clustering effectiveness measure and visualisation that would otherwise be an ideal choice, and so this chapter seeks to address those limitations. In doing so, this chapter develops the SpecVCMV algorithm that will then be used in section 6.2. As such, this chapter is a seemingly distant (but necessary) tangent to the stated purpose of this thesis.

Clustering algorithms are automated systems for the grouping of *unlabelled* data points based on some predefined similarity or distance measure. For example, the length and width of a sample of flowers' sepals and petals could be measured, as was the case in the popular Iris flower dataset (Fisher, 1936). A clustering algorithm would seek to determine whether a number of distinct species of flowers exist within the sample, independently of any labelling performed by experts. This process is distinct to classification, where the sample would contain labelled data points, allowing a classification system to learn what properties are typical of each group and therefore classify additional unlabelled data points with some degree of accuracy. Unlike classification, one of the difficulties in clustering is that there is no label to measure how well the items have been clustered. This chapter will begin by discussing clustering validity indexes, moving on to clustering visualisations, and addressing some of the major visualisation methods along with their limitations. This chapter will go on to propose and test a new visualisation method, Visual Cluster (Membership) Validity, which can be used in future sections to test the effectiveness of clustering annual financial company information.

6.1.1 Cluster Validity Indexes

There are many metrics that can be used for comparing the effect of clustering. One of the most popular is the Davies-Bouldin index (DB) (Davies & Bouldin, 1979), which measures the within-cluster scatter, the between-cluster separation, and returns the maximum ratio of within-cluster scatter versus between-cluster separation. A lower value indicates smaller within-cluster scatter or larger between-cluster separation hence a low DB indicates better clustering. Prior to the Davies-Bouldin index, a popular method was the Dunn Index (DI) (Dunn, 1974), which can measure the maximum distance across clusters, the mean distance between all pairs of points in a cluster, the distance from all points in the cluster to the mean or some other measure of maximal within-cluster distance, and builds a ratio against minimal inter-cluster distance using the same formulations. There are, in fact, so many metrics that at this point Hubert (1985) is often cited, “We will not try to review this literature comprehensively since that task would require the length of a monograph”. One of the issues that arise with so many methods for measuring cluster validity is that, as noted in Bezdek & Pal (1998), even when the same clustering algorithm is used on the same dataset, the optimum number of clusters varies depending on the cluster validity measurement that is used.

6.1.2 Cluster Validity Visualisations

One of the issues with such cluster validity measurements is that by reducing an entire dataset to a single number, much information is lost (Hathaway & Bezdek, 2003), and it becomes difficult to understand the underlying structure of the data. Visualisations such as scatter plots can be used when the data is 2 or 3-dimensional, but as the number of dimensions increase so does the difficulty in perceiving the clusters. To overcome these limitations, various clustering visualisation methods have been proposed which can project hyper-dimensional data as abstractions of the data, for example visualising pairwise distance rather than the data itself. In many ways the work of Huang et al. (2001), Bezdek & Hathaway (2002), Chen & Liu (2003) and Hathaway & Bezdek (2003) have spearheaded this field. This section will begin by outlining a number of the visualisation methods that are in use today, before discussing the motivations for developing an alternative.

The Visual Assessment of (Cluster) Tendency (VAT) method (Bezdek & Hathaway, 2002) begins by calculating the pairwise dissimilarity matrix between each data point and every other data point. The dissimilarity matrix is sorted by way of Minimum Spanning Trees (MST) (Prim, 1957) and the resulting pairwise distances are displayed as greyscale pixels with the maximum distance between points shown as white pixels and the minimum distance (zeros along the main diagonal as these represent the distance between a point and itself) shown as black pixels. The use of MST makes the VAT method excel at visualising clusters of odd shapes, such as concentric circles. Figure 6-1 shows such a dataset, with the points joined by lines in the order that the MST algorithm sorted them, and the diagram also shows the resulting VAT visualisation. The VAT visualisation makes it apparent that the distance between point 20 and 30 is relatively small because the colour of the VAT diagram at x=20, y=30 is dark grey, whereas the distance between point 20 and point 60 is high because the colour of the VAT diagram at x=20, y=60 is white. The black diagonal band from top-left to bottom-right is representative of each point being near to its immediate neighbours but comparatively far from its more distant neighbours. The two apparent blocks along the top-left to bottom-right diagonal is indicative of two clusters, and is caused by the MST algorithm "jumping" to the inside circle while ordering the data points.

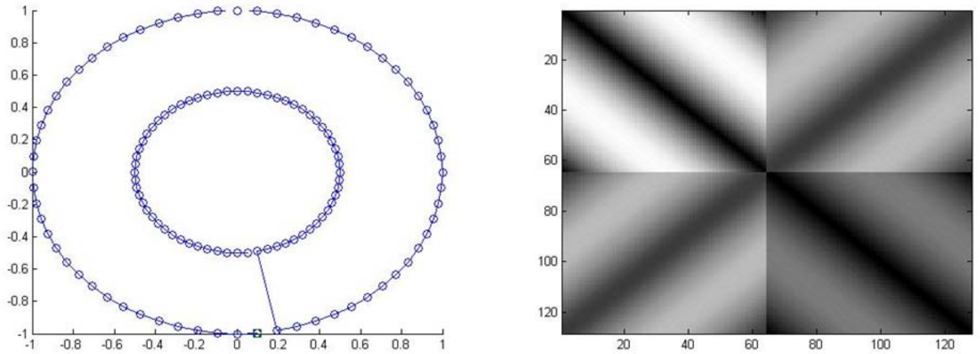


Figure 6-1 - VAT of Concentric Circle Dataset

The MST algorithm however is susceptible to the “chaining effect”, whereby the path of least distance does not necessarily maintain cluster homogeny. The dataset used in Figure 6-2 below is two Gaussian distributions centred around (-2, -2) and (2, 2), with the points joined in the order the MST algorithm sorted them, and the VAT algorithm which appears to identify a spurious third cluster. This a result of the MST algorithm doubling back to data points missed in the previous pass. Had the two Gaussian clusters been in multi-dimensional space, the spurious cluster would appear as a legitimate cluster and it would not be immediately possible to identify that this had occurred.

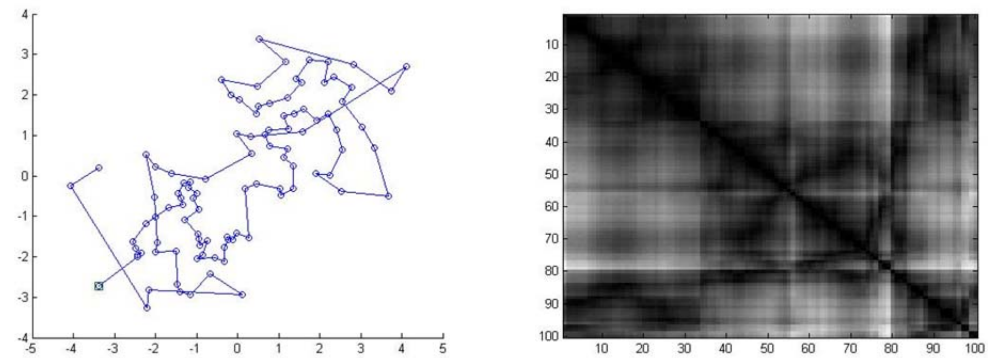


Figure 6-2 - VAT of Volumetric Cloud Dataset

Visual Cluster Validity (VCV) (Hathaway & Bezdek, 2003) is a generalisation and extension of the VAT algorithm that begins by outsourcing the clustering of the data to an external clustering algorithm of the user's choice. In doing so, the VCV method is confined to visualising the results of a clustering algorithm and does not cluster the data itself. For the external algorithm, Hathaway & Bezdek (2003) used Fuzzy c-Means (FCM) and Fuzzy c-Regression Models (FCRM) to illustrate the ability to visualise volumetric clouds as well as clusters that are identified using a linear or non-linear regression function which seek define clusters by findings lines of best fit that minimise squared error.

Hathaway & Bezdek's VCV algorithm specifies that each cluster must be sorted using MST so that similar clusters are located close together on the resulting visualisation, and so if a fuzzy clustering method has been used then the cluster membership must be "hardened". Furthermore if fuzzy clustering has been used, the VCV algorithm then sorts the data points within each cluster by cluster membership to each point's nearest cluster. Like VAT, the VCV algorithm visualises pairwise distance, but the distance function (which defines the colour of the pixel) is specified as the minimum sum of distances to each cluster prototype ($R^*_{ik} = \min_{1\leq j\leq c}\{d_{ji} + d_{jk}\}$) rather than the Euclidean distance between the two points. This method has gone on to produce variations, such as the "relational" VCV method (Ding & Harrison, 2007) which can be used in more specific circumstances.

The VCV algorithm can be summarised as a comparison to VAT as shown in Table 6-1:

	VAT	**VCV**
Initial Clustering Algorithm	**None**	**User's choice**
Extra-Cluster Sorting	**None**	**Minimum Spanning Trees of cluster prototypes in Euclidean space**
Intra-Cluster Sorting	**Minimum Spanning Trees of data points in Euclidean space**	**Winning cluster degree of membership (if fuzzy)**
Distance Measure (Visualisation of)	**Euclidean pair-wise distance**	**Euclidean distance to cluster prototype**

Table 6-1 - Summary of VAT and VCV

In example Figure 6-3 below, fuzzy c-means clustering has been used on three volumetric clouds, but the number of clusters specified for the clustering algorithm to find is 4 (c=4), and as a result the cluster in the top left corner has been spuriously split into two micro-clusters. Due to the algorithm visualising the distance to the cluster prototype (in this case the cluster centroid), the visualisation shows just three clusters and demonstrates that the first two clusters are in fact the product of one larger cluster.

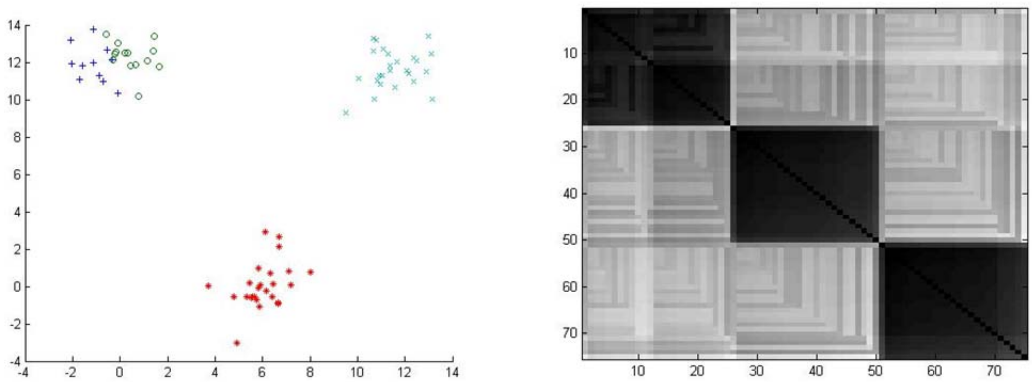


Figure 6-3 - Clustered and VCV for 3 Volumetric Clouds (c=4)

By visualising the distance to the cluster prototype, rather than the pairwise distance, VCV is able to display linear and non-linear regression clusters, such as in the example in Figure 6-4 in which two intersecting lines are identified by the FCRM algorithm.

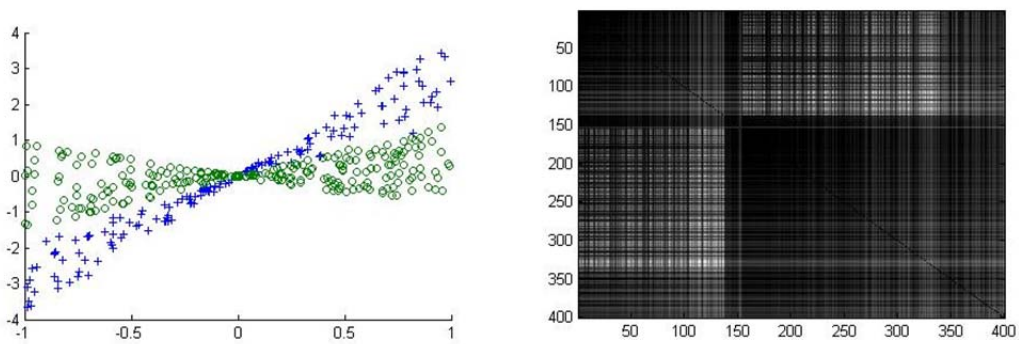


Figure 6-4 - Clustered Data and VCV for 2 Intersecting Lines (c=2)

However the VCV algorithm has some underlying assumptions that can theoretically lead to situations where the results are misleading. The first assumption in VCV is caused by the choice of distance measure. There are times when a data point that is "far" from its cluster prototype is indicative of a poorly clustered point; however this is not always the case. In the example given in Figure 6-5, two points are highlighted that are of equal distance from their cluster centroids. These would have very different cluster memberships when clustered using FCM because one of those two points is quite close to the competing cluster, while one of those points is far away from the competing cluster. However the VCV algorithm, which visualises distance, would not communicate the different cluster memberships calculated by the underlying FCM algorithm.

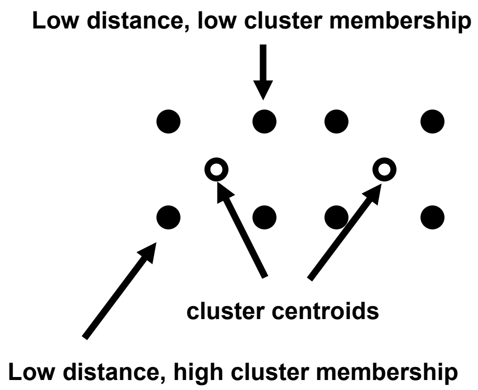


Figure 6-5 - Distance does not always equate to Cluster Membership

The second assumption is caused by the algorithm with which clusters are sorted. VCV's cluster sorting method, while based on Minimum Spanning Trees, arbitrarily keeps the cluster with arbitrary label "1" as the first cluster (Hathaway & Bezdek, 2003). In doing so, VCV may split similar clusters due to the order that the clusters are selected by the algorithm. In the example given in Figure 6-6 two very nearby clusters in the two left of the diagram are split because one of them happened to be arbitrarily labelled "1".

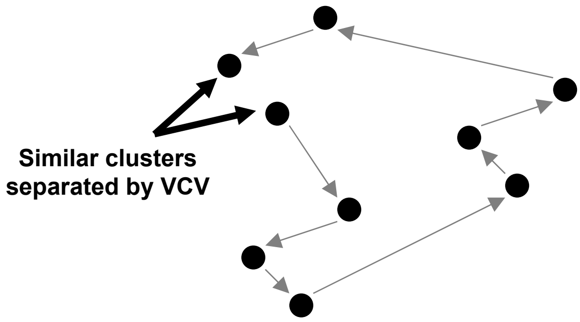


Figure 6-6 - Similar clusters separated by VCV

The third assumption in the VCV algorithm is that the cluster prototypes are used for sorting and are assumed to be in Euclidean space, but this may not always be a valid assumption. In situations where the external clustering algorithm returns non-Euclidean parameters (such as FCRM, which returns polynomial parameters), cluster similarity cannot be measured by simply calculating Euclidean distance of cluster prototype parameters, as is done by the VCV algorithm.

For example in the diagram Figure 6-7, three clusters of data have been successfully identified as $y = 1x + 3$, $y = 1x + 1$ and $y = 0x + 1$. It is apparent that cluster $y = 1x + 3$ and $y = 1x + 1$ are more similar than $y = 0x + 1$. However when these cluster prototypes, (1,3), (1,1) and (0, 1) respectively, are treated as if they're in Euclidean space as shown on the right hand side, the VCV algorithm treats $y = 1x + 3$ to be less similar with $y = 1x + 1$ and $y = 0x + 1$ to be more similar. Therefore there may in theory be situations in which the clusters are sorted incorrectly due to the assumption that cluster prototypes are Euclidean in nature.

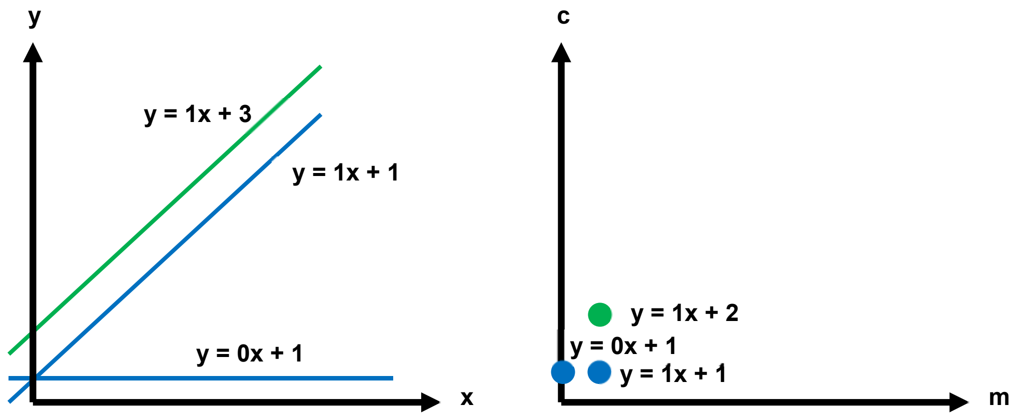


Figure 6-7 - Polynomial parameters presented in Euclidean Space

To demonstrate the impact of these assumptions, a set of data is crafted with two volumetric clouds centred around (5, 2.5) and (10, 2.5) and shown in Figure 6-8, with the resulting VCV visualisation (c=5). In particular note the dark area in the bottom left corner of the visualisation showing a small distance between the top left and bottom right clusters caused caused by the VCV algorithm spuriously separating the left hand volumetric cloud in the original dataset.

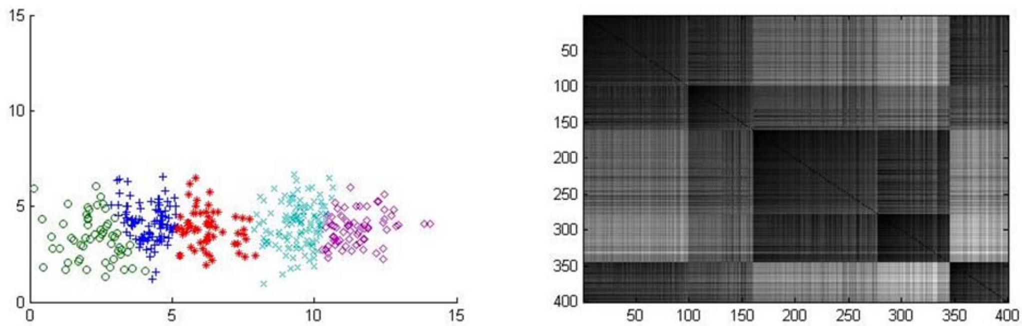


Figure 6-8 - Clustered Data and VCV for 2 Volumetric Clouds (c=5)

Finally the VCV algorithm requires the user to choose the external clustering method, and in high-dimensionality datasets this can be difficult as the data often cannot be perceived directly.

The SpecVAT algorithm (Wang, et al., 2008) uses the external clustering algorithm Spectral Clustering (Shi & Malik, 1997) to map the data points into an abstracted space, then uses the VAT algorithm outlined earlier in this section on the points in the spectrally mapped abstraction. Spectral Clustering is a well-accepted mechanism to perform clustering, and works by using an affinity function (usually an exponential function of Euclidean distance) to calculate an "affinity matrix" between pairwise points. The affinity function ensures that neighbouring points have a pairwise affinity approaching 1 while points that are not neighbouring have a pairwise affinity approaching 0, essentially creating a mathematical graph of the "connectedness" of each point to every other point. Graph theory then allows the eigenvectors of the normalised Laplacian matrix to be calculated, and taking the first k parameters of each eigenvector has the effect of mapping each point into a k-dimension space in which highly connected data points are located nearby, even if they were distant in the original Euclidean space. This means that odd shapes, such as the concentric circles example used in Figure 6-1 are mapped into volumetric clouds in the spectral space, and can then be visualised using the VAT algorithm.

The use of Spectral Clustering is an excellent method for avoiding the need for the user to choose the external clustering algorithm, particularly when the data cannot be perceived directly. However the use of the VAT algorithm introduces VAT's limitations, including the chaining phenomenon, which in k-dimensional volumetric clouds now becomes difficult to identify. In the example shown in Figure 6-9, three volumetric clouds centred around (0, 0), (0, 3) and (3, 0) are clustered using Self-Tuning Spectral Clustering (Zelnick-Manor & Perona, 2004) (which uses k-means on the points in the abstracted space), the abstracted space in k-dimensions, and the resulting SpecVAT visualisation are shown, demonstrating that the SpecVAT algorithm has identified a number of spurious clusters.

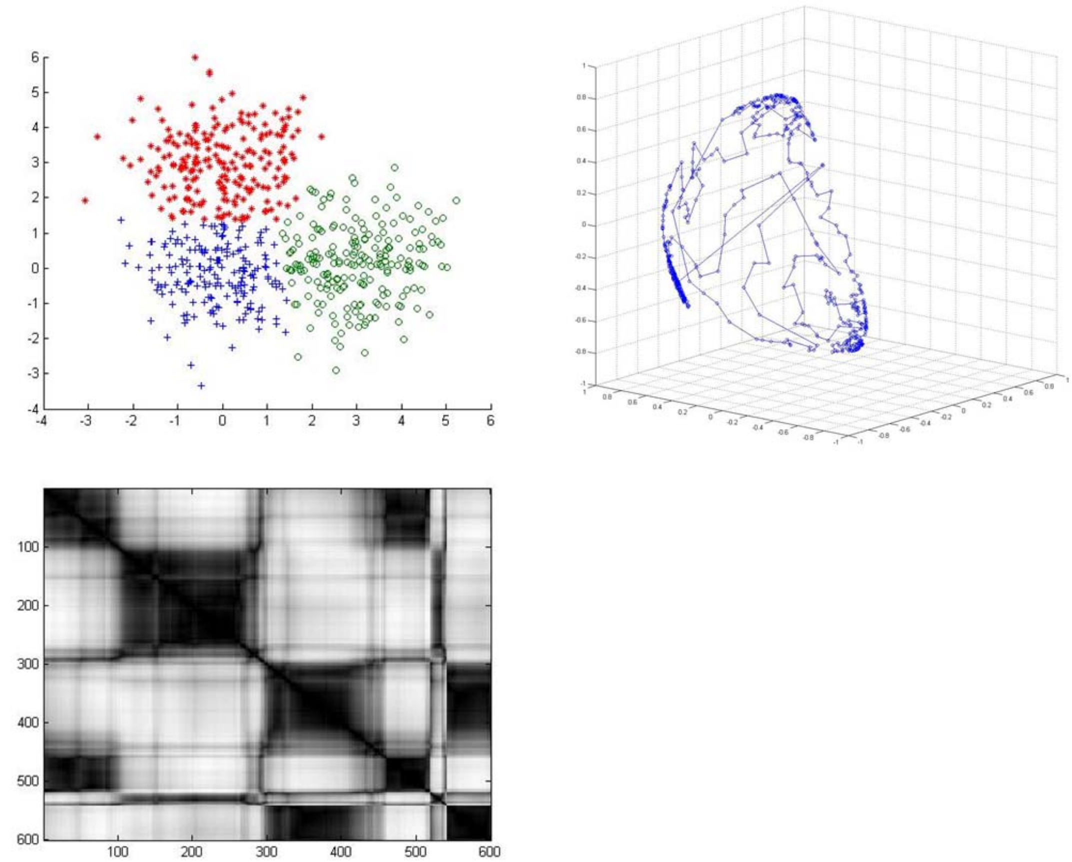


Figure 6-9 – Clustered data (top left), VAT sorted eigenvectors transformed to eigenspace (top right), and resulting SpecVAT for three volumetric clouds (bottom left, with c=3, K=7)

To address these issues, it is tempting to simply apply the improved VCV algorithm as opposed to VAT, but to do so would re-introduce the assumptions found in the VCV algorithm that have already been demonstrated, so an alternative is proposed in this thesis.

6.1.3 Spectral Visual Cluster (Membership) Validity

Spectral Visual Cluster (Membership) Validity (SpecVCMV) incorporates a number of adaptations to the VCV algorithm to address the assumptions outlined in the previous section.

The structure of Hathaway & Bezdek's VCV algorithm is maintained, specifically "(s1) [step 1] the clusters themselves are (possibly) reordered; and then (s2) the data in each cluster are reordered" (Hathaway & Bezdek, 2003). Step 0 (s0) is inferred to be computing the U membership (U_{oq}, where o is a data point and q is a cluster), and the cluster prototypes (V) as the cluster centre in Euclidean space or parameters of a when applying an external clustering algorithm such as fuzzy c-means (FCM) or fuzzy c-regression models (FCRM). A more detailed explanation of the above algorithm can be found in Hathaway & Bezdek (2003).

It is therefore proposed in this study that the external clustering's own cluster membership (confidence) be used as the measure of distance instead of the Euclidean distance to cluster prototypes. Thus the limitation outlined in the previous section is overcome; however this adds the additional requirement that the external clustering algorithm be fuzzy. Within Spectral Clustering, the clustering itself is typically performed using k-means on the spectrally mapped data points, but Fuzzy C-Means (FCM) allows cluster membership (U_{oq}) to be used instead. Therefore the pair-wise dissimilarity from Hathaway & Bezdek (2003) can be redefined as $R_{ik}^{*} = \min_{1\le j\le c}\left\{\left(1-U_{ji}\right)^{2}\left(1-U_{jk}\right)^{2}\right\}$.

It is argued that a visualisation of clustering confidence is a stronger indication or successful or unsuccessful clustering than to visualise distance, which has been shown in some circumstance to mask potential issues. However it is now assumed that poor clustering on behalf of the external clustering algorithm will be reflected in the clustering confidence. That is to say that if the external clustering algorithm returns highly confident results on poor clustering outcomes, this high degree of confidence will be reflected in the visualisation.

The cluster sorting algorithm, designed to separate clusters that are different and unite clusters that are similar, can now be considered. As the external clustering algorithm may return cluster prototypes that are not in Euclidean space, it is necessary to consider non-Euclidean based sorting algorithms. Cluster similarity can be measured as the sum of cluster membership, that is if cluster i has a high sum of cluster membership to cluster j, and cluster j has a high sum of cluster membership to cluster i, cluster i and j are similar. This can be expressed as: Cluster p contains objects (o) after hardening, and each data-point has U membership to every cluster q, thus represented as U_{oq}. Summing the U_{oq} membership for cluster p results in U_{pq}, the extra-cluster U membership between clusters p and q. That is, $U_{pq} = \sum_{o \in p} U_{oq}$.

This algorithm generates a dissimilarity matrix of generated clusters, and this allows the VAT algorithm, based on Prim's MST, to be used to order the clusters. For easy comparison, a similar format to Bezdek & Hathaway (2002) is used in the following algorithm, however P is used as the cluster dissimilarity matrix to differentiation from the original algorithm which used data-point dissimilarity.

Step 1 Set $K = \{1,2,\dots,c\}; I = J = \emptyset; P[0] = (0,\dots,0)$

Step 2 Select $(i,j) \in \mathrm{argmin}_{p \in K, q \in K}\{U_{pq}\}$

Set $P(1) = i; I = \{i\};$ and $J = K - \{i\}$

Step 3 For $r = 2,\dots,c$:

Select $(i,j) \in \mathrm{argmax}_{p \in I, q \in J}\{U_{pq}\}$

Set $P(r) = j;$ Replace $I \leftarrow I \cup \{j\};$ and $J \leftarrow J - \{j\}$

Next r

With cluster sorting performed (s1), the sorting of the data-points within each cluster needs to be performed (s2). Assuming that fuzzy clustering has been used, the VCV algorithm uses cluster membership, but it is instead proposed that "distance to cluster prototype" from Hathaway & Bezdek (2003) can be reintroduced, but as a sorting rather than visualisation dimension as a way of highlighting microclusters within the larger clusters. To do this, is it necessary to calculate proximity in the spectral mapping space. Hathaway & Bezdek (2003) define the distance in a two-dimensional space (when c=2) to be $d_{ik} = ((v_{i1} - x_{k1})^2 + (v_{i2} - x_{k2})^2)^{0.5}$.

At this point, the algorithm can be thought of as the VCMV component of the SpecVCMV algorithm. Incorporating Spectral Clustering into this and the algorithm can be formatted similarly to that of SpecVAT (Wang, et al., 2008) for easy comparison:

Input: **D** = $[d_{ij}]$: an $n \times n$ scaled matrix of pair-wise dissimilarities, with $1 \geq d_{ij} \geq 0; d_{ij} = d_{ji}; d_{ii} = 0$, for $1 \leq i$ and $j \leq n$ where k is the number of eigenvectors used (the dimension of the embedding subspace), which is the number of clusters to find.

(1): Compute a local scale $\sigma_i = d(o_i, o_K) = d_{iK}$ where o_K is the K-th nearest neighbour of o_i.

(2): Construct the weighting matrix $\boldsymbol{W} \in \Re^{n \times n}$ by defining $w_{ij} = \exp(\frac{-d_{ij}d_{ji}}{\sigma_i \sigma_j})$ for $i \neq j$ and $w_{ii} = 0$.

(3): Let $\boldsymbol{M}$ be a diagonal matrix with $m_{ii} = \sum_{j=1}^{n} w_{ij}$ (i.e. the (i,i) element of $\boldsymbol{M}$ is the sum of $\boldsymbol{W}$'s i-th row), and construct the matrix $\boldsymbol{L'} = \boldsymbol{M}^{\frac{-1}{2}} \boldsymbol{W} \boldsymbol{M}^{\frac{-1}{2}}$ which is a normalised version of the Laplacian matrix.

(4): Choose $v_1, v_2, \ldots, v_k$, the k largest eigenvectors of $\boldsymbol{L'}$ to form the matrix $\boldsymbol{V} = [v_1, \ldots, v_k] \in \Re^{n \times k}$by stacking the eigenvectors in columns.

(5): Normalise the rows of V with unit the Euclidean norm to generate $\boldsymbol{V'} \in \Re^{n \times k}$, i.e., $v'_{ij} = \frac{v_{ij}}{\|v_{i:}\|}$.

(6): For $i = 1,2, \ldots, n$ let $u_i \in \Re^k$be the vector corresponding to the i-th row of $\boldsymbol{V'}$ and treat it as a new instance in the k-dimensional embedding space (corresponding to original o_i), then apply the FCM algorithm with k clusters to $\boldsymbol{V'}$, treating each element u_i as a distinct data point to be clustered to obtain the cluster prototype $\boldsymbol{P}$ in spectrally mapped space, and the cluster memberships $\boldsymbol{U}$ for each data point.

(7): Using the Euclidean distance between each point u_i in $\boldsymbol{V'}$ and each cluster prototype $\boldsymbol{P}$, construct the sorting distance vector $s' = s'_i = \min_{1<c<k}(\|u_i - p_c\|)$.

(8): Using the cluster memberships in $\boldsymbol{U}$, construct the pairwise dissimilarity matrix $\boldsymbol{D}'$ between objects by defining $d'_{ij} = \min_{1 \le c \le k} \left\{ (1 - U_{ci})^2 + \left(1 - U_{cj}\right)^2 \right\}$

(9): Construct a new pairwise dissimilarity matrix $\boldsymbol{P}'$ between clusters by defining $p'_{ij} = \sum_{o \in i} U_{oj}$, and then apply the VAT algorithm to $\boldsymbol{P}'$ to obtain the new cluster order $\widetilde{\boldsymbol{P}}'$.

(10): In the order of clusters found in $\widetilde{\boldsymbol{P}}'$, identify the objects in $\boldsymbol{D}'$ to be sorted when $U_{ik} = \max_{1 \le k \le c}(U_k)$, and sort the datapoints within each cluster by s', resulting in the reordered dissimilarity matrix $\widetilde{\boldsymbol{D}}'$ and its corresponding greyscale image $I(\widetilde{\boldsymbol{D}}')$**.**

Output: Spectrally-mapped and reordered dissimiliarity matrix $\widetilde{\boldsymbol{D}}'$ and its corresponding greyscale image $I(\widetilde{\boldsymbol{D}}')$.

Similar to Wang, et al. (2008), Otsu's (1979) threshold selection method can be used on the output of SpecVCMV, $I(\widetilde{\boldsymbol{D}}')$, to determine the optimum number of clusters. The "goodness measure" (GM) is defined as as $GM(k) = \sigma^2{}_B(T_k^*)$ where $\sigma^2{}_B = \omega_1\omega_2(\mu_2 - \mu_1)^2$, $\mu_1 = \frac{\mu(T)}{\omega(T)}$, $\mu_2 = \frac{\mu_L - \mu(T)}{1 - \omega(T)}$,$\omega_1 = \sum_{l=1}^{T} p_l$, $\omega_2 = \sum_{l=T+1}^{L} p_l$, and $T_k^* = \operatorname{argmax}_{1 \le T \le L} \sigma^2{}_B(T)$, and so is maximising between-class variance in the histogram of $I(\widetilde{\boldsymbol{D}}')$ for each threshold (T). The optimum number of clusters is calculated according to $c = \operatorname{argmax}_{1 \le k} GM(k)$.

To aid in a more general understanding of the proposed VCMV algorithm and how it can be hybridised with Spectral Clustering, Table 6-2 compares VAT, VCV, the proposed VCMV, SpecVAT, how SpecVCV may be implemented, and the proposed SpecVCMV.

	VAT	VCV	VCMV
Clustering	None	External (e.g. k-means, FCM, FCRM)	Fuzzy (e.g. FCM, FCRM)
Inter-Cluster Sorting	None	Arbitrarily keep 1st cluster, calculate pair-wise distance in Euclidean space, sort remainder using MST as if cluster prototypes are Euclidean	Calulate pair-wise distance by sum of cluster membership, 1st cluster is one of two most distant clusters, sort remainder by MST
Intra-Cluster Sorting	Start with one of the two most distant points, sort using MST	If fuzzy clustering used, sort by cluster membership	Sort by Euclidean distance to cluster prototype
Visualise on	Euclidean pair-wise distance	Euclidean distance to cluster prototype	Cluster membership

Table 6-2 - Summary of VAT, VCV and VCMV

	SpecVAT	SpecVCV	SpecVCMV
Mapping to Spectral Space	Self-Tuning Spectral Clustering	Self-Tuning Spectral Clustering	Self-Tuning Spectral Clustering
Clustering	None	k-means from Spectral Clustering	Fuzzy Volumetric Cloud (e.g. FCM)
Inter-Cluster Sorting	None	Arbitrarily keep 1st cluster, using k-means cluster prototypes calculate pair-wise distance in Euclidean space, sort remainder using MST	Calulate pair-wise distance by sum of cluster membership, 1st cluster is one of two most distant clusters, sort remainder by MST
Intra-Cluster Sorting	Start with one of the two most distant points in Spectral space, sort using MST	None	Sort by Euclidean distance to FCM cluster prototype in Spectral Space
Visualise on	Euclidean pair-wise distance in Spectral Space	Euclidean distance to cluster prototype in Spectral Space	Cluster membership

Table 6-3 - Summary of SpecVAT, SpecVCV and SpecVCMV

It is of course necessary to demonstrate the effectiveness of the SpecVCMV algorithm. While there is no improvement in the other examples used so far, using the two examples that were chosen to highlight the limitations of the VCV best demonstrates the improvements of the proposed visualisation. Applying SpecVCMV to the example used in used in Figure 6-8 with c=5 results in a visualisation that is much more representative of the underlying FCM clustering and is shown in Figure 6-10. The light colour, particularly in the bottom right of each cluster block, successfully communicates the lack of low cluster memberships returned by the underlying

FCM algorithm for many of the points, caused by the inappropriate number of clusters selected for the purpose of exposing the assumptions built into the VCV algorithm.

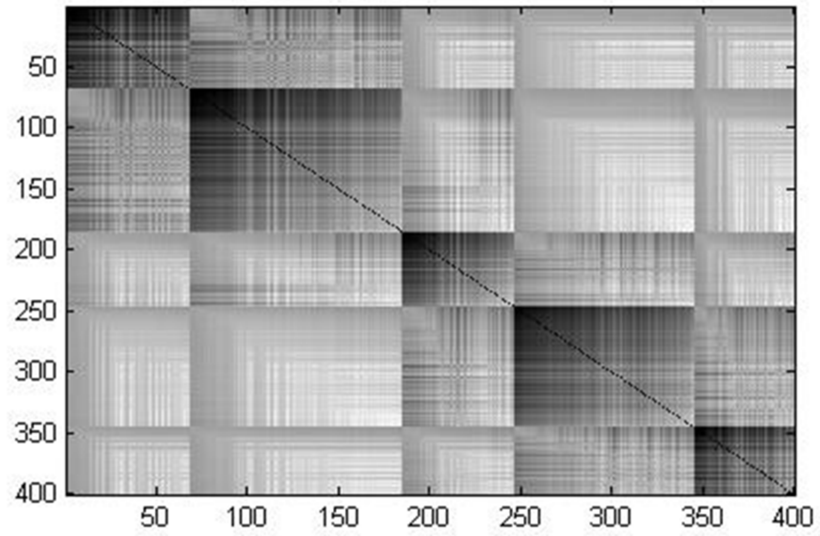
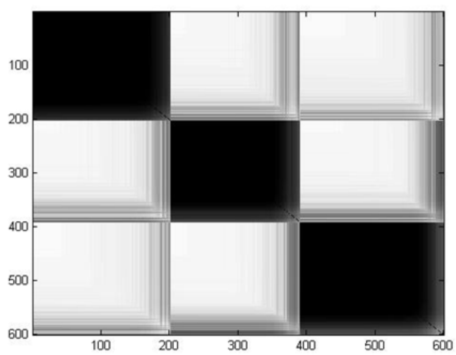


Figure 6-10 - SpecVCMV sorted eigenvectors (c=5)

Applying SpecVCMV to the example in Figure 6-9, now shown in Figure 6-11 shows that cluster homogeneity is now maintained, grey is shown in the visualisation when the FCM clustering algorithm indicates indecision in cluster membership, and the ADNC algorithm has successfully identified the number of clusters in the underlying data.



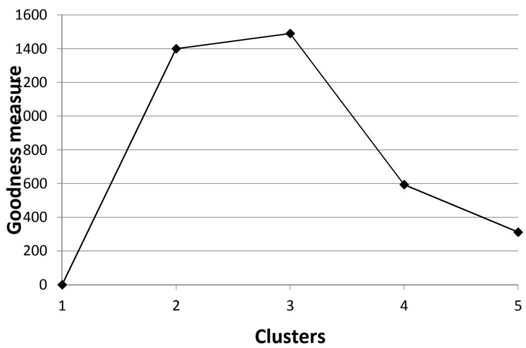


Figure 6-11 - SpecVCMV sorted eigenvectors (c=3) and GM for 1≤c≤5

When there is a large class separability there is little to no benefit in using SpecVCMV over SpecVAT as the only visible difference will be the cluster ordering, this makes it difficult to demonstrate the superiority of the technique using a dummy dataset. In cases where the clusters are more ambiguous, such as the examples used here, the cluster membership confidence is more clearly communicated, and the grouping technique increases confidence that microclusters have been identified.

The datasets used in Wang et al. (2008) have been replicated here. Originally from Zelnik-Manor & Perona (2004), the first 6 “synthetic” datasets can be downloaded from the Internet, while the “real” datasets are available through the UCI Machine Learning Repository. Wang et al. (2008) describes the real datasets as follows:

> *a) R-1: Breast-cancer database includes 699 instances, each of which has 9 attributes and belongs to one of 2 classes. Since there are 16 instances that contain a single missing attribute value, we removed them and used the remaining 683 instances for our experiment. b) R-2: This data set was used in [4]. It contains single light- source Face images of 3 different individuals, each seen under 585 viewing conditions. Each original image was down-sampled to 30×40 pixels, thus providing in total 1755 images with 1200 dimensions (i.e., 30 × 40). c) R-3: Genetic data set is originally from the work in [18], which is a 194 × 194 matrix consisting of pair-wise dissimilarities from a set of 194 human gene products that were clustered into three protein families. d) R-4: Iris data set contains 3 physical classes, 50 instances each, where each class refers to a type of iris plant and the attributes of each instance include 4 numeric values. e) R-5: Voting data set consists of 435 US House of Representatives members’ votes on 16 key votes (267 democrats and 168 republicans). Votes were numerically encoded as 0.5 for “yea”, -0.5 for “nay” and 0 for “unknown disposition”, so that the voting record of each congressman is represented as a ternary-valued vector in R16. f) R-6: Wine data set contains the results of a chemical analysis of*

wines grown in the same region, but derived from 3 different cultivars. The analysis determines the quantities of 13 constituents found in each of three types of wines. The total number of instances is 178.

Figure 6-12, Figure 6-13, Figure 6-14 and Figure 6-15 now demonstrate the SpecVCMV algorithm on those same datasets, in comparison to the SpecVAT algorithm.

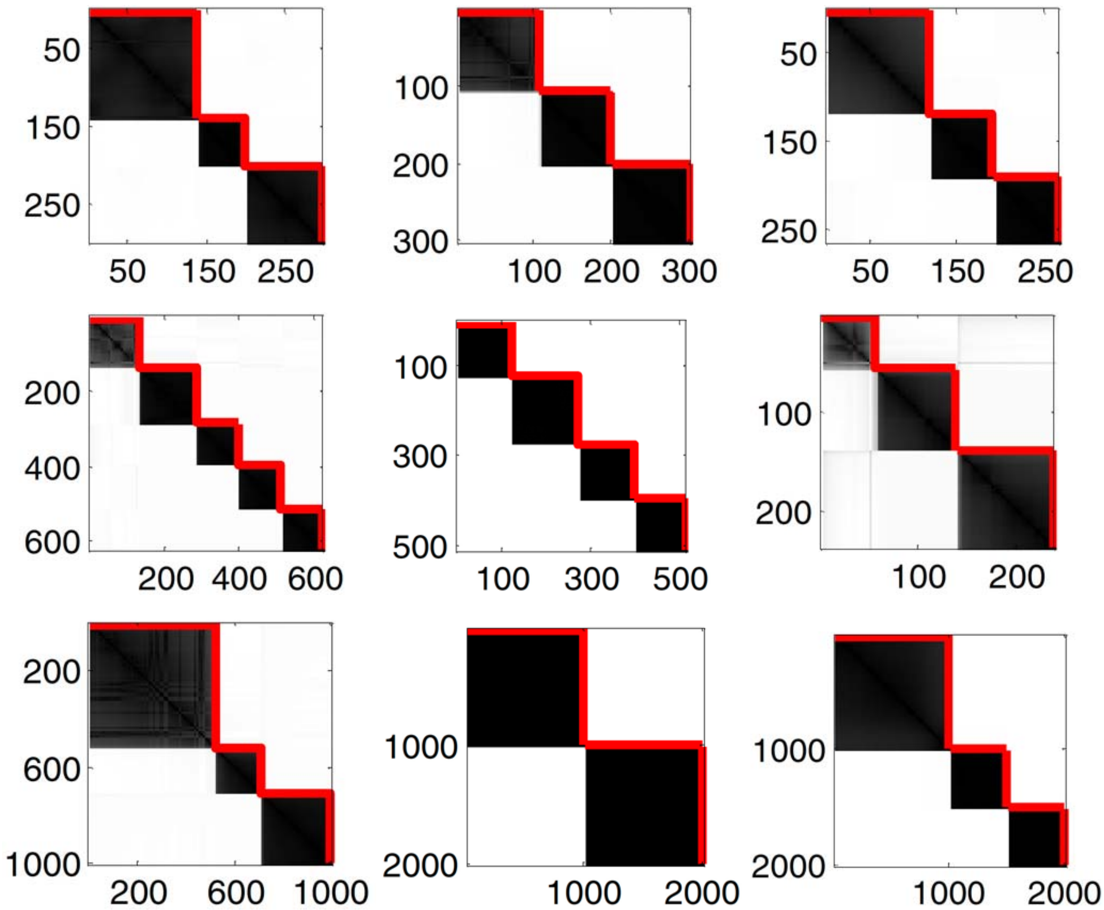


Figure 6-12 - SpecVAT images of synthetic data S-1 through S-9 (Wang, et al., 2008) (note red are present in Wang but are not part of the algorithm)

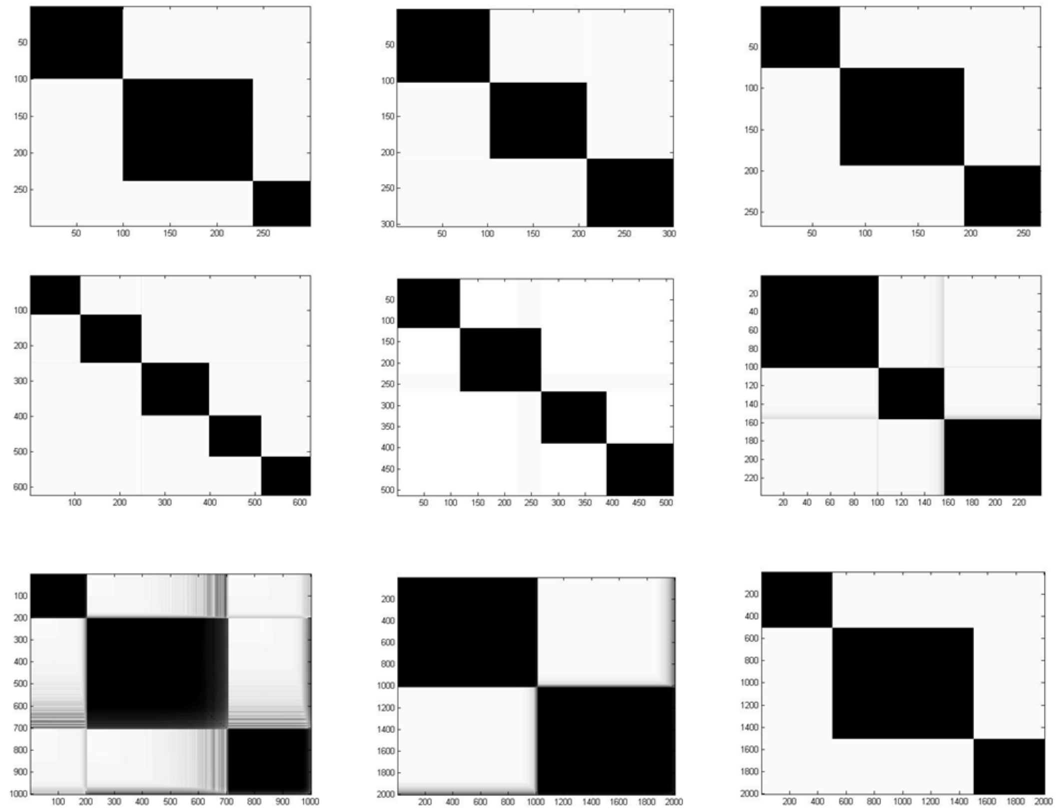


Figure 6-13 - SpecVCMV images of synthetic data S-1 through S-9

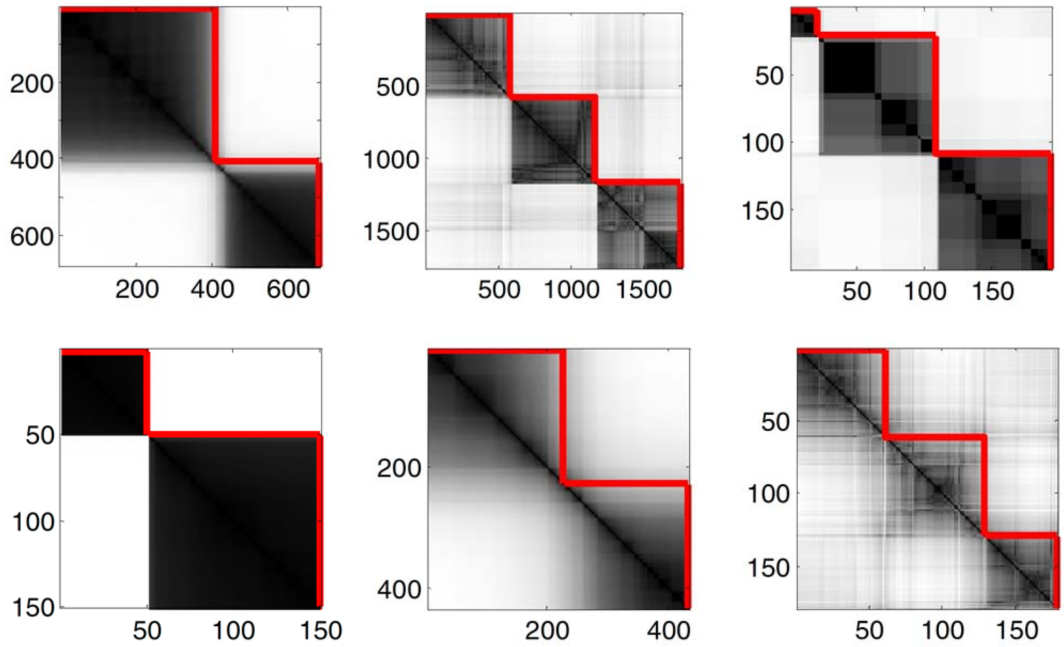


Figure 6-14 - SpecVAT images of real data R-1 through R-6 (Wang, et al., 2008)

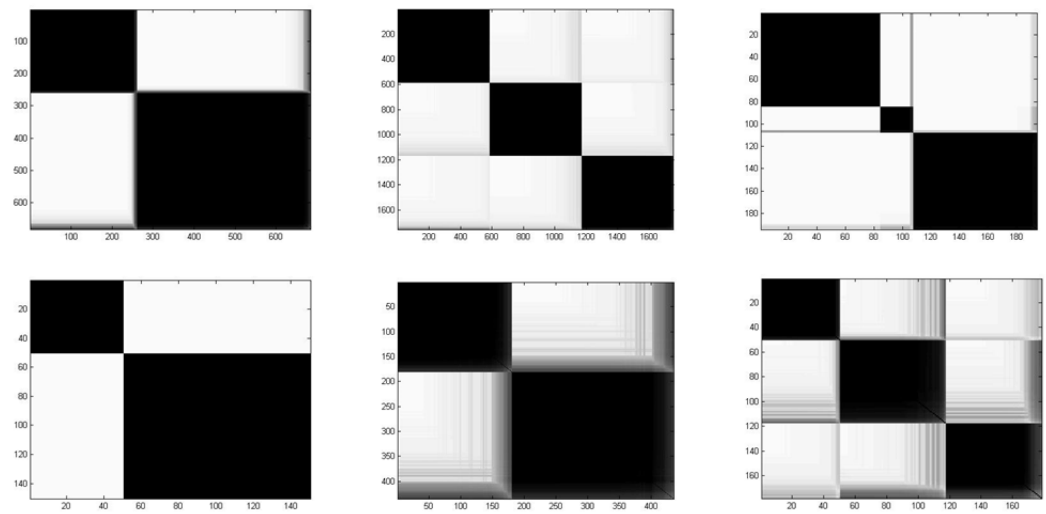


Figure 6-15 - SpecVCMV images of real data R-1 through R-6

Particular attention should be paid to R-4 and R-5 in which the SpecVCMV visualisations are comparatively crisp, though the large sections of grey indicate some low clustering confidence on a number of datapoints. This demonstrates that there are some data points in these sets that have a high Euclidean distance to their respectively cluster prototypes, but that the external clustering algorithm is confident in the choice of cluster nevertheless. R-3 contains a grey lines in the middle and bottom right clusters, the SpecVCMV algorithm is communicating that there are some data points that have a high cluster membership to a competing cluster, allowing for identification of possible miss-classified data points from the Spectral Clustering algorithm.

The chaining phenomenon noted earlier becomes more problematic as the number of clusters increase, however the above cases demonstrate no more than three clusters. However as the number of clusters increase the chaining phenomenon becomes more of an issue but it becomes increasingly difficult to visualise as the eigenspace dimensionality also increases. A good indicator of the chaining phenomenon is when the number of dark blocks in the visualisation is larger than the parameter c, as is the case with Figure 6-16's UCI "image segmentation" dataset (Frank & Asuncion, 2010) – which contains 7 underlying classes. This 19 parameter, 2310 instance dataset is highly complex. As can be seen in the figure, a number of spurious divisions have taken place as a result of chaining, creating a very noisy visualisation. By comparison, the SpecVCMV algorithm visualisation more clearly shows how the clustering algorithm has segmented the data.

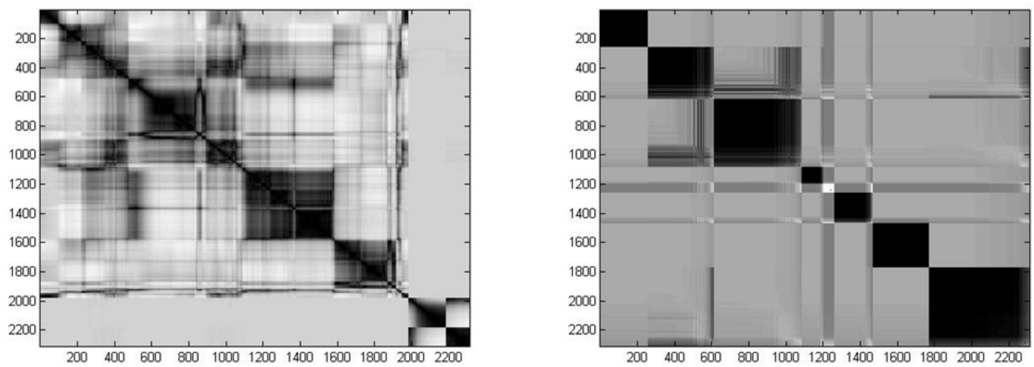


Figure 6-16 - SpecVAT and SpecVCMV for "image segmentation" dataset (c=7)

6.1.4 Conclusion

The SpecVAT algorithm (Wang, et al., 2008) is a powerful method for visualising hyper-dimensional data, however the algorithm introduces limitations found in the minimum spanning trees (MST) algorithm. Using VCV, the generalisation and extension of VAT is the natural progression for this line of work, but as demonstrated the VCV method builds in underlying assumptions about the data that may not hold true. This section has demonstrated that an extension of the VCV algorithm, VCMV, addresses those assumptions and when combined with Spectral Clustering, SpecVCMV represents a powerful visualisation method that can successfully visualise the clustering of both synthetic and real datasets.

6.2 The Effect of Objective Clustering

At this stage it is worth re-capping the purpose of grouping companies together. As has been discussed in section 3.2.5, there is a need to identify companies that behave similarly on the basis that such companies are likely to exhibit similar symptoms of failure. Certainly the literature review in chapter 2 showed that researchers tend to build classification models only within a particular industry. This kind of stratification shows that it is commonly accepted that company similarity is a useful quality in a classification model.

To utilise a grouping technique it is necessary to answer the question, what makes a company "similar" in a given year versus another company in a given year? The concept of similarity is defined by the *distance function*. That is, if the distance function were defined to be related to the letters in the company code, the companies ALU and ALW may be considered more "similar" than to the company ZCP. Clearly such a distance function is not going to achieve the goal of finding companies that are financially similar. Interested readers are directed to the ugly duckling theorem (Watanabe, 1985), where the clustering of three one-dimensional data points {0, 1, 1} can be performed by grouping data points based on their location within the set, or alternatively by their value, with equally valid (but opposite) cluster results. The need to select a useful distance function, and therefore the selection of appropriate inputs, is highlighted. Fortunately, chapter 4.3 has provided a cross-section of factors that have demonstrated their relationship to corporate health for each dataset in use.

There is a multitude of clustering techniques available, from simple algorithms such as the k-means algorithm through to intelligent systems approaches such as Winner-Takes-All Neural Networks or Self-Organising Feature Maps. Section 6.1 for example utilised Fuzzy C-Means, Fuzzy C-Regression Models as well as Spectral Clustering (which typically uses k-means on the spectrally mapped data). One of the issues in clustering this dataset is its size. For example, Spectral Clustering requires the affinity matrix and Laplacian matrix to be calculated from a dissimilarity matrix. A moderately sized 50,000 company-years stored as 64-bit floating point numbers would require 18.63 gigabytes of random access memory (RAM) just to store the pairwise distance matrix. Furthermore, such a clustering technique would only identify similarity between company-years at given instants in time, instead of clustering the behaviour of a company over time.

To address these issues, Deboeck & Kohonen (1998) propose mapping each company-year of data to a cell by training a Neural Network based Self-Organising Map, the first-level SOM, then using the co-ordinates from the first-level SOM at different time periods to train a second-level SOM. In doing so, the clustering of companies becomes based on similar movement across the first-level SOM rather than on a single point in time. This methodology will be outlined in the following section.

6.2.1 Clustering with a Two-Level SOM

As noted in section 4.2.4, Self-Organising Maps are a type of unsupervised Neural Network that uses a “Winner Takes Most” algorithm to strengthen the synapses that activate the winning output neuron (and its neighbours) for each set of inputs, such that similar sets of inputs activate the same or nearby output neurons. Being a Neural Network, each case is presented to the Self-Organising Map individually and the learning algorithm applied, rather than performing matrix operations on the entire dataset. This allows large datasets to be processed without requiring enormous computational resources, making Neural Networks a suitable choice for an unsupervised learning algorithm.

However, as outlined in Deboeck & Kohonen (1998), “it has been found that an analysis of one year’s financial statements is insufficient to give a reliable picture of the state of the company”. It is argued that it is necessary to use data from multiple years, and that this can be performed “by using two SOMs in a hierarchy” (Deboeck & Kohonen, 1998).

> *The first-level SOM is trained with yearly financial statements, so that for a given year a company can be positioned on the first-level SOM based on its financial statement for that year. The second-level SOM is then trained with the company’s coordinates on the first-level SOM during two or three consecutive years, as illustrated in [Figure 6-17]. This way, each unit on the second-level SOM corresponds to a trajectory on the first-level SOM, capturing one typical pattern of change in a company’s financial statements from year to year.* (Deboeck & Kohonen, 1998)

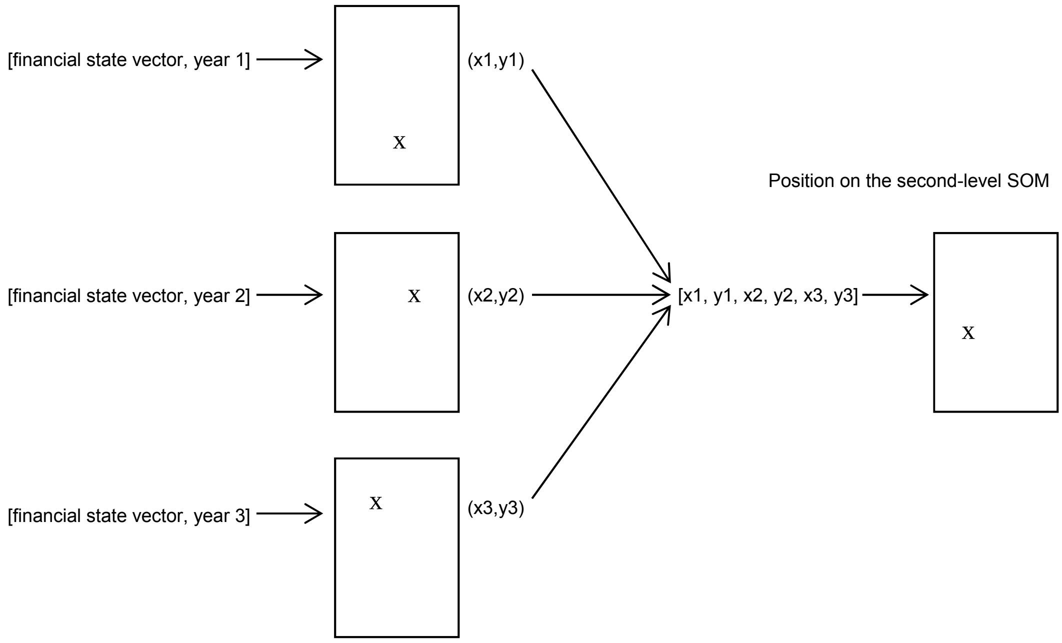


Figure 6-17 – Deboeck-Kohonen Multi Level SOM (the letter x is used to show a location)

Deboeck & Kohonen (1998) found that the second-level map exposed information that was not captured in the first-level map, and that there were areas from which failing companies generally did not leave. In summary, it was found that it is possible "to recognize different patterns of corporate behaviour", and so this chapter will utilise the Deboeck-Kohonen Multi-Level Self-Organising Map (DK-ML-SOM) as the basis of an objective clustering algorithm.

As demonstrated in section 4.3, the optimised subset of factors for each of the Genetic Programming and the Neural Network classification algorithms differs, and both the Genetic Programming and Neural Network algorithms were shown to increase classification accuracy when using a reduced subset of factors. Therefore two DK-ML-SOM's were trained for each dataset, one using the GP-optimised factors and one using the NN-optimised factors. The SOM's output neurons were then clustered using Spectral Clustering, visualised with SpecVCMV with the optimum number of clusters determined using the ADNC algorithm, with the cases assigned to clusters according to which output was most activated by the SOM. The DK-ML-SOM clustering methodology used requires a company to be clustered according to its movement over 3 years. To allow comparisons of the effect of objective clustering against alternative grouping methods outlined in this chapter, a dataset that contains only company-years with 3 consecutive years of data is used in these experiments, representing 81.1% of available Compustat data and 60.9% of available Aspect Data.

In order to fairly compare the effects of the DK-ML-SOM, the commonly used "industry grouping" method will be used as the basis for comparison. There are many industry classification schemes used across the globe. The United States, and the Compustat Legacy Global dataset for instance, primarily uses NAICS, the North American Industry Classification System. NAICS uses an industry-specific description to define which companies will be grouped together. For example, "The Agriculture, Forestry, Fishing and Hunting sector comprises establishments primarily engaged in growing crops, raising animals, harvesting timber, and harvesting fish and other animals from a farm, ranch, or their natural habitats", while "The Mining, Quarrying, and Oil and Gas Extraction sector comprises establishments that extract naturally occurring mineral solids, such as coal and ores; liquid minerals, such as crude petroleum; and gases, such as natural gas" (Census Bureau, 2007).

Not being based in the United States, the Aspect dataset uses GICS, the Global Industry Classification Standard. GICS is developed by Standard & Poor's, and for the purposes of this chapter serves an equally useful purpose in grouping companies by industry.

Both NAICS and GICS are taxonomical systems that rely on the manual assignment of companies to a particular industry code, based on information such as reported sources of revenue. These systems therefore represent an opportunity to test the hypotheses outlined in section 3.2.5.

It is also necessary to test the effect of both objective clustering of company-years and the industry groupings assigned through NAICS or GICS against the classification accuracy of not performing any grouping at all. As outlined in 3.2.5, dividing the data in any way has the side effect of reducing the number of cases available for training of the classifier and if the groups to classify are too small then the classification system may be unable to establish relationships between dependent and independent variables. Therefore the very act of dividing the dataset potentially has an impact on classification ability that requires investigation.

Once clustering has taken place, whether by the objective DK-ML-SOM or by industry grouping, the use of the forward best-first search outlined in 4.3.6 will be used to ensure that comparisons are drawn between the best combination of factors for each cluster or group, rather than assuming that the best factors before dividing the data would remain the best choice of factors in each subdivision.

6.2.2 Experimental Results

The first-level DK-ML-SOM's were initialised using normalised factors that had resulted in the highest in-sample validation accuracy from section 4.3. Hexagonal map units were used which were initialised using the largest eigenvectors of the input dataset, and the SOM was trained using the Batch training algorithm (Kohonen, 1995).

Using the GP-optimised factors on the Compustat dataset this resulted in a 42 by 110 map. This map is visualised in Figure 6-18 with the Euclidean distance from each unit to its neighbours shown in the U-matrix, along with the values for each input factor for each output neuron on the map.

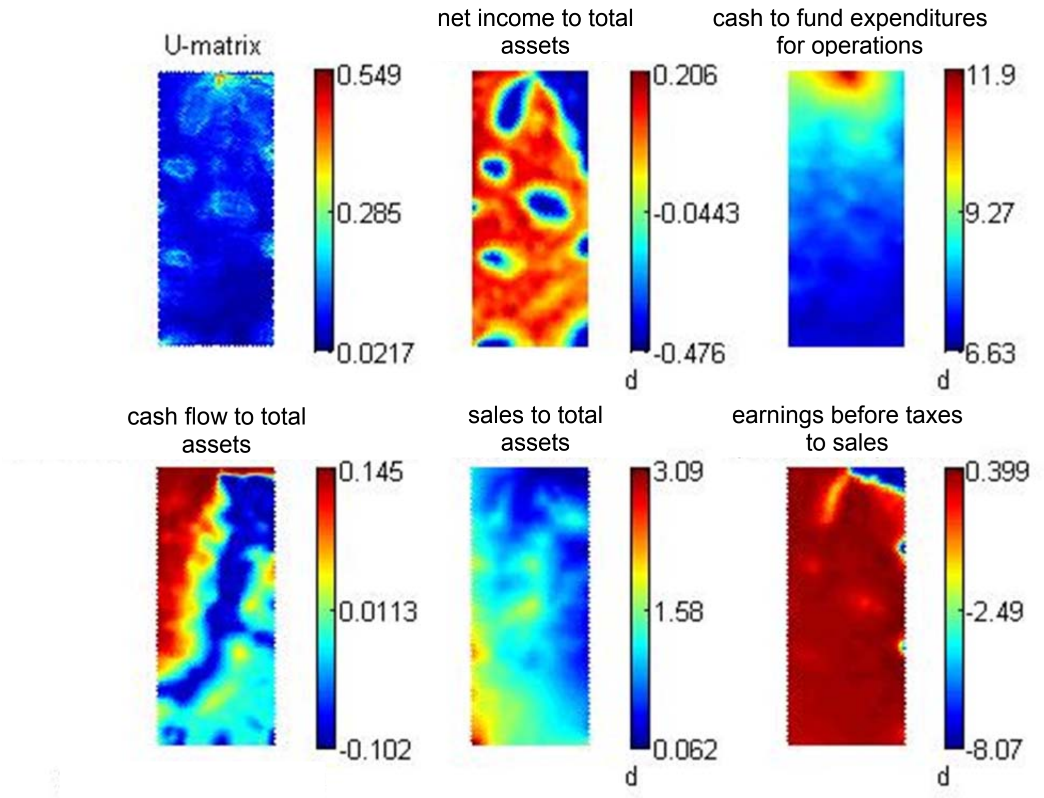


Figure 6-18 – Level 1 SOM of GP-optimised Factors on Compustat Dataset

The above map can be interpreted in quadrants: The top left quadrant generally has high cash flow to total assets, the top right quadrant generally has low net income to total assets and low earnings before taxes to sales, the bottom left quadrant has high sales to total assets, and the bottom right quadrant has moderate sales to total assets with moderate cash flow to total assets.

The output coordinates (L1-x, L1-y), the coordinates from 12 months prior (L1-x minus1, L1-y minus1), and the coordinates from 24 months prior (L1-x minus2, L2-y minus2) were then trained on the second level DK-ML-SOM, resulting in the Figure 6-19, with the clustering and SpecVCMV visualisation shown in Figure 6-20. It is expected that the visualisations of how the input factors are mapped to the output neurons are similar for each year, due to the fact that while individual companies may move around the map the overall profile of the market doesn't drastically change from year to year.

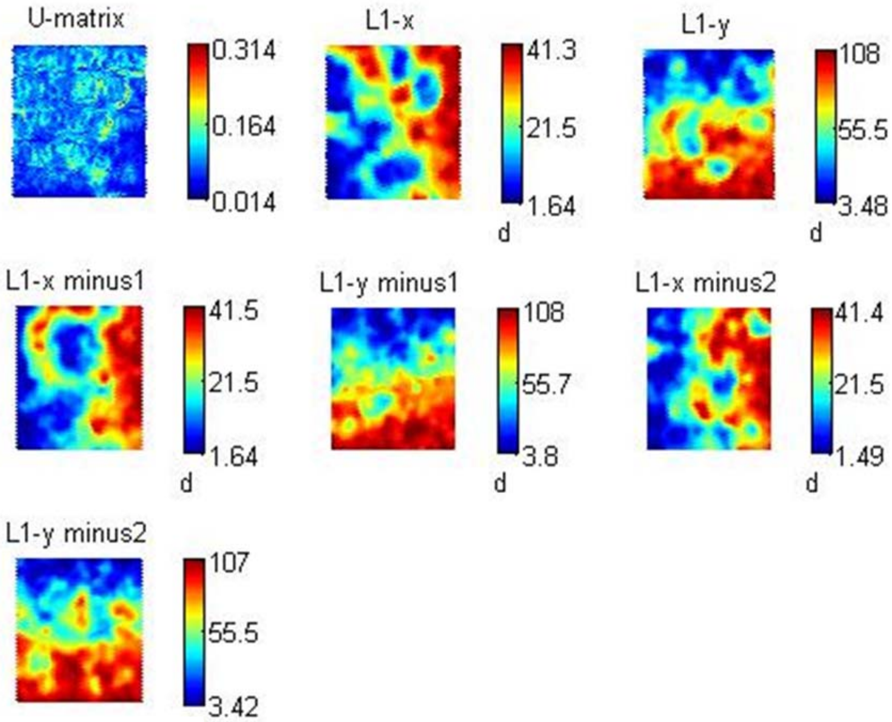


Figure 6-19 - Level 2 SOM of GP-optimised Factors on Compustat Dataset

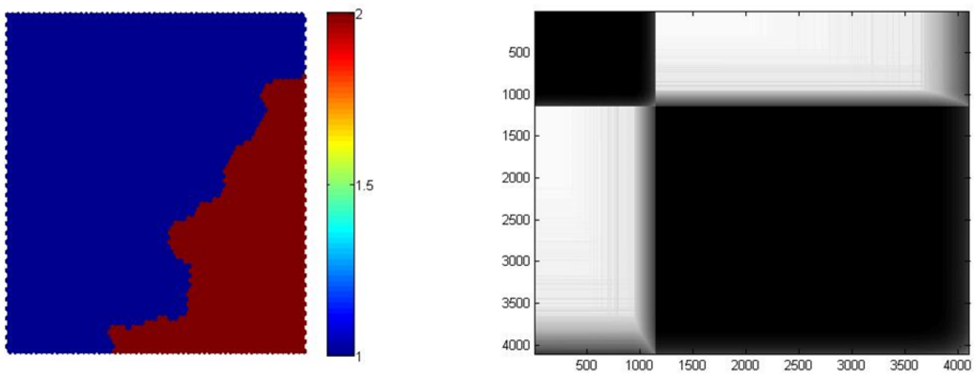


Figure 6-20 - Level 2 Cluster Map & SpecVCMV of GP-optimised Factors on Compustat Dataset

The SpecVCMV algorithm has found a fairly clear clustering, though the results are now much more abstract. A company in the bottom right corner of the level 2 SOM has had coordinates around (41, 107) for 3-years running (low net income to total assets, low cash to fund expenditures from operations, low cash flow to total assets, medium sales to total assets and high earnings before taxes to sales), while a company in the bottom left corner has had coordinates around (1, 108) for 3 years running (high net income to total assets, high sales to total assets and earnings before taxes to sales, low cash to fund expenditures from operations and low cash flow to total assets), companies in the other areas of the level 2 map represent companies that fluctuate more around the level 1 map over the 3 year period. The SpecVCMV algorithm has identified a band of map units (approximately 200 units) for which the fuzzy c-means algorithm is indicating indecision. While this could identify a poor choice of clusters, the ADNC algorithm showed that clusters greater than 2 created greater indecision in the fuzzy c-means algorithm and resulted in a SpecVCMV histogram with less contrast.

Each company-year was then assigned to one of the two clusters by determining the best matching map unit and allocating the unit's cluster number, the resulting clusters were trained separately using the Genetic Programming algorithm within the forward best-first accuracy based factor search methodology outlined in section 4.3, the numerical results for which are available in Appendix Z. The Genetic Programming algorithm achieved an in-sample accuracy of 76.5% and 84.5% on each cluster, with 77.4% and 83.4% on the out-of-sample. Combined, the clustered GP-optimised Compustat dataset achieved a weighted accuracy of 78.9% with 79.2% on the out-of-sample.

Similarly the Neural Network-optimised factors were clustering using the DK-ML-SOM, resulting in the following maps (Figure 6-21 and Figure 6-22), cluster map and SpecVCMV (Figure 6-23).

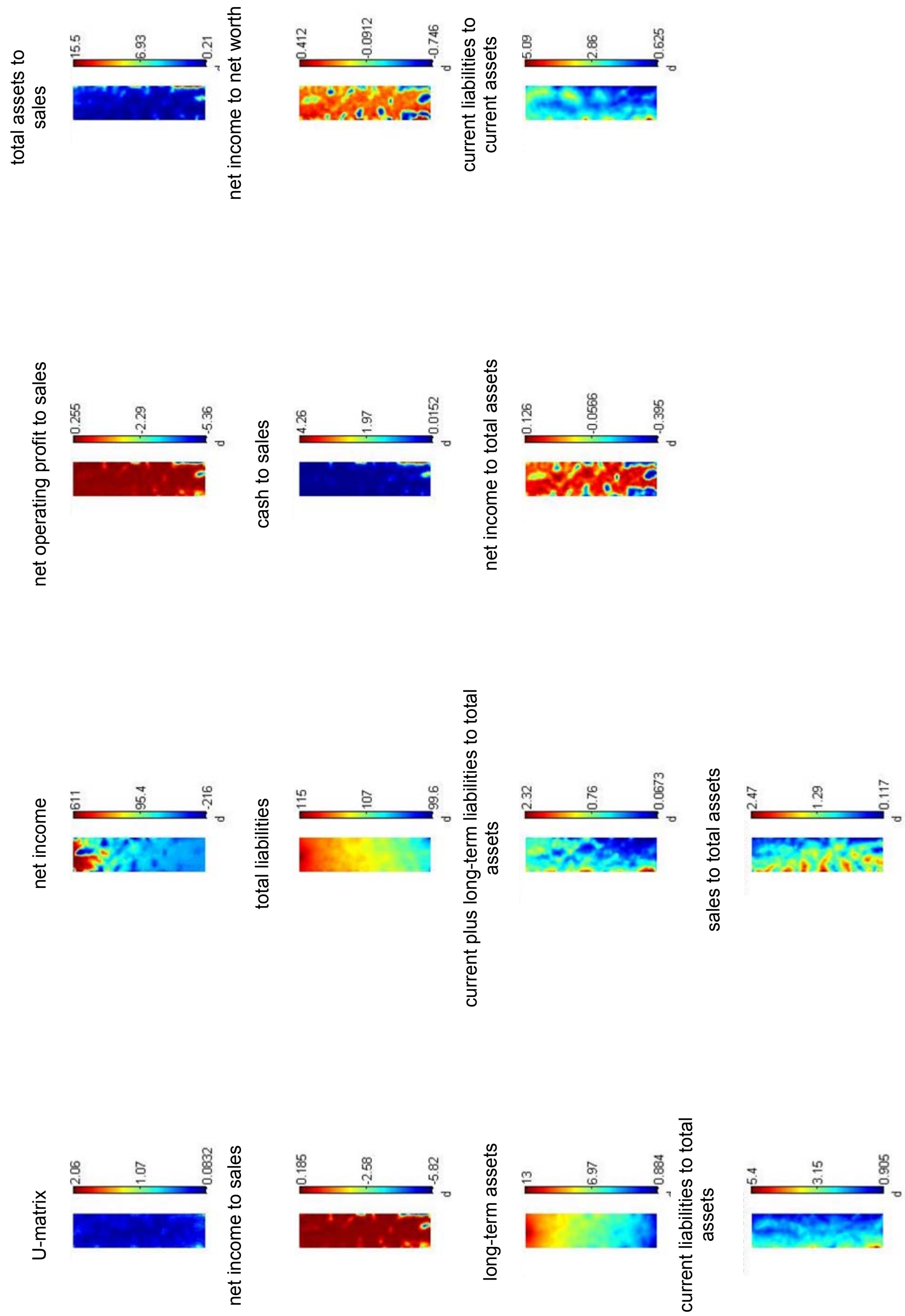


Figure 6-21 - Level 1 SOM of NN-optimised Factors on Compustat Dataset

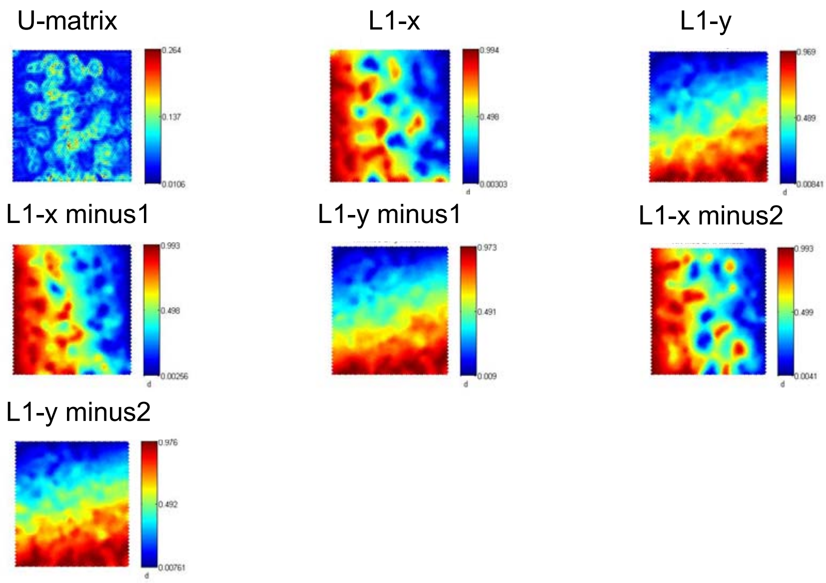


Figure 6-22 - Level 2 SOM of NN-optimised Factors on Compustat Dataset

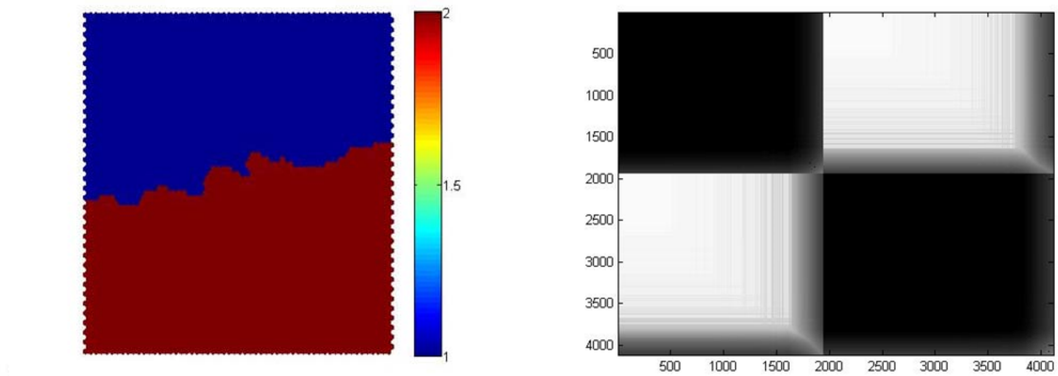


Figure 6-23 - Level 2 Cluster Map & SpecVCMV of NN-optimised Factors on Compustat Dataset

The clustered cases were then presented to the Neural Networks based best-first forwards search, found in Appendix AA, resulting in an in-sample accuracy of 89.5% and 79.1% with 73.5% and 80.6% respectively on the out-of-sample. Combined, the clustered NN-optimised Compustat dataset achieved a weighted accuracy of 84.0% with 77.3% on the out-of-sample.

As the ADNC algorithm determined the optimum number of clusters to be 2 in both the GP and NN-optimised models, it was decided to use NAICS industry codes to also divide the data into two groups to provide the best comparison. The most natural division in the data appeared to be along the line of manufacturing versus non-manufacturing, so company-years with a NAICS code of 31 through 33 were included in group 1 and the rest were placed into group 2. When using Genetic Programming in a best-first forward search, an accuracy of 80.8% (manufacturing) and 80.5% (non-manufacturing) was achieved on the in-sample data, with 78.5% and 67.6% on the out-of-sample. Combined, the grouped GP-optimised Compustat dataset achieved a weighted accuracy of 80.6% with 71.6% on the out-of-sample. When using Neural Networks in a best-first forward search, an accuracy of 84.2% and 81.1% was achieved with 72.2% and 68.9% on the out-of-sample, resulting in a net weighted accuracy of 82.2% with 73.8%. In both the Genetic Programming and Neural Network best-first forward searches, the clustered data outperformed the industry grouped data on the out-of-sample dataset. Both the Genetic Programming and Neural Network based best-first forward search results are found in Appendix BB.

Finally for comparison purposes, the unclustered and ungrouped data was re-optimised using the same best-first forward search, with Genetic Programming resulting in an in-sample accuracy of 77.6% with 74.5% ($p<0.01$) on the out-of-sample, and with Neural Networks resulting in 77.7% and 72.0% ($p<0.01$). Interestingly the Genetic Programming unclustered and ungrouped results here outperformed the dataset that was grouped by industry, showing that at least in some scenarios the act of grouping by industry can inhibit the performance of a classification system. The results of this best-first forward search can be found in Appendix CC.

The results are summarised in the table Table 6-4:

Method	Dataset	In-Sample	Out-of-Sample
Genetic Programming	Unclustered & Ungrouped	77.6	74.5
Genetic Programming	Cluster 1	76.5	77.4
Genetic Programming	Cluster 2	84.5	83.4
Genetic Programming	Manufacturing	80.8	78.5
Genetic Programming	Non-Manufacturing	80.5	67.6
Neural Networks	Unclustered & Ungrouped	77.7	72.0
Neural Networks	Cluster 1	89.5	73.5
Neural Networks	Cluster 2	79.1	80.6
Neural Networks	Manufacturing	84.2	82.2
Neural Networks	Non-Manufacturing	81.1	68.9

Table 6-4 - Summary of Objective Clustering on Compustat Dataset

Within the Aspect dataset, the first-level DK-ML-SOM's were initialised using normalised factors that had resulted in the highest in-sample validation accuracy using Genetic Programming from chapter 4.3, resulting in a 20 by 97 map as shown in Figure 6-24.

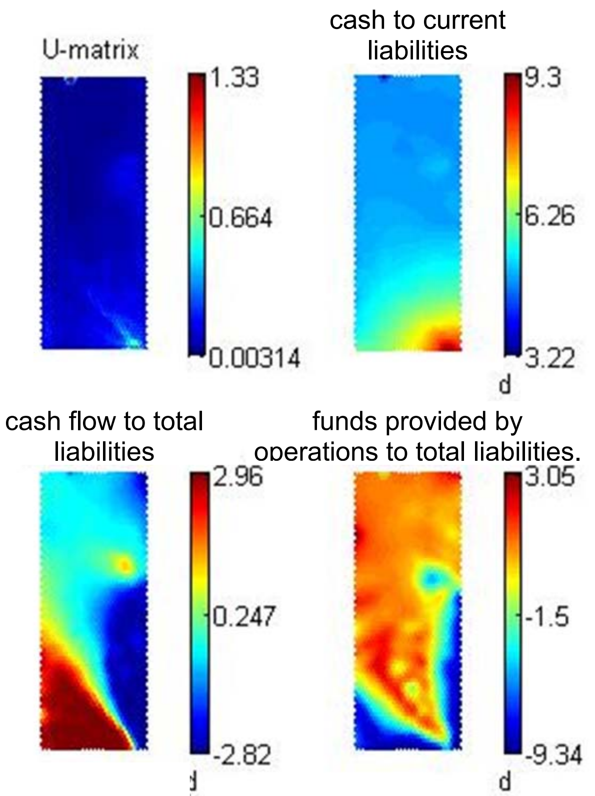


Figure 6-24 – Level 1 SOM of GP-optimised Factors on Aspect Dataset

In turn the 3-year map coordinates were used to train the second level SOM, resulting in the Figure 6-25 and Figure 6-26's map, clusters, and SpecVCMV.

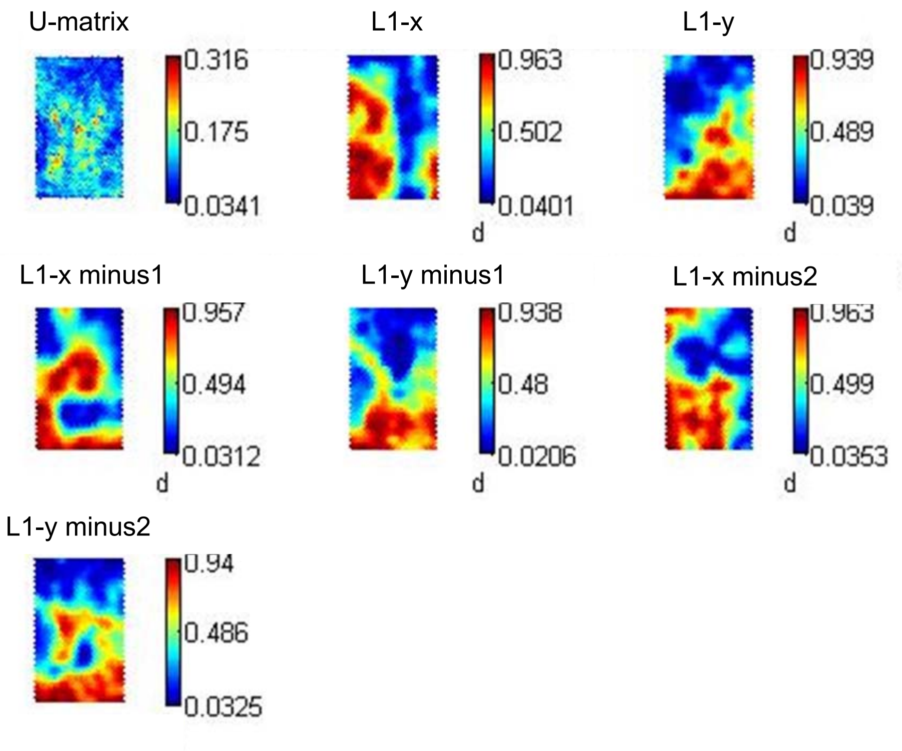


Figure 6-25 - Level 2 SOM of GP-optimised Factors on Aspect Dataset

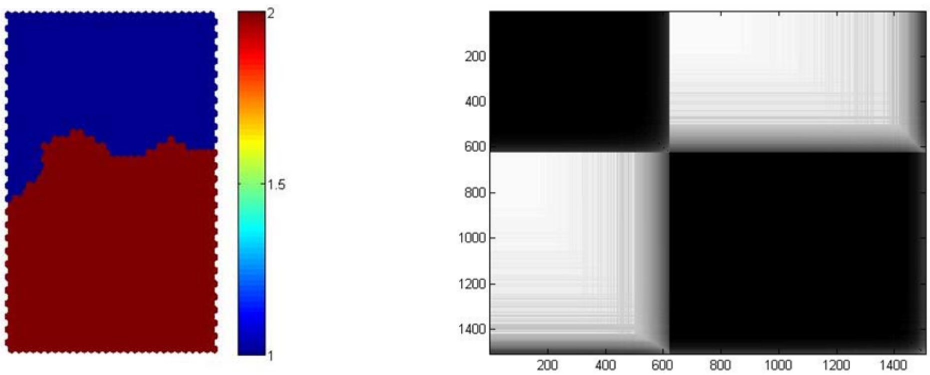


Figure 6-26 - Level 2 Cluster Map & SpecVCMV of GP-optimised Factors on Aspect Dataset

The clustered data was then optimised using the GP best-first forwards search, the results for which can be found in Appendix DD, resulting in an in-sample accuracy of 81.6% and 71.6% on each of the two clusters, and an out-of-sample accuracy of 65.3% and 68.4% respectively. When combined, this results in a weighted net accuracy of 75.9% with 67.1% on the out-of-sample.

The process was repeated using the NN-optimised factors, resulting in the Figure 6-27, Figure 6-28 and Figure 6-29's first and second level map, and second level clusters with SpecVCMV.

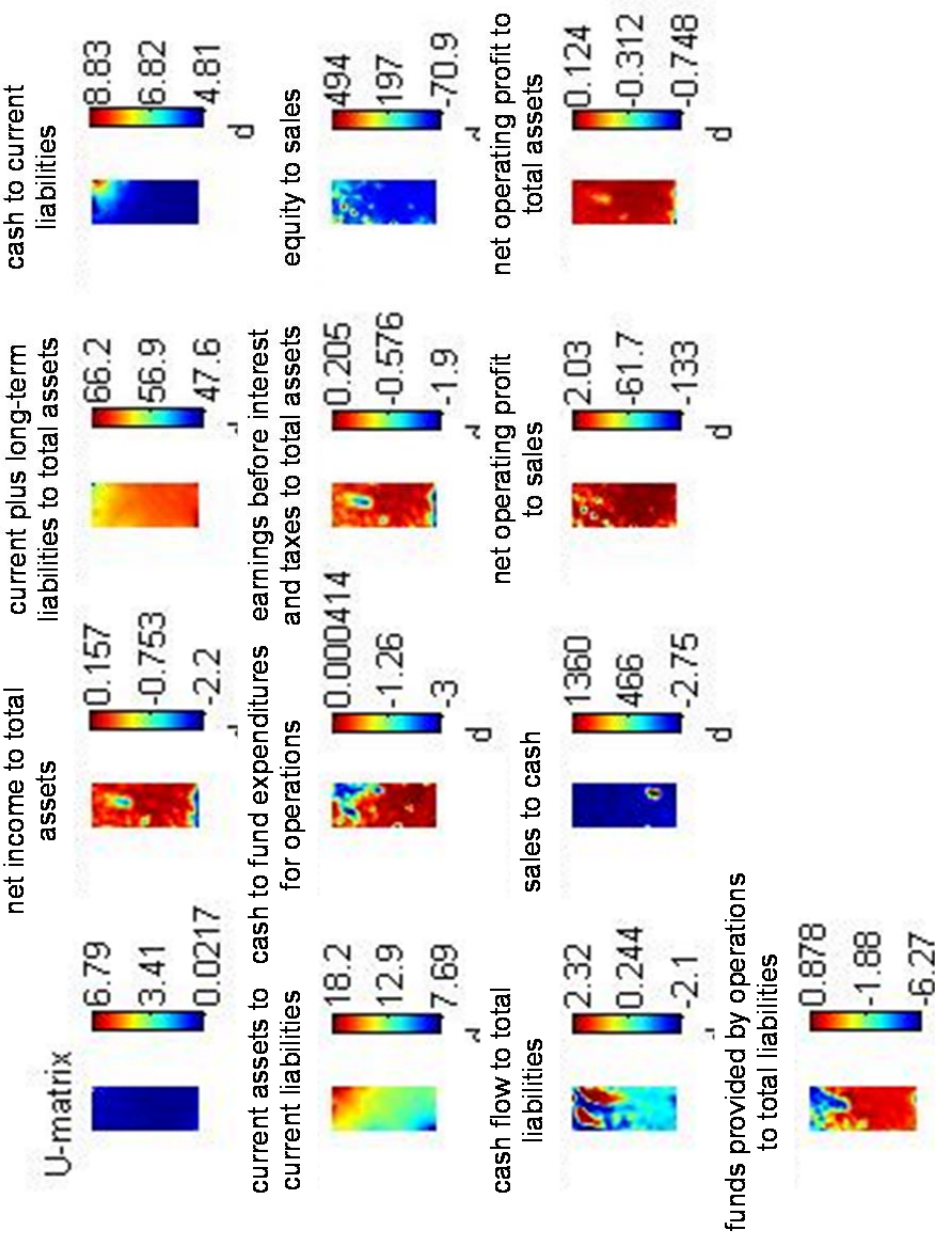


Figure 6-27 – Level 1 SOM of NN-optimised Factors on Aspect Dataset

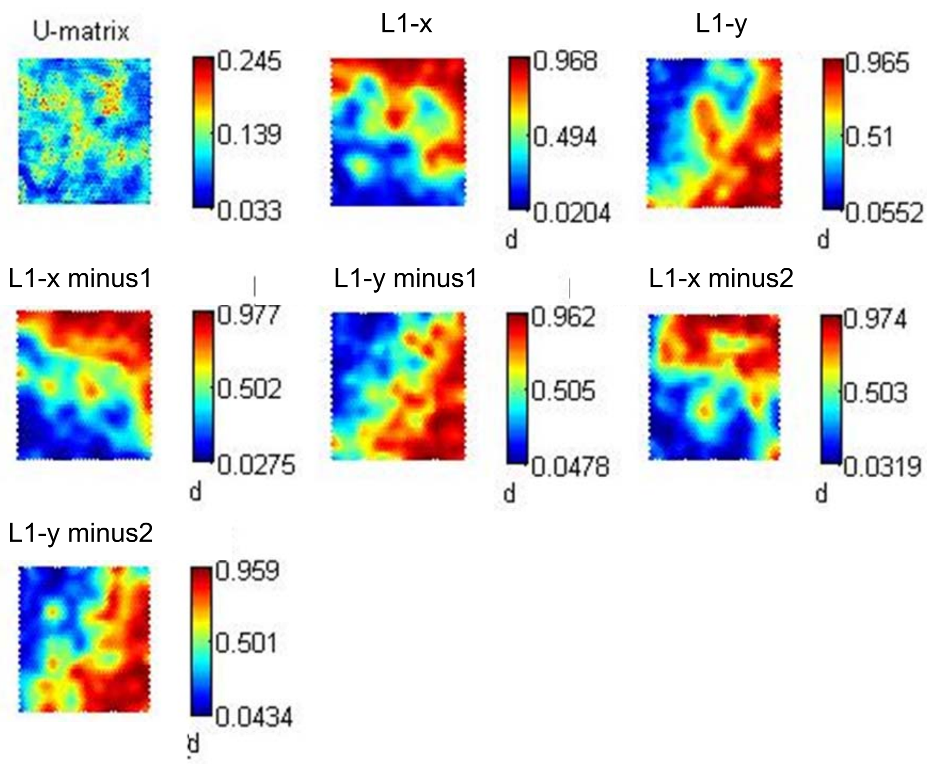


Figure 6-28 – Level 2 SOM of NN-optimised Factors on Aspect Dataset

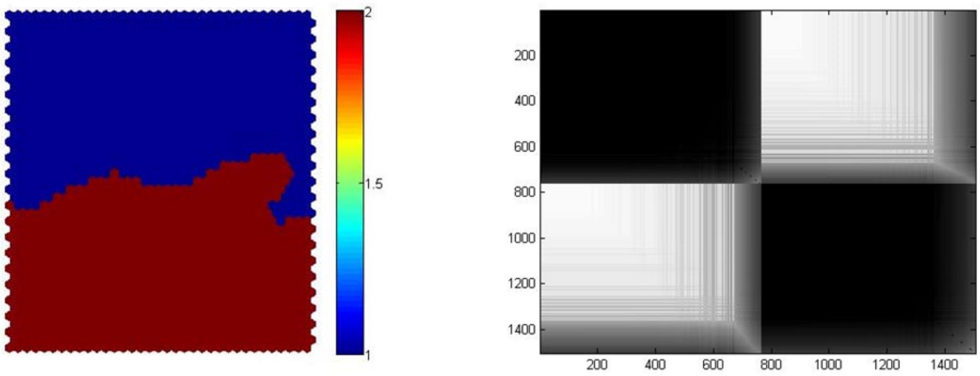


Figure 6-29 - Level 2 Cluster Map & SpecVCMV of NN-optimised Factors on Aspect Dataset

Once company-years had been clustered according to the cluster of their best matching SOM unit, each cluster was processed using the forward best-first algorithm, see Appendix EE, yielding an in-sample accuracy of 74.7% and 71.9% on the in-sample set and 62.2% and 60.7% on the out-of-sample. Combining the two clusters results gives a weighted in-sample accuracy of 73.4% with 61.5% on the out-of-sample.

Again the ADNC algorithm found the greatest SpecVCMV contrast with 2 clusters, so the Aspect dataset used for these experiments was similarly divided into manufacturing and non-manufacturing company-years by using the GICS sectors provided in the Aspect Data. Again the forward best-first algorithm was applied to both grouping using both Genetic Programming and Neural Networks, the results for which can be found in Appendix FF and Appendix GG. Using Genetic Programming, the industry grouped data achieved 71.7% and 70.8% on the in-sample with 59.1% and 65.5% on the out-of-sample, resulting in a combined weighted accuracy of 71.2% with 62.7% on the out-of-sample, a large decrease in comparison to the objectively clustered result. When using Neural Networks, the grouped data achieved 65.5% and 70.4% on the in-sample with 56.8% and 59.4%, resulting in a combined weighted accuracy of 68.3% on the in-sample and 58.3%, again a decrease in comparison to the objectively clustered data.

Finally the forward best-first algorithm is applied to the same dataset that has not been clustered or grouped, using both Genetic Programming and Neural Networks, the results for which can be found in Appendix HH. The best in-sample accuracy achieved was 72.9% with 70.8% on the out-of-sample when using Genetic Programming ($p<0.01$), and 69.3% with 61.9% on the out-of-sample when using Neural Networks ($p<0.01$). While the out-of-sample results are approximately equal to that of the clustered dataset when using Neural Networks, the clustered dataset's out-of-sample accuracy is noticeably increased for Genetic Programming when using the clustered dataset.

The results are summarised in Table 6-5.

Method	Dataset	In-Sample	Out-of-Sample
Genetic Programming	Unclustered & Ungrouped	72.9	60.8
Genetic Programming	Cluster 1	81.6	65.3
Genetic Programming	Cluster 2	71.7	68.4
Genetic Programming	Manufacturing	71.7	59.1
Genetic Programming	Non-Manufacturing	70.8	65.5
Neural Networks	Unclustered & Ungrouped	69.3	61.9
Neural Networks	Cluster 1	74.7	62.2
Neural Networks	Cluster 2	71.9	60.7
Neural Networks	Manufacturing	65.5	56.8
Neural Networks	Non-Manufacturing	70.4	59.4

Table 6-5 - Summary of Objective Clustering on Aspect Dataset

6.2.3 Conclusion

This chapter has found that for both datasets using both Genetic Programming and Neural Networks, objectively clustering data using a combination of Doeboeck-Kohonen Multi-Level Self-Organising Maps and Spectral Visual Cluster Membership Validity significantly improves out-of-sample accuracy when the forward best-first search for an objective factor set is used.

Furthermore this chapter has found that there is little to no benefit to out-of-sample accuracy when arbitrarily dividing the dataset into manufacturing versus non-manufacturing companies based on the NAICS or GICS industry classification. It is possible that a different arbitrary division of the data such as retail versus non-retail, or greater divisions of the data such as mining versus utilities versus construction versus manufacturing etc. would have yielded a higher out-of-sample accuracy than the manufacturing versus non-manufacturing grouping used in this section, but a true application of domain knowledge for grouping of companies should also consider company asset size, number of employees, geographical location and so-on, which creates an enormous burden of work and was the very motivation for introducing an objective clustering procedure.

7. Results

This thesis has shown that across two datasets, both Genetic Programming and Neural Networks are powerful classification methods. It has found that out-of-sample classification accuracy can be increased by first performing a non-deterministic embedded accuracy-based factor search, then refining those factors using a forward best-first accuracy-based factor search. It has been demonstrated that including share market information or macroeconomic data does not reliably increase out-of-sample accuracy. It has proposed a new method of clustering and visualisation which are effective on both synthetic and real additional datasets, and used that method in combination with the Deboeck & Kohonen Multi-Level Self-Organising Map to objectively cluster company-years, finding that this methodology is superior in improving classification accuracy to both the arbitrary grouping of company-years by industry as well as leaving the data entirely ungrouped.

It is now necessary to test the effectiveness of the methodology outlined in the previous chapters, and analyse the outcomes to draw broader conclusions that can be related back to bankruptcy failure theory.

7.1 Using the Methodology to Predict Failure

While this thesis has built an effective model for bankruptcy classification, it has not yet shown that the methodologies proposed are useful for bankruptcy prediction. As noted in chapter 4.3, the data used in the experiments thus far has been divided randomly into thirds, an in-sample training set, an in-sample validation set (though the Neural Networks Cascade-Correlation algorithm allows these to be combined), and an out-of-sample “applied” set. This was done because it allowed experiments to be performed without possible changes in the environment to impact on out-of-sample accuracy.

But it is probable that changes in the environment mean that what once was an effective predictor of failure is no longer an optimum choice. For example the codification of Accounting Standards by the Australian Accounting Standards Board's (AASB) in the *Corporations Act 2001* may change the effectiveness of a classification algorithm, or the increasing cost of credit as a result of the Global Financial Crisis 2008 may mean that previously high leverage had a lower impact on a firm's likelihood of failure. As a proxy for determining whether a classification algorithm used on current data will be useful for future (currently unavailable) data, it is common to reserve the out-of-sample "applied" dataset from the most recent data, and therefore the in-sample training and validation data includes only the least recent data. If a classification algorithm performs well on randomly divided data but poorly on sequentially divided data, then it can be reasonably concluded that the underlying relationship between the independent and the dependent variables in the most recent data is very different to the relationship between the independent and dependent variables in the least recent data – highlighting the difficulty in predicting something in a fast-changing environment.

Conversely similar classification performance demonstrates that the underlying relationships between independent and dependent variables do not change quickly, and that classification models developed earlier are likely to be able to predict future cases.

Therefore this chapter will implement the methods found to improve accuracy on a dataset from which the most recent third of cases have been reserved for the out-of-sample set. It will begin by performing the non-deterministic and best-forward factor search to identify key variables, using those factors as the inputs for the Deboeck & Kohonen Multi-Level Self-Organising Map, then using the proposed Spectral Visual Cluster Membership Validity method to cluster cases. The out-of-sample accuracy will then be compared to the results from chapter 6.2.

7.1.1 Removal of Low-Contributory Factors

As outlined in section 4.3.5, this section uses a Genetic Programming environment configured to perform 10 runs with a stopping criteria of 80 Generations Without Improvement (GWI), then doubling the GWI and allowing the model to continue doubling the GWI until such as time as the model has converged. While this was shown to be unnecessary due to the most accurate solutions being found within a small GWI, it was important to maintain the choice of parameters so that the results could be fairly compared with those outlined in section 6.2. The following factors survived the evolutionary process in the Compustat dataset with the full results available in Appendix HH:

- cash to total assets
- net income to total assets
- net income
- net income to net worth
- cash flow to total liabilities
- current liabilities to total assets
- net income to sales
- total assets to sales
- sales to total assets
- net worth to sales
- cash to current liabilities
- current assets to current liabilities
- current liabilities to current assets

The following factors survived the evolutionary process in the Aspect dataset with the full results available in Appendix II:

- cash flow to total assets
- net income to sales
- net income to net worth
- current liabilities to total assets
- current plus long-term liabilities to total assets
- cash to total assets
- cash to current liabilities
- retained earnings to total assets
- earnings before interest and taxes to total assets
- cash flow to current liabilities
- current liabilities to equity
- earnings before taxes to equity
- total liabilities
- book value to total liabilities

- net operating profit to sales
- sales to net worth
- net operating profit to total assets
- income from operations to total assets

7.1.2 Best-First Forward Search

As outlined in section 4.3.6 the best-first search was then performed using Genetic Programming on the Compustat dataset (Appendix JJ), Neural Networks on the Compustat Dataset (Appendix KK), Genetic Programming on the Aspect Dataset (Appendix LL) and Neural Networks on the Aspect Dataset (Appendix MM), with the graphical results shown in Figure 7-1 and Figure 7-2.

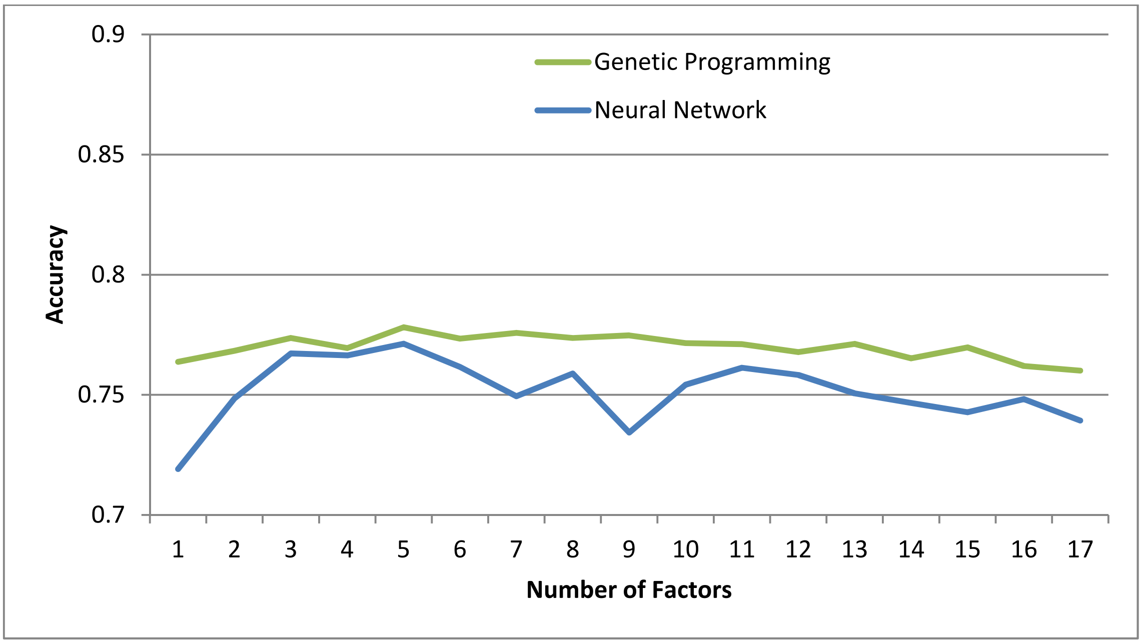


Figure 7-1 - Number of Factors versus Accuracy for Compustat Dataset (Sequential Division)

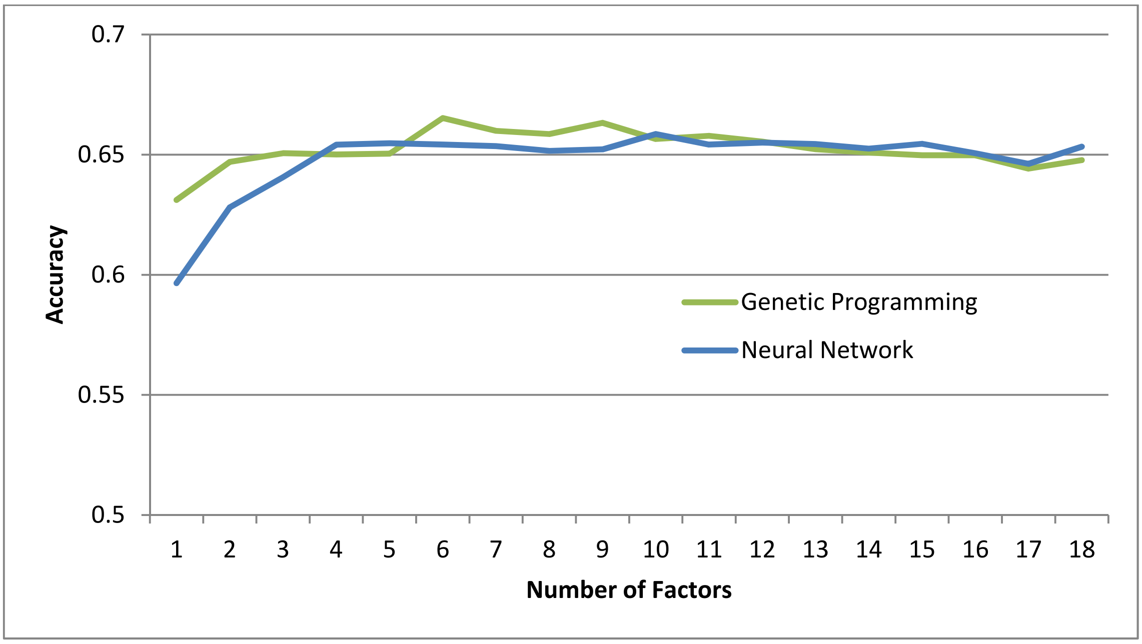


Figure 7-2 - Number of Factors versus Accuracy for Aspect Dataset (Sequential Division)

On the Compustat dataset, Genetic Programming achieved 77.8% with 71.3% on the out-of-sample, while Neural Networks achieved 77.1% with 72.6% on the out-of-sample. On the Aspect dataset, Genetic Programming achieved 66.5% with 66.7% on the out-of-sample while Neural Networks achieved 65.9% with 64.8% on the out-of-sample.

7.1.3 Clustering with a Two-Level SOM

The methodology identified in section 6.2.1 is applied using the factors identified in the previous section, resulting in the level 1 maps, level 2 maps, cluster maps and SpecVCMV found in Figure 7-3 through Figure 7-14.

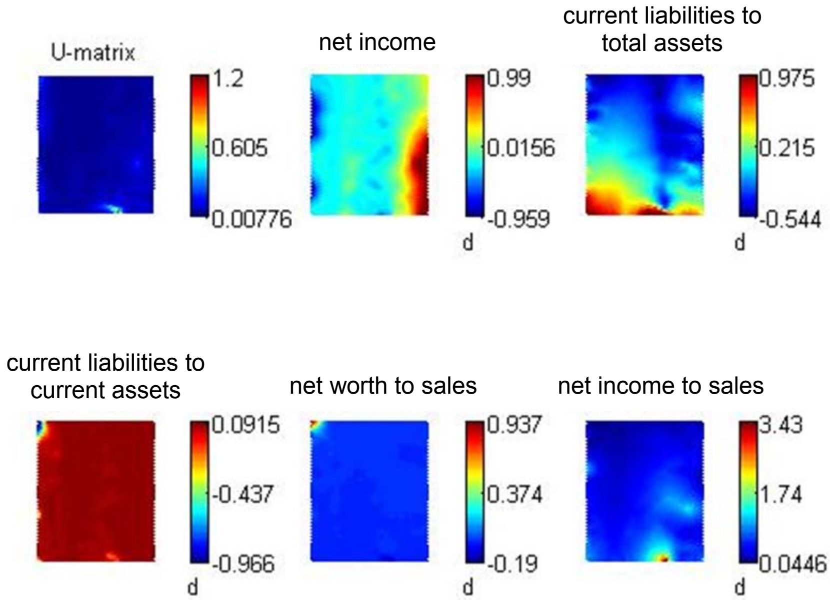


Figure 7-3 – Level 1 SOM of GP-optimised Factors on Compustat Dataset (Sequential Division)

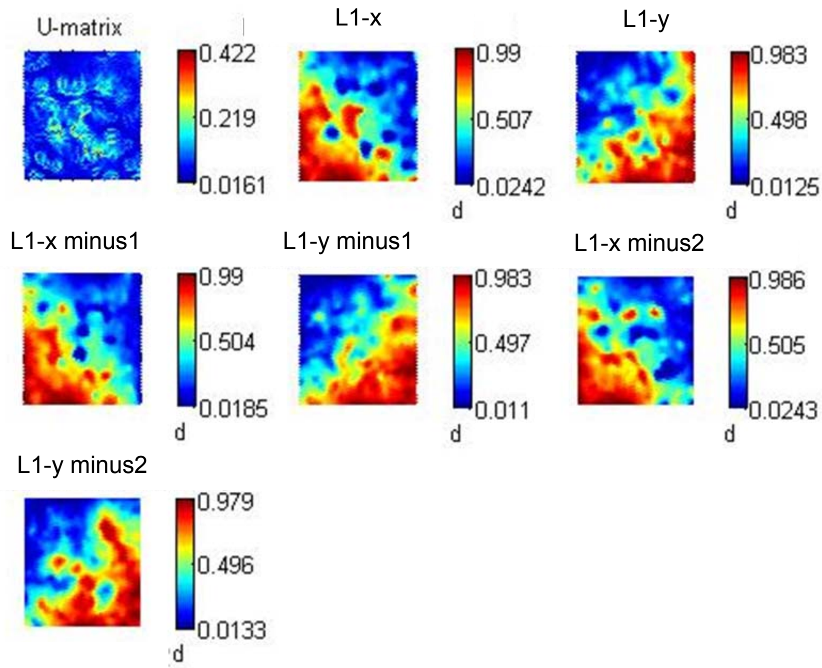


Figure 7-4 - Level 2 SOM of GP-optimised Factors on Compustat Dataset (Sequential Division)

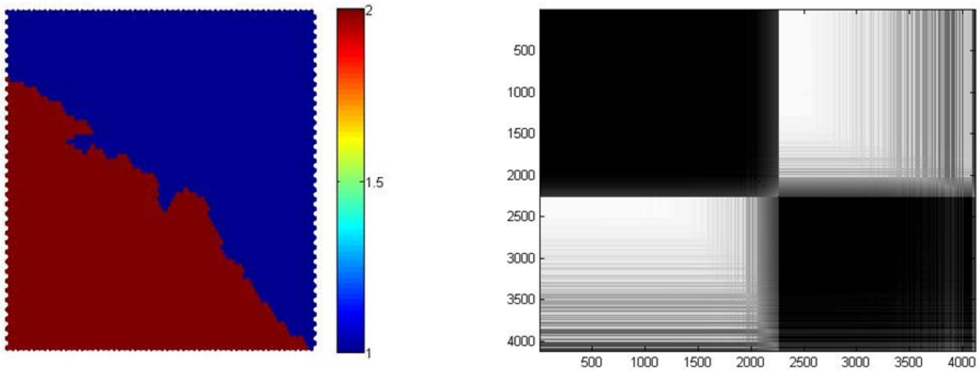


Figure 7-5 - Level 2 Cluster Map & SpecVCMV of GP-optimised Factors on Compustat Dataset (Sequential Division)

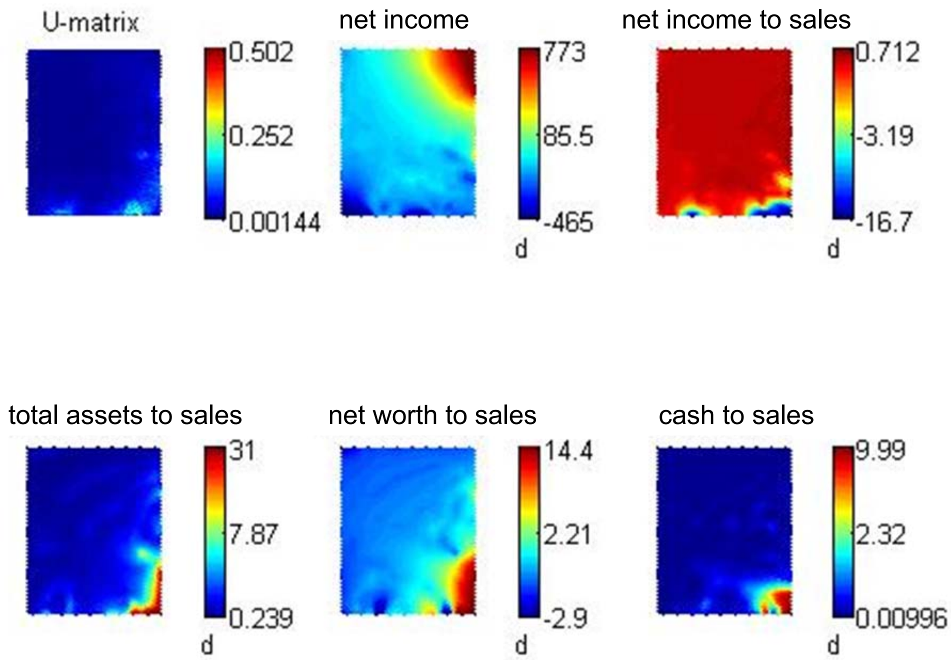


Figure 7-6 – Level 1 SOM of NN-optimised Factors on Compustat Dataset (Sequential Division)

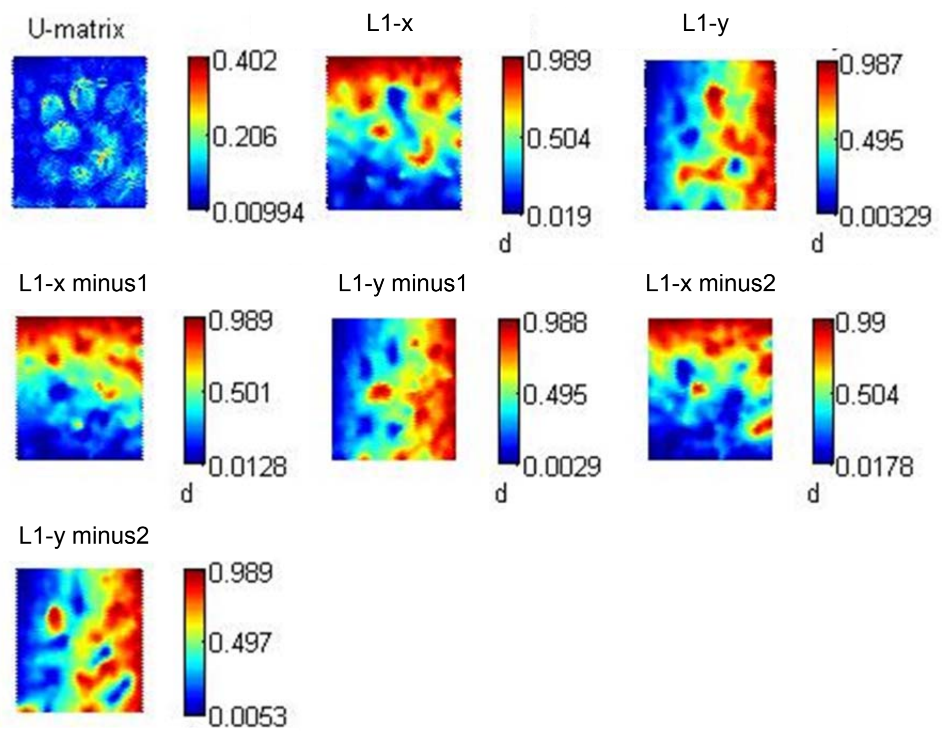


Figure 7-7 - Level 2 SOM of NN-optimised Factors on Compustat Dataset (Sequential Division)

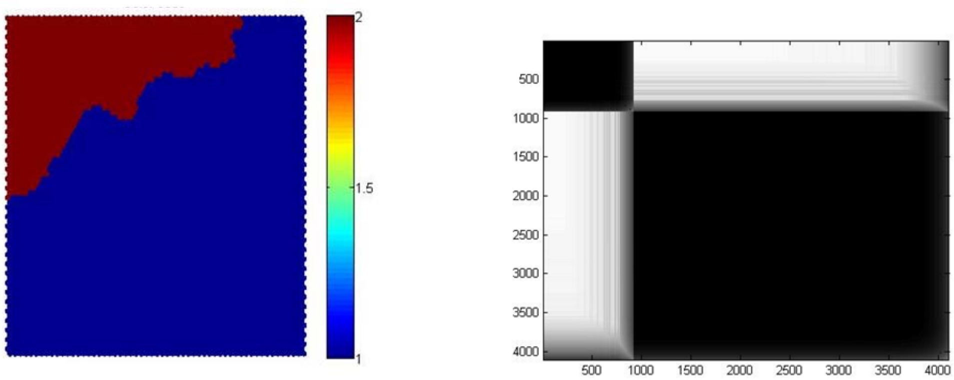


Figure 7-8 - Level 2 Cluster Map & SpecVCMV of NN-optimised Factors on Compustat Dataset (Sequential Division)

In particular, note the small size of one of the clusters when using Neural Networks on the Compustat data (Sequential Division) in Figure 7-8.

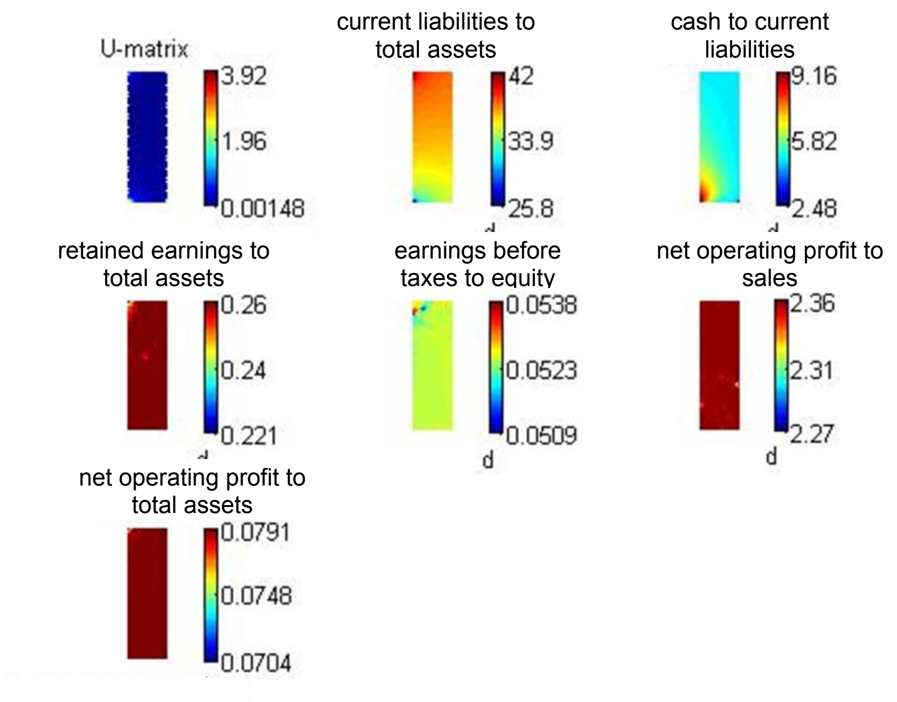


Figure 7-9 – Level 1 SOM of GP-optimised Factors on Aspect Dataset (Sequential Division)

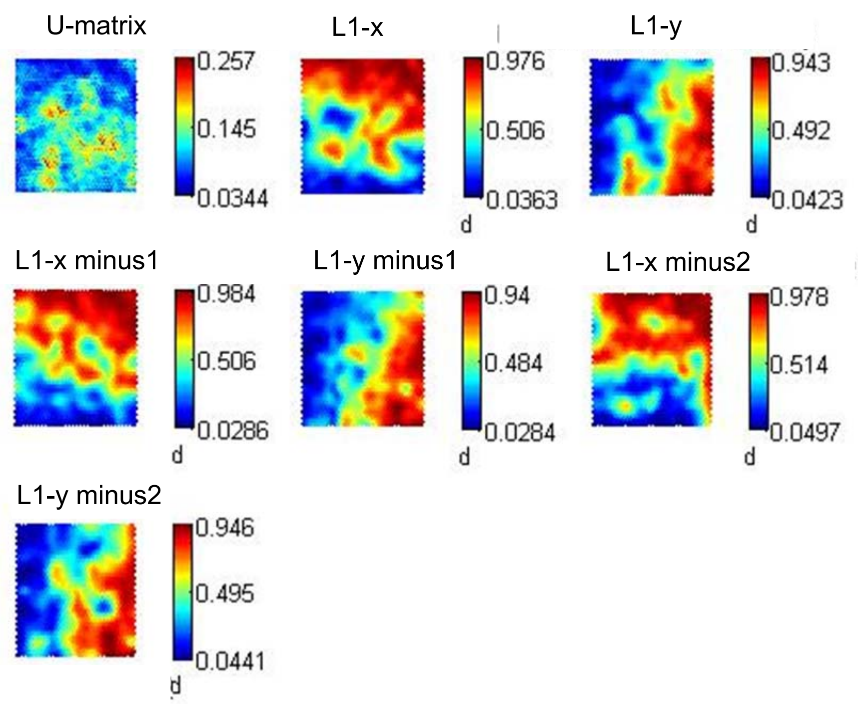


Figure 7-10 - Level 2 SOM of GP-optimised Factors on Aspect Dataset (Sequential Division)

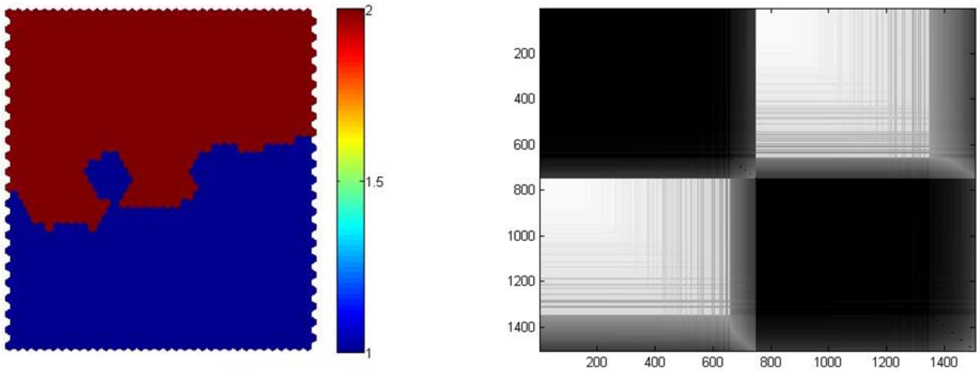


Figure 7-11 - Level 2 Cluster Map & SpecVCMV of GP-optimised Factors on Aspect Dataset (Sequential Division)

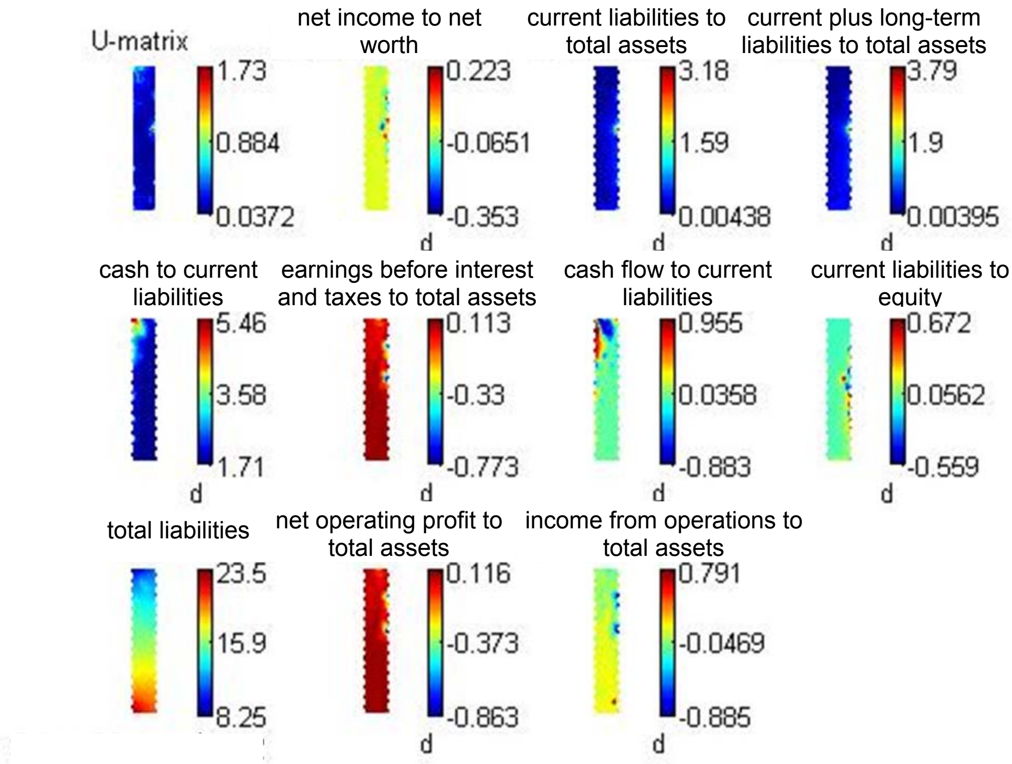


Figure 7-12 – Level 1 SOM of NN-optimised Factors on Aspect Dataset (Sequential Division)

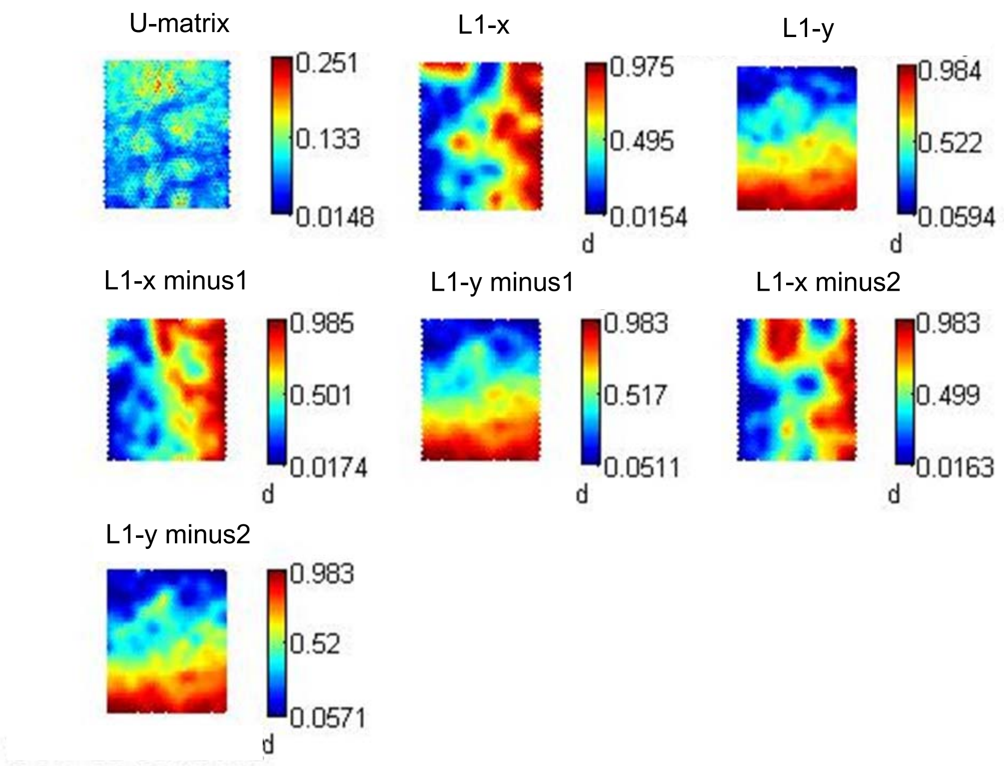


Figure 7-13 - Level 2 SOM of NN-optimised Factors on Aspect Dataset (Sequential Division)

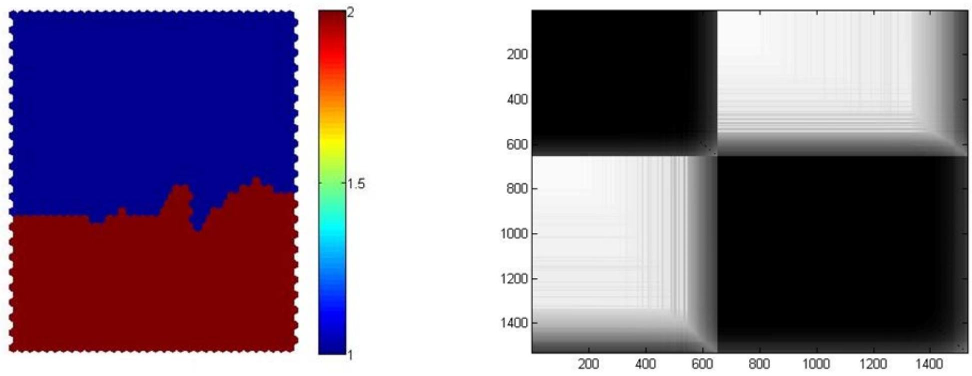


Figure 7-14 - Level 2 Cluster Map & SpecVCMV of NN-optimised Factors on Aspect Dataset (Sequential Division)

Following the methodology outlined in 6.2.1, each cluster of the data requires the forward best-first search to be re-performed, the numerical results for which can be found in Appendix NN,

Appendix OO, Appendix PP and Appendix QQ. Of particular note is the clustering that has occurred using the Neural Networks methodology on the Compustat dataset, in which cluster 2 is a small cluster (approximately 30% of the cases) and contains just 0.03% of failure cases. This has resulted in the Neural Network model simply classifying all cases in cluster 2 as non-failure. As the out-of-sample set for cluster 2 contains no failure cases at all, this results in both the within and out-of-sample accuracies approaching 100%. However the out-of-sample accuracy for cluster 1 has decreased. This is directly contrary to the findings of Deboeck & Kohonen (1998) which identified regions in the SOM from which failing companies do not emerge, this study found a region in the SOM from which healthy companies never emerge.

A table of results from this chapter is shown in Table 7-1 and Table 7-2.

Method	Dataset	In-Sample	Out-of-Sample
Genetic Programming	Cluster 1	77.1	73.1
Genetic Programming	Cluster 2	89.1	79.9
Neural Networks	Cluster 1	76.2	65.4
Neural Networks	Cluster 2	100.0	100.0

Table 7-1 - Summary of Corporate Failure Prediction on Compustat Dataset

Method	Dataset	In-Sample	Out-of-Sample
Genetic Programming	Cluster 1	72.6	61.6
Genetic Programming	Cluster 2	76.8	73.6
Neural Networks	Cluster 1	73.3	59.7
Neural Networks	Cluster 2	77.1	65.1

Table 7-2 - Summary of Corporate Failure Prediction on Aspect Dataset

Combining the results of the clusters, using the GP methodology on the Compustat dataset resulted in an in-sample weighted accuracy of 82.3% with 76.3% on the out-of-sample, using the NN methodology results in a 84.7% in-sample accuracy with 79.2% on the out-of-sample. On the Aspect dataset, when using Genetic Programming the weighted combined in-sample accuracy achieved 74.9% on the in-sample and 67.3% on the out-of-sample, whereas with Neural Networks the in-sample achieved 75.0% with 62.0% on the out-of-sample.

In comparison to the accuracy on the randomly divided dataset used in chapter 6.2, using the GP methodology on the Compustat data in this sequentially divided dataset resulted in a 2.9% reduction in accuracy on the out-of-sample, while the Neural Network methodology is 1.9% higher on the out-of-sample. On the Aspect data, the GP methodology on this dataset is 0.2% higher, and the NN methodology is 0.5% higher. Higher out-of-sample accuracies than the randomly divided dataset is a very unexpected outcome, and is possibly caused by the sequential division of data having the side effect of ensuring each company is equally represented in the in-sample and out-of-sample datasets.

Until this point in the thesis all of the results obtained have been comparative analysis, such as examining the effect on classification accuracy by adding share market data or comparing classification algorithms for example. These aspects of the research have been undertaken with the underlying goal being to build a greater understanding of the overall relationship between bankruptcy and the available information. It had been assumed that if the corporate environment was changing, sequentially divided data may prevent the detection of new relationships between the available data and corporate failure. This chapter, on the other hand, which uses the previously tested methodology on sequentially divided data, is examining whether or not the methodology can genuinely be used to predict a company's survivability. That is to say that by reserving the "most recent" portion of data, the model can be trained on hypothetical "past" information and the out-of-sample set becomes hypothetical "future" information, such that the predictive performance of the model is calculated. It is therefore appropriate to present the above results in a confusion matrix (Kohavi & Provost, 1998), in particular because it highlights the difficulty of dealing with rare event detection (Choe, et al., 2000). Four confusion matrices are shown in Table 7-3, Table 7-4, Table 7-5 and Table 7-6, one for each of the two datasets with each of the two tested methodologies.

		Predicted		
		Failure	Non-Failure	Accuracy
Actual	Failure	182	81	69.2%
	Non-Failure	4548	25738	85.0%
	Precision	3.8%	99.7%	84.8%

Table 7-3 - Confusion Matrix for Genetic Programming on Compustat Dataset (Seq Division)

		Predicted		
		Failure	Non-Failure	Accuracy
Actual	Failure	185	78	70.3%
	Non-Failure	5036	25250	83.4%
	Precision	3.5%	99.7%	83.3%

Table 7-4 - Confusion Matrix for Neural Network on Compustat Dataset (Seq Division)

		Predicted		
		Failure	Non-Failure	Accuracy
Actual	Failure	243	83	74.5%
	Non-Failure	1822	3449	65.4%
	Precision	11.8%	97.7%	66.0%

Table 7-5 - Confusion Matrix for Genetic Programming on Aspect Dataset (Seq Division)

		Predicted		
		Failure	Non-Failure	Accuracy
Actual	Failure	209	117	64.1%
	Non-Failure	1030	4237	80.4%
	Precision	16.9%	97.3%	79.5%

Table 7-6 - Confusion Matrix for Neural Network on Aspect Dataset (Seq Division)

Certainly the final accuracy, particularly of the Compustat dataset using the Genetic Programming methodology of 84.8%, is a result that demonstrates this particularly methodology is able to achieve classifications well beyond that of the naïve model. Looking at just the actual failure cases, the methodology was able to successfully classify 69.2% of cases, and of the non-failure cases the methodology achieved 85.0% accuracy. However, the sheer number of non-failed cases versus failed cases means that a 15% error rate on the non-failed cases results in approximately 25 times as many classified non-failed cases for every correctly identified failure case, or a positive predictive value (PPV) of 3.8%. That is, just 3.8% of all cases classified as failed will actually fail. In the Aspect dataset, the larger number of failed cases relative to non-failed cases mean the precision is higher, but even using the Neural Network methodology for which the net unweighted accuracy was 79.5%, the PPV of 16.9% means that for the successfully identified failure case there will be more than 6 times as many incorrectly predicted to be a failure.

Low PPV values are common in studies where there is a low prevalence of the condition being predicted. For example an algorithm that is 90% accurate on 10,000 cases of class 1 (9,000 correct with 1000 incorrectly classified as class 2) and 90% accurate on 100 cases of class 2 (90 correct with 10 incorrectly classified as class 1) will result in the 1000 cases incorrectly classified as class 2 overshadowing the 100 cases correctly classified as class 2, a precision of 8.3%. For example in Kerlikowske et al. (1993), 31,000 women underwent mammography screening, and a PPV of 9% was reported. To overcome this limitation, many classification systems (including both Genetic Programming and Neural Networks), are in fact classifying cases with a continuous output variable, then using an arbitrary threshold to harden the continuous output variable into a binary output variable. This threshold can be adjusted to favour type-I or type-II error. This thesis assumes that misclassifying a non-failing company as high risk is of lower cost than misclassifying a failing company as low risk, this is especially true of a creditor, and so the threshold used is acceptable. Provided the limitations of rare event detection are acknowledged, low PPV values do not necessary invalidate the usefulness of the classification system.

This section has therefore shown that the classification methodology outlined throughout this thesis can be successfully applied to a bankruptcy prediction problem. Having done so, the outcomes of the classification methodology can be examined to gain insights into corporate failure itself.

7.2 Analysis of Classifiers and their Classifications

Using the classification methodology outlined in chapter 6.2 and validated as a useful predictive system in section 7.1, there is an opportunity for deeper analysis into when and why the classification system is successful, and when it is not. In doing so, this chapter aims to build a greater understanding of the limitations of the methodology and gain some useful insights into any relationships between corporate failure and the successful classification of corporate failure.

To this end, the output from the Genetic Programming classification methodology outlined in section 7.1 will be analysed, as this methodology results in (comparatively) simple mathematical algorithms from which insight can be gained. The Neural Network training algorithm results in highly complex multi-dimensional visualisations that make it very difficult to gain meaningful insight using the methodology, so this chapter focuses on the outputs from the Genetic Programming algorithm.

7.2.1 Compustat Dataset Cluster 1

While there are many companies within the Compustat dataset that could have been selected within this cluster, it was opted to analyse Bethlehem Steel Corporation, which filed for bankruptcy protection in 2001, due to its size providing more than adequate secondary sources that examine the company's failure in more detail.

While the 4^{th} year prior to failure was found in cluster 2, incorrectly classified, the final 3 years of the company prior to bankruptcy were located in cluster 1 and were correctly classified using the algorithm that is documented in Appendix RR. The output of the algorithm from its 3 input parameters can be visualised in Figure 7-15 with net income to net worth on the x axis, net income to sales on the y, multiple values of net worth to sales shown as parameter n across multiple plots, and the model's output value displayed on the z axis (where z>0.5 indicates classified failure). Furthermore the 3 cases from the final 4 years of the company which are located in this cluster can be added as data labels. In doing so, an understanding of how the model perceives the underlying relationships can be built.

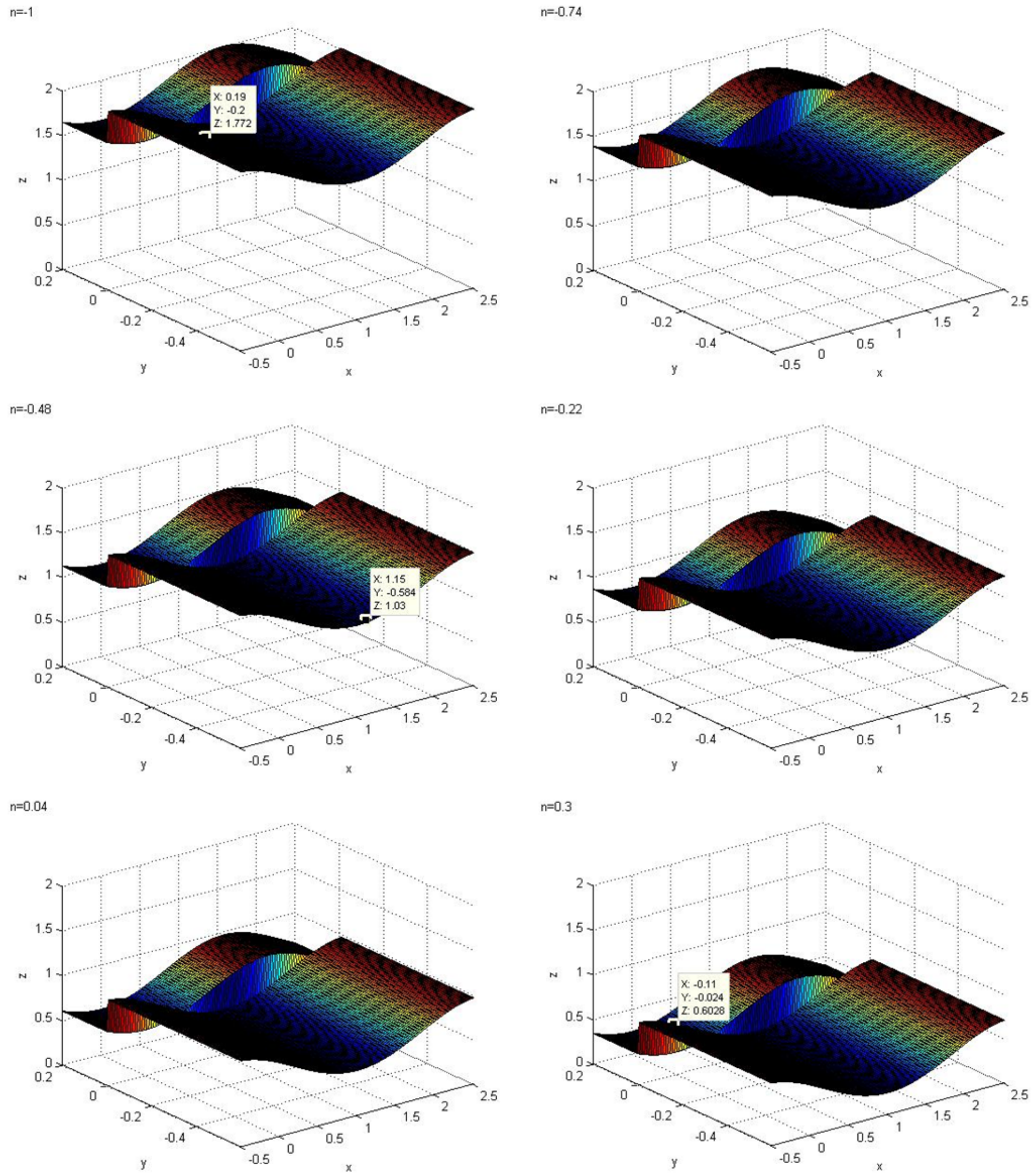


Figure 7-15 - Compustat Cluster 1 Failure Classification Surface

It is immediately apparent that n, net worth to sales, is the dominating factor as anything with $n<0.04$ is immediately classified as a failure across this domain used for these cases. However assuming net worth to sales to be greater than 0.04, as net income to sales changes from a positive to a negative value this has the effect of inversing the effect of net income to net worth. Thus when net income to sales is negative (as it is in the examples here) the model considers net income to net worth values around 1 to be least likely to fail. As net income to net worth deviates away from 1, the model finds an increasing likelihood of failure.

There is now an opportunity to examine secondary sources that may shed light on the causes of Bethlehem Steel's bankruptcy. Warren's detailed analysis of Bethlehem Steel identified issues as far back as the 1960's: "New circumstances in international trade, in technology, and in the trends and structure of demand henceforward would provide a less satisfactory business environment" (Warren, 2008, p. 181). While the late 1960's saw reinvestment in shipbuilding and repairs, these efforts were late and net income continued to decline. Various crises went unmanaged, such as the 1973 energy crisis, court ordered fines for illegal billing practices, and a favourable U.S. government offer to buy ships being rejected in the late 1980's. While there were many indicators of success, "the trend of net income, though it followed an uneven course year to year, was downward overall" (p. 196), and is perceived by the model. The 1980's saw managerial issues taking place with inexperienced staff at the helm, increasing environmental standards and overseas competition from newer more advanced plants increasing. "Net sales were down by 28 percent and there was a deficit on a previously unequalled scale: nearly \$1.5billion" (p. 217), again this fall in net sales was perceived by the model. In response, management halved the workforce which saw a positive net profit in 1987 through 1989, new investments in modernisation initiatives, and the paring down of many operations. In the early 1990's, Bethlehem engaged a project to equip with electric furnaces, which while successful went \$145 million over budget (p. 243), and the president of competitor Nucor on the topic of the new furnaces is quoted as saying “It’s a joke, we’ll kill them” (The Morning Call, 2010). By 1994, Bethlehem was again engaging in retrenchment to reduce costs, and by 1995 many production centres had closed. “In the ten years through 1999, on net sales of $44.94 billion, only four years showed a net income; the net loss over the whole period was $1.92 billion”. By comparison U.S. Steel, in the same period seven years showed a positive net income (Warren, 2008, p. 246). In hindsight, CEO of Bethlehem, Walter Williams said in an interview with Fortune, “We were all stuck with our basic steelmaking—just too much to write off and too much to shut down” (Loomis, 2004). Macroenvironmental forces and increased competition meant that by 2000 revenues were falling well faster than reductions in cost, resulting in a net loss of $118 million. “By the third quarter of 2001, Bethlehem shipments of steel products of all kinds were 117,000 tons less than the levels of the third quarter of 2000, average prices were some $40 a ton lower, and sales, $160 million lower. A net third quarter deficit of $35 million in 2000 had become a loss of $134 million a year later, or more than $1.4 million a day. Such a drain on company resources could not long be sustained.” (Warren, 2008, p. 261)

Bethlehem Steel is in fact a classic example of an Argenti Type 3 failure, as there were management issues many years prior to failure, a change in the market to which management does not respond, a big project, and a two-phase failure in which the first crash waterlogs the company a number of years before the failure that puts the company into bankruptcy (Argenti, 1976, p. 162). While Argenti (1976) also identifies creative accounting, no evidence for such was found in Bethlehem Steel. This is interesting because the three contributory factors in the model use ratios that are likely targets for window dressing, however the shape of the classificatory surface shows that this particular model identifies deviations from an optimal value, making creative accounting a much more difficult prospect than if linear Discriminant Analysis had been used as the predictor.

Given the long period of time over which Bethlehem Steel descended into bankruptcy, it is interesting to look at the classifications for Bethlehem Steel in other years. In fact, of the 4 "non-failure" years in which this company was classified in this cluster, 3 of them were incorrectly classified as failure. This indicates that at least some of the error in the algorithm for this cluster is due to companies exhibiting failure symptoms but simply failing over a longer period of time than the failure horizon allowed for.

7.2.2 Compustat Dataset Cluster 2

Cluster 2 failed companies tended to be smaller than those found in cluster 1, making the analysis of secondary sources more difficult. For the purpose of analysing cluster 2, Cone Mills Corporation was selected which was a textile company that failed in 2003. What makes this particular example interesting is that while in the 2^{nd} and 3^{rd} year out the company was correctly classified as a failing company, in the final year of available data the cluster 2 algorithm incorrectly classified the company as non-failure. The output from the algorithm located in Appendix SS is shown in Figure 7-16, with the three year company-years found in this cluster shown on the plot with the x axis representing net income (in millions), the y-axis representing current liabilities to current assets, and the z-axis again showing the model's output value.

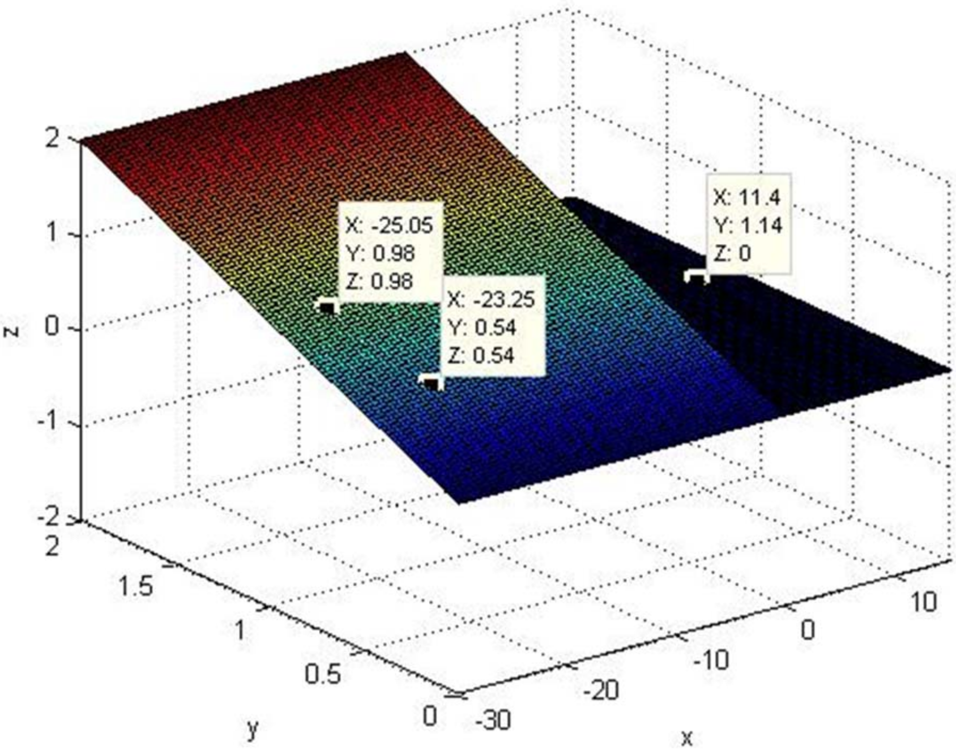


Figure 7-16 – Compustat Dataset Cluster 2 Failure Classification Surface

The surface shows that companies with a positive net income in cluster 2 are immediately considered non-failed, and those with a negative net income are classified as failed if current liabilities to current assets > 0.5. This has led to the misclassification of the company in the year immediately prior to its failure, due to it having a positive net income in that year. The error surface shows that a company in this cluster that misrepresents its net income to be positive when in fact it should have been negative would be misclassified as non-failed.

After some years of mixed results as a private company, Cone Mills Corporation returned to public offering in 1992, while simultaneously withdrawing from corduroy manufacturing to boost net profits (Gray, 1992). However by just 1995, with sales peaking at $910 million, the company had started posting net losses. A restructuring effort was undertaken in 1998 in an attempt to recover the situation, and the company defended against a takeover by Summit Capital Corp by using a shareholder rights plan. It wasn't until 2002 before the restructure meant that the company posted another profitable year ($7.2 million), and things on the surface appeared to have recovered. CEO John Bakane is quoted as saying "2002 fulfils the commitment that we made last year to return Cone Mills to profitability and achieves our $50 million EBITDA goal set two years ago", while the CFO, Gary Smith, said "We continue to operate with daily liquidity of more than $20 million" (Textile News, 2003). Indeed as noted earlier, the positive net income in the year of its failure caused the classification system to move Cone Mills into the "non-failed" category. However in just 2003, Cone Mills was unable to make a $4.1 million bond interest payment. Bakane said, "While we returned the company to profitability in 2002, the events of this year have been such that we simply cannot support our present capital structure in the fact of current market conditions" (Beltran, 2003). W.L. Ross & Co, who went on to buy Cone Mills and some of its liabilities for $46 million on the condition that Cone Mills entered bankruptcy, blamed the trade liberalisation pact of that year with Vietnam, citing a large influx of cheap denim (Beltran, 2003), however Cone Mills financials show write-downs in excess of $123 million in the company's final quarter, not including goodwill (Wharton Research Data Services, 2003), indicating that the profit and loss statements have been overstating the company's viability.

Like Bethlehem Steel, Cone Mills Corporation appears to be an Argenti Type 3 failure (Argenti, 1976, p. 162), there is certainly evidence of changes to the marketplace that management did not respond to, a two-stage crash, rising leverage (long-term debt rose to over $177 million in 2001) and there is some evidence of creative accounting. However, through all of this the most accurate predictors of failure for the company-years in cluster 2 was simply net income and current liabilities and current assets, so the model did not find the cash flow issues or evidence of creative accounting to be a predictor.

7.2.3 Aspect Dataset Cluster 1

While the dataset contained 3 years of One.Tel's data, only one had three years of consecutive data and this one was incorrectly classified as a non-failure in cluster 1, making it a particularly interesting case to analyse. For the purpose of context, all 3 years of One.Tel data are shown on the plot in Figure 7-17, though only the one located at x=0.115 and y=0.9 was available to the classification model. In the diagram the x-axis represents cash flow to total assets, the y-axis represents cash to current liabilities, and the z-axis represents the output of the algorithm.

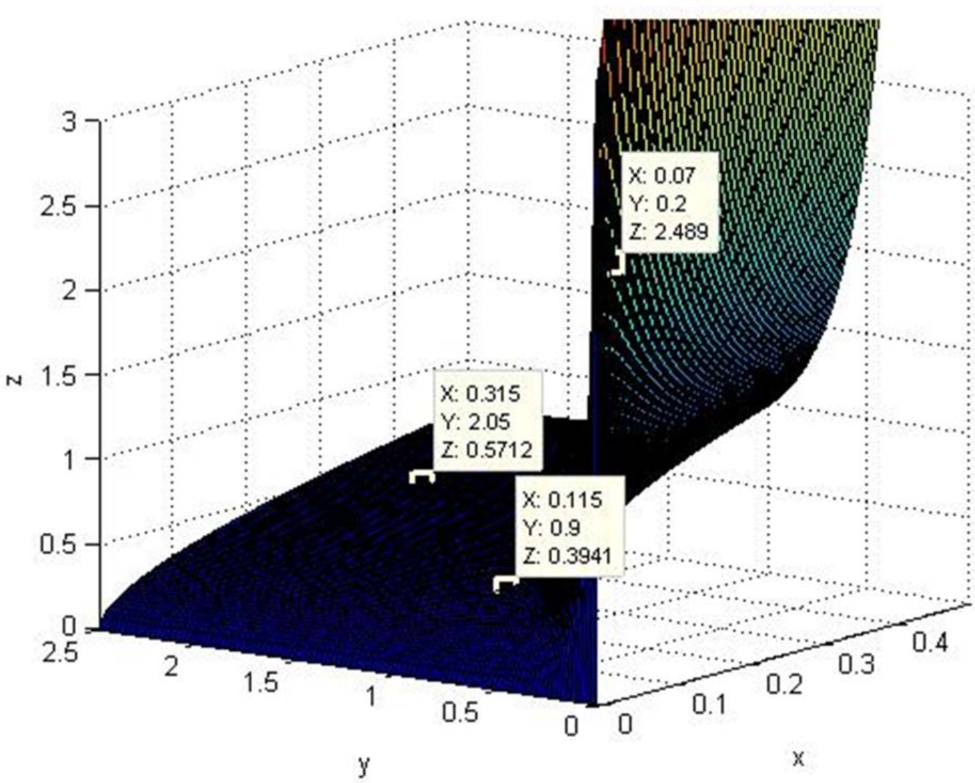


Figure 7-17 - Aspect Dataset Cluster 1 Failure Classification Surface

The above diagram indicates that companies in this cluster are considered likely to survive when they have a high cash to current liabilities: as cash to current liabilities decreases below approximately 0.4 the likelihood of failure climbs exponentially. One.Tel's cash to current liabilities 3 years prior to failure was approximately 0.19 and the resulting position on the diagram is striking. A more linear relationship also exists for cash flow to total assets, showing that as cash flow increases relative to current assets, the likelihood of failure increases. One.Tel had a cash flow to total assets ratio of 0.31 two years prior to its failure, which pushes its failure score over 0.5 and would have resulted in a failure classification. However in the year immediately prior to its failure One.Tel had increased their cash to current liabilities to 0.89, and increased their total assets relative to their cash flow which decreased the ratio cash flow to total assets to 0.11, resulting in the incorrect non-failure classification.

Launched in 2005, One.Tel began under agreement with Optus for network services, SIM cards and so on, meaning that One.Tel's gross profit came in the form of the margin it made from reselling Optus services. By the start of 1997 One.Tel had over 160,000 customers, revenue approaching $150 million and a net profit after tax of over $3.5 million (Monem, 2011). One.Tel soon floated on the Australian Stock Exchange and engaged in a strategy to expand internationally, purchasing spectrum to build its own mobile network while its share price continued to rise. As later documented in the New South Wales Supreme Court, One.Tel's strategy was to be the dominant mobile phone provider, securing $710 million from News Ltd and PBL in exchange for shares (Australian Securities and Investments Commission v Rich, 2009), and having Lucent Technologies build an Australian mobile network for $1.15 billion and a European mobile network for $20 billion. By November 1999, One.Tel's market capitalisation was over $5 billion (Barry, 2002, p. 417), and had operating profits after tax approaching $7 million (Monem, 2011). For the 1999-2000 financial year One.Tel reported losses of over $290 million with $335 million in cash reserves (Monem, 2011). Part of this was due to ASIC requiring One.Tel to declare $173 million of advertising and customer acquisition costs that had been hidden in the balance sheet in breach of Corporations Law (Barry, 2002). Monem (2011) argues that creative accounting existed from as early as 1998, as the annual report "claimed that the company was cash flow positive from normal operations", but that "after 1996-97, One.Tel's cash flow from operations was never positive", that "One.Tel was not collecting cash fast enough to finance its aggressive corporate expansion" and provides evidence that the ratio of cash collected to incremental sales revenue had dropped to 55% by 2000, by making "ever-increasing cash payments to suppliers and to a growing pool of employees". Monem (2011) goes on to document decreasing sales revenue per customer from $933 in the financial year ending 1997, down to $508 by the year ending 1999, while cash paid to employees and suppliers increased by 51% over the same period, stating "Clearly One.Tel was pricing its services even below it's 'cash costs' for at least 1998-99 and 1999-2000". It is documented that Jodee Rich and Mark Silbermann did not review finance journals, trial balances or ledgers (Australian Securities and Investments Commission v Rich, 2009), and the enormous loss in the financial year ending 2000 is justified in part by a change in accounting policy, "Previously the Company followed a method of accounting, in compliance with Australian GAAP, where these costs were deferred over periods appropriate to the nature of each asset, whereas in the UK these costs are generally written off as incurred. The Company has decided to adopt a policy acceptable under UK GAAP and from the current year these costs are written off as incurred" (Connect 4, 2000). "Thus, One.Tel's operating profits reported in all the past years were largely due to non-conservative accounting policy choices" (Monem, 2011). In mid-2000, the company's cash flows were under pressure due to being unable to send any bills for six weeks as a result of problems in the implementation of GST, call centres were "under siege" from angry customers which was failing to answer up to 80% of callers, with staff turnover "running at 300 per cent a year" (Barry, 2002, p. 213). The "Next Gen" network was behind schedule but launched in 1999 regardless, and an agreement with Telstra meant that customers who did not have One.Tel coverage would automatically roam onto Telstra's network, who in turn would charge One.Tel full retail prices. Combined with "Free Time", which allowed One.Tel customers to call each other for free, even when roaming, meant that One.Tel's profit per customer continued to fall. "Even towards the end, in April 2001, one-third of calls were being carried for free, almost half were still roaming on Telstra, and a good proportion of the rest were being charged at uncommercial rates. For much of the time, the company was either collecting no revenue from calls or giving the revenue away to its biggest rival. Meanwhile it was paying out millions of dollars on dealer commissions, advertising, marketing, sales, handset subsidies and administration—before it even began to think about paying the interest and repayments on the $1.15 billion cost of the network and the $523 million it had spent on spectrum." (Barry, 2002, p. 260). In May 2001, One.Tel entered receivership with its share price closing at 16c.

One.Tel is an excellent example of a company which engaged in overtrading, in all three of McRobert & Hoffman's (1997) dimensions: Physical, as the Next-Generation network was launched before completion; human, as evidenced by the enormous staff turnover and an inability to man their call centres; and financial as evidence by their poor cash to current liabilities ratio as indicated on Figure 7-17.

One.Tel is also an example of an Argenti Type 2 company, noting that "Type 2 proprietors are super-salesmen; they are leaders of men, flamboyant, loquacious, restless and bubbling with ideas. The scale of their ambition is almost pathological. They never accept advice, they 'know it all'" (Argenti, 1976, p. 158), and this was certainly the case for One.Tel. There is evidence for one man rule, Barry (2002, p. 248) argues that "despite its supposedly democratic nature, One.Tel had always been a thinly disguised autocracy in which Jodee made all the key decisions". Argenti (1976) cites creative accounting as turnover increases but profits do not, resulting in overtrading and failure.

7.2.4 Aspect Dataset Cluster 2

To demonstrate the final cluster, Auto Group Limited was selected because it appears that the company may have been intentionally overstating net profit and this had a considerable effect on the company's viability. Of the six years of data available for Auto Group, the first two years were incorrectly classified as failed, while the final four years were correctly classified as failed.

On the diagrams in Figure 7-18 the x-axis represents current plus long-term liabilities to total assets, the y-axis represents earnings before interest and taxes to total assets, n (for which multiple values are shown) represents net operating profit to sales, and the models output value is shown on the z-axis.

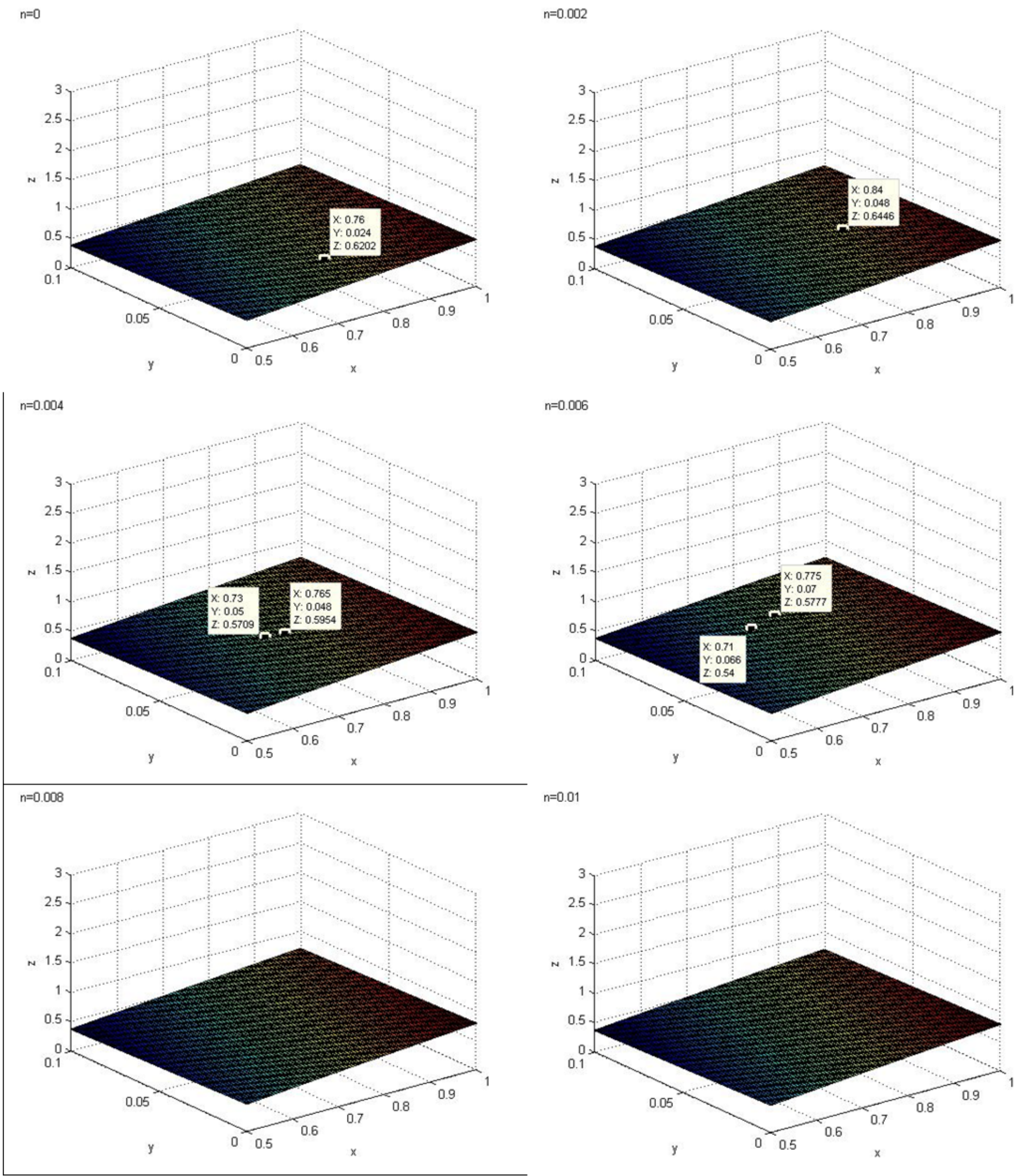


Figure 7-18 - Aspect Dataset Cluster 2 Failure Classification Surface

While it is difficult to see in the above diagrams due to the small range of n used, as n increases the surface moves linearly downward along the z axis, indicating that small values of profit to sales make companies more likely to fail. The surface also shows that for all values of n, failure increases linearly as liabilities to assets increase, and decreases as earnings before interest and taxes to assets increases, none of which is particularly surprising. It appears that in order for Auto Group Limited to be classified as a non-failed company, it would have had to increase profit to sales to 0.1, decrease long-term liabilities to total assets to 0.68, or increase to earnings before interest and taxes to total assets to 0.095.

Being a smaller company, though still Australia's largest auctioneer (Australian Associated Press, 2006), the number of secondary sources that examine the failure of Auto Group Limited is comparatively small. Established in 1988 and listed on the Australian Stock Exchange in 1993, by 2005 Auto Group was selling "over 80,000 vehicles per annum, employing almost 600 full time employees and operating in every mainland state of Australia" (Connect 4, 2005). In 1998, Auto Group published a net profit after tax of \$1.9 million, increasing to \$2.5 million in 1999. 2002 saw profit decrease to \$1.8 million, decreasing a further 62% to \$682,000 in 2003 while revenue decreased by less than 1%. Richard Moffitt, in the Chairman's annual report noted "aggressive new vehicle marketing and pricing strategies undertaken by most, if not all local distributors forced significant pressures upon used vehicle demand and consequently pricing". 2004 saw a three-fold net profit increase to \$2.4 million, on an increase in revenue of just 7% after the acquiring of Frankston Mitsubishi and Bayside Honda and Kia in Brisbane. But 2005 saw profits down to lowest levels yet, just \$562,000 even though revenue increased another 17%. Borrowing costs nearly doubled, and the chairman noted in the annual report, "a large over supply of used vehicles and a significant drop in demand diminished both sales volumes & gross margins" (Connect 4, 2005). 2005 saw the group acquire more dealerships including Cranbourne Suzuki and Sydney City Kia & Mitsubishi as well as a controlling interest in National Finance Choice. October 2005 saw an agreement between A.P. Eagers to explore a merger (A.P. Eagers, 2005). Certainly evidence that a chain of acquisitions is present, a highlighted failure cause in Ross & Kami (1973). This was quickly followed by the statement “While it is too early to know for certain, it is possible that irregularities may exist in those financial statements and that they may have a material impact on Auto Group's financial position as stated in its 2005 Annual Report” (RWE Australian Business News, 2005). The company was placed into administration and, “following preliminary investigations into the 2005 Financial Statements by the newly appointed Chief Financial Officer of Auto Group, the board of Auto Group has formed the view that the consolidated net profit of Auto Group and its controlled entities for the year ended 30 June 2005 was overstated in the 2005 Financial Statements by approximately $4.5 million before tax”. Then “following the appointment of SJ Parbery and MJ Robinson of PPB as administrators - in the opinion of the company's board the company and its subsidiaries are insolvent or likely to become insolvent” (Investogain Limited, 2012). In 2008 the liquidators advised that they would not pursue potential claims for breaches of the Corporations Act (Investogain Limited, 2012), unfortunately denying this research the opportunity to investigate further the nature of the net profit overstatement.

The failure of Auto Group could be weakly classified as a Type 3 failure (Argenti, 1976), as the benefit of hindsight shows the acquisition of dealerships when the company was already facing financial difficulty. However at the time it is possible that management were simply unaware of the profit overstatements, or it could be that the acquisitions were an attempt to bring the company back to profitability at a time when there was demonstrable creative accounting occurring. There is certainly evidence that the accounting irregularities fed back into the management decisions made by the directors, who opted to pursue a merger based on this information.

Regardless, it is interesting to note that while the irregular 2005 financial year showed a better net operating profit to sales (0.0025) than the 2000 financial year, it was not sufficiently better to prevent the model from classifying 2005 as an imminent failure risk. In fact, the 2005 financial year had the highest model output of all six years of data (0.64), primarily due to having the highest liabilities to total assets of any of the six years.

7.2.5 Conclusion

This section has demonstrated that the relationships between financial statements and the model's expectation of failure are often non-linear, which explains the increased accuracy of Genetic Programming over other classification systems shown in chapter 4.2. In particular, this section demonstrates that there is value in an analysis of how a classification came to be made as it allows further and deeper analysis into what aspects of a company's financial statements should be given attention, in addition to making a prediction on the company's risk of failure within the given failure horizon.

This section has shown that the causes of bankruptcy are certainly found in the cases selected, and it has been shown that symptoms of failure generally appear in a company's financial statements many years before its ultimate failure. While this chapter serves as evidence to support Argenti (1976), it falls short of proving a causal relationship. While there would be value in collecting data identifying underlying causes of bankruptcy such as “one man rule”, “creative accounting” and so on, to do so while a company remains operational is difficult if not impossible to someone without insider information. However if models such as those presented in this chapter can be used to identify at-risk companies, and those companies can be further investigated by potential investors or creditors for underlying causes of failure, then external parties will be better informed and empowered to make sound financial decisions.

8. Conclusion

8.1 Summary of Results

This thesis has set out to answer a number of questions, originally outlined in section 3.2, which are reconsidered here with a presentation of the relevant findings from each chapter.

Section 3.2.1 questioned the effect of a lack or quality of accounting information and the possibility of accounting manipulation, considering whether or not information from the stock market might be more useful than financial data alone. However chapter 5 found that the inclusion of share market information had mixed results in improving the classification accuracy of the model. Furthermore, chapter 7 showed that reasonable predictive accuracies could be attained using the methodology outlined in this thesis, and section 7.2.4 demonstrated that even in cases where net profit is known to be overstated, a predictive model is able to accurately detect these anomalies and identify failing companies. In particular this finding is in contrast to Clarke et al. (1997) in which it is claimed that financial ratios are an ineffective mechanism for the prediction of failure, particularly when companies are acting to intentionally misrepresent their financial position.

Section 3.2.3 questioned whether overtrading could be detected in a classification model and asked “What impact does cash have on a model’s ability to predict failure?” This research found cash-related ratios such as cash to sales, cash to total assets, etc. were found in both datasets using the heuristic factor search in chapter 4.3, but perhaps more importantly chapter 7.2 found that the only factor with an exponential impact on model output was cash flow to total assets, then demonstrated the position of One.Tel which clearly engaged in overtrading on this model. While the other three predictive models analysed in chapter 7.2 did not include cash-related factors in order to successfully predict corporate failure, there was also no evidence of overtrading in the cases analysed for these models.

Section 3.2.4 considered gearing, but unfortunately the typical measures for gearing such as debt to equity were excluded due to a lack of data in both datasets so this question is left unaddressed.

Section 3.2.5 proposed an examination of the scope of forces across an organisation and it was therefore proposed that there may be value in incorporating macroeconomic factors and that there may also be value in performing an objective clustering of companies rather than simply using industry classifications. In order to test these, chapter 5.2 examined the effect of macroeconomic data but did not find that it assisted the classification model. However chapter 6 developed an improved cluster visualisation technique which was then used in section 6.2, finding that objective clustering of company-years is superior to grouping by industry and not clustering at all.

One interesting result for which no question was specifically posed, was that in three of the four predictive models analysed in section 7.2, the factors found to be useful were generally income or sales related. For example cluster 1 on the Compustat dataset achieved 77.1% accuracy (with 73.1% on the out-of-sample) using net income to net worth, net income to sales, and net worth to sales. The theoretical sources identified in chapter 3 tend to treat profitability as a late symptom of failure, while other factors such as availability of cash are given more importance, but this research shows that income and profitability were more critical to the firms’ risk of failure in these datasets.

8.2 Contributions

This thesis has demonstrated that while there is much literature on corporate failure prediction, such research can be improved both in accuracy and the underlying contribution to corporate failure theory by performing objective factor selection as well as objective clustering on the cases that will be classified when using a supervised learning algorithm. To do so this research has proposed a new cluster visualisation technique, which has been demonstrated to overcome assumptions that are built into other visualisation methods, and the effectiveness of the visualisation has been shown on both synthetic and real datasets.

It has also been demonstrated that while additional information, such as share price data, may intuitively contain useful information that is not available in end of year financial statements, the inclusion of such data may in some situations hinder the predictive systems ability to find the most accurate solution, and there is often an associated decrease in complete data availability when considering additional informational sources.

Finally this research has performed an analysis on the resulting predictive models by examining classifications made with it, and therefore related the successful or unsuccessful classification of companies to the underlying causes of bankruptcy identified in the literature.

8.3 Limitations and Directions for Future Research

This research has utilised many factors for which data was consistently available, however many factors such as debt ratios had to be excluded due to a lack of complete data, which in turn prevented this study from addressing all posed questions. It would be worthwhile applying the demonstrated methodology to alternative datasets that have more complete data availability to determine whether better out-of-sample accuracy can be achieved.

While this research opted to use a fixed failure horizon in the determination of a failure or non-failure classification, it would be interesting to apply the methodology using different horizons of failure and examining whether some factors are more useful than others in a short-term prediction environment versus a long-term prediction environment.

The cluster visualisation methodology proposed in this study was performed on the map units from the Self-Organising Map as it was beyond reasonable processing limits to apply the clustering algorithm to the data itself. There is value in using a more computationally efficient external clustering methodology than Spectral Clustering and potentially clustering company-years directly (both single year and multi-year).

The small case studies undertaken provide some insight into whether or not the typical causes of failure were present, as well as how the model determined the failure score, however there is an opportunity to do further analysis on the financial statements across multiple years to support or refute the findings of this thesis.

Appendix A Optimised Normalisations Applied

CompuStat Data

Name	Norm	Grad	Offset
Working capital to total assets	logistic	11.5279	0.0685
Retained earnings to current assets	logarithmic	2.0616	-0.0949
Earnings before interest and taxes to current assets	logistic	0.9347	-0.7184
Sales to total assets	logistic	2.559	0.2203

Aspect Data

Name	Norm	Grad	Offset
Working capital to total assets	logistic	3.736	0.0227
Retained earnings to current assets	logistic	2.3682	-0.3087
Earnings before interest and taxes to current assets	logistic	1.6552	0.9427
Sales to total assets	logistic	0.2001	-4.6394

Appendix B 5-Fold Cross Validation Results

CompuStat Data

	rand1		rand2		rand3		rand4		rand5		Mean	
	In-Sample	Out-of-Sample	In-Sample	Out-of-Sample	In-Sample	Out-of-Sample	In-Sample	Out-of-Sample	In-Sample	Out-of-Sample	In-Sample	Out-of-Sa
Unnormalised												
SVM	71.3	69.7	76.5	68.8	71.6	73.6	71.0	74.8	72.1	69.3	72.5	71.2
LR	70.9	69.8	72.5	66.4	72.3	73.8	66.1	75.0	73.3	71.6	71.0	71.3
DA	54.5	55.0	56.5	52.1	47.1	48.1	50.5	58.0	44.0	49.7	50.5	52.6
GP	76.5	72.1	**75.5**	**72.7**	**73.1**	**75.6**	**73.3**	**74.4**	**75.1**	**75.7**	74.7	74.1
NN	76.5	72.1	74.2	68.7	71.8	74.2	71.9	74.5	73.6	73.6	73.6	72.6
Normalised												
SVM	68.5	67.7	70.1	67.6	68.0	62.3	70.4	71.1	62.3	64.3	67.9	66.6
LR	71.5	72.0	72.4	70.5	73.8	75.8	68.5	75.0	76.6	73.8	72.6	73.4
DA	68.3	67.8	68.0	62.8	73.3	70.7	61.8	71.9	69.5	65.8	68.2	67.8
GP	76.2	71.7	76.4	73.0	74.3	75.9	71.3	75.1	75.8	74.7	74.8	74.1
NN	74.6	69.2	73.9	69.3	75.7	75.8	72.0	75.8	76.7	73.8	74.6	72.8
Optimised Normalised												
SVM	69.8	68.9	65.8	65.6	66.9	64.4	67.7	69.1	61.8	64.0	66.4	66.4
LR	75.7	71.4	**74.2**	**70.8**	**75.4**	**75.8**	**70.9**	**74.4**	**77.5**	**74.4**	74.8	73.4
DA	72.6	71.6	74.6	69.8	76.9	75.2	68.3	75.7	73.5	74.7	73.2	73.4
GP	76.3	71.9	76.0	73.9	76.4	68.0	74.2	75.3	76.9	72.8	75.9	72.4
NN	78.5	69.6	76.8	71.5	78.5	75.7	75.5	75.8	76.6	73.9	77.2	73.3

Aspect Dataset

	rand1		rand2		rand3		rand4		rand5		Mean	
	In-Sample	Out-of-Sample	In-Sample	Out-of-Sample	In-Sample	Out-of-Sample	In-Sample	Out-of-Sample	In-Sample	Out-of-Sample	In-Sample	Out-of-Sa
Unnormalised												
SVM	60.12	60.44	60.14	61.02	60.77	59.76	60.88	59.93	61.77	59.12	60.74	60.05
LR	61.48	60.54	60.65	57.36	61.73	57.94	58.64	60.47	61.11	61.60	60.72	59.58
DA	52.22	55.06	53.34	51.92	52.37	54.27	54.27	53.93	52.95	53.58	53.03	53.75
GP	64.87	62.28	66.71	61.81	64.74	64.23	65.17	64.05	67.53	61.68	65.80	62.81
NN	62.69	63.65	63.04	60.04	63.20	62.09	64.02	63.10	63.13	62.72	63.22	62.32
Normalised												
SVM	59.85	58.04	60.09	59.78	61.77	58.11	58.43	58.80	62.20	56.94	60.47	58.33
LR	63.49	63.16	65.43	62.67	62.01	59.96	63.72	63.70	63.33	63.25	63.60	62.55
DA	58.35	59.99	59.27	57.78	57.56	59.27	60.75	59.60	57.41	61.81	58.67	59.69
GP	64.97	64.47	67.62	61.95	64.60	63.37	64.90	64.03	65.66	65.28	65.55	63.82
NN	64.64	64.50	65.55	62.04	65.95	62.73	64.88	64.61	65.39	63.87	65.28	63.55
Optimised Normalised												
SVM	55.46	56.62	60.94	58.65	60.32	60.44	61.64	62.49	56.83	61.13	59.04	59.87
LR	64.36	64.21	65.33	62.78	65.41	62.71	64.26	63.52	64.06	62.73	64.68	63.19
DA	63.38	62.94	65.88	60.47	63.31	62.46	62.96	64.97	61.93	66.15	63.49	63.40
GP	65.92	63.17	66.65	60.67	65.44	62.98	65.91	65.34	67.08	62.84	66.20	63.00
NN	65.07	65.10	67.04	60.30	67.13	62.75	66.42	66.07	65.88	66.07	66.31	64.06

Appendix C Factors with Data in Compustat Dataset

- * cash flow to sales
- * cash flow to total assets
- * cash flow to net worth
- cash flow to total debt
- * net income to sales
- * net income to total assets
- * net income to net worth
- net income to total debt
- * current liabilities to total assets
- * long-term liabilities to total assets
- * current plus long-term liabilities to total assets
- * cash to total assets
- quick assets to total assets
- * current assets to total assets
- * working capital to total assets
- * cash to current liabilities
- quick assets to current liabilities
- * current assets to current liabilities
- * cash to sales
- accounts receivable to sales
- inventory to sales
- quick assets to sales
- current assets to sales
- working capital to sales
- * net worth to sales
- * total assets to sales
- * cash to fund expenditures for operations
- defensive assets to fund expenditures for operations
- defensive assets minus current liabilities to fund expenditures for operations.
- * retained earnings to total assets
- * earnings before interest and taxes to total assets
- * sales to total assets.
- inventory to net working capital
- current assets to total debt
- total debt to equity
- * fixed assets to equity
- cash flow to current liabilities
- * current liabilities to equity
- equity and long-term debt to fixed assets
- fixed assets to sales
- * earnings before taxes to sales
- * earnings before taxes to equity
- earnings before taxes plus depreciation to total debt.
- total debt to total assets
- * Net income
- dividends
- * non-cash current assets
- * long-term assets
- * total liabilities.
- net quick assets to inventory
- * cash flow to total liabilities
- * book value to total liabilities
- net worth to long-term liabilities
- net worth to fixed assets
- net operating profit to interest
- sales to inventory
- sales to accounts receivable
- sales to working capital
- sales to current assets minus inventories
- * sales to cash
- * net operating profit to sales
- sales to fixed assets
- * sales to net worth
- * long-term liabilities to current assets
- sales to total capital
- net available for total capital to sales
- * log tangible assets
- interest coverage
- working capital to long-term debt
- book equity to total capital
- * current liabilities to current assets
- funds provided by operations to total liabilities.
- * log of total assets
- quick assets to total liabilities
- cash plus short term investments plus net receivables to current liabilities
- current liabilities plus long-term debt to total assets
- current liabilities to total debt

* present after removal of missing data

Appendix D Factors with Data in Aspect Dataset

- * cash flow to sales
- * cash flow to total assets
- * cash flow to net worth
- cash flow to total debt
- * net income to sales
- * net income to total assets
- * net income to net worth
- net income to total debt
- * current liabilities to total assets
- * long-term liabilities to total assets
- * current plus long-term liabilities to total assets
- * cash to total assets
- quick assets to total assets
- * current assets to total assets
- * working capital to total assets
- * cash to current liabilities
- quick assets to current liabilities
- * current assets to current liabilities
- cash to sales
- * accounts receivable to sales
- inventory to sales
- quick assets to sales
- current assets to sales
- working capital to sales
- * net worth to sales
- * total assets to sales
- * cash to fund expenditures for operations
- * defensive assets to fund expenditures for operations
- * defensive assets minus current liabilities to fund expenditures for operations.
- * retained earnings to total assets
- * earnings before interest and taxes to total assets
- * sales to total assets.
- inventory to net working capital
- current assets to total debt
- total debt to equity
- * fixed assets to equity
- * cash flow to current liabilities
- * current liabilities to equity
- equity and long-term debt to fixed assets
- * fixed assets to sales
- * equity to sales
- * earnings before taxes to sales
- * earnings before taxes to equity
- earnings before taxes plus depreciation to total debt.
- total debt to total assets
- * cash to total assets
- * Net income
- dividends
- * non-cash current assets
- * long-term assets
- * total liabilities.
- net quick assets to inventory
- * cash flow to total liabilities
- * book value to total liabilities
- net worth to fixed assets
- sales to inventory
- sales to accounts receivable
- * sales to working capital
- sales to current assets minus inventories
- * sales to cash
- * net operating profit to sales
- sales to fixed assets
- * sales to net worth
- * net operating profit to total assets
- net operating profit to total debt
- * long-term liabilities to current assets
- * log tangible assets
- interest coverage
- earnings to debt
- working capital to long-term debt
- book equity to total capital
- * current liabilities to current assets
- * funds provided by operations to total liabilities.
- * log of total assets
- quick assets to total liabilities
- current liabilities plus long-term debt to total assets
- * income from operations to total assets
- income from operations plus taxes plus interest expense to total assets
- current liabilities to total debt

* present after removal of missing data

Appendix E Input Impacts for Compustat Data

Factor	Average Impact	Maximum Impact
current plus long-term liabilities to total assets	1.1140	1.1140
total liabilities.	0.1007	0.1007
log of total assets	0.0000	0.0000
cash to total assets	0.0000	0.0000
net income to total assets	26.2691	26.2691
Net income	22.3531	23.9426
net income to net worth	20.3685	20.3685
cash to fund expenditures for operations	6.7705	6.7705
book value to total liabilities	0.0000	0.0000
cash flow to total assets	0.0305	0.0305
cash flow to total liabilities	0.0000	0.0000
log tangible assets	0.0000	0.0000
cash flow to net worth	0.0000	0.0000
current liabilities to total assets	0.7015	0.9404
long-term liabilities to total assets	0.0000	0.0000
current liabilities to equity	0.0000	0.0000
net income to sales	15.0076	22.3998
total assets to sales	5.5896	5.5896
earnings before interest and taxes to total assets	0.0000	0.0000
sales to total assets.	0.7772	0.7772
earnings before taxes to sales	25.5443	26.1653
net operating profit to sales	24.3035	24.3035
net worth to sales	0.0000	0.0000
earnings before taxes to equity	0.0000	0.0000
sales to net worth	4.9611	9.8443
cash to current liabilities	0.0000	0.0000
cash to sales	4.8641	4.8641
sales to cash	0.0000	0.0000
current assets to total assets	0.0000	0.0000
long-term assets	14.1413	14.1413
cash flow to sales	0.0000	0.0000
fixed assets to equity	0.0000	0.0000
non-cash current assets	0.0000	0.0000
retained earnings to total assets	0.0000	0.0000
working capital to total assets	0.0000	0.0000
current assets to current liabilities	0.0000	0.0000
long-term liabilities to current assets	0.0000	0.0000
current liabilities to current assets	0.2003	0.2003

Appendix F Input Impacts for Aspect Data

Factor	Average Impact	Maximum Impact
cash flow to sales	0.0000	0.0000
cash flow to total assets	0.0000	0.0000
cash flow to net worth	0.0000	0.0000
net income to sales	0.0000	0.0000
net income to total assets	5.6848	9.3675
net income to net worth	0.0000	0.0000
current liabilities to total assets	4.9173	6.2280
long-term liabilities to total assets	0.0000	0.0000
current plus long-term liabilities to total assets	1.8953	1.8953
cash to total assets	1.4372	1.4372
current assets to total assets	0.0000	0.0000
working capital to total assets	0.0000	0.0000
cash to current liabilities	8.8947	15.6154
current assets to current liabilities	4.5936	6.4643
accounts receivable to sales	0.0000	0.0000
net worth to sales	0.0000	0.0000
total assets to sales	0.0000	0.0000
cash to fund expenditures for operations	5.5746	9.1386
defensive assets to fund expenditures for operations	0.0000	0.0000
defensive assets minus current liabilities to fund expenditures for operations.	0.0000	0.0000
retained earnings to total assets	10.6884	12.4457
earnings before interest and taxes to total assets	0.0890	0.0890
sales to total assets.	0.0000	0.0000
fixed assets to equity	0.0000	0.0000
cash flow to current liabilities	0.0000	0.0000
current liabilities to equity	0.0000	0.0000
fixed assets to sales	0.0000	0.0000
equity to sales	1.3610	1.3610
earnings before taxes to sales	0.0000	0.0000
earnings before taxes to equity	0.0000	0.0000
cash to total assets	2.2437	4.1109
Net income	0.0000	0.0000
non-cash current assets	0.0000	0.0000
long-term assets	0.0000	0.0000
total liabilities.	0.0000	0.0000
cash flow to total liabilities	7.3956	7.3956
book value to total liabilities	0.0000	0.0000
sales to working capital	0.0000	0.0000
sales to cash	0.4975	0.4975
net operating profit to sales	1.3531	1.3531
sales to net worth	1.8338	1.8338
net operating profit to total assets	8.9077	10.3130
long-term liabilities to current assets	0.0000	0.0000
log tangible assets	0.0000	0.0000
current liabilities to current assets	0.0000	0.0000
funds provided by operations to total liabilities.	6.0390	6.0390
log of total assets	0.0000	0.0000
income from operations to total assets	0.0000	0.0000

Appendix G In-Sample and Out-of-Sample Accuracy for Experiments in Section 4.3

			In-Sample			Out-of-Sample		
Dataset	Algorithm	Factorset	Accuracy	Failed	Non-Failed	Accuracy	Failed	Non-Faile
CompuStat	GP	Best-First	76.4	73.5	79.3	73.8	68.0	79.4
CompuStat	GP	Surviving	75.1	78.6	71.7	71.4	70.1	72.6
CompuStat	GP	Beaver	75.7	69.4	82.0	73.8	66.0	81.6
CompuStat	GP	Altman	72.6	73.4	71.7	72.1	72.2	72.0
CompuStat	GP	All Available	74.5	62.2	88.0	71.3	59.8	87.8
CompuStat	NN	Best-First	75.1	83.3	66.8	70.4	66.3	77.3
CompuStat	NN	Surviving	75.0	84.1	65.8	70.1	65.6	78.3
CompuStat	NN	Beaver	72.2	81.5	62.8	69.8	65.9	76.1
CompuStat	NN	Altman	73.1	84.5	61.7	67.1	62.6	76.4
CompuStat	NN	All Available	74.0	78.6	69.4	69.8	66.8	74.1
Aspect	GP	Best-First	68.4	77.6	58.4	65.3	70.5	59.7
Aspect	GP	Surviving	66.2	50.2	83.1	63.3	38.5	60.0
Aspect	GP	Beaver	66.9	61.2	73.0	63.4	55.0	72.3
Aspect	GP	Altman	64.6	72.6	57.5	61.1	63.5	58.6
Aspect	GP	All Available	66.0	57.7	74.9	63.2	63.5	65.0
Aspect	NN	Best-First	66.0	76.7	55.2	64.0	61.2	68.7
Aspect	NN	Surviving	59.7	52.9	66.4	52.3	52.2	52.4
Aspect	NN	Beaver	58.8	61.2	56.5	57.5	57.2	57.9
Aspect	NN	Altman	61.2	85.4	37.0	60.4	57.0	70.3
Aspect	NN	All Available	65.8	60.2	71.4	62.5	63.3	61.8

Appendix H Validation Accuracy for Best-First Search Using Genetic Programming on Compustat Dataset

1	earnings before taxes to sales	74.58
2	net income to total assets	76.01
3	cash to fund expenditures for operations	76.08
4	sales to total assets.	76.04
5	cash flow to total assets	76.39
6	total liabilities.	75.52
7	total assets to sales	75.50
8	long-term assets	75.90
9	current plus long-term liabilities to total assets	75.89
10	current liabilities to current assets	75.39
11	cash to sales	75.23
12	Net income	75.03
13	net operating profit to sales	75.30
14	net income to sales	75.17
15	net income to net worth	75.16
16	current liabilities to total assets	74.77
17	sales to net worth	75.13

Appendix I Validation Accuracy for Best-First Search Using Neural Networks on Compustat Dataset

1	net income	73.01
2	net operating profit to sales	73.47
3	total assets to sales	74.43
4	net income to sales	74.01
5	total liabilities	73.85
6	cash to sales	73.78
7	net income to net worth	72.81
8	long-term assets	71.75
9	current plus long-term liabilities to total assets	72.31
10	net income to total assets	74.28
11	current liabilities to current assets	73.64
12	current liabilities to total assets	74.17
13	sales to total assets.	75.10
14	earnings before taxes to sales	74.96
15	cash flow to total assets	74.52
16	cash to fund expenditures for operations	74.66
17	sales to net worth	74.96

Appendix J Validation Accuracy for Best-First Search Using Genetic Programming on Aspect Dataset

1	net income to total assets	63.74
2	cash to current liabilities	66.26
3	cash flow to total liabilities	68.14
4	cash to total assets	68.01
5	current plus long-term liabilities to total assets	67.75
6	sales to net worth	68.33
7	funds provided by operations to total liabilities.	67.28
8	current liabilities to total assets	67.50
9	retained earnings to total assets	68.35
10	cash to fund expenditures for operations	67.21
11	net operating profit to sales	67.63
12	net operating profit to total assets	66.70
13	equity to sales	66.91
14	current assets to current liabilities	66.72
15	sales to cash	66.61
16	earnings before interest and taxes to total assets	66.13

Appendix K Validation Accuracy for Best-First Search Using Neural Networks on Aspect Dataset

1	current plus long-term liabilities to total assets	59.23
2	funds provided by operations to total liabilities.	60.77
3	current assets to current liabilities	64.17
4	cash flow to total liabilities	65.24
5	sales to cash	64.97
6	net operating profit to sales	64.94
7	equity to sales	64.98
8	net income to total assets	64.59
9	net operating profit to total assets	65.79
10	cash to current liabilities	65.58
11	cash to fund expenditures for operations	65.65
12	earnings before interest and taxes to total assets	65.96
13	sales to net worth	65.48
14	current liabilities to total assets	63.43
15	cash to total assets	61.49
16	retained earnings to total assets	59.65

Appendix L Input Impacts for Compustat Data without Share Market Information

Factor	Average Impact	Maximum Impact
current plus long-term liabilities to total assets	19.5509	19.5509
total liabilities.	0.0000	0.0000
log of total assets	0.0000	0.0000
cash to total assets	0.0000	0.0000
net income to total assets	16.6235	23.6853
Net income	24.0203	25.7024
net income to net worth	11.1084	19.8879
cash to fund expenditures for operations	0.2653	0.2653
book value to total liabilities	0.0000	0.0000
cash flow to total assets	0.0000	0.0000
cash flow to total liabilities	0.0000	0.0000
log tangible assets	0.0000	0.0000
cash flow to net worth	0.0000	0.0000
current liabilities to total assets	21.3722	21.3722
long-term liabilities to total assets	0.0000	0.0000
current liabilities to equity	0.0000	0.0000
net income to sales	23.1147	26.5832
total assets to sales	0.0000	0.0000
earnings before interest and taxes to total assets	0.0000	0.0000
sales to total assets.	0.7228	1.4456
earnings before taxes to sales	22.5712	24.4147
net operating profit to sales	24.8536	25.8183
net worth to sales	15.9702	15.9702
earnings before taxes to equity	0.0000	0.0000
sales to net worth	0.0000	0.0000
cash to current liabilities	0.0000	0.0000
cash to sales	0.0000	0.0000
sales to cash	0.0000	0.0000
current assets to total assets	0.0000	0.0000
long-term assets	0.0000	0.0000
cash flow to sales	0.0000	0.0000
fixed assets to equity	0.0000	0.0000
non-cash current assets	0.0000	0.0000
retained earnings to total assets	18.6213	18.6213
working capital to total assets	0.0000	0.0000
current assets to current liabilities	0.0000	0.0000
long-term liabilities to current assets	0.0000	0.0000
current liabilities to current assets	0.0033	0.0067

Appendix M Input Impacts for Aspect Data without Share Market Information

Factor	Average Impact	Maximum Impact
cash flow to sales	0.0000	0.0000
cash flow to total assets	0.0000	0.0000
cash flow to net worth	0.0000	0.0000
net income to sales	0.0000	0.0000
net income to total assets	0.0000	0.0000
net income to net worth	0.0000	0.0000
current liabilities to total assets	0.0000	0.0000
long-term liabilities to total assets	1.3378	1.3378
current plus long-term liabilities to total assets	2.7932	2.7932
cash to total assets	0.7369	1.4739
current assets to total assets	1.1116	1.1116
working capital to total assets	2.5420	2.5420
cash to current liabilities	11.7603	16.9791
current assets to current liabilities	0.0000	0.0000
accounts receivable to sales	0.0000	0.0000
net worth to sales	0.0000	0.0000
total assets to sales	0.0000	0.0000
cash to fund expenditures for operations	18.2083	20.1963
defensive assets to fund expenditures for operations	18.1255	19.0433
defensive assets minus current liabilities to fund expenditures for operations.	17.7577	17.8143
retained earnings to total assets	10.6384	13.9324
earnings before interest and taxes to total assets	0.1815	0.1815
sales to total assets.	0.0000	0.0000
fixed assets to equity	0.0000	0.0000
cash flow to current liabilities	1.7019	1.7019
current liabilities to equity	0.0000	0.0000
fixed assets to sales	0.0000	0.0000
equity to sales	0.0000	0.0000
earnings before taxes to sales	0.0000	0.0000
earnings before taxes to equity	0.0000	0.0000
cash to total assets	0.0000	0.0000
Net income	0.0000	0.0000
non-cash current assets	0.0000	0.0000
long-term assets	0.0000	0.0000
total liabilities.	0.0000	0.0000
cash flow to total liabilities	0.0000	0.0000
book value to total liabilities	0.0000	0.0000
sales to working capital	0.0000	0.0000
sales to cash	2.4888	3.4528
net operating profit to sales	12.8921	12.9399
sales to net worth	0.0000	0.0000
net operating profit to total assets	6.7612	11.1859
long-term liabilities to current assets	0.0000	0.0000
log tangible assets	0.0000	0.0000
current liabilities to current assets	0.0000	0.0000
funds provided by operations to total liabilities.	0.0000	0.0000
log of total assets	0.0000	0.0000
income from operations to total assets	0.0000	0.0000

Appendix N Validation Accuracy for Best-First Search Using Genetic Programming on Compustat Dataset with Market Data

1	earnings before taxes to sales	75.60
2	net income to total assets	76.01
3	cash to fund expenditures for operations	75.87
4	close price	76.26
5	market cap	76.23
6	sales to total assets.	77.16
7	net income to sales	76.01
8	volume	76.66
9	variance	75.75
10	net worth to sales	76.44
11	beta	76.47
12	net operating profit to sales	76.34
13	Net income	77.11
14	net income to net worth	76.27
15	current liabilities to current assets	76.27
16	current liabilities to total assets	76.09
17	return	76.27
18	current plus long-term liabilities to total assets	75.71
19	retained earnings to total assets	74.56

Appendix O Validation Accuracy for Best-First Search Using Neural Networks on Compustat Dataset with Market Data

1	close price	73.78
2	current liabilities to total assets	76.90
3	net income to total assets	78.34
4	retained earnings to total assets	79.47
5	cash to fund expenditures for operations	79.91
6	beta	80.04
7	current plus long-term liabilities to total assets	80.22
8	net operating profit to sales	80.05
9	return	80.26
10	current liabilities to current assets	79.75
11	variance	79.28
12	earnings before taxes to sales	79.82
13	net income to sales	79.19
14	net worth to sales	79.25
15	sales to total assets.	79.24
16	Net income	78.89
17	market cap	78.61
18	volume	78.96
19	net income to net worth	79.01

Appendix P Validation Accuracy for Best-First Search Using Genetic Programming on Compustat Dataset without Market Data

1	earnings before taxes to sales	75.60
2	net income to total assets	76.01
3	cash to fund expenditures for operations	75.87
4	current liabilities to total assets	76.09
5	current liabilities to current assets	76.25
6	net income to sales	77.42
7	net income to net worth	76.23
8	net worth to sales	76.82
9	sales to total assets.	76.27
10	Net income	76.27
11	current plus long-term liabilities to total assets	76.27
12	net operating profit to sales	75.47
13	retained earnings to total assets	74.33

Appendix Q Validation Accuracy for Best-First Search Using Neural Networks on Compustat Dataset without Market Data

1	Net income	73.23
2	net operating profit to sales	73.84
3	net worth to sales	75.85
4	net income to sales	76.66
5	earnings before taxes to sales	75.00
6	net income to net worth	75.21
7	current liabilities to total assets	72.26
8	net income to total assets	74.51
9	current plus long-term liabilities to total assets	76.37
10	cash to fund expenditures for operations	76.78
11	current liabilities to current assets	78.19
12	retained earnings to total assets	75.28
13	sales to total assets.	75.49

Appendix R Validation Accuracy for Best-First Search Using Genetic Programming on Aspect Dataset with Market Data

1	net operating profit to sales	65.25
2	sales to cash	68.27
3	current plus long-term liabilities to total assets	68.27
4	cash to fund expenditures for operations	68.88
5	current assets to total assets	68.47
6	retained earnings to total assets	68.31
7	net operating profit to total assets	68.65
8	cash to current liabilities	68.65
9	defensive assets minus current liabilities to fund expenditures for operations.	68.70
10	variance	68.50
11	volume	68.18
12	defensive assets to fund expenditures for operations	67.91
13	cash flow to current liabilities	68.54
14	beta	68.59
15	return	68.36
16	close	67.93
17	long-term liabilities to total assets	67.48
18	market cap	67.52
19	cash to total assets	67.54
20	working capital to total assets	66.79
21	earnings before interest and taxes to total assets	66.63

Appendix S Validation Accuracy for Best-First Search Using Neural Networks on Aspect Dataset with Market Data

1	close price	61.87
2	working capital to total assets	65.19
3	net operating profit to total assets	68.00
4	long-term liabilities to total assets	69.18
5	cash flow to current liabilities	69.16
6	cash to current liabilities	69.28
7	retained earnings to total assets	69.83
8	cash to fund expenditures for operations	70.67
9	earnings before interest and taxes to total assets	69.88
10	market cap	69.76
11	return	69.22
12	defensive assets minus current liabilities to fund expenditures for operations.	69.56
13	net operating profit to sales	69.52
14	beta	69.70
15	volume	68.62
16	defensive assets to fund expenditures for operations	69.64
17	variance	70.05
18	cash to total assets	68.68
19	sales to cash	69.00
20	current plus long-term liabilities to total assets	69.34
21	current assets to total assets	68.78

Appendix T Validation Accuracy for Best-First Search Using Genetic Programming on Aspect Dataset without Market Data

1	net operating profit to sales	65.25
2	sales to cash	68.27
3	current plus long-term liabilities to total assets	68.27
4	cash to fund expenditures for operations	68.88
5	current assets to total assets	68.47
6	retained earnings to total assets	68.31
7	net operating profit to total assets	68.65
8	cash to current liabilities	68.65
9	defensive assets minus current liabilities to fund expenditures for operations.	68.70
10	earnings before interest and taxes to total assets	68.11
11	cash flow to current liabilities	67.27
12	working capital to total assets	68.25
13	long-term liabilities to total assets	67.91
14	cash to total assets	66.91
15	defensive assets to fund expenditures for operations	66.75

Appendix U Validation Accuracy for Best-First Search Using Neural Networks on Aspect Dataset without Market Data

1	net operating profit to total assets	61.83
2	cash to current liabilities	67.32
3	current plus long-term liabilities to total assets	68.13
4	net operating profit to sales	68.85
5	cash to total assets	68.57
6	cash to fund expenditures for operations	69.40
7	working capital to total assets	69.44
8	defensive assets minus current liabilities to fund expenditures for operations.	68.78
9	defensive assets to fund expenditures for operations	69.00
10	earnings before interest and taxes to total assets	69.57
11	cash flow to current liabilities	70.30
12	sales to cash	67.71
13	long-term liabilities to total assets	67.27
14	current assets to total assets	67.39
15	retained earnings to total assets	66.87

Appendix V Validation Accuracy for Best-First Search Using Genetic Programming on Compustat Dataset with Macroeconomic Data

1	earnings before taxes to sales	74.76
2	net income to total assets	75.12
3	cash to fund expenditures for operations	75.86
4	net income to net worth	76.19
5	cash flow to total assets	76.88
6	Dow Jones Index	76.20
7	GDP Growth	76.12
8	sales to total assets.	75.90
9	net operating profit to sales	75.90
10	net income to sales	75.01
11	real effective exchange rate	76.20
12	current liabilities to current assets	75.59
13	current liabilities to total assets	75.42
14	consumer price index	75.03
15	total assets to sales	75.03
16	long-term assets	74.59
17	cash to sales	75.78
18	interest rate	75.03
19	current plus long-term liabilities to total assets	75.03
20	sales to net worth	75.03
21	total liabilities	75.03
22	net income	74.38

Appendix W Validation Accuracy for Best-First Search Using Neural Networks on Compustat Dataset with Macroeconomic Data

1	Net income	73.01
2	net operating profit to sales	73.47
3	total assets to sales	74.43
4	net income to sales	74.01
5	total liabilities.	73.85
6	cash to sales	73.78
7	net income to net worth	72.81
8	long-term assets	71.75
9	current plus long-term liabilities to total assets	72.31
10	net income to total assets	74.28
11	real effective exchange rate	74.29
12	sales to net worth	74.09
13	sales to total assets.	73.49
14	dow jones industrial index	74.60
15	cash to fund expenditures for operations	74.62
16	current liabilities to current assets	74.28
17	current liabilities to total assets	75.61
18	earnings before taxes to sales	74.93
19	cash flow to total assets	74.84
20	GPD Growth	75.18
21	consumer price index	73.53
22	interest rate	74.94

Appendix X Validation Accuracy for Best-First Search Using Genetic Programming on Aspect Dataset with Macroeconomic Data

1	net income to total assets	63.92
2	cash to current liabilities	66.81
3	cash flow to total liabilities	68.15
4	cash to total assets	67.94
5	net operating profit to sales	67.35
6	cash to total assets	67.24
7	GDP Growth	67.26
8	consumer price index	67.37
9	current assets to current liabilities	67.00
10	cash to fund expenditures for operations	67.62
11	current plus long-term liabilities to total assets	67.54
12	current liabilities to total assets	67.26
13	equity to sales	67.00
14	sales to net worth	67.39
15	interest rate	66.89
16	sales to cash	66.56
17	australian all ordinaries	66.51
18	retained earnings to total assets	66.81
19	net operating profit to total assets	66.79
20	earnings before interest and taxes to total assets	66.43
21	real effective exchange rate	66.46
22	funds provided by operations to total liabilities.	66.07

Appendix Y Validation Accuracy for Best-First Search Using Neural Networks on Aspect Dataset with Macroeconomic Data

1	current plus long-term liabilities to total assets	59.23
2	funds provided by operations to total liabilities.	60.77
3	current assets to current liabilities	64.17
4	cash flow to total liabilities	65.24
5	sales to cash	64.97
6	net operating profit to sales	64.94
7	equity to sales	64.98
8	retained earnings to total assets	64.65
9	cash to current liabilities	64.85
10	net operating profit to total assets	65.01
11	cash to fund expenditures for operations	65.65
12	current liabilities to total assets	63.65
13	net income to total assets	63.92
14	earnings before interest and taxes to total assets	62.81
15	consumer price index	61.30
16	sales to net worth	62.12
17	real effective exchange rate	61.64
18	australian all ordinaries index	62.20
19	cash to total assets	61.76
20	cash to total assets	61.86
21	interest rate	61.13
22	GDP growth	61.95

Appendix Z Validation Accuracy for Best-First Search Using Genetic Programming on Cluster 1 & 2 on Compustat Dataset

earnings before taxes to sales	75.06
cash to fund expenditures for operations	75.99
net income to sales	76.51
net income to net worth	76.28
cash flow to total assets	75.78
total assets to sales	76.11
net operating profit to sales	75.33
Net income	75.78
cash to sales	75.24
total liabilities.	75.76
current liabilities to total assets	75.83
long-term assets	75.50
sales to net worth	75.68
net income to total assets	74.48
current liabilities to current assets	74.96
sales to total assets.	74.48
current plus long-term liabilities to total assets	74.33

earnings before taxes to sales	81.71
net income to net worth	82.34
Net income	82.71
current liabilities to total assets	82.34
net income to total assets	84.47
long-term assets	82.82
net income to sales	82.97
sales to net worth	83.45
cash flow to total assets	82.58
current liabilities to current assets	82.58
current plus long-term liabilities to total assets	82.71
total assets to sales	82.91
net operating profit to sales	82.55
cash to sales	82.77
sales to total assets.	82.17
total liabilities.	82.61
cash to fund expenditures for operations	81.98

Appendix AA Validation Accuracy for Best-First Search Using Neural Networks on Cluster 1 & 2 on Compustat Dataset

earnings before taxes to sales	81.28
cash flow to total assets	83.06
current liabilities to current assets	83.67
net income to sales	83.69
total assets to sales	84.21
current plus long-term liabilities to total assets	84.98
cash to sales	86.89
total liabilities.	88.06
net operating profit to sales	88.64
cash to fund expenditures for operations	88.21
net income to net worth	89.50
net income to total assets	87.73
Net income	88.49
long-term assets	87.97
current liabilities to total assets	86.98
sales to net worth	85.82
sales to total assets.	80.86

current plus long-term liabilities to total assets	72.06
Net income	74.91
sales to total assets.	78.22
net income to sales	79.10
cash flow to total assets	78.86
total liabilities.	78.93
cash to sales	78.32
net operating profit to sales	78.75
sales to net worth	78.44
long-term assets	78.58
net income to net worth	78.30
total assets to sales	78.40
earnings before taxes to sales	77.51
cash to fund expenditures for operations	77.94
current liabilities to total assets	77.37
current liabilities to current assets	76.68
net income to total assets	76.38

Appendix BB Validation Accuracy for Best-First Search Using Genetic Programming on Manufacturing and Non-Manufacturing Companies on Compustat Dataset

current plus long-term liabilities to total assets	77.34
net income to total assets	78.50
current liabilities to total assets	80.81
total liabilities.	80.78
current liabilities to current assets	80.54
long-term assets	80.48
net income to net worth	79.51
earnings before taxes to sales	79.97
Net income	78.31
cash to fund expenditures for operations	78.24
sales to net worth	78.50
sales to total assets.	79.86
net operating profit to sales	78.95
net income to sales	78.10
total assets to sales	77.09
cash to sales	77.49
cash flow to total assets	77.10

earnings before taxes to sales	80.47
current liabilities to current assets	80.48
net income to total assets	80.48
current liabilities to total assets	80.47
cash to fund expenditures for operations	80.49
sales to total assets.	80.47
cash to sales	80.48
cash flow to total assets	80.36
sales to net worth	80.21
net income to net worth	80.48
long-term assets	80.47
Net income	80.21
current plus long-term liabilities to total assets	80.21
total liabilities.	80.01
net operating profit to sales	80.24
net income to sales	78.28
total assets to sales	78.69

Appendix CC Validation Accuracy for Best-First Search Using Genetic Programming and Neural Networks on Unclustered and Ungrouped Compustat Dataset

earnings before taxes to sales	76.68
net income to total assets	76.66
cash to fund expenditures for operations	76.84
current plus long-term liabilities to total assets	77.62
sales to total assets.	77.61
cash flow to total assets	77.59
long-term assets	77.21
net income to sales	77.15
current liabilities to total assets	77.46
cash to sales	77.29
current liabilities to current assets	76.35
sales to net worth	77.48
net operating profit to sales	77.34
total liabilities.	76.45
total assets to sales	76.93
Net income	76.45
net income to net worth	75.97

Net income	73.23
net operating profit to sales	74.96
total assets to sales	75.62
net income to sales	75.64
cash to sales	75.62
net income to net worth	75.19
total liabilities.	75.29
earnings before taxes to sales	75.67
current plus long-term liabilities to total assets	75.23
sales to total assets.	76.96
sales to net worth	77.65
long-term assets	77.17
net income to total assets	75.48
current liabilities to total assets	76.67
current liabilities to current assets	76.69
cash to fund expenditures for operations	76.41
cash flow to total assets	75.78

Appendix DD Validation Accuracy for Best-First Search Using Genetic Programming on Cluster 1 & 2 on Aspect Dataset

net operating profit to sales	80.59
net income to total assets	80.88
cash to total assets	81.57
funds provided by operations to total liabilities.	81.57
retained earnings to total assets	81.32
sales to cash	81.19
cash to fund expenditures for operations	81.24
sales to net worth	80.92
current plus long-term liabilities to total assets	81.59
cash to current liabilities	81.07
earnings before interest and taxes to total assets	80.39
current assets to current liabilities	80.32
net operating profit to total assets	80.62
cash flow to total liabilities	80.32
current liabilities to total assets	80.32
equity to sales	80.07
cash to total assets	79.80

cash to current liabilities	66.78
earnings before interest and taxes to total assets	69.67
current plus long-term liabilities to total assets	70.76
current assets to current liabilities	70.76
net income to total assets	70.88
sales to net worth	71.48
sales to cash	70.88
funds provided by operations to total liabilities.	71.68
cash flow to total liabilities	71.31
current liabilities to total assets	70.76
net operating profit to sales	71.35
net operating profit to total assets	71.09
retained earnings to total assets	70.90
equity to sales	70.88
cash to total assets	70.18
cash to fund expenditures for operations	69.87
cash to total assets	69.45

Appendix EE Validation Accuracy for Best-First Search Using Neural Networks on Cluster 1 & 2 on Aspect Dataset

current plus long-term liabilities to total assets	65.06
net operating profit to total assets	69.37
current assets to current liabilities	70.93
cash to current liabilities	72.51
cash to fund expenditures for operations	74.22
cash to total assets	74.46
net operating profit to sales	73.45
cash flow to total liabilities	73.12
current liabilities to total assets	74.73
retained earnings to total assets	74.20
equity to sales	73.53
cash to total assets	74.41
sales to cash	73.79
net income to total assets	73.10
funds provided by operations to total liabilities.	71.86
earnings before interest and taxes to total assets	72.80
sales to net worth	71.54

cash to current liabilities	64.91
earnings before interest and taxes to total assets	65.90
current plus long-term liabilities to total assets	66.71
retained earnings to total assets	67.43
net operating profit to total assets	69.54
sales to cash	71.27
sales to net worth	70.84
net operating profit to sales	71.18
current liabilities to total assets	71.91
cash flow to total liabilities	70.32
current assets to current liabilities	68.90
equity to sales	68.55
cash to total assets	68.15
cash to fund expenditures for operations	67.49
net income to total assets	68.06
cash to total assets	68.38
funds provided by operations to total liabilities.	67.23

Appendix FF Validation Accuracy for Best-First Search Using Genetic Programming on Manufacturing and Non-Manufacturing Companies on Aspect Dataset

net operating profit to sales	69.47
funds provided by operations to total liabilities.	70.62
current plus long-term liabilities to total assets	70.74
cash flow to total liabilities	70.73
current liabilities to total assets	70.69
retained earnings to total assets	70.51
earnings before interest and taxes to total assets	71.73
current assets to current liabilities	70.80
cash to current liabilities	70.08
net operating profit to total assets	70.15
net income to total assets	70.01
sales to net worth	70.66
equity to sales	71.53
cash to fund expenditures for operations	70.07
cash to total assets	68.65
sales to cash	69.10

net income to total assets	66.49
current liabilities to total assets	69.92
cash flow to total liabilities	70.25
equity to sales	70.78
retained earnings to total assets	70.80
current assets to current liabilities	70.78
cash to total assets	70.78
sales to net worth	70.78
net operating profit to total assets	69.58
current plus long-term liabilities to total assets	70.18
cash to total assets	70.78
funds provided by operations to total liabilities.	68.67
cash to fund expenditures for operations	69.84
sales to cash	69.58
cash to current liabilities	69.96
net operating profit to sales	68.98
earnings before interest and taxes to total assets	68.30

Appendix GG Validation Accuracy for Best-First Search Using Neural Networks on Manufacturing and Non-Manufacturing Companies on Aspect Dataset

current liabilities to total assets	59.94
retained earnings to total assets	61.73
earnings before interest and taxes to total assets	61.85
net operating profit to total assets	62.79
funds provided by operations to total liabilities.	65.36
current assets to current liabilities	65.52
sales to cash	65.36
net operating profit to sales	64.02
current plus long-term liabilities to total assets	63.13
cash flow to total liabilities	64.50
net income to total assets	62.55
sales to net worth	60.78
cash to current liabilities	62.77
equity to sales	60.66
cash to total assets	59.74
cash to fund expenditures for operations	61.08

net operating profit to total assets	60.75
current liabilities to total assets	66.20
cash to total assets	67.40
retained earnings to total assets	68.61
funds provided by operations to total liabilities.	68.95
earnings before interest and taxes to total assets	69.27
sales to cash	69.70
current plus long-term liabilities to total assets	69.35
cash to total assets	69.36
net income to total assets	70.38
equity to sales	69.61
cash flow to total liabilities	68.66
net operating profit to sales	69.15
cash to current liabilities	68.70
sales to net worth	68.27
current assets to current liabilities	68.37
cash to fund expenditures for operations	68.01

Appendix HH Input Impacts for Compustat Dataset (Sequential Division)

Factor	Average Input	Maximum Input
current plus long-term liabilities to total assets	0	0
total liabilities.	0	0
log of total assets	0	0
cash to total assets	0.00164	0.00328
net income to total assets	25.55919	25.97919
Net income	26.12919	27.66356
net income to net worth	0.97473	0.97473
cash to fund expenditures for operations	0	0
book value to total liabilities	0	0
cash flow to total assets	0	0
cash flow to total liabilities	0.00656	0.00656
log tangible assets	0	0
cash flow to net worth	0	0
current liabilities to total assets	0.05574	0.05574
long-term liabilities to total assets	0	0
current liabilities to equity	0	0
net income to sales	21.42287	25.28689
total assets to sales	1.91939	2.93963
earnings before interest and taxes to total assets	0	0
sales to total assets.	1.45438	1.73074
earnings before taxes to sales	0	0
net operating profit to sales	0	0
net worth to sales	4.3959	4.3959
earnings before taxes to equity	0	0
sales to net worth	0	0
cash to current liabilities	0.44124	0.44124
cash to sales	0	0
sales to cash	0	0
current assets to total assets	0	0
long-term assets	0	0
cash flow to sales	0	0
fixed assets to equity	0	0
non-cash current assets	0	0
retained earnings to total assets	0	0
working capital to total assets	0	0
current assets to current liabilities	2.46238	2.46238
long-term liabilities to current assets	0	0
current liabilities to current assets	1.4085	1.98745

Appendix II Input Impacts for Aspect Dataset (Sequential Division)

Factor	Average Impact	Maximum Impact
cash flow to sales	0	0
cash flow to total assets	0.82529	0.82529
cash flow to net worth	0	0
net income to sales	10.6382	10.6382
net income to total assets	0	0
net income to net worth	2.57258	4.12164
current liabilities to total assets	2.84101	7.19615
long-term liabilities to total assets	0	0
current plus long-term liabilities to total assets	4.81008	11.57592
cash to total assets	1.87549	3.49908
current assets to total assets	0	0
working capital to total assets	0	0
cash to current liabilities	8.69881	10.90395
current assets to current liabilities	0	0
accounts receivable to sales	0	0
net worth to sales	0	0
total assets to sales	0	0
cash to fund expenditures for operations	0	0
defensive assets to fund expenditures for operations	0	0
defensive assets minus current liabilities to fund expenditures for operations	0	0
retained earnings to total assets	9.80461	11.75483
earnings before interest and taxes to total assets	1.65357	3.30714
sales to total assets.	0	0
fixed assets to equity	0	0
cash flow to current liabilities	0.08754	0.08754
current liabilities to equity	5.30233	7.4393
fixed assets to sales	0	0
equity to sales	0	0
earnings before taxes to sales	0	0
earnings before taxes to equity	7.3822	9.26781
cash to total assets	0	0
Net income	0	0
non-cash current assets	0	0
long-term assets	0	0
total liabilities.	2.00188	2.00188
cash flow to total liabilities	0	0
book value to total liabilities	5.65332	9.14174
sales to working capital	0	0
sales to cash	0	0
net operating profit to sales	11.76429	11.76429
sales to net worth	0.46462	0.46462
net operating profit to total assets	7.94372	10.52194
long-term liabilities to current assets	0	0
log tangible assets	0	0
current liabilities to current assets	0	0
funds provided by operations to total liabilities.	0	0
log of total assets	0	0
income from operations to total assets	1.27113	2.44036

Appendix JJ Validation Accuracy for Best-First Search Using Genetic Programming on Compustat Dataset (Sequential Division)

net income	76.37
current liabilities to total assets	76.84
current liabilities to current assets	77.37
net worth to sales	76.95
net income to sales	77.81
cash to current liabilities	77.34
current assets to current liabilities	77.58
cash to total assets	77.37
sales to net worth	77.47
book value to total liabilities	77.15
total assets to sales	77.11
working capital to total assets	76.78
net income to total assets	77.12
cash flow to total liabilities	76.52
sales to total assets	76.98
cash to sales	76.20
net income to net worth	76.01

Appendix KK Validation Accuracy for Best-First Search Using Neural Networks on Compustat Dataset (Sequential Division)

net income	71.91
net income to sales	74.85
total assets to sales	76.72
net worth to sales	76.64
cash to sales	77.13
book value to total liabilities	76.16
sales to net worth	74.94
net income to net worth	75.89
net income to total assets	73.43
cash to total assets	75.42
cash flow to total liabilities	76.13
cash to current liabilities	75.83
current liabilities to total assets	75.06
current assets to current liabilities	74.66
working capital to total assets	74.28
current liabilities to current assets	74.82
sales to total assets.	73.93

Appendix LL Validation Accuracy for Best-First Search Using Genetic Programming on Aspect Dataset (Sequential Division)

net operating profit to total assets	63.11
current liabilities to total assets	64.70
net operating profit to sales	65.06
earnings before taxes to equity	65.00
cash to current liabilities	65.04
retained earnings to total assets	66.53
cash flow to total assets	65.99
cash flow to current liabilities	65.86
current plus long-term liabilities to total assets	66.32
income from operations to total assets	65.65
total liabilities.	65.79
current liabilities to equity	65.55
sales to net worth	65.23
net income to sales	65.08
book value to total liabilities	64.97
earnings before interest and taxes to total assets	64.97
cash to total assets	64.42
net income to net worth	64.77

Appendix MM Validation Accuracy for Best-First Search Using Neural Networks on Aspect Dataset (Sequential Division)

current plus long-term liabilities to total assets	59.65
earnings before interest and taxes to total assets	62.81
total liabilities.	64.06
current liabilities to total assets	65.41
net income to net worth	65.47
current liabilities to equity	65.42
cash flow to current liabilities	65.35
income from operations to total assets	65.16
net operating profit to total assets	65.22
cash to current liabilities	65.86
retained earnings to total assets	65.42
net income to sales	65.50
earnings before taxes to equity	65.44
book value to total liabilities	65.25
cash to total assets	65.45
cash flow to total assets	65.06
sales to net worth	64.62
net operating profit to sales	65.33

Appendix NN Validation Accuracy for Best-First Search Using Genetic Programming on Cluster 1 & 2 on Compustat Dataset (Sequential Division)

net income to sales	72.39
net income to net worth	75.85
net worth to sales	77.13
net income	76.54
current assets to current liabilities	76.80
cash to total assets	76.80
working capital to total assets	76.88
sales to net worth	77.01
sales to total assets	76.52
current liabilities to current assets	76.33
cash to current liabilities	76.70
net income to total assets	76.83
cash flow to total liabilities	75.75
cash to sales	75.45
book value to total liabilities	76.10
total assets to sales	75.83
current liabilities to total assets	76.08

net income	88.02
current liabilities to current assets	89.09
sales to total assets.	88.50
cash to total assets	88.17
book value to total liabilities	88.25
cash to current liabilities	88.27
net worth to sales	88.14
sales to net worth	88.26
total assets to sales	88.05
net income to sales	88.02
current liabilities to total assets	88.02
cash flow to total liabilities	88.06
net income to total assets	87.97
net income to net worth	88.19
current assets to current liabilities	88.02
working capital to total assets	88.02
cash to sales	87.93

Appendix OO Validation Accuracy for Best-First Search Using Neural Networks on Cluster 1 on Compustat Dataset (Sequential Division)

net income to total assets	70.34
cash to current liabilities	74.15
cash flow to total liabilities	74.62
net worth to sales	75.29
net income to sales	75.34
sales to net worth	75.31
Net income	74.80
net income to net worth	74.91
book value to total liabilities	75.52
cash to sales	74.57
cash to total assets	75.58
total assets to sales	76.15
current liabilities to total assets	74.83
sales to total assets.	75.57
current liabilities to current assets	75.64
working capital to total assets	75.48
current assets to current liabilities	74.48

Appendix PP Validation Accuracy for Best-First Search Using Genetic Programming on Cluster 1 & 2 on Aspect Dataset (Sequential Division)

cash to current liabilities	67.52
cash flow to total assets	72.57
current plus long-term liabilities to total assets	72.07
sales to net worth	72.45
net operating profit to sales	72.01
current liabilities to equity	72.07
net income to net worth	72.34
earnings before taxes to equity	71.69
cash to total assets	71.69
earnings before interest and taxes to total assets	71.74
total liabilities.	71.75
retained earnings to total assets	70.97
net operating profit to total assets	71.03
cash flow to current liabilities	71.43
income from operations to total assets	71.61
net income to sales	71.36
current liabilities to total assets	70.39
book value to total liabilities	69.45

net operating profit to sales	74.25
current plus long-term liabilities to total assets	75.16
net operating profit to total assets	75.70
book value to total liabilities	75.27
current liabilities to total assets	75.43
earnings before interest and taxes to total assets	75.22
retained earnings to total assets	75.33
cash flow to current liabilities	76.72
sales to net worth	75.22
current liabilities to equity	75.06
cash flow to total assets	75.97
net income to sales	75.00
total liabilities.	75.00
cash to current liabilities	75.27
cash to total assets	75.16
net income to net worth	75.00
earnings before taxes to equity	74.57
income from operations to total assets	73.45

Appendix QQ Validation Accuracy for Best-First Search Using Neural Networks on Cluster 1 & 2 on Aspect Dataset (Sequential Division)

current liabilities to total assets	63.40
earnings before interest and taxes to total assets	65.24
net operating profit to total assets	67.05
total liabilities.	68.74
net income to sales	69.69
book value to total liabilities	69.63
income from operations to total assets	70.52
current plus long-term liabilities to total assets	71.66
cash flow to current liabilities	71.13
cash to total assets	70.88
sales to net worth	71.86
cash flow to total assets	72.25
retained earnings to total assets	72.53
current liabilities to equity	72.25
net operating profit to sales	72.44
cash to current liabilities	73.28
net income to net worth	71.72
earnings before taxes to equity	70.72

net operating profit to total assets	69.76
book value to total liabilities	72.44
total liabilities.	73.69
net income to net worth	75.08
net operating profit to sales	76.54
cash flow to current liabilities	76.64
cash to current liabilities	75.86
net income to sales	73.83
sales to net worth	75.22
income from operations to total assets	75.22
cash to total assets	76.31
cash flow to total assets	75.53
current plus long-term liabilities to total assets	77.12
earnings before taxes to equity	74.78
retained earnings to total assets	75.59
current liabilities to equity	76.68
earnings before interest and taxes to total assets	75.73
current liabilities to total assets	74.98

Appendix RR Best Validation Accuracy Genetic Programming Algorithm for Clusters 1 & 2 on Compustat Dataset (Sequential Division)

```
f[0]=f[1]=f[2]=f[3]=f[4]=f[5]=f[6]=f[7]=0;
L0: f[0]+=v[17];
L1: f[0]=sin(f[0]);
L2: cflag=(f[0] < f[1]);
L3: f[0]-=f[0];
L4: f[0]=cos(f[0]);
L5: if (!cflag) f[0] = f[1];
L6: f[0]-=v[7];
L7: f[0]=cos(f[0]);
L8: f[0]=cos(f[0]);
L9: f[0]-=v[23];
```

```
f[0]=f[1]=f[2]=f[3]=f[4]=f[5]=f[6]=f[7]=0;
L0: f[0]+=1.248895406723023f;
L1: f[0]=-f[0];
L2: f[0]*=0.2877938747406006f;
L3: f[0]+=v[6];
L4: f[2]+=f[0];
L5: f[0]-=f[0];
L6: cflag=(f[0] < f[2]);
L7: f[0]+=v[38];
L8: if (cflag) f[0] = f[3];
```

Appendix SS Best Validation Accuracy Genetic Programming Algorithm for Clusters 1 & 2 on Aspect Dataset (Sequential Division)

```
f[0]=f[1]=f[2]=f[3]=f[4]=f[5]=f[6]=f[7]=0;
L0: f[0]-=v[2];
L1: f[0]/=v[13];
L2: f[1]+=f[0];
L3: f[0]/=v[13];
L4: cflag=(f[0] < f[1]);
L5: cflag=(f[0] < f[0]);
L6: f[0]=-f[0];
L7: f[0]/=v[13];
L8: f[0]+=f[0];
L9: tmp=f[2]; f[2]=f[0]; f[0]=tmp;
L10:  tmp=f[0]; f[0]=f[0]; f[0]=tmp;
L11:  f[3]+=f[0];
L12:  tmp=f[3]; f[3]=f[0]; f[0]=tmp;
L13:  f[0]+=f[2];
L14:  cflag=(f[0] < f[0]);
L15:  tmp=f[0]; f[0]=f[0]; f[0]=tmp;
L16:  f[0]-=0.002621650695800781f;
L17:  if (!cflag) f[0] = f[0];
L18:  f[0]*=f[1];
L19:  f[0]=sqrt(f[0]);
L20:  f[3]-=f[0];
L21:  f[0]+=f[2];
L22:  f[0]=-f[0];
L23:  if (!cflag) f[0] = f[1];
L24:  cflag=(f[0] < f[3]);
L25:  if (cflag) f[0] = f[3];
L26:  f[0]-=f[3];
L27:  f[0]+=f[3];
L28:  f[0]=cos(f[0]);
L29:  if (!cflag) f[0] = f[3];
L30:  if (cflag) f[0] = f[0];
L31:  f[0]/=f[3];
L32:  if (cflag) f[0] = f[1];
L33:  f[0]/=v[13];
L34:  if (!cflag) f[0] = f[0];
L35:  f[0]=fabs(f[0]);
L36:  if (!cflag) f[0] = f[0];
L37:  f[0]+=f[0];
L38:  tmp=f[2]; f[2]=f[0]; f[0]=tmp;
L39:  f[0]*=f[1];
L40:  cflag=(f[0] < f[0]);
L41:  if (cflag) f[0] = f[3];
L42:  f[0]=-f[0];
L43:  f[0]=-f[0];
L44:  cflag=(f[0] < f[1]);
L45:  f[0]-=v[2];
L46:  if (!cflag) f[0] = f[0];
L47:  f[0]=fabs(f[0]);
L48:  f[0]=sqrt(f[0]);
```

```
f[0]=f[1]=f[2]=f[3]=f[4]=f[5]=f[6]=f[7]=0;
L0: f[0]-=v[9];
L1: f[0]=-f[0];
L2: f[0]=sqrt(f[0]);
L3: f[3]-=f[0];
L4: f[0]=sqrt(f[0]);
L5: f[0]*=f[3];
L6: f[0]/=1.258495330810547f;
L7: f[0]+=0.002621650695800781f;
L8: f[0]=fabs(f[0]);
L9: f[0]-=v[22];
L10:  f[0]-=v[40];
```

References

A.P. Eagers, 2005. *A.P. Eagers Limited reaches agreement to merge with Auto Group Limited.* [Online]
Available at: http://www.apeagers.com.au/pdfs/docs-1069219-v1-APE%20Announcement%20071205%20_5_.pdf

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

Aizerman, A., Braverman, E. & Rozoner, L., 1964. Theoretical foundations of the potential function method in pattern recognition learning. *Automation and Remote Control,* Volume 25, pp. 821-837.

Ali, S. & Smith-Miles, K., 2006. A meta-learning approach to automatic kernel selection for support vector machines. *Neurocomputing,* 70(1-3), pp. 173-186.

Altman., E., 1993. *Corporate Financial Distress and Bankruptcy.* 2nd ed. New York: John Wiley & Sons.

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

Altman, E., Haldermen, R. & Narayanan, P., 1977. Zeta Analysis - A New Model to Identify Bankruptcy Risk of Corporations. *Journal of Banking and Finance,* 1(1), pp. 29-54.

Altman, E., Marco, G. & Varetto, F., 1994. Corporate Distress Diagnostic: Comparisons Using Linear Discriminant Analysis and Neural Networks (The Italian Experience). *Journal of Banking and Finance,* 18(3), pp. 505-529.

American Bankruptcy Institute, 2005. *Quarterly Business Bankruptcy Filings for 1994-2004.* [Online]
Available at:
http://www.abiworld.org/ContentManagement/ContentDisplay.cfm?ContentID=13742
[Accessed 31 May 2005].

Anandarajan, M., Lee, P. & Anandarajan, A., 2001. Bankruptcy Prediction of Financially Stressed Firms: An Examination of the Predictive Accuracy of Artificial Neural Networks. *International Journal of Intelligent Systems in Accounting, Finance & Management,* 10(2), pp. 69-81.

Andrés, J., Landajo, M. & Lorca, P., 2012. Bankruptcy prediction models based on multinorm analysis: An alternative to accounting ratios. *Knowledge-Based Systems,* Volume 30, pp. 67-77.

Argenti, J., 1976. *Corporate Collapse: the causes and symptoms.* s.l.:McGraw-Hill Book Company.

Argenti, J., 1983. Predicting corporate failure. *Accountants Digest,* Volume 138.

Atiya, A., 2001. Bankruptcy prediction for credit risk using neural networks: a survey and new results. *IEEE Transactions on Neural Networks,* 12(4), pp. 929-935.

Australian Associated Press, 2006. Auto Group steers into administration. *The Age*, 13 February.

Australian Securities and Investments Commission v Rich (2009).

Barmash, I., 1973. *Great business disasters.* New York: Ballantine Books.

Barniv, R. & Raveh, A., 1989. Identifying Financial Distress: A New Nonparametric Approach. *Journal of Business Finance and Accounting,* 16(3), pp. 361-383.

Barry, P., 2002. *Rich Kids.* Sydney: Bantom Books.

Beaver, W., 1966. Financial Ratios as Predictors of Failure. *Journal of Accounting Research,* 4(3), pp. 71-111.

Beaver, W., 1968. Market Prices, Financial Ratios, and the Prediction of Failure. *Journal of Accounting Research,* 6(2), pp. 179-193.

Beltran, L., 2003. The Dealmakers: Ross makes play for Cone Mills. *The Wall Street Journal: Market Watch*, 16 September.

Bezdek, J. & Hathaway, R., 2002. VAT: A Tool for Visual Assessment of (Cluster) Tendency. *Proceedings of the 2002 International Joint Conference on Neural Networks,* Issue 3, pp. 2225-2230.

Bezdek, J. & Pal, N., 1998. Some New Indexes of Cluster Validity. *IEEE Transactions on Systems, Management and Cybernetics,* 28(3), pp. 301-315.

Blum, M., 1974. Failing Company Discriminant Analysis. *Journal of Accounting Research,* 12(1), pp. 1-25.

Boritz, J. & Kennedy, D., 1995. Effectiveness of Neural Network Types for Prediction of Business Failure. *Expert Systems with Applications,* 9(4), pp. 503-512.

Calderon, T. & Cheh, J., 2002. A Roadmap for Future Neural Networks Research in Auditing and Risk Assessment. *International Journal of Accounting Information Systems,* 3(4), pp. 203-236.

Census Bureau, 2007. *2007 NAICS Definitions: 11 Agriculture, Forestry, Fishing and Hunting.* [Online]
Available at: http://www.census.gov/naics/2007/def/NDEF11.HTM#N11

Chauhan, N., Ravi, V. & Chandra, D., 2009. Differential evolution trained wavelet neural networks: Application to bankruptcy prediction in banks. *Expert Systems with Applications,* 36(4), p. Differential evolution trained wavelet neural networks: Application to bankruptcy prediction in banks.

Chen, H. et al., 2011. A novel bankruptcy prediction model based on an adaptive fuzzy k-nearest neighbor method. *Knowledge-Based Systms,* 24(8), pp. 1348-1359.

Chen, K. & Liu, L., 2003. Validation and refining clusters via visual rendering. *Proceedings of the International Conference on Data Mining,* pp. 501-504.

Chen, K. & Shimerda, T., 1981. An Empirical Analysis of Useful Financial Ratios. *Financial Management,* 10(1), pp. 51-60.

Choe, W., Ersoy, O. & Bina, M., 2000. Neural network schemes for detecting rare events in human geonomic DNA. *Bioinformatics,* 16(12), pp. 281-288.

Clarke, F., Dean, G. & Oliver, K., 1997. *Corporate Collapse: Regulartory, Accounting and Ethical Failure.* s.l.:Cambridge University Press.

CME Group Index Services, 2011. *Dow Jones Industrial Average.* [Online]
Available at: http://www.djaverages.com/?view=industrial&page=overview

Coats, P. & Fant, L., 1993. Recognising Financial Distress Patterns Using a Neural Network Tool. *Financial Management,* 22(3), pp. 142-155.

Connect 4, 2000. *Annual Reports Collection: One.Tel Limited,* Pyrmont: Thomson Reuters.

Connect 4, 2005. *Annual Reports Collection: Auto Group Ltd,* Pyrmont: Thomson Reuters.

Davies, D. & Bouldin, D., 1979. A Cluster Seperation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 1(2), pp. 224-227.

Deakin, E., 1972. A Discriminant Analysis of Predictors of Business Failure. *Journal of Accounting Research,* 10(1), pp. 167-180.

Deboeck, G. & Kohonen, T., 1998. *Visual Explorations in Finance with Self-Organizing Maps.* New York: Springer.

Dimitras, A., Slowinski, R., Susmaga, R. & Zopounidis, C., 1999. Business Failure Prediction Using Rough Sets. *European Journal of Operational Research,* 114(2), pp. 263-280.

Ding, Y. & Harrison, R., 2007. Relational Visual Cluster Validity (RVCV). *Pattern Recognition Letters,* Volume 28, pp. 2071-2079.

Dunn, J., 1974. Well Separated Clusters and Optimal Fuzzy Partitions. *Cybernetics and Systems,* 4(1), pp. 95-104.

Edmister, R., 1972. An Empirical Test of Financial Ratio Analysis for Small Business Failure Prediction. *Journal of Financial and Quantitative Analysis,* 7(2), pp. 1477-1494.

Elam, R., 1975. The Effect of Lease Data on the Predictive Ability of Financial Ratios. *The Accounting Review,* 50(1), pp. 25-43.

Eljelly, A., 2004. Liquidity - profitability tradeoff: An empirical investigation in an emerging market. *International Journal of Commerce and Management,* 14(2), pp. 48-61.

Etheridge, H. & Sriram, R., 1997. A Comparison of the Relative Costs of Financial Distress Models. *Etheridge, H. & Sriram, R. (1997), "A Comparison of the Relative Costs of Financial DistInternational Journal of Intelligent Systems in Accounting, Finance & Management,* 6(3).

Everitt, B., 1977. *The Analysis of Contingency Tables.* London: Chapman and Hall.

Fahlman, S. & Lebiere, C., 1990. The Cascade-Correlation Learning Architecture. *Advances in Neural Information Processing Systems,* Volume 2, pp. 524-532.

Fine, T., 1999. *Feedforward Neural Network Methodology.* New York: Springer-Verlag.

Fine, T., 1999. *Feedfoward Neural Network Methodology.* New York: Springer-Verlag.

Fisher, R., 1936. The Use of Multiple Measurements in Taxonomic Problems. *Annals of Human Genetics,* 7(2), pp. 179-188.

Fisher, R., 2011. The use of multiple measurements in taxonomic problems. *Annals of Human Genetics,* 7(2), pp. 179-188.

Frank, A. & Asuncion, A., 2010. [Online]
Available at: http://archive.ics.uci.edu/ml

Friedrichs, F. & Igel, C., 2005. Evolutionary tuning of multiple SVM parameters. *Neurocomputing,* Volume 64, pp. 107-117.

Fusaro, P. & Miller, R., 2002. *What Went Wrong at Enron: Everyone's Guide to the Largest Bankruptcy in U.S. History.* New Jersey: John Wiley & Sons.

Gallagher, T. & Andrew, J., 2007. *Financial Management: Principals and Practice.* 4th ed. s.l.:Freeload Press.

Gluyas, R., 2012. Nine's lenders call in advisers. *The Australian*, 13 September.

Gray, M., 1992. Cones Mills Corp. to cut back on corduroy manufacturing. *Daily News Record*, 11 February.

Guyon, I., Gunn, S., Nikravesh, M. & Zadeh, L., 2006. *Feature Extraction: Foundations and Applicaitons.* s.l.:Springer.

Hathaway, R. & Bezdek, J., 2003. Visual Cluster Validity for Prototype Generator Clustering Models. *Pattern Recognition Letters,* Issue 24, pp. 1563-1569.

Holland, J., 1975. *Adaptation in Natural and Artificial Systems.* Ann Arbor: University of Mishigan Press.

Huang, Z., Cheung, D. & Ng, M., 2001. An empirical study on the visual cluster validation method with Fastmap. *Proceedings of the Seventh International Conference on Database Systems for Advanced Applications,* pp. 84-91.

Hubert, L. & Arabie, P., 1985. Comparing Partitions. *Journal of Classification,* 2(1), pp. 193-218.

Investogain Limited, 2012. *AUTO GROUP LIMITED (ALO).* [Online]
Available at: http://www.delisted.com.au/company/auto-group-limited

Johnson, C., 1970. Ratio Analysis and the Prediction of Firm Failure. *Journal of Finance,* 25(5), pp. 1166-1168.

Kane, G., Richardson, F. & Graybeal, P., 2010. Recession-Induced Stress and the Prediction of Corporate Failure. *Contemporary Accounting Research,* 13(2), pp. 631-650.

Kane, G., Richardson, F. & Meade, N., 1998. Rank Transformations and the Prediction of Corporate Failure. *Contemporary Accounting Research,* 15(2), p. 145.

Kerlikowske, K. et al., 1993. Positive Predictive Value of Screening Mammography by Age and Family History of Breast Cancer. *Journal of the American Medical Association,* 270(20), pp. 2444-2450.

Kharbanda, O. & Stallworthy, E., 1985. *Corporate Failure: Prediction, Panacea and Prevention.* United Kingdom: McGraw-Hill Book Company.

Kohavi, R. & Provost, F., 1998. Glossary of terms. *Editorial for the Special Issue on Applications of Machine Learning and the Knowledge Discovery Process,* Volume 30, pp. 2-3.

Kohonen, T., 1995. *Self-Organizing Maps.* 2nd ed. Berlin: Springer-Verlag.

Kotler, P., Armstrong, G., Brown, L. & Adam, S., 1998. *Marketing.* 4th ed. Sydney, Australia: Prentice Hall.

LeCun, Y., 1985. Une procedure d'apprentissage pour reseau a seuil assymetrique. *Cognitivia,* Volume 85, pp. 599-604.

Lee, K., Booth, D. & Alam, P., 2005. A comparison of supervised and unsupervised neural networks. *Expert Systems with Applications,* 29(1), pp. 1-16.

Lee, K., Han, I. & Kwon, Y., 1996. Hybrid Neural Network Models for Bankruptcy Predictions. *Decision Support Systems,* 18(1), pp. 63-72.

Lensberg, T., Eilifsen, A. & McKee, T., 2006. Bankruptcy Theory Development and Classification via Genetic Programming. *European Journal of Operations Research,* 169(2).

Libby, R., 1975. Accounting Ratios and the Prediction of Failure: Some Behavioural Evidence. *Journal of Accounting Research,* 13(1), pp. 150-162.

Liu, H. & Motoda, H., 1998. *Feature Selection for Knowledge Discovery and Data Mining.* s.l.:Kluwer Academic Publishers.

Liu, J., 2004. Macroecoomic deteminants of corporate failure: evidence from the UK. *Applied Economics,* Volume 36, p. 939.945.

Loomis, C., 2004. The Sinking of Bethlehem Steel. *Fortune*, 5 April.

McKee, T., 2000. Developing a Bankruptcy Prediction Model via Rough Sets Theory. *International Journal of Intelligent Systems in Accounting, Finance & Management,* 9(3), pp. 159-173.

McKee, T. & Lensberg, T., 2002. Genetic Programming and Rough Sets: A Hybrid Approach to Bankruptcy Classification. *European Journal of Operational Research,* 18(2), pp. 436-451.

McLean, B. & Elkind, P., 2003. *The Smartest Guys in the Room.* New York: Penguin Group.

McRobert, A. & Hoffman, R., 1997. *Corporate Collapse: An Early Warning System for Lenders, Investors and Suppliers.* s.l.:McGraw-Hill Book Company.

Mehrotra, S., 1992. On the Implementation of a Primal-Dual Interior Point Method. *SIAM Journal on Optimization,* Volume 2, pp. 575-601.

Menard, S., 2002. *Applied Logistic Regression Analysis.* 2nd ed. California: Sage Publications.

Miller, D., 1990. *The Icarus Paradox: How Exceptional Companies Bring About Their Own Downfall,.* New York: HarperCollins.

Min., S., Lee, J. & Han, I., 2006. Hybrid genetic algorithms and support vector machines for bankruptcy prediction. *Expert Systems with Applications,* 31(3), pp. 652-660.

Mitchell, T., 1997. *Machine Learning.* s.l.:McGraw Hill.

Monash University, n.d. *Library Catalogue,* s.l.: s.n.

Monem, R., 2011. The One.Tel Collapse: Lessons for Corporate Governance. *Australian Accounting Review,* 21(4), pp. 340-351.

Morningstar Australasia Pty Limited, 2012. *Australian All Ordinaries.* [Online]
Available at: http://www.morningstar.com.au/

Morningstar Australiasia Pty Limited, 2006. *Aspect Financials,* Sydney: s.n.

Nam, C., Tong, K., Nam, P. & Hoe, L., 2008. Bankruptcy Prediction Using a Discrete-Time Duration Model Incorporating Temporal and Macroeconomic Dependencies. *Journal of Forecasting,* Volume 27, pp. 493-506.

Nikolic, N., Zarkic-Joksimovic, N., Stojanovski, D. & Joksimovic, I., 2013. The application of brute force logistic regression to corporate credit scoring models: Evidence from Serbian financial statements. *Expert Systems with Applications,* 40(15), pp. 5932-5944.

Noguchi, Y., 2011. *Why Borders Failed While Barnes & Noble Survived.* [Online]
Available at: http://www.npr.org/2011/07/19/138514209/why-borders-failed-while-barnes-and-noble-survived
[Accessed 2012].

Odom, M. & Sharda, R., 1990. A neural network model for bankruptcy prediction. *Proceedings of the IEEE International Conference on Neural Networks,* pp. 163-168.

Ohlson, J., 1980. Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research,* 18(1), pp. 109-131.

Otsu, N., 1979. A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man and Cybernetics.*

Palacios, L., 2007. [Online]
Available at: http://wrds-web.wharton.upenn.edu/wrds/support/Additional%20Support/WRDS%20Presentations/_000user2007/merging_crsp_compustat.pdf

Parker, D., 1985. *Technical Report TR-47 - Learning logic: Casting the cortex of the human brain in silicon,* Massachusetts: Cambridge.

Patton, M., 2002. *Qualitative Research & Evaluation Methods.* 3rd ed. California: Sage Publications.

Pawlak, Z., 1982. Rough sets. *International Journal of Information and Computer Sciences,* Volume 11, pp. 344-356.

Prim, R., 1957. Shortest connection networks and some generalizations. *Bell Systems Technical Journal,* Volume 36, pp. 1389-1401.

Probst, G. & Raisch, S., 2005. Organizational crisis: The logic of failure. *Academy of Management Executive,* Volume 19, pp. 90-105.

Prosoft, 1999. *Rose2 – Rough Set Data Explorer, User's Guide 2.0..* s.l.:Prosoft.

Queen, M. & Roll, R., 1987. Firm Mortality: Using Market Indicators to Predict Survival. *Financial Analysts Journal,* 43(3), pp. 9-26.

Ribeiro, B. et al., 2012. Enhanced default risk models with SVM+. *Expert Systems with Applications,* 39(11), pp. 10140-10152.

Rojas, R., 1996. *Neural Networks: A Systematic Introduction.* Berlin: Springer-Verlag.

Rösch, D. & Scheule, H., 2005. A Multifactor Approach for Systematic Default and Recovery Risk. *The Journal of Fixed Income,* 15(2), pp. 63-75.

Rose, P., Andrews, W. & Giroux, G., 1982. Predicting Business Failure: A Macroeconomic Perspective. *Journal of Accounting, Auditing, and Finance,* pp. 20-31.

Ross, J. & Kami, M., 1973. *Corporate Management in Crisis: Why the Mighty Fall.* New Jersey: Prentice Hall.

Rubin, D., 1987. *Multiple Imputation for Nonresponse in Surveys.* New York: J. Wiley & Sons.

RWE Australian Business News, 2005. Around the Traps ... with the ferret. *RWE Australian Business News*, 26 October.

Salmi, T. & Martikainen, T., 1994. A Review of the Theoretical and Empirical Basis of Financial Ratio Analyais. *Finnish Journal of Business Economics,* 94(4), pp. 426-448.

Salzberg, S., 1997. On Comparing Classifiers: Pitfalls to Avoid and a Recommended Approach. *Data Mining and Knowledge Discovery,* Issue 1, pp. 317-328.

Scholkopf, B., Simard, P., Smola, A. & Vapnik, V., 1998. Prior Knowledge in Support Vector Kernels. *Advances in Neural Information Processing Systems,* Volume 10.

Shanks, G., 2002. Guidelines for Conducting Positivist Case Study Research in Information Systems. *Australian Journal of Information Systems,* Issue December, pp. 76-85.

Shi, J. & Malik, J., 1997. Normalized Cuts and Image Segmentation. *IEEE Transactions on Patern Analysis and Machine Intelligence,* Volume 22, pp. 888-905.

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

Shoebridge, N., 2011. Nine looks for leeway to pay $2.7bn debt. *Financial Review*, 18 November.

Słowiński, R., 1992. *Intelligent Decision Support: Handbook of Applications and Advances of the Rough Set Theory.* The Netherlands: Kluwer Academic Publishers.

Smith, R. & Winakor, A., 1935. *Changes in the financial structure of unsucessful corporations.* Illinois: Bureau of Business Research.

Stone, M., 1974. Cross-validatory Choice and Assessment of Statistical Predictors. *Journal of the Royal Statistical Society B,* 36(2), pp. 111-147.

Textile News, 2003. Cone Mills: 2002 Best Year Since 1994. *Textile News*, March.

The MathWorks, Inc, 2011. *Bioinformatics Toolbox Documentation,* Massachusetts: s.n.

The Morning Call, 2010. Forging America:The History of Bethlehem Steel. *The Morning Call*.

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

Tsai, C. & Cheng, K., 2012. Simple instance selection for bankruptcy prediction. *Knowledge-Based Systems,* Volume 27, pp. 333-342.

Varetto, F., 1999. Genetic Algorithms Applications in the Analysis of Insolvency Risk. *Journal of Banking & Finance,* 22(10-11), pp. 1421-1439.

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

Viscovery Software GmbH, 2007. *Viscovery SOMine.* Austria: s.n.

Wang, L. et al., 2008. SpecVAT: Enhanced Visual Cluster Analysis. *2008 Eigth International Conference on Data Mining,* pp. 638-647.

Wang, Z., 2004. Financial Ratio Selection for Default-Rating Modelling: A Model-Free Approach and its Empirical Performance. *Journal of Applied Finance,* 14(1).

Warren, K., 2008. *Bethlehem Steel: Builder and Arsenal of America.* Pittsburgh: University of Pittsburgh Press.

Watanabe, S., 1985. *Pattern Recognition: Human and Mechanical.* New York: John Wiley & Sons.

Whalley, J., 2012. Hanging by a thread, can Nine Entertainment get a deal?. *Herald Sun*, 16 October.

Wharton Research Data Services, 2003. *North America Monthly Updates - Fundamentals Quarterly.* [Online]
Available at: http://wrds-web.wharton.upenn.edu/wrds/ds/compm/fundq/index.cfm?navGroupHeader=Compustat%20Monthly%20Updates&navGroup=North%20America

Wilcox, J., 1973. A Prediction of Business Failure Using Accounting Data. *Journal of Accounting Research,* Issue 11, pp. 163-180.

Wilson, N., Summers, B. & Hope, R., 2000. Using Payment Behaviour Data for Credit Risk Modelling. *International Journal of the Economics of Business,* 7(3), pp. 333-347.

Wilson, R. & Sharda, R., 1994. Bankruptcy Prediction Using Neural Networks. *Decision Support Systems,* 11(5).

Wong, N., Ray, P., Stephens, G. & Lewis, L., 2012. Artificial immune systems for the detection of credit card fraud: an architecture, prototype and prelininary results. *Information Systems Journal,* 22(1), pp. 53-76.

World Bank Group, 2012. *Data - The World Bank.* [Online]
Available at: http://data.worldbank.org/

Wu, C., 2004. Using Non-Financial Information to Predict Bankruptcy: A Study of Public Companies in Taiwan. *International Journal of Management,* 21(2).

Yang, Z., Platt, M. & Platt, H., 1999. Probabilistic Neural Networks in Bankruptcy Prediction. *Journal of Business Research,* Issue 44, pp. 67-74.

Zelnick-Manor, L. & Perona, P., 2004. Self-Tuning Spectral Clustering. *Proc. Advances in Neural Information Process Systems,* Issue 17, pp. 368-347.

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

Zhu, X. & Davidson, I., 2007. *Knowledge Discovery and Data Mining: Challenges and Realities.* Pennsylvania: Information Science Reference.