Modelling Construction Business Performance

Literature Review [Project: 2005-017-A]

Editor: Michael Falta

The research described in this report was carried out by Michael Falta and Steve Su

Project Leader: Natalie Gallery

Team Members: Dayv Carter, Robert Czernkowski, Michael Falta, Roger Frith, Natalie Gallery, Bob Johnson, Stewart Jones, Shelley Lockton, Cameron Murphy, Neal Ryan, Dale Rylko, Jason Smith, Steve Su, Roger Willett

Researchers: Robert Czernkowski, Michael Falta, Natalie Gallery, Stewart Jones, Steve Su, Roger Willett

Distribution List

Cooperative Research Centre for Construction Innovation
Authors

Disclaimer

The Client makes use of this Report or any information provided by the Cooperative Research Centre for Construction Innovation in relation to the Consultancy Services at its own risk. Construction Innovation will not be responsible for the results of any actions taken by the Client or third parties on the basis of the information in this Report or other information provided by Construction Innovation nor for any errors or omissions that may be contained in this Report. Construction Innovation expressly disclaims any liability or responsibility to any person in respect of any thing done or omitted to be done by any person in reliance on this Report or any information provided.

Please direct all enquiries to:

Chief Executive Officer
Cooperative Research Centre for Construction Innovation
9th Floor, L Block, QUT, 2 George St
Brisbane Qld 4000
AUSTRALIA
T: 61 7 3864 1393
F: 61 7 3864 9151
E: enquiries@construction-innovation.info
W: www.construction-innovation.info

Table of Contents

List of Tables

List of Figures

1. PREFACE

This literature review is organised in two sections due to a change of research staff during phase 1 of the project. Section 3 is a review with focus on variable selection for business failure analysis in the construction industry and has been compiled by Michael Falta. Section 4 is a review on methods for modelling financial distress with no particular focus on an industry sector. It has been compiled by Steve Su.

2. EXECUTIVE SUMMARY

Table 1 displays the major factors of failure for businesses in the construction industry that have been found and investigated in the open literature, with the exception of Clift (2006). These partly challenge the results stated in documents that the Queensland Building Services Authority (industry partner) has provided. Whether it is simply the different size of construction businesses considered (journal articles usually use the Dun & Bradstreet database containing large US construction companies, in comparison to Australian and Queensland building firms) that accounts for these differences, or the reason lies in varying business environments of particular countries, is one line of research that will be further pursued.

Table 1 Major factors of business failure in the construction industry.

Factor	Financial/Non-financial	Reference	Comment
Accounting ratios	Financial	Freear (1980);	Based on Dun & Bradstreet data;
Insufficient profits	Financial	Clift (2006); Koksal & Arditi (2004); Russell & Zhai (1996); Kangari (1987);	Organisational nature; Company level;
Operating expenses	Financial	Koksal & Arditi (2004);	Organisational nature; Company level;
Burdensome institutional debt	Financial	Koksal & Arditi (2004); Russell & Zhai (1996); Kangari (1987);	Organisational nature; Company level;
Interest rates	Financial	Russell & Zhai (1996); Russell & Jaselskis (1992); Kangari (1987);	Environmental nature; Company level;
Industry weakness	Both	Koksal & Arditi (2004);	Environmental nature; Company level;
Managerial incompetence	Both	Koksal & Arditi (2004); Dun & Bradstreet (1973);	
Age of business	Non- financial	Kale & Arditi (1999); Russell & Jaselskis (1992); Lynch (2003);	Organisational nature; Company level;
Size of business	Both	Lynch (2003);	Company level;
Cash flow	Financial	Langford, Iyagba & Komba (1993);	Project level;
Claim awareness; complaints	Non-financial	Clift (2006); Langford, Iyagva & Komba (1993);	Company level; Project level;
Compliance with licensing requirements	Both	Clift (2006);	

Also deducible from Table 1 is the different character of financial and non-financial factors. Their likely impact, in time, to a possible business failure has not been considered however. This observation leads to our second line of research – model development considering (construction industry) business failures not as an observation at particular points in time but as a *process that may lead to a failure eventually*.

Further, a comparison of the mathematical models used for analyses reported in both literature reviews below reveals that the variety and sophistication of applications in other industry sectors, such as banking and insurance, is far more advanced than what has been used to understand construction businesses. This comment is important as the construction industry characteristics are quite different from other industry sectors. From this, the third line of research that we will pursue includes an investigation of the applicability of modern modelling methods such as, for example, neural networks. Note that Altman's *Z*-score is the standard against which the performance of alternative approaches to predict business failure have to be measured.

Producing a framework that incorporates all three lines of research is the ultimate goal and, based on this literature review, it will be an important piece of intellectual contribution to (a) understanding construction industry business failure, (b) classifying the usefulness of mathematical models applied to businesses in the construction industry, and (c) flagging of those businesses that are at risk of a failure prior to a possible, catastrophic event.

3. BUSINESS FAILURE IN THE CONSTRUCTION INDUSTRY

3.1 Introduction

Searching academic databases for records on ‘business failure’, ‘business distress’ or ‘bankruptcy’ yields a large body of studies on qualitative, empirical, theoretical and simulation aspects. It is a central part of this research to distil from this large quantity of potentially relevant reports and methodologies those which can both flag and predict business failure in the construction industry. An additional search term, such as, ‘construction’, ‘construction industry’ or ‘contractor’ yields a much smaller number of hits, many of which emphasize the construction industry’s distinctive characteristics.

We scientists need first to understand the subject of investigation and the environment in which it lives. To do so, an analysis of existing successful and failed approaches to particular research questions is helpful before embarking on new territory. This guides the structure of the following report for we first review papers that specifically report on aspects of business failure in the construction industry followed by, (a) an overview of promising candidates borrowed from other disciplines and industries, and (b) a possible novel approach. An Australian (Queensland) perspective on the topic will also drive this investigation as most of the published research has been applied to the US and UK construction industries.

3.2 Some Construction Industry Characteristics

The (building and) construction industry is an important part of national economies of industrialised countries (>6% of Australian GDP (2004-05) with >7.6% of labour force, 2003-04; 7.4% of UK GDP (2000); and ~5% of US GDP (2002) with ~5.4% of labour force, 2003-04). In Australia, the construction industry is grouped into the following three areas of activity (Australian Bureau of Statistics, 2006):

- residential building (e.g., houses and flats);
- non-residential building (e.g., offices, shops and hotels); and
- engineering construction (e.g., roads, bridges, water mains and sewerage).

The construction industry is typically characterized by (e.g., Kale & Arditi, 1999):

- slow technological changes and slow changes in innovation processes;
- large numbers of participants with a small number of big firms being awarded the majority of jobs;
- easy entry to the construction business;
- start-ups needing comparatively low working capital at commencement;
- influenced by macroeconomic trends in that

    booms attract new companies that may not be sufficiently qualified, and during

    recessions there is overbidding and competition;
- one-off nature of projects;
- high capital intensity;
- temporary nature and duration of investor-contractor-subcontractor relationships; and

- fragmented nature of industry structure and of the construction process organisation.

3.3 Preliminaries

We need consensus about the definition of a 'business failure'. A frequently cited definition is:

> "a business that ceases operations following assignment or bankruptcy ceases operations with losses to creditors after such actions as foreclosure or attachment; voluntarily withdraws leaving unpaid debts; is involved in court actions such as receivership, reorganisation or rearrangement; or has voluntarily compromised creditors." Dun and Bradstreet Corporation (2006).

In a contractor evaluation setting, Russell and Zhai (1996) note that "contractor failure is defined as the termination of a contractor's operation", the latter of which is usually invoked by the project owner as part of the contractor's non-performance clause.

3.4 Business Failure Analysis in the Construction Industry

Several studies (e.g., Koksal & Arditi, 2004; Russell & Zhai, 1996; Kangari, 1987; Freear, 1980) use the Dun and Bradstreet data as a source for empirical (US) data. Investigations on business failure in the UK make use, for example, of questionnaires (Hall, 1994) and Extel Services (Mason & Harris, 1979).

Predominantly, the models used are the Z-score (Altman, 1968), a discriminant analysis based model, and ratio analysis (Abidali & Harris, 1995; Langford, Iyagba & Komba, 1993; Mason & Harris, 1979); Hall (1994) employs a logit regression.

In a further line of rather conceptual and descriptive analyses on contractor businesses, firms were subdivided into (a) their technical and financial capabilities (Kangari, 1987), and (b) environmental and organisational determinants (Koksal & Arditi, 2004). Both are somewhat similar subdivisions as environmental determinants are macroeconomic factors and natural forces (that cannot be influenced by the business manager) and organisational determinants are human, organisational and financial capital factors. The latter study extended the failure analysis in that failure was recognised as a process consisting of symptoms (indicators) and outcomes: symptoms of failure (performance factors) are driven by determinants; and outcomes (failure or survival/success) are driven by symptoms. This suggests a more complex model than simply applying a Z-score is appropriate.

Interestingly, only around 40% of failure determinants have been identified as being of an organisational nature (Koksal & Arditi, 2004), with the remainder being due to environmental factors. In the first category, insufficient capital and lack of business knowledge account for more than two thirds of failures and are mainly driven by insufficient profits, heavy operating expenses and burdensome institutional debt. The single most relevant environmental factor is industry weakness driven by insufficient profits and heavy operating expenses. Notably, only 0.8% of the organisational factors are allocated to over-expansion, which somewhat challenges the attention given to this factor by Australian regulators. Russell and Zhai (1996) and Kangari (1987) rate bad debts and insufficient profits as major causes of constructor failure with high interest rates, loss of market, no consumer spending and no future as further important indicators.

The main symptoms above suggest that a connection between the business' age and its likelihood of failure may also exist, as a not yet established firm's lack of experience and average performance would improve over time (March, 1991). This age dependency of firms and failure has been investigated by Kale and Arditi (1999) using Dun and Bradstreet data. They find:

> "the exchange relationships between a construction firm and its clients, sureties, subcontractors, vendors, and financial institutions and others take place in an established environment. In this established environment, newly established companies which lack legitimacy will be much more exposed to environmental selection processes."

The study reveals an age-dependent failure pattern in which the number of failed companies increases during the first three to five years (adolescence period) and decreases thereafter.

The above analyses treat a construction business as a homogeneous entity; however, (larger) firms may have various divisions (e.g., Building and Civil Engineering, Property and Estate Agency) where comparisons with appropriate industry sectors would be more indicative, in order to understand their behaviour. Furthermore, construction businesses have been analysed according to the type of projects they are engaged in (Langford, Iyagba & Komba, 1993). On a project level, the following indicators of cost and time overruns, delayed payments to subcontractors and suppliers, claims awareness and broken loans covenants are relevant. Russell (1991, Table 2) gives an overview of some failed constructors, the project failure causes and costs involved.

A further line of thought is to introduce a banking sector risk or reliance classifier. According to Abidali and Harris (1995), such a measure would predict that the more years a company is classed as being at risk, the more likely it would be to fail.

Further, apart from the Z-score on financial quantities, which is only able to indicate business failure faithfully within two years of potential failure, Abidali and Harris (1995) advocate the use of the A-score on managerial characteristics. This ties into considering failure as a process as, among other matters, it signals business weaknesses during a much longer period of time, compared to the Z-score, before failure occurs. Major indicators of failure identified include: autocratic chief executives, the same individual acting as chief executive and chairman, company boards comprising of too many non-contributing directors, lack of engineering skills, lack of a strong financial director, insufficient managerial skills, incomplete accountancy systems, defective bidding systems, and poor marketing skills. Managerial errors in decision-making, such as too much reliance on short term loans, over-trading, suffering losses in projects, and the acquisitions of a potentially failing firm, also need to be considered as potential risk factors of business failure.

3.5 Prediction of Business Failure in the Construction Industry

All models mentioned above can certainly be utilised for prediction purposes regarding construction business failures and some have been tested for this capability via holdout periods. Prediction in a more mathematical perspective, however, has been addressed in Russell and Zhai (1996) and Russell and Jaselskis (1992). The first study employs a random coefficient method to describe the stochastic dynamics of a construction firm's future position, trend and volatility in comparison to variables, such as, prime interest rates, new construction value in place, net worth/total assets, gross profit/total assets and net working capital/total assets. The model yielded satisfactory predictive capabilities based on a sample of 49 failed and 71 non-failed companies. Russell and Jaselskis (1992) emphasise the importance of the project level. The authors use a binary discrete choice model in order to predict the probability of failure for a given construction project prior to awarding a contract. The methodology is a logistic regression approach for which the parameters are estimated using maximum likelihood estimation. Complimentary to the definitions given in Section 1.2, they define failure

> "as a significant breach of the contractor's legal responsibilities to the owner (for example, bankruptcy or material breach of contract related to meeting desired project objectives such as cost, schedule, and quality)."

Special focus is therefore given to the contract administrators. Main variables of the model are derived from how well the project owner evaluates its contractors and the cost monitoring before and during a project, the level of support received by the project manager from the contractor's management during the project and the "early involvement of the project manager measured as a percent of the anticipated duration". It may be noted that all except the last variable are qualitative and results are therefore sensitive to their formulation.

3.6 Further Aspects on Variable Selection

Representatives of the Building Services Authority (Queensland, Australia) and the Building Commission (Victoria, Australia) suggested, via personal communication, that a leading indicator of future contractor behaviour is their rate of growth: a sudden expansion of a contractor's business activities is treated suspiciously, as it is assumed that rapid growth will result in a contractor being unable to pay current liabilities, and will therefore accept new contracts in order to meet existing short-run financial obligations. The Building Services Authority (Queensland, Australia) asserted that in such circumstances the firms' sales turnover are monitored more frequently. Furthermore, construction firms that are associated with overseas parents are usually exposed to a different risk profile than firms that only operate nationally. If an appropriate pooling is performed, a different set of variables will be needed to analyse firms in both groups to produce results at a similar quality level.

Clift (2006) identified continued trading losses (inadequate costing and debt collection) and poor accounting practice as main causes of failure through analysing the Building Services Authority (Queensland, Australia) database for the financial years 2003-04 and 2004-05. Considering that business failure is regarded herein as a process, the common early warning signs for business failure identified by Clift (2006) are of additional importance. They include:

- reduced profits or accumulated losses over previous two or more years;
- reliance on the Deed of Covenant and Assurance to meet BSA's financial licensing requirements;
- formal complaints about non-payment on behalf of subcontractors and suppliers; and
- delays in supplying required financial reviews for license renewal.

Lynch (2003) investigated approximately 15,000 licensees of the Building Services Authority (Queensland, Australia) between 1986 and 1996 for their survival prospects and found the following four variables to be relevant:

- business strategy: generalists fail, on average, less frequently than alterations specialists or new dwelling specialists;
- business size: Lynch (2003, p.246) proposes that "within each primary strategy, builders either consciously or unconsciously make trade-offs between growth and survival." and

concludes that small size builders are “persistently associated with higher survival prospects than large size [builders].”;

- business age: pursuing conservatively any business strategy will not improve survival chances through time; and
- temporary business exit: builders who seize business activities during periods of time may increase their survival probability, but never reduce it.

Overall, factors associated with business distress/failure that have been identified in prior studies inform the development of a suitable model to analyse and predict construction business performance in the Queensland context.

BUSINESS DISTRESS MODELLING METHODS

3.7 Introduction

Modelling business failures is a topic covered extensively in accounting, finance, decision sciences and operation research literature. Other disciplines such as statistics and bioinformatics have also discussed and presented algorithms to tackle the classification problem. This literature review is designed to give a broad overview of the mathematical methods that have the potential to model business distress for builders in the Queensland Building Services Authority database. Additionally, Zopounidis and Dimitras (1998) give a fairly comprehensive overview of the major works in this area, covering analysis such as single ratio analysis, discriminant analysis, logit and probit methods, recursive partitioning algorithm, survival analysis, neural networks and multi-criteria decision aid methods. Balcaen and Ooghe (2005) is the most recent literature review to date but is not comprehensive in the sense it only covers a limited range of statistical techniques.

This literature review has two major components which address the typical issues faced by researchers in modelling business failures: identifying the important variables and the choice of modelling technique.

3.8 Variable Selection

Hamer (1983) stated that variables for statistical modelling should be selected on the basis of minimizing the cost of data collection and maximizing the model applicability. Indeed, while Courtis (1978) identified 79 variables useful in predictive studies which were grouped into profitability ratios, managerial performance ratios and solvency ratios, it is impractical to use all of them in financial distress modelling. In practice, the variables used in business failure prediction literature are mainly a subset of financial ratios and occasionally include macroeconomic variables and other qualitative factors. A quick glance throughout various studies reveals there is a lack of consistency regarding to which variables should be used. This inconsistency is not surprising from a statistical point of view and there are a number of plausible reasons. Firstly, the variable selection for most studies is naturally limited by availability. Secondly, when there are many variables, it is usually preferable to reduce the number of variables by some kind of simplification procedure and it is well known that even a slight change in data can sometimes lead to a different set of variables being chosen. Lastly, many of the studies can be differentiated by the different industry and statistical methods they employ, therefore, the differences in variable selection is an expected phenomenon.

Dimitras, Zanakis and Zopunidis (1996) illustrated the above statements succinctly. In that paper, they investigated 47 studies from Journal of Banking and Finance, Journal of Business Finance and Accounting, Journal of Accounting Research, Omega, Decision Sciences, Journal of Finance and European Journal of Operation research across 12 different countries (Australia, Canada, Finland, France, Greece, Israel, Italy, Japan, Sweden, Holland, England and United States). While the most frequently used financial ratios are perceived to be Working Capital/Total Assets, Total Debt/Total Assets, Current Assets/Current Liabilities, Earnings before Interest and Taxes /Total Asset and Net Income/Total Assets, as shown in Table 2 there are no consistency between different studies. Other studies on business distress such as Tirapat and Nittayagasetwat (1999), Shah and Murtaza (2000) also used a subset of variables listed in Table 2. Still there are others such as Ooghe and Verbaere (1982) that uses additional variables such as amounts payable within one year for sales and services rendered over current working assets or Stein and Ziegle (1984) that uses other variables such as transfer credits/credit turnover which is not covered in Table 2.

In addition to financial ratios and their transformations such as taking logarithms, some literature works have also advocated the use of entropy from information theory. Pany (1979) used entropy analysis to examine the failed bank's financial volatility and other study as Lev (1971) also found failed firms have higher entropy values.

The concept of entropy comes from information theory, which in turn, is largely concerned with measuring the message of potential information as a decreasing function of an event, defined by log(1/p) where p is the probability of an event. For example, if there are 95 red balls and 5 white balls, an information regarding to the next ball drawn will be red will have a value of $\log(1/0.95) \approx 0.513$. However if instead, an information is being received that the next ball would be white, this will have a value of $\log(1/0.05)\approx 2.9957$. This difference reflects the fact that it is more informative to know that the next ball is white as it is expected that a red ball would usually be drawn. An entropy is the average or expected information and in the context of this example it is $0.95 \log(1/0.95)+0.05 \log(1/0.05)$.

In Pany (1979), the entropy used to analyse the financial variables is defined as:

$$-\sum_{i=1}^{n} q_i \frac{\log(q_i)}{p_i} \qquad (1)$$

where:

p_i= individual account balances for prior period as a percentage of total account balance in a financial statement.

q_i= individual account balances for current period as a percentage of total account balance in a financial statement.

The concept of entropy is useful not only in providing additional measures to the volatility of financial performances. The general concept of entropy can also be applied different context such as in classification which was used in constructing an expert system in Messier and Hansen (1988).

There are also some studies in multi-criteria decision aid methods that advocate the inclusion of qualitative information such as quality of management, technical capacity, market share, social importance which can be very important considerations. These are often ignored in financial distress modelling, partly because of then difficulty of measuring these items objectivity and using financial ratios alone often gives quite a high success of rate of classification. The inclusion of the multi-criteria decision aid methods in this literature review is designed to show that there are methods which incorporate subjective information quite successfully and can predict financial distress with remarkable accuracy.

In all, the key message from all of these prior studies is that the selection of the important variables is usually dependent on the data and it is necessary to exercise sound statistical techniques to choose the appropriate variables that can give an adequate picture of the organisational financial health. This point is also iterated in Balcaen and Ooghe (2005). There are several approaches:

- Conduct a correlation analysis and remove highly correlated redundant variables from the correlation matrix.

- The stepwise approach which involves removing the least significant results from regression analysis of the general unrestricted model in logit and probit models as discussed in Miller (1984). This does have potential problems as there is usually only one simplification path, so an omission of an important variable at the start of the process would cause the retaining of many other variables to proxy its role, resulting in a model that retain too many variables.
- The optimal regression approach which tries almost every combination of variables to give a simpler model that has the least information loss from the full model. A discussion on this subject can be found in Coen, Gomme and Kendall (1969).
- The General to Specific (Gets) modelling (Hendry 1995; Hoover & Perez 1999; Hendry 2000; Hendry & Doornik 2001) which was claimed by these authors to be a superior simplification regime than either stepwise approach or optimal regression approach since these methods do not check the congruence of reductions of the full model, resulting in unreliable inferences. The congruence here refers to no misspecifications of the statistical model. In this approach, the congruence of the original model is tested and this is maintained and checked throughout the selection process. The search strategies involved here requires consideration of different reduction paths and removal of either a block or a single variable to ensure the final model is the simplest congruent model possible. In the event of more than one model being identified, this approach will use encompassing tests[1] to resolve the choice. Hendry also demonstrated the accuracy of this approach through simulation and have refuted common statistical criticisms associated with p-value model reduction algorithm on his www.pcgive.com website.
- Multivariate data reduction method which usually involves principle component analysis where a series of different linear combinations of financial ratios is constructed in such a way that the information loss of multivariate data is minimized (Johnson & Wichern 1982).

[1] Encompassing is an econometrics concept as explained in Hoover and Perez Hoover and Perez (1999), a model is said to encompasses another if it conveys all the information conveyed by the other model. For example, consider a case where there is a general model G that uses all the unique variables of A and B and they all have the same dependent variable. If A is a valid restriction of the model G (e.g. based on the F test) and mode B is not, then model A encompasses model B and we know everything about model G from model A.

Table 2 Financial ratios used by countries in the Dimitras (1996) study. This table is designed to illustrate the lack of consistency of variable selection between different studies rather than to give a comprehensive view of the range of variables used in business distress modelling.

	Australia	Canada	Finland	France	Greece	Israel	Italy	Japan	Sweden	Holland	UK	US	Total
WC/TA				1	5			1			5	4	16
TD/TA		1	1		5							8	15
CA/CL	2	1	1		2	1						5	12
EBIT/TA	1		3		1				1		1	5	12
NI/TA	1		1		2							7	11
CF/TD			2	1								6	9
QA/CL			1								5	3	9
CF/S			6	1								1	8
RE/TA	1				1					1	1	3	7
S/TA		1	2	1						1		2	7
GP/TA					6								6
NI/SE					1		1			1	3		6
Cash/TA								1			2	2	5
PBT/S											5		5
S-TP/TC											5		5
Inv/S									1		2	1	4
QA/TA											1	3	4
TA/GNP											2	2	4

WC= Working Capital, TD= Total Debt, CA= Current Asset, CL= Current Liabilities, EBIT= Earnings before interest and tax, GP=Gross Profit, SE=Share holder's Equity, TP=Trading Profit, TC= Total Capital, Inv=Inventory, GNP=Gross National Product, NI=Net Income, CF=Cash Flow, QA=Quick Asset, PBT=Profit Before Tax, TA=Total Assets

3.9 Modelling Financial Distress

The mathematical financial distress models discussed in the literature are extensive and it is not possible to cover all of them. In addition, there are masses of classification models used in bioinformatics that could also be adapted in this context. This section attempts to provide a list of representative works to demonstrate the variety of techniques that can be used.

Perhaps the simplest way of modelling financial distress is to examine the financial ratios, using the financial ratio individually to calculate a cut off score for each ratio on the basis of minimizing misclassification errors. In Beaver (1966), the ratios found to have the highest discrimination powers are Cash flow/Total Debts, Net Income/Total Assets and Total Debts/Total Assets. Even though these ratios were found to give good predictions, academics (Edmister 1972) have criticised this approach as it can be difficult to determine the financial health of the firm when different ratios give contradicting results especially a single ratio cannot contain full information on the financial status of the firm.

The realization that different variables can be used conjointly to measure the financial health of the firm leads to many different financial distress modelling as shown in Table 3. This table shows some of the most frequent methods appeared in the literature in this subject but it is by no means an exhaustive list. This literature only covers technique that have been found useful and have been applied to financial distress modelling. It excludes analysis such as clustering (Schmidt 1984; Stein & Ziegler 1984) which was found to be a poor technique in identifying financially distressed firms. Other techniques such as Bayesian dimensional scaling (Oh & Raftery 2001) are also excluded, since while they can be useful, they have not been used specifically in the financial distress modelling context.

Table 3 A compendium of business distress models.

<table>
<tr><th></th><th>More Objective</th><th>Less Objective</th></tr>
<tr><td>Parametric</td><td>

- **Probit Analysis** (Grablowsky & Talley 1981; Izan 1984)
- **Logit Analysis** (Martin 1977; Schmidt 1984; Srinivasan & Kim 1987; Tam and Kiang 1992; Tirapat and Nittayagasetwat 1999; Charitou, Neophytou and Charalambous 2004; Jones and Hensher 2004; Lussier 2005)
- **Discriminant Analysis** (Takahashi & Kurokawa 1948; Altman, Haldeman & Narayana 1977; Altman 1984; Izan 1984; Micha 1984; Stein & Ziegler 1984; Taffler 1984; Frydman, Altman & Kao 1985; Leeuwen 1985; Srinivasan & Kim 1987; Wood & Piesse 1988; Laitinen 1991; Luoma & Laitinen 1991; Laitinen 1992; Tam & Kiang 1992; Altman 2000; Ganesalinggam & Kumar 2001; Altman 2002)
- **Survival Analysis** (Lane, Looney & Wansley 1986; Luoma & Laitinen 1991)
- **Time Series** (Kahya & Theodossiou 1999)

</td><td></td></tr>
<tr><td>Non parametric</td><td>

- **Multi-Dimensional Scaling** (Molinero & Ezzamel 1991; Neophytou & Molinero 2004)
- **Principle Component Analysis** (Takahashi & Kurokawa 1948; Ganesalinggam & Kumar 2001)
- **Linear Programming Classification** (Freed & Glover 1981; Freed & Glover 1981; Bajgier and Hill 1982; Srinivasan & Kim 1987; Gutpa, Rao & Bagvhi 1990; Koehler & Erenguc 1990; Rubin 1990; Lam & Moy 2003)
- **Classification trees and data driven expert systems** (Frydman, Altman & Kao 1985; Srinivasan & Kim 1987; Messier & Hansen 1988; Shaw & Gentry 1988; Salcedo-Sanz, Fernández-Villacañas, Segovia-Vargas & Bousoño-Calzón 2005)
- **Neural Networks** (Tam & Kiang 1992; Patuwo, Hu & Hung 1993; Altman, Marco & Varetto 1994; Lee, Han & Kwon 1996; Serrango-Cinca 1996; Charalambous, Charitou & Kaourou 2000; Shah & Murtaza 2000; Charitou, Neophytou & Charalambous 2004)

</td><td>

- **Multi-criteria Decision Aid Methods**
  - **Utility based approaches, e.g. *Preference Disaggregation* UTADIS** (Zopounidis & Doumpos 1999), **MHDIS** (Doumpos & Zopounidis 1999)
  - **Rough set theory** (Slowinski & Zopounidis 1995; Salcedo-Sanz, Fernández-Villacañas, Segovia-Vargas & Bousoño-Calzón 2005)
  - **Outranking Relations, e.g. *ELECTRE*** (Dimitras, Zopounidis & Hurson 1995)
- **User driven Expert Systems** (Duchessi & Belardo 1987; Srinivasan & Kim 1987; Duchessi, Shawky & Seagle 1988; Elmer & Borowski 1988; Srinivasan & Ruparel 1990)

</td></tr>
</table>

Table 3 classifies the financial distress models into four different categories. The parametric methods usually require some assumption on the distributional form of the variables, usually multivariate normal as in the case of discriminant analysis. The objectivity of the method is determined on the criteria whether the method requires additional input from the user beyond the data presented. For example, the ELECTRE III multi-criteria methods require the user to build a profile of limits to classify the firms into different categories in conjunction to linear programming and Bayesian techniques normally requires some knowledge of the prior probabilities. The disadvantage of parametric methods is that financial ratios are usually not normal, casting doubt to the validity of the statistical results. Non parametric methods are designed to circumvent these problems but may be less stable than parametric methods, with a few additional observations changing the results dramatically as often in the case of classification and regression trees. The Bayesian and multi-criteria methods can be criticised for being too subjective, but they are able to incorporate prior experience of the decision maker which can often boost the classification performance of the model. The choice of the technique used therefore, depends heavily on the circumstances. In the case where the decision maker has very little knowledge of the financial distress characteristics, the more objective methods may provide valuable insights to the possible patterns. In other cases, such as in the case of a bank loan manager, it may be preferable to use the more subjective method utilising the wealth of prior experiences to build a good model on financial distress. Regardless of the approach, a good financial distress model must have low misclassification errors acceptable to the decision maker and be efficient in the sense it minimizes the amount of information required.

The rest of the literature review will discuss an outline of each of the method in Table 3 with a special emphasis on its application to financial distress modelling with references to the literature. However it will not discuss the empirical results such as the variables used and the misclassification error since the purpose of this literature review is to highlight methods that have been found to be useful rather than commenting on the success of these methods which is very much dependent on the individual circumstances.

3.10 *Business Failure Modelling Methodology*

3.10.1 Probit and Logistic regression

Probit and logistic regression methods can give the probability of a firm being financially distressed based on the attributes or characteristics of the firm. These methods are mostly used as additional analysis in many of the business distress modelling to highlight the superiority of their new methods. An extension to the use of simple logit model is the use of mixed logit model in Jones and Hensher (2004).

In a simple probit model, the explanatory variables or the attributes of the firm $X_1,X_2,\ldots X_p$ and the dependent variable W (taking values of 0 or 1 representing healthy and distressed firm) can be written into a linear model:

$$W=\beta_0+ \beta_1X_1+\ldots \beta_pX_p+\varepsilon \qquad (2$$

$$E(W)=Y=\beta_0+ \beta_1X_1+\ldots \beta_pX_p \qquad (3)$$

In probit modelling, a firm is classified as being financially distressed if Y exceed a threshold Y*, otherwise it is healthy. Using normality assumption, the probability that Y* is less or equal to Y can be computed from the following:

$$P(Y \geq Y^*) = F(Y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\beta_0+\beta_1 X_1+\ldots\beta_p X_p} e^{-t^2/2} dt \tag{4}$$

$$F^{-1}(Y)=\beta_0+\beta_1 X_1+\ldots\beta_p X_p \tag{5}$$

W is known as the normal equivalent deviate or normit, probit is normit + 5, this is to adjust Y being negative when $F(Y) < 0.5$. The probit model, using the ordinary least square estimate is therefore:

$$Z=\beta_0^* +\beta_1 X_1+\ldots\beta_p X_p \tag{6}$$

In particular, $Z= F^{-1}(Y)+5$ and β_0^* is β_0+5. The expression in (5) has unequal variance σ_ε^2 in terms of its error, so an adjusted weighted least square regression[2] is carried out to obtain the final probit model in (6).

$$Z^*=\beta_0^{**} + \beta_1^* X_1+\ldots\beta_p^* X_p \tag{7}$$

Probit regressions are less frequently used than the logit model, perhaps owing to the greater availability of logit model in computer packages.

In similar fashion, let the independent variables or the attributes of the firm $X_1,X_2,\ldots X_p$ and the dependent variable W (taking values of 0 or 1 representing healthy and distressed firm), the logit model takes the following form in (7).

$$P(W = 1) = \frac{1}{1+e^{\beta_0+\beta_1 X_1+\ldots\beta_p X_p}} \tag{8}$$

Equivalently:

$$\ln\left(\frac{P(W = 1)}{1-P(W = 1)}\right) = \beta_0 + \beta_1 X_1 + \ldots\beta_p X_p \tag{9}$$

The logit model was used as a viable business distress modelling method in Tirapat and Nittayagasetwat (1999) and investigated the business distress classification accuracy under various logit probability cut off points for both the training sample and hold out samples. They also included macroeconomic variables in addition to the financial attributes of the firm.

It is well known that logit models are sensitive to multicollinearity[3] and this is particularly serious in financial distress modelling since the independent variables are often financial

[2] This is found by dividing the depend and independent variables in (5) by the standard deviation of the error term and the coefficients are re-estimated using the least squares technique to give the coefficients in (6).
[3] Other assumptions include equal variance of residuals, non correlated errors. Sometimes it is necessary to transform the data using logarithm or otherwise to achieve these.

ratios which share the same denominator or numerator. This problem is prevalent in many of the works on business failure modelling and is likely lead to poor model performances in light of the new data.

A recent development in logit model is mixed logit models (McFadden & Train 2000) originated from the hedonic models developed in Cardell and Dunbar (1980), Boyd and Melman (1980). The basic idea of the mixed logit model is that the business distress alternatives or categories such as distressed and insolvent may be correlated and heteroscedastic so instead of using the expression in (1), the following expression is used:

$$W=\beta_0+ \beta_1X_1+\ldots \beta_pX_p+\eta+\varepsilon \quad (10)$$

$$\eta= \alpha_1X_{1u}+\ldots \alpha_sX_{su} \quad (11)$$

The first part $\beta_0+ \beta_1X_1+\ldots \beta_pX_p$ is independently and identically distributed among the alternatives and individual firms and the second part $\eta+\varepsilon$ is correlated among the alternatives and heteroscedastic. η is made up series of unobserved parts $(X_{1u},\ldots X_{su})$ which corresponds to the observed parts $(X_1,X_2,\ldots X_s)$. η can be interpreted as a random term with zero mean with a statistical distribution which is dependent on the underlying parameters and observed data for each alternative and individual firm. Furthermore, the mixed logit model assumes an extreme value distribution such as Gumbel for ε with η taking more general distributions such as triangular, normal and log normal. The conditional probability of choosing an alternative i given η is:

$$P(i|\eta)=\exp(\beta_0+ \beta_1X_1+\ldots \beta_pX_p+\eta)_i/\Sigma_j(\exp(\beta_0+ \beta_1X_1+\ldots \beta_pX_p+\eta)_j \quad (12)$$

Consequently, the probability of choosing an alternative or P(i) is:

$$P(i)=\int P(i|\eta)f(\eta|\theta)d\eta \quad (13)$$

Where θ is the parameter of the distribution f. Expression (11) give the mixed logit model with $P(i|\eta)$ being a mixture of logits and f as the corresponding distribution. The specification of the random components η usually involves identifying the mean and standard deviation of a particular β associated with an attribute of an alternative outcome. This was the strategy adopted in Jones and Hensher (2004) in their application of mixed logit model in financial distress modelling. While Jones and Hensher (2004) showed the performance of this model is better than the traditional multinomial logit model, the interpretation of the coefficients in these models are more complex and requires careful interpretation of the random components than the traditional linear models.

3.10.2 Principle component analysis

Principle component analysis is not a classification technique but an exploratory technique which can be used to identify characteristics of the financially distressed and healthy firms. The foundation of principle component analysis originated from Pearson (1901) and Hotelling (1933). This technique aims to take a set of p variables $X_1,X_2,\ldots X_p$ such as firm's financial and non financial characteristics and find a linear combination of these to produce uncorrelated indices $Z_1,Z_2,\ldots Z_p$. Each of the Z_i, for $i=1,2,3,\ldots p$ represents a dimension of the data and they are ordered so that $var(Z_1)\geq var(Z_2)\geq\ldots var(Z_p)$. The main use of the principle component analysis is to reduce the dimensionality of the data to a few of the Z_i, so that the multivariate data set is adequately described by these indices.

Mathematically, the variables X_i are normalized with zero mean and unit variance, and the aim is to find $Z_i=a_{i1}X_1+ a_{i2}X_2+\dots a_{ip}X_p$ such that $Var(Z_i)$ is maximized under the constraint[4] $a_{i1}^2 + a_{i2}^2 + \dots a_{ip}^2 = 1$ with Z_i being uncorrelated to each other. The principle components Z_i can then be plotted against each other to gauge if there are discerning patterns between distressed and healthy firms in relation to their principle component Z_i scores. A combination of Z_i can then be chosen to classify firms into their financial status. This technique was used in Takahashi and Kurokawa (1948) and more recently in Ganesalinggam and Kumar (2001). The principle component scores Z_i can also be used as variables in other statistical analysis such as logistic regression to enhance the classification performance of other statistical models.

3.10.3 Discriminant analysis

Perhaps the most popular technique in business distress modelling is the two group discriminant analysis. Due to its availability on standard computing packages, almost every literature on business failure covered this technique. Among the most cited work in this area are those by Altman et. al (Altman, Haldeman & Narayana 1977; Altman 1984; Altman, Marco & Varetto 1994; Altman 2000; Altman 2002). Despite the violations of the statistical assumptions of discriminant analysis in business distress modelling[5], discriminant analysis is still a widely used method as it can usually provide a fairly good classification. The statistical significance of the results, however, would need to be revised in the light of these violations.

Perhaps the simplest discriminant analysis is the linear discriminant analysis devised by R.A. Fisher as a way of distinguishing between groups. This analysis seeks to find a function of linear combination of X_i variables (e.g. i-th financial attributes of the firm with p attributes in total) that can separate healthy and distressed firms. The output of Fisher's discriminant analysis is a set of linear function as shown in (13).

$$Z=a_1X_1+a_2X_2+\dots a_pX_p \tag{14}$$

In (13), the $a_1,a_2,\dots a_p$ are chosen to maximise the F ratio (Fisher 1936), which is the between groups variation M_b divided by within group variation M_w. In this manner the discriminant function is the one that maximizes the variance between the groups and minimizes the variance within each group. Usually, under this method, when group sizes are equal, the cut off value to classify the firms is the mean of the two centroids (for two-group discriminant analysis). If the groups are unequal, the cut off is the weighted mean.

Among the most frequently used discriminate functions in the literature are those derived from probability distributions. These are an alternative to the Fisher's discriminant functions and the steps are detailed as follows:

---

[4] $Var(Z_i) = a_{i1}^2 + a_{i2}^2 + \dots a_{ip}^2$

[5] For example: non multivariate normal variables, lack of independence between different attributes of the same firm and unequal within group variable variances between failed and healthy firms. If the within group variable variances are not the same between failed and healthy firms, then it is necessary to use quadratic discriminant analysis.

Let the attributes of the firm be: $s=s_1,s_2,\ldots s_p$ and based on these attributes it is desirable to classify the firm into classes $c_1,c_2,\ldots c_k$. The probability that the firm with attributes in s to belong to class c_j is:

$$P(c_j \mid s) = \frac{P(s \mid c_j)P(c_j)}{P(s)} \tag{15}$$

Expression (14) used Bayes rule, with $P(c_j)$ represents the probability finding a firm to come from class j. As is usual with discriminant analysis, the $s_1,s_2,\ldots s_p$ are assumed to come from a multivariate normal distribution $N(\mu_j,\Sigma_j)$ where μ_j is a vector of means and Σ_j is the dispersion matrix. The probability of observing s given it is in class j is given below.

$$P(s \mid c_j) = \frac{1}{\sqrt{2\pi}^p}\frac{1}{\sqrt{\det \Sigma_j}}\exp\left(-\frac{1}{2}(s-\mu_j)'\Sigma_j^{-1}(s-\mu_j)\right) \tag{16}$$

The goal is to maximise the probability in which a certain firm belongs to a particular class j, this means to maximise $P(c_j|s)$. The maximum of $P(c_j|s)$ is attained if:

$$d_j(s)<d_r(s),\ \forall r \neq j \tag{17}$$

where:

$$d_j(s) = -\frac{1}{2}\log(\det \Sigma_j) - \frac{1}{2}(s-\mu_j)'\Sigma_j^{-1}(s-\mu_j) + \log(P(c_j)) \tag{18}$$

In particular, the function d_j is the quadratic discriminant function of s. A linear discriminant function is attained by assuming Σ_j to be identical for all classes:

$$d_j(s) = s'\Sigma^{-1}\mu_j - \frac{1}{2}(\mu_j)'\Sigma^{-1}(\mu_j) + \log(P(c_j)) \tag{19}$$

with estimates of μ_j, Σ_j being obtained from the sample.

The value of cut off can be determined by the analyst or the same strategy shown above in the case of Fisher's discriminant analysis could be used. Altman, Haldeman and Narayana (1977) devised the following "optimal" cut off score for their discriminant model:

$$\text{Cut off}_1 = \ln\frac{q_1C_1}{q_2C_2} \tag{20}$$

where, q_1 and q_2 are the prior probabilities of bankrupt and non bankrupt firms and C_1 and C_2 are the costs of type I and type II errors respectively. In particular, Altman, Haldeman and Narayana (1977) argued that the type 1 and type II error costs can be calculated as follows.

$$C_1 = 1 - \frac{\text{Amount of loan losses recovered}}{\text{Gross loan losses}} \qquad (21)$$
$$C_2 = \text{effective loan interest - opportunity cost for bank}$$

The discrimination functions described above are highly sensitive to outliers and can fail in case of heavy tailed distributions. There have since been some new developments which utilise non parametric methods (Epanechnikov 1969; Ghosh & Chaudhuri 2005) to try to surpass the normality assumption imposed by the above analysis and their applicability to financial distress modelling is yet to be seen.

3.10.4 Multidimensional Scaling

Multidimensional scaling (Torgerson 1952; Kruskal 1964; Kruskal 1964) constructs a map to show the relationship between objects, using a table of distances. In the case of financial distress modelling (Molinero & Ezzamel 1991; Neophytou & Molinero 2004), the idea is to calculate the distances between financial attributes of pair-wise firms and then project them on to a "map" to examine whether there is a pattern between failed and successful firms. The procedure is as follows:

1. Calculate the distance, δ_{ij} between firms i and j in relation to n financial attributes.

$$\delta_{ij} = \sqrt{(F_{i1} - F_{j1})^2 + (F_{i2} - F_{j2})^2 + \ldots + (F_{in} - F_{jn})^2} \qquad (22)$$

The $F_{i1,}$ for example, represents the first attribute of firm i. The attributes may be financial ratios or other characteristics of the firm. Expression (21) would generate a square distance matrix Δ with each element being δ_{ij}. Δ is usually called the dissimilarity matrix, with dimension $p \times p$, where p is the number of firms.

2. Create a rank order matrix Γ based on Δ, with the largest distance having a value p(p-1)/2 and the smallest having a value of 1.
3. Randomly generate a configuration of firm co-ordinates $x_{iw}, x_{2w}, \ldots x_{pw}$ for w=1,2,3…t dimensions. The number of dimensions t is usually chosen by the user, one way is to use the concept of STRESS which is described below. Alternatively the eigenvalues of the dissimilarity matrix can be computed to find out how much variation is explained by dimensions. Usually the number of dimensions that explains more than 80% of variation in the multivariate data is chosen.
4. Calculate the distance d_{ij} between firm i and j in the reduced dimension t. This creates matrix Θ.

$$d_{ij} = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \ldots + (x_{it} - x_{jt})^2} \qquad (23)$$

5. Compare Θ with Γ to create disparities $\hat{d}_{ij}$. The entry that has a smaller rank in Γ should also have a smaller value in Θ. If this is not the case, the adjacent entries of the Θ that violate the order would be replaced by their mean. This process is iterated until a set of disparities is found that does not violate the rank order in Γ.
6. The goodness of fit between the configuration distances and the disparities is measured by a suitable statistic. Kruskal's STRESS 1 formula is perhaps the most commonly used. It measures the degree of "stress" required to get x_{ij} into δ_{ij}. STRESS would improve as the

number of dimensions increases; therefore the number of dimensions can be chosen in such a way that there is minimal improvement in STRESS level.

$$\text{STRESS1} = \frac{\sum (d_{ij} - \hat{d}_{ij})}{\sum \hat{d}_{ij}^2} \tag{24}$$

7. Compute a new improved configuration. The formula usually used to find the new co-ordinate x_i^* for a particular dimension is shown in expression (24).

$$x_i^* = x_i + \left[\frac{\sum_{j=1}^{p-1} \left(d_{ij} - \hat{d}_{ij} \right) \left(x_j - x_i \right)}{d_{ij}} \right] \tag{25}$$

8. Once a set of new configuration is calculated, the process returns to 4 and it is iterated until the STRESS1 cannot be further reduced.

The outcome of this analysis is a set of co-ordinates for p firms in t dimensions. These can be used to map out how failed and healthy firms are related. It is often desirable to keep the number of dimensions t to 2 or 3, so the results can be displayed graphically. It is not always possible to do this and in the case of Neophytou and Molinero (2004), six dimensions were chosen. These dimensions are then examined on a pairwise basis on two dimensional plots. An example of healthy and unhealthy firms plotted in dimension 1 and 3 from Neophytou and Molinero (2004)[6] is shown in Figure 1. This is a good example showing how firms can be classified quite effectively using the multidimensional scaling technique.

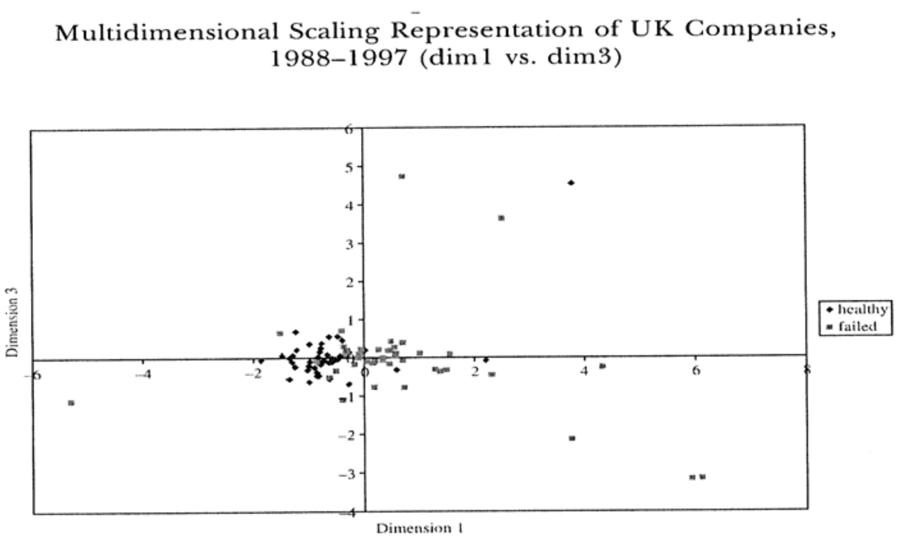

Figure 1 Multidimensional scaling result from Neophytou and Molinero (2004) using dimension 1 and 3.

[6] The authors also examined the significance of these dimensions using logistic regression analysis, which leads them to choose dimension 1,3 and 4 as the basis for classifying firms.

3.10.5 Time Series Analysis

Time series analysis is a particularly attractive tool in business distress modelling as it can incorporate multi period information and account for serial correlation across firms' attributes over time. Theodossiou (1993) describes the time series behaviour of the healthy and failed firms through k-th order vector autoregressive (VAR) model:

$$X_{i,t}=A_{f,s}+A_h+X_{i,t-1}B_1+\ldots+X_{i,t-k}B_k+\epsilon_{i,t}, \text{ for } s=1,2,\ldots,m \text{ and } i=1,2,3,\ldots p \qquad (26)$$

$$A_{f,s}=0, \text{ for healthy firms and } s>m \text{ for failed firms} \qquad (27)$$

The $X_{i,t}$ are the attributes such as financial ratios and B_i is a p by p matrix of VAR coefficients. $\epsilon_{i,t}$ is the error vector with the following properties.

$E(\epsilon_{i,t})=0$, $E(\varepsilon'_{i,t}\,\epsilon_{i,t})=\Sigma$ and $E(\varepsilon'_{j,r}\,\epsilon_{i,t})$ as long as $i\neq j$ or r $r\neq t$ is true, i.e. the error is uncorrelated across firms and time. A_h is a vector of intercept for healthy firms. $A_{f,s}$ is the vector of deviations of corresponding attributes in A_h vector for failed firm at s years prior to failure. $A_{f,s}$ is interpreted as the permanent shifts in mean structure of the variables due to financial distress. By construction, $A_{f,s}$ is zero for all the attributes of healthy firms and zero for failed firms extracted prior to the starting point of shift in the distribution of $X_{i,t}$ from the healthy population to the failed population ($s>m$). $A_{f,s}$ is also denoted as A_f.

Using these preliminary formulations, Theodossiou (1993) developed a time series cumulative sum model (CUSUM). This model is designed to provide a signal of firm's deteriorating condition as soon as:

$$C_{i,t}=\min(C_{i,t-1}+Z_{i,t}-K,0)< -L \text{ for } K, L>0 \qquad (28)$$

In expression (27), $C_{i,t}$ and $Z_{i,t}$ are cumulative and annual time series performance scores for the i-th firm at time t. K and L are sensitivity parameters. The score $Z_{i,t}$, as discussed in Kahya and Theodossiou (1999)[7] is shown below.

$$Z_{i,t}=\beta_0+(X_{i,t}-A_h-X_{i,t-1}B_1-\ldots-X_{i,t-k}B_k)\,\beta_1=\beta_0+(A_{f,s}+\epsilon_{i,t})\,\beta_1 \qquad (29)$$

$$\beta_0=(0.5/D)A_f\Sigma^{-1}A'_f=D/2 \qquad (30)$$

$$\beta_1=(-1/D)\,\Sigma^{-1}A'_f \qquad (31)$$

$$D^2=A_f\Sigma^{-1}A'_f \qquad (32)$$

The quantity D is the Mahalanobis generalised distance of the error terms which can be interpreted as random components of the variables for financially healthy and distressed firms. Among the population of healthy firms, the annual performance score $Z_{i,t}$ has a mean of

[7] Theodossiou (1993), Kahya and Theodossiou (1999), Kahya, Ouandlous and Theodossiou (2001) did not give consistent formula for β_0 and β_1. The correct formulae are given here.

D/2 and in the case of financially distressed population[8], the mean is –D/2. The CUMSUM model in (27) means that a healthy firm would typically have positive $Z_{i,t}$ scores with $C_{i,t}$ equals to 0. A typical failing firm would have $Z_{i,t}$ scores falling below K, resulting in cumulative negative $C_{i,t}$ scores. Once $C_{i,t}$ falls below –L, a change in firm's financial condition is signalled. Interpreted in this manner, K and L determine the occurrence and detection of a change in financial condition of the firm respectively. A larger K would lower the probability of misclassifying the firm as healthy but increase the probability of misclassifying a healthy firm as failed. The opposite effect holds for L.

To find the optimal values of K and L, first define:

P_f=Pr($C_{i,t}$>-L| firm failed and s=1), and

P_h=Pr($C_{i,t}$≤-L| firm is financially healthy)

The optimal values of K and L are found by solving the EC or expected cost function:

$$\underset{K,L}{\mathrm{Min}}\, EC = w_f P_f(K,L) + (1 - w_f) P_h(K,L) \tag{33}$$

The weights w_f are determined by the user with w_h=1-w_f. Usually this is calculated by:

$$w_f = \frac{\pi_f c_f}{\pi_f c_f + \pi_h c_h} \tag{34}$$

$$w_h = \frac{\pi_h c_h}{\pi_f c_f + \pi_h c_h} \tag{35}$$

The variables c_h and c_f are misclassification costs of healthy and failed firms. The corresponding prior probabilities of financially distressed and healthy firms from the population are π_f and π_c respectively. In the absence of prior information equal weights (π_f=0.5,π_c=0.5) are usually used. The process of finding optimal K and L usually requires randomly dropping one healthy firm and one distressed firm and re-estimates all the CUSUM model parameters and reclassified to their new scores to derive P_f and P_h and subsequently values for K and L. A series of w_f is then generated and the combination of K, L and w_f that leads to the lowest EC is chosen.

Once the optimal K and L are found, they are substituted into (27). The CUMSUM score $C_{i,t}$ over the years for a particular firm is then calculated recursively by (27) and a firm is classified as financially distressed as soon as its $C_{i,t}$ score falls below -L.

3.10.6 Linear Programming

One of the earliest works in using linear programming methods for classification came from Freed and Glover (1981; 1981) which was used by Mahmood and Lawrence (1987), Gutpa, Rao and Bagvhi (1990) in classifying financial health of the firm. There are many different ways in which the linear programming can be set up to achieve this purpose. The objective

[8] This result is proved in Kahya and Theodossiou (1999).

function may for example be based on: maximising minimum distances of misclassification (Freed & Glover 1981), optimize the sum of distances (Bajgier and Hill 1982), minimize the sum of interior[9] distances (Freed & Glover 1986), minimize the sum of deviations (Freed & Glover 1986) or even a 'hybrid' of minimize the difference between exterior[10] and interior distances (Glover, Keene & Duea 1988; Glover 1990). Other variations based on minimizing the number of classifications using mixed integer programming have also been proposed in Banks and Prakash (1991), Koehler and Erenguc (1990).

These early research can be problematic. As demonstrate in Xiao (1993), the maximising minimum distances (MMD) and minimizing sum of deviations (MSD) models do not work in every case in the sense it is possible to get multiple optimal solutions from the linear programming models which suggest different classifications. This occurs even when the two groups are well separated. This is not desirable since it is important to be able to make clear decision of the financial status of the firm. Xiao (1993) then went on to demonstrate the conditions when these methods would fail and generally recommend MSD over MMD models. Theoretical results aside, the MSD models also appear to classify better in simulation studies than MMD models (Bajgier & Hill 1982; Freed & Glover 1986; Joachimsthaler & Stam 1988).

The most recent development in the use of linear programming to solve classification problems arise from the works of Lam, Choo and Moy (1996), Lam and Moy (2003). The authors in these works provide a simple linear programming technique which can classify better than MSD models or Fisher's linear discriminant function in a number of simulations where there is overlap between groups. The basic model considered in their papers is described below.

LP1:

$$\text{Minimize } \sum_{i \in G_1} (d_i^+ + d_i^-) + \sum_{i \in G_2} (e_i^+ + e_i^-) \tag{36}$$

such that:

$$\sum_{j=1}^{q} \left(a_{ij} - {}_1\overline{a_j}\right) w_j + d_i^- - d_i^+ = 0, \ i \in G_{1,} \tag{37}$$

$$\sum_{j=1}^{q} \left(a_{ij} - {}_2\overline{a_j}\right) w_j + d_i^- - d_i^+ = 0, \ i \in G_{2,} \tag{38}$$

$$\sum_{j=1}^{q} \left({}_1\overline{a_j} - {}_2\overline{a_j}\right) w_j \geq 1, \tag{39}$$

[9] Interior distances refer to distances within groups
[10] Exterior distances refer to distances between groups.

where G_1 represents group 1, G_2 represents group 2, a_{ij} represents the j-th attributes such as current asset ratio for group i and there are q attributes, ${}_1\overline{a_j}$ represents the average value of the j-th attributes for all firms in group 1, ${}_2\overline{a_j}$ represents the average value of the j-th attributes for all firms in group 2. The variables to evaluate from expressions (35), (36), (37) and (38) are: w_j represents the weights applied to the mean adjusted attributes, d_i^+ represents the positive deviation for group 1, d_i^- represents the negative deviation for group 1, e_i^+ represents the positive deviation for group 2 and e_i^- represents the negative deviation for group 2.

These formulations seek to minimize the positive and negative deviations for group 1 and 2 from the classification weights w_j. Expression (38) is designed to avoid unacceptable solutions as discussed in Markowski and Markowski (1985) and restricts the two group mean classification scores being greater than or equal to 1. Lam and Moy (2003) claimed that their model is not unlike the philosophy of Fisher's discriminant analysis in maximizing a ratio of the between group deviations to the within group deviations. They also claimed that LP1 does not produce unacceptable nor improper solutions as long as the mean vector of attributes of group 1 is not identical to that of group 2.

Once the weights wj are found, the cut off value c of the classification can be determined in two ways:

$$\text{Minimize} \sum_{i=1}^{n} h_i \tag{40}$$

such that:

$$S_i + h_i \geq c,\ i \in G_{2,} \tag{41}$$

$$S_i + h_i \leq c,\ i \in G_{2,} \tag{42}$$

Or in a mixed integer model:

$$\text{Minimize} \sum_{i=1}^{n} z_i \tag{43}$$

such that:

$$S_i + Mz_i \geq c,\ i \in G_{2,} \tag{44}$$

$$S_i + Mz_i \leq c,\ i \in G_{2,} \tag{45}$$

In both cases the S_i is the score calculated from the weights obtained in LP1 for firms i=1,2,…n. In the first case, h_i is set to be greater or equal to zero and the objective is to minimize the sum of deviations from the cut off value. In the second case, z_i is binary with M being set to a large number and the objective is to minimize the total number of misclassified objects. Lam and Moy (2003) extends LP1 into piecewise linear programming model with set of weights putting greater emphasis on observations that can be clearly distinguished by various different discriminant methods and demonstrated that this technique can give quite good classification results over existing methods such as MSD and Fisher's discriminant analysis. The practical differences in results however are quite small, differing from 1-2% in most cases.

3.10.7 Survival Analysis

Survival analysis using the Cox proportional hazard model appeared in Luoma and Laitinen (1991) as well as Lane, Looney and Wansley (1986). While it has not been extensively used in the business distress modelling literature, it is nevertheless a viable statistical technique. The main concepts of the Cox proportional hazard model are described below.

Let T be the random variable representing time to failure of a bank. $S(t)=P(T>t)$ is the survivor function, representing the probability that a firm will live beyond t time units. The distribution function of time to failure is $F(t)=1-S(t)$ with density function being $f(t)=-S'(t)$. The hazard function, the probability that the firm will not fail at the next instant, given it has not failed at time t is:

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t < T < t + \Delta t \mid T > t)}{\Delta t} = \frac{-S'(t)}{S(t)} \qquad (46)$$

Usually, it is desirable to evaluate h(t) (Cox & Oakes 1984), rather than F(t) or f(t), when h(t) is found F(t) and f(t) can be found using:

$$\begin{aligned} F(t) &= 1 - \exp\left[-\int_0^t h(u)du \right] \\ f(t) &= F'(t) \end{aligned} \qquad (47)$$

The most commonly used survival analysis is the Cox proportional hazard model:

$$h(t|z)=\exp(\beta' z)h_0(t) \qquad (48)$$

In expression (47), the z represents a vector of attributes such as financial ratios with the β' being the transpose of coefficients similar in a normal regression model. The exponential function is chosen for $\beta' z$ since this would simplify the regression coefficients β. If z is all 0, then $h(t|z)=h_0(t)$, this is the baseline hazard function, often assumed to be non parametric and left unspecified. Remarkably, even though the baseline function is not specified, the Cox proportional hazard model can still be estimated by method of partial likelihood (Cox 1972), this has an advantage of not having to make incorrect assumptions about the form of the baseline hazard[11].

In addition to expression (47), it is possible to use the fact that $S_0(t) = \exp\left[-\int_0^1 h_0(u)du\right]$, to derive the following survival function as in (48).

$$S(t|z) = S_0(t)^{\exp(\beta' z)} \tag{49}$$

The Cox model would require identifying failed and successful firms with observations made to the attributes of the firms at some k periods before failure. The fitted survival function in expression 2) will give the probability that certain financial institution possessing certain attributes would survive for m ($m \leq k$) periods into the future. To classify the firm into financial healthy and distressed categories, Lane, Looney and Wansley (1986) based the probability cut offs on the proportion of non failed banks, in much same way as was done in the case of logistic regression in Martin (1977).

3.10.8 Multi-criteria decision aid methods (MCDA methods)

Financial distress classification is characterized by multiple criteria (usually financial ratios) which may give conflicting results. Very often in practice, there is a complex evaluation process that is subjective and requires the expertise of the decision maker. The MCDA methods are designed for these types of problems and they are well suited to financial distress modelling. An important contribution of the MCDA is that they can incorporate both quantitative and qualitative information thereby allowing their models to achieve as much as possible with all the available information.

4.4.8.1 Utility-based approaches

This approach requires, in a nutshell, the decision maker's preference to be modelled into a utility function with the optimal decision taken at the maximum of the utility function. Multi-attribute Utility Theory (MAUT) is an extension of the traditional utility theory in the multivariate case. The utility function in MAUT, is defined in such a way that if $u(a_i) > u(a_f)$, a_i is preferred to a_f and if a_i is indifferent to a_f then $u(a_i) = u(a_k)$. Usually, the utility function is made of linear combinations of marginal utilities:

$$u(a_i) = w_1 u_1(g_{i1}) + w_2 u_2(g_{i2}) + ... w_n u_n(g_{in}),$$

where $w_1, w_2, \ldots w_n$ represent the criteria tradeoffs the decision maker is willing to take and $g_1, g_2, \ldots g_n$ represent the criteria needed to make the decision. In the context of business distress modelling, the set $a_1, a_2, \ldots a_k$ represents the set of firms to be used.

The determination of the additive utility function requires cooperation between the decision analyst and the decision maker to decide the form of the utility function and the criteria tradeoffs using interactive techniques such as the mid point value (Keeney & Raiffa 1993).

[11] Although it would be less efficient than using a maximum likelihood estimates for a correctly specified hazard regression model. However, often the correct hazard regression model is never known in practice.

This is usually a very time consuming process. The preference disaggregation analysis (PDA) is developed to combat this problem. The PDA disaggregates the global preference of the decision maker to determine the criteria aggregation model that underlies the preference result. Instead of estimating the global utility model directly as in MAUT, a linear interpolation approach is taken to estimate the global utility model indirectly. There are a wide range of variations of PDA methods in financial distress modelling and two of the most commonly referred technique **UTADIS** (**UT**ilités **A**dditives **DIS**criminantes) and **MHDIS** (**M**ulti-group **H**ierachical **DIS**crimination) are discussed here.

In UTADIS, for each business distress criterion (e.g. financial ratio), the decision maker is asked to classify firms into different classes of business distress. Once this is completed, a linear interpolation procedure is applied to derive the marginal utility function and the global utility function is evaluated. The firms $\{a_1, a_2, a_3, \ldots a_k\}$ are then classified into in classes $c_1, c_2, \ldots c_q$ in such a way that c_1 is preferred to c_2, c_2 is preferred to c_3 and so forth. The rule for classification is shown in expression (49) with $t_1 > t_2 > \ldots > t_{q-1}$ being the classification threshold.

$$\begin{array}{ll} u(a_i) \geq t_1 & \forall a_i \in c_1 \\ t_1 > u(a_i) \geq t_2 & \forall a_i \in c_2 \\ \ldots\ldots\ldots\ldots\ldots\ldots & \ldots\ldots\ldots\ldots\ldots \\ u(a_i) < t_{q-1} & \forall a_i \in c_q \end{array} \tag{50}$$

In the context of expression (49) there are two possible errors, the firm might be underestimated by the decision maker, resulting an underestimating error where the developed model assigns a firm to a lower rank than its real ranking or the overestimating error which assigns a lower ranked firm to a higher rank. The underestimation error is defined as $\sigma_i^+ = \max\{0, t_d - u(a_i)\}, \forall a_i \in c_i, d = 1,2,3,\ldots q-1$ and the over estimation error is given by $\sigma_i^- = \max\{0, -t_d + u(a_i)\}, \forall a_i \in c_i, d = 2,3,\ldots q$. To minimize the these errors it is necessary to calculate the appropriate classification threshold by using a linear programming technique and the details can be found in Zopounidis and Dimitras (1998) and Zopounidis and Doumpos (1999).

MHDIS is an extension of UTADIS method. MHDIS classifies the state of business failures through a hierarchical procedure, starting by discriminating the group c_1 from all the other groups $\{c_2, c_3, \ldots c_q\}$ and then proceed to discriminate between firms belonging to other groups. At each stage of this hierarchical process, two additive utility functions are developed for the classification of the firms as shown in expression (50).

$$u_k(a_i) = \sum_{j=1}^{n} u_{ki}(g_{ij}), u_{\sim k}(a_i) = \sum_{j=1}^{n} u_{\sim ki}(g_{ij}), \tag{51}$$

Usually, both functions are restricted to [0,1] and the function u_k measures the utility for the decision maker to assign a firm into c_k. The second function $u_{\sim k}$ corresponds to the classification into the set of groups $c_{\sim k} = \{c_{k+1}, c_{k+2}, \ldots, c_q\}$. In the context of business distress modelling, there are usually two classes, healthy and unhealthy firms and the utility functions in expression (50) would represent the utility functions for healthy and unhealthy firms respectively. Using these utility functions, the rules used to perform the classifications of the firms are as follows:

$$
\begin{aligned}
&\text{If } u_1(a_i) > u_{\sim 1}(a_i) \text{ then } a_i \in c_1 \\
&\text{Else if } u_2(a_i) > u_{\sim 2}(a_i) \text{ then } a_i \in c_2 \\
&\text{........} \\
&\text{Else if } u_{q\text{-}1}(a_i) > u_{\sim (q\text{-}1)}(a_i) \text{ then } a_i \in c_{q-1} \\
&\text{Else if } a_i \in c_q
\end{aligned}
\tag{52}
$$

The optimal additive utility function model is found using three linear programming procedures. At each stage m of the hierarchical discrimination process (m=1,2,3,…q-1) two linear and one mixed integer programming are solved to find the optimal pair of utility functions in the sense that the total number of misclassification and clarity of distinction between the groups. In the first stage, the magnitude of the classification errors in distance term is minimized using a linear programming approach (LP1). Then, a mixed integer problem (MIP) is solved to minimize the total number of misclassifications that occur after the solution of LP1, while retaining the correct classifications. Finally, a second linear programming is solved to maximise the clarity of the classification obtained from the solutions of LP1 and MIP. A detailed description of this method can be found in Doumpos and Zopounidis (1999), Zopounidis and Doumpos (2000).

4.4.8.2 Rough set theory

The rough set theory is another useful tool in classification problem. This theory was introduced by Pawlak (1982) and since then there has been application of this method in business distress modelling (Slowinski & Zopounidis 1995; Dimitras, Slowinski, Susmaga & Zopounidis 1999). In rough set theory, every firm has two types of attributes. The condition attributes describe the characteristics of the firm (e.g. financial ratios, usually discretized into discrete intervals) and the decision attributes define the group classification of the firm. Firms with the same condition attributes are described as being "indiscernible", and this concept forms the main basis for the rough set theory. Any set of all indiscernible firms is an elementary set and a union of elementary sets is referred to as crisp or precise otherwise it is a rough set (vague, imprecise). A rough set can be approximated by a pair of crisp sets, known as the lower and upper approximation. The lower approximation includes firms that are known to belong to the set and the upper approximation includes firms that could possibly belong to the set. The quality of approximation of classification ζ is defined as the sum of all the firms under lower approximation over the total number of firms.

To ensure that an efficient set of attributes is chosen for classification, rough set theory begin by finding subset of attributes that provide the same quality of classification (ζ) as the whole set of attributes, these subset of attributes are called reducts and there are usually several reducts. The set of attributes that appears in all the reducts is called the core, which means it cannot be excluded in the analysis without reducing the quality of classification. It is also possible that a given situation will have no core.

Once these reducts are found, a reduct is chosen to develop the decision rules for classification. This reduct would always consists the core, with the smallest number of attributes and not miss any of the attributes considered by decision maker to be important in classifying the firms into different business distress categories. The decision rule is based on the logical argument that if the condition attributes across different groups of firms are found to be indiscernible then only an approximate classification can be given, otherwise an exact grouping can be found. The procedures to develop decision rules can be found in the rough set theory literature (Grzymala-Busse 1992; Slowinski & Stefanowski 1992; Stefanowski & Vanderpooten 1994).

In predicting a new firm into business distress categories, the rough set theory could either successfully give a classification or the following situations may occur:

1. The new firm matches an approximate rule or several rules indicating different business distress classes.
2. The new firm did not match any of the existing rules.

In the first situation, the decision maker is informed of the strength of classification rules (measured by number of firms satisfying the condition attributes and belonging to the suggested business distress class). In the second case, the valued closeness relation (VCR) by Slowinski (1993) can be used. This involves applying indifference, strict difference and veto thresholds on particular attributes used in concordance and discordant tests. Firstly, the concordance procedure would find a set of attributes affirming firm z is close to different rules and assess their relative importance. Secondly, a discordance procedure which is to find attributes of the firm z not in agreement with the first procedure, to calculate the possible reduction of the level of concordance. This is the same type of tests that is also used in outranking relation procedures and the concepts are outlined in 4.4.8.3. In fact, the traditional methods of rough set theory has lead to other developments, most notably the development of a preference model such as ELETRE TRI which was also found to be effective in modelling business failures.

4.4.8.3 Outranking relations approach

The outranking relations approach started from the development of ELECTRE (Elimination Et Choix Traduisant la REalité) by Bernard Roy. An outranking relation is a binary relation where the decision maker assesses the outranking strength between firms a_i and a_f. The assessment of the strength is based on whether there are sufficient evidence through the coalition of criteria to determine that a_i is at least as good as a_f, with no other evidence to refute this statement. There are usually two stages in outranking relations approach: the first step is to rank the firms while the second step may involve further analysis on the outranking relations to obtain the best alternatives, or to sort them into categories, or to rank them from the most preferred to least preferred scale.

The ELECTRE TRI is probably the most frequently used method in this category for business distress modelling and a brief outline of the method is described here. ELECTRE TRI aims to sort a given set of firms $A=\{a_1, a_2, a_3, \ldots a_k\}$ into ordered categories (from worst to best) $c_1, c_2, \ldots c_q$. To define the categories, ELECTRE TRI uses reference profiles $r_1, r_2, \ldots r_{q-1}$. The reference profile r_i is used to evaluate the condition the firm must met for a given criteria as not to be classified into the c_i group. For an individual firm α, the concordance index $c_j(\alpha, r_i)$, (an increasing function) is calculated to affirm that for criterion j, firm α is as good as r_i as shown in expression (52).

$$\begin{array}{ll} \text{if } g_j(\alpha) \le g_j(r_i) - p_j(r_j) \text{ then} & c_j(\alpha, r_i) = 0 \\ \text{if } g_j(\alpha) - p_j(r_j) < g_j(\alpha) \le g_j(r_i) - q_j(r_j) \text{ then} & 0 < c_j(\alpha, r_i) \le 1 \\ \text{if } g_j(\alpha) > g_j(r_i) - q_j(r_i) \text{ then} & c_j(\alpha, r_i) = 1 \end{array} \quad (53)$$

In expression (52), the $g_j(.)$ is the criterion function for j-th criterion, the indifference threshold $q_j(r_j)$ defines the maximum accepted difference between $g_j(r_i)$ and $g_j(\alpha)$, to reach a conclusion that there α is indifferent to r_i. The preference threshold function, $p_j(r_j)$ is the maximum accepted difference between the $g_j(r_i)$ and $g_j(\alpha)$, to signal a difference in preference between α and r_i.

In most cases there will be more than one criterion, the global concordance index is then $\gamma(\alpha, r_i) = \frac{\sum_{j=1}^{n} w_j c_j(\alpha, r_i)}{\sum_{j=1}^{n} w_j}$ where w_j is the weight for criterion j. It is also possible to calculate the discordance index, (a decreasing function) which is shown in expression (53).

$$\begin{array}{ll} \text{if } g_j(\alpha) > g_j(r_i) - p_j(r_j) \text{ then} & d_j(\alpha, r_i) = 0 \\ \text{if } g_j(\alpha) - v_j(r_j) < g_j(\alpha) \le g_j(r_i) - p_j(r_j) \text{ then} & 0 < d_j(\alpha, r_i) \le 1 \\ \text{if } g_j(\alpha) \le g_j(r_i) - v_j(r_i) \text{ then} & d_j(\alpha, r_i) = 1 \end{array} \tag{54}$$

The only new variable in expression (53) is $v_j(r_j)$ which is the veto threshold for criterion j. This is the minimum accepted difference between $g_j(r_i)$ and $g_j(\alpha)$ and it represents a totally different preference between r_i and α according to criterion j. The decision maker usually needs to define the preference, veto and indifferent thresholds as well as the weights w_j from prior experience.

Once the concordance and discordance indices are found, they can then be used to construct a credibility index as shown in (54).

$$\sigma_s(\alpha, r_i) = \begin{cases} \gamma(\alpha, r_i) & \text{if } d(\alpha, r_i) \le \gamma(\alpha, r_i) \\ \gamma(\alpha, r_i) \prod_j \frac{1 - d(\alpha, r_i)}{\gamma(\alpha, r_i)} & \text{if } d(\alpha, r_i) > \gamma(\alpha, r_i) \end{cases} \tag{55}$$

The credibility index is use to give an overall outranking relation where $\sigma_s(\alpha, r_i) \ge \lambda \Leftrightarrow$ α outranks r_i, with λ being the cut off level between 0.5 to 1. The ranking relationships are then defined in (55).

$$\begin{array}{lcl} \alpha \text{ is indifferent to } r_i & \Leftrightarrow & \alpha \text{ outranks } r_i \text{ and } r_i \text{ outranks } \alpha \\ \alpha \text{ is preferred to } r & \Leftrightarrow & \alpha \text{ outranks } r_i \text{ and no } r_i \text{ outranks} \\ r_i \text{ is preferred to } \alpha & \Leftrightarrow & \text{no } \alpha \text{ outranks } r_i \text{ and } r_i \text{ outranks } \alpha \\ \alpha \text{ is incomparable to } r_i & \Leftrightarrow & \text{no } \alpha \text{ outranks } r_i \text{ and no } r_i \text{ outranks } \alpha \end{array} \tag{56}$$

In categorising the firms, firm α is compared to the worst profile r_1 and if α outranks r_i then it is compared to r_2 until:

1. α outranks r_i and (r_{i+1} outranks α or α is indifferent to r_{i+1}). Here, the firm is assigned to c_{i+1}

2. α outranks r_i and α is incomparable to $r_{i+1}, r_{i+2} \ldots r_{i+k}$ but r_{i+k+1}outranks α. Here, the optimist approach is to assign the firm into c_{i+k+1} with the pessimistic approach assign the firm to c_{i+1}.

The use of pessimistic and optimistic approach depends on whether a more a prudent policy is required or whether it is desirable to favour better rankings in the light of other qualities. The value λ not only affects the sorting but also the degree of optimism, where a lower value indicates a more optimistic approach. When $\lambda=1$, both approaches will give the same classification. The application of ELECTRE III method can be found in the works of Dimitras, Zopounidis and Hurson (1995).

3.10.9 Expert Systems

Expert systems are "computer programs that use specialised symbolic reasoning to solve difficult problems"(Luconi, Malone & Morton 1986). Symbolic reasoning is a set of rules resulting from logic and learning to produce a reasonable answer to a particular problem. Expert systems may be user defined (built by human expert themselves) or data driven (built by a computer learning algorithm such as inductive learning with some human expert interventions).

4.4.9.1 User-driven expert systems

The user driven expert systems are almost entirely developed by experts based on their prior experience and knowledge (Duchessi & Belardo 1987; Elmer & Borowski 1988). The success of these expert systems depends heavily on the ability of the expert to correctly identify financially distressed firms through a list of criteria. For example, a rule might be:

IF the earning trend is positive AND the current ratio trend is up

THEN there will be no loan default

ELSE there will be a loan default

The user defined expert system will then search for instances of mismatches and matches in the database and the rules can be accepted, rejected or modified based on the outcome. While some works developed the expert systems entirely through their experts, others like Srinivasan and Ruparel (1990) have combined other mathematical technique such as Analytical Hierarchical Process (AHP) (Saaty 1980; Saaty & Alexander 1989; Saaty 1992; Saaty 1992) to resolve conflicting conclusions that could be reached by the experts in assessing business failures.

Saaty developed a 1-9 "intensity of importance" scale where 1 means two factors contribute equally to the objective and 9 means the evidence of one favouring the other is of the highest possible validity. The odd numbers 3, 5 and 7 represent preference level from slightly, strongly to very strongly with the even numbers 2, 4 and 6 represent some kind of compromise between the scales.

To demonstrate the AHP, assume that the financial status of the firm can be classified into A (successful), B (possibly failure) and C (High chance of failure). In the case of criteria X (financial debt status), the relative importance of these classifications for a given firm (calculated based on some prior experience of the human expert) is shown in the following matrix.

	A	B	C
A	1	2	8
B	1/2	1	6
C	1/8	1/6	1

This matrix is a preference or relative importance matrix. For example, for criteria X, A is slightly more important than B so it was given a 2. On the contrary, B is less important than A and therefore it was given 1/2. Then, each element of the preference matrix is divided by its corresponding column sums to give the following:

	A	B	C
A	0.61538	0.63158	0.53333
B	0.30769	0.31579	0.40000
C	0.07692	0.05263	0.06667

The row sums of the above matrix will give 1.78, 1.02, 0.96 and scaling these numbers so that they sum to 1 give 0.593432299, 0.341160594 and 0.065407108 respectively. These numbers can be interpreted as the priority vector. The financial debt status has most relevance to classify the firm into A, followed by B and C. In addition to evaluating the priorities, it is necessary to check for consistency or the logic of these priorities. This is calculated by finding the consistency ratio: consistency index of data divided over consistency index by random chance. The data consistency index is defined as $(\lambda_{max}\text{-}n)/(n\text{-}1)$ (Saaty 1980) where n is the number of row of the preference matrix and λ_{max} can be found by the following procedure:

1. Find the vector (3×1) resulting from matrix multiplication between the preference matrix (3×3) and priority vector (3×1).
2. Scale the vector in step 1 by dividing through its corresponding priority vector
3. Approximate λ_{max} by the mean of vector in step 2.

The average consistency index of random chance for 3×3 preference matrix is 0.58 (Saaty & Vargas 1982 p. 25). In the above example the consistency ratio is 0.015. Generally, Saaty and Vargas (1982) argued that a consistency ratio above 0.1 means the pair wise judgements are random and another round of comparison may be required.

For another criteria Y (Poor Reputation), the expert may use a different grouping. For example:

	A	B	C	Priority Vector
A	1	1/4	1/6	0.08694791
B	4	1	1/3	0.27371757
C	6	3	1	0.63933452

Also, the importance of criteria X and Y can also be modelled:

	X	Y	Priority Vector
X	1	1/8	1/9
Y	8	1	8/9

This means the priority weighting for each of the criterion in assessing the financial health of the firm can now be found as follows:

	X	Y
A	0.593432299	0.08694791
B	0.341160594	0.27371757
C	0.065407108	0.63933452

The weight for category A is $1/9\cdot(0.593432299)+8/9\cdot(0.08694791)= 0.08694791$, similarly for categories B and C the weights are 0.27371757 and 0.63933452 respectively. In this example, this firm would be classified into group C. Usually, the preference matrix at each criterion are based on some characteristic of the firm developed by the human expert, so the classification of the firm is on a case by case basis. The AHP can be used as a stand alone technique in Srinivasan and Kim (1987) or embedded as part of the expert system in Srinivasan and Ruparel (1990).

While the use of AHP in financial distress expert systems can resolve conflicting criteria in determining the financial status of the firm and have been successfully implemented in some expert systems, it does require substantial negotiation between the expert and the computer analysts to build a workable system. Especially changing economic conditions do require changing the expert systems and it can be an expensive exercise. Sometimes the experts also cannot see what could happen under different scenarios as they may not have prior experience in those areas, this would limit the applicability of expert system. Also, the very subjective nature of these expert systems can make them unattractive to some organizations, as too much reliance is placed on the ability of the experts rather than from the evidence of the data. For these reasons, recent expert systems in financial distress modelling have moved towards a data driven approach.

4.4.9.2 Classification trees

Classification Trees or Recursive Partitioning Algorithm (RPA) are very much like an expert system without human interventions. It involves building many different nodes, with each node representing a rule until a classification decision is made. RPA usually have binary splits in the financial distress modelling literature (Frydman, Altman & Kao 1985) and an example is shown in Figure 2 with B represents bankruptcy and NB represents no bankruptcy.

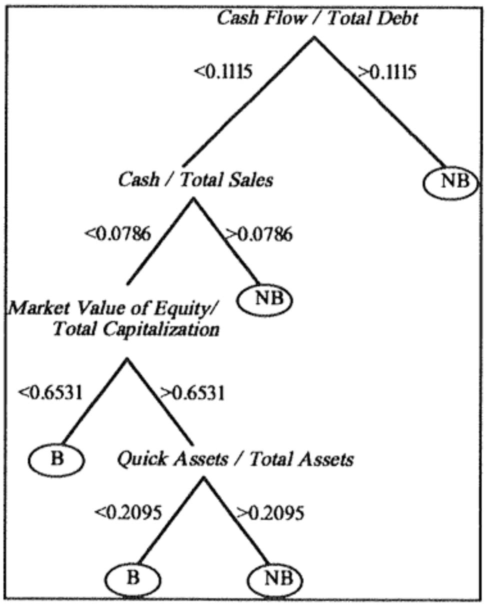

Figure 2 An RPA tree from Frydman, Altman and Kao (1985).

The criterion for the splitting rule is usually based on minimizing the misclassification risk in financial distress modelling. The risk of misclassification at each node t is defined as: $r(t) = (c_{12} + c_{21})p_1 p_2 \dfrac{n_2(t)n_1(t)}{p(t)m_1 m_2}$ (57)

m_1, m_2	Total number of firms in failed (1) and non failed (2) firms
$n_1(t)$, $n_2(t)$	Number of firms in each group at node t
c_{21}	Cost of misclassifying a firm in group 1 while it is in group 2
c_{12}	Cost of misclassifying a firm in group 2 while it is in group 1
p_1, p_2	Prior probabilities of a firm being a member of 1 or 2
$p(t)$	Probability of classifying a firm at node t

This measure is used rather than the usual entropy or gini index (Breiman, Friedman, Olshen & Stone 1984) since in financial distress modelling, the training data set often over present rare, failed cases rather than being proportional to the population. This is quite a natural restriction since in most cases, there are usually more financially sound firms than distressed ones.

The usual process of RPA can be thought of as purification process, designing a set of rules so the mass of "impure" firms can be purified into the correct categories. The idea is to split the data in such a way that the descendent subsets are purer than the present subset. A splitting rule would usually maximise the decrease in the impurities of the two sub samples compared to the impurity of the parent sample. To obtain the best splitting rule, each variable is examined and the best split by the variable is selected and this is repeated in the next step. The splitting stops when splitting does not lead to any decrease in impurity of the tree T_1. T_1 usually over fits the data so it is often desirable to prune this to obtain a smaller tree T_2 which has a less cross validated classification error.

The pruning process from a full tree T_1 to a smaller tree usually involves minimizing $R=R(T)+\gamma size(T)$: where R(T) is usually the number of misclassifications on the training set or deviance of the partition with size(T) represents the total number of nodes or leaves. The constant γ is usually set to be positive to penalize a bigger tree. A range of γ is then chosen for R to create a series of smaller trees t_i from T_1 using the full sample. Among these smaller trees t_i, T_2 is found by using a v-fold cross validation scheme, i.e. the t_i that has the lowest re-substitution classification error. The process is as follows. Firstly, the sample data can be divided to v parts with approximately equal sizes. Observations in v-1 groups are then used to generate a new ancillary tree by minimising the corresponding R and level of γ from t_i. Once the ancillary tree is generated, the observations in the single remaining group are classified by the ancillary tree, and the error in classification calculated. This procedure is repeated v times, each time with a different group left out, and the overall average of the cross validation errors is calculated for tree t_i. This process is then repeated for all t_i trees. The tree t_i with the lowest average cross validation is then chosen as T_2. While this is perhaps most frequently used pruning method in business failure modelling (Frydman, Altman & Kao 1985; Srinivasan & Kim 1987), there are also other ways of pruning such as such as shrinking (Gelfland, Ravishankar & Delp 1991) which can also be used.

4.4.9.3 Data-driven expert systems

The data driven expert systems uses machine learning rules such as inductive learning (Messier & Hansen 1988; Shaw & Gentry 1988) and genetic programming (Salcedo-Sanz, Fernández-Villacañas, Segovia-Vargas & Bousoño-Calzón 2005). Neural network is also a special machine learning mechanism which is covered in a separate section.

There are many types of inductive learning algorithms and in particular, AQ (Michalski 1983) and ID3 (Quilan 1983) have been used in the financial distress modelling literature. The AQ algorithm involves developing a set of IF… THEN rules based on the positive and negative examples. In a group of financially distressed firms, the negative examples will be firms that are not financially distressed. Generally, for each given group, AQ algorithm find a set of rules using the attributes of the firms (financial ratios, trend of sales etc) that cover all the positive examples and no negative examples using the least attributes. The generation of the rules can be done via several strategies, such as the dropping condition rule, adding alternative rule, closing interval rule (Michalski 1983 p. 106). A pseudo code of the AQ rule is provided below:

AQ Algorithm Pseudo Code

For each category:

rule-list = {}

P = set of +ve examples

N = set of -ve examples

repeat until P is empty {

  set the list of elements in P as p

  generate a list of maximally general rules which match P but none of N

  choose a single **best** rule b e.g. covers the most elements

    in P, is the shortest, simplest rule, etc

  rule-list := put b into the rule-list

  delete all p from P covered by b

}

Once P is empty: Output the rule-list

An example of AQ algorithm rules developed in Shaw and Gentry (1988) is shown below:

Avg inventory ≥ \$7000 and net worth ≤ \$47000 → Low risk

\$3700 ≤ net worth ≤ \$48000 and inventory > \$8000 → Moderate risk

Financial rating = H,A and total debt ≥ \$26000→ High risk

A more popular machine learning technique is the ID3 (Iterative Dichtomiser) algorithm by Quilan (1983) this uses measures of entropy and the minimal entropy rule to build a decision tree. The decision tree is made up conditions with discrete outcomes to either a new condition or a conclusive classification. The algorithm is detailed as follows:

For a firm that can be classified into n sets, $c_1, c_2, \ldots c_n$, let the probability of a firm being classified into class c_i as $p(c_i)$, then the entropy (measure of uncertainty) of classification Φ is:

$$\Phi = \sum_{i=1}^{n} -p(c_i)\log_2 p(c_i) \qquad (58)$$

At any stage of the split, the attribute chosen will be the one that yields the smallest entropy. The split is conducted iteratively until all firms are classified correctly.

The splitting of factor attributes such as “sales trend is a) increasing, b) decreasing or c) stable” is straight forward. However, sometimes in the case of an integer valued or even continuous attributes, it is necessary to determine the value of the split. For example, for attribute A there might be a range of ordered values from $a_1, a_2, \ldots a_k$. The split may take place at the minimum entropy with two subsets $\{a_1, a_2, \ldots a_j\}$ and $\{a_{j+1}, \ldots a_k\}$with the value of split taken at half way point of a_j and $a_{j+1.}$ This technique was used by Messier and Hansen (1988) in which they produced some quite successful decision trees in classifying the financial status of the hold out firms.

4.4.9.4 Genetic programming

Genetic programming creates a computer program by breeding a population of computer programs to solve problems. This type of programming is inspired by the biological genetic operations and Darwin’s “Survival of the fittest” concept. A general description of the Genetic programming which can also be found at www.genetic-programming.com or www.genetic.programming.org is as follows:

1. Generate an initial population of random computer programs composed of primitive functions (such as adding two numbers, assign 1 when both outputs are positive and 0 otherwise) and terminals (variables such as financial ratios) of the problem.
2. Repeat the following steps until the termination criterion is satisfied.
   a. Execute a program in the population and assign it a fitness value according to how well it solves the problem.
   b. Repeatedly do the following steps on the population until the termination criterion is satisfied.
      i. Reproduction: Copy an existing program.
      ii. Crossover: Create one or more new offspring programs by recombining randomly chosen parts from two selected programs.
      iii. Mutation: Create a new offspring program by randomly mutating a randomly chosen part of one selected program.
      iv. Architecture-altering operations: This involves changing the architecture of the program such as adding and deleting subroutines and arguments.
3. The single best computer program in the population produced during the run is considered the solution to the problem and evaluated on a test set. If the run is successful, the result maybe a solution.

From the above procedure it is necessary to determine the set of functions, terminals, fitness criteria, parameters and variables for controlling the run and the termination criterion. In the context of a business failure prediction problem, Salcedo-Sanz, Fernández-Villacañas, Segovia-Vargas and Bousoño-Calzón (2005) evolves a decision tree to distinguish between financially healthy and distressed firms using financial ratios (terminals of the problem).

In their work, genetic programming presents the solution using a decision tree. Each node of the tree has one of the following functions {+, -, OR-TH, AND-TH, NOT-TH, IFLTE}: “+”adds two numbers, “-“ substracts them, “OR-TH” returns 1 when at least one of the input are positive and -1 when the inputs are all negative. “AND-TH” returns 1 when both inputs are positive and -1 when both inputs are negative. “NOT-TH” returns 1 when its single input is negative and -1 when positive. Lastly, “IFLTE” is a macro with two input parameters (a and b) and two conditional nodes (c and d). When $a \geq b$, the macro evaluate c (maybe a function leading to another subtree) otherwise it is evaluate to d. In their decision tree, a positive value returns a healthy firm and a negative value signals a financially distressed firm.

As in conformity with biostatistics terms, a false positive FP is a firm classified as being positive from the tree but in fact has failed. The concepts of false positives (FP) false negative (FN) correct negatives (CN) and correct positives (CP) is then used to generate the fitness function. Two fitness functions in particular have been examined in Salcedo-Sanz, Fernández-Villacañas, Segovia-Vargas and Bousoño-Calzón (2005).

$$f_1 = 1 - \alpha_1\left(\frac{CN}{N_{neg}}\right) - \beta_1\left(\frac{CN}{CN+FN}\right) \tag{59}$$

$$f_2 = 1 - \frac{\alpha_1}{2}\left(\frac{CN}{N_{neg}} + \frac{CP}{N_{pos}}\right) - \frac{\beta_1}{2}\left(\frac{CN}{CN+FN} + \frac{CP}{CP+FN}\right) \tag{60}$$

$$f_3 = 1 - \left(\frac{CP+CN}{N_{pos} + N_{negative}}\right) \tag{61}$$

The first fitness function, expression (58), comprises of recall $\left(\frac{CN}{N_{neg}}\right)$ and precision $\left(\frac{CN}{CN+FN}\right)$ of the failed firms. In expression (59), the coefficients α_1 and β_1 can take values in the range [0,1]. The term $\left(\frac{CN}{N_{neg}} + \frac{CP}{N_{pos}}\right)$ is the recall of failed and successful firms and the term $\left(\frac{CN}{CN+FN} + \frac{CP}{CP+FN}\right)$ is the precision. The coefficients α_1 and β_1 can therefore act as weights in which the precision or the recall can be accentuated or suppressed. Salcedo-Sanz, Fernández-Villacañas, Segovia-Vargas and Bousoño-Calzón (2005) used α_1=0.8 and β_1=0.2 in their study. The recall measures the ability of the classification tree to extract the cases of interest where as the precision measures the quality of extracted cases. As it is desirable to maximise the recall and precision, the aim is to minimize (58) or (59). Similarly, in expression (60), the goal is to maximise number of CP+CN or equivalently minimize (60). The error in classification is calculated using v-fold cross validation as was done in the case of classification trees. The paper also compared the classification error of genetic programming with SVM[12] (support vector machine) (Burges 1998; Scholkopf & Smola 2002) and found genetic programming usually has a superior performance.

[12] SVM separate data using kernel functions to separate m dimensional data so the classification boundary between different groups are maximised.

4.4.9.5 Neural networks

Neural network attempts to find patterns of business failures by emulating the biological functions of the human brain. It requires a set of training data to train the computer to identify patterns before developing a stable model that can be used to classify firms into different financial distress categories. The key features of a generic neural network model are described in Figure 3. Figure 4 outlines the process of neuron activation to produce the output and the training of the neural network is shown in Figure 5.

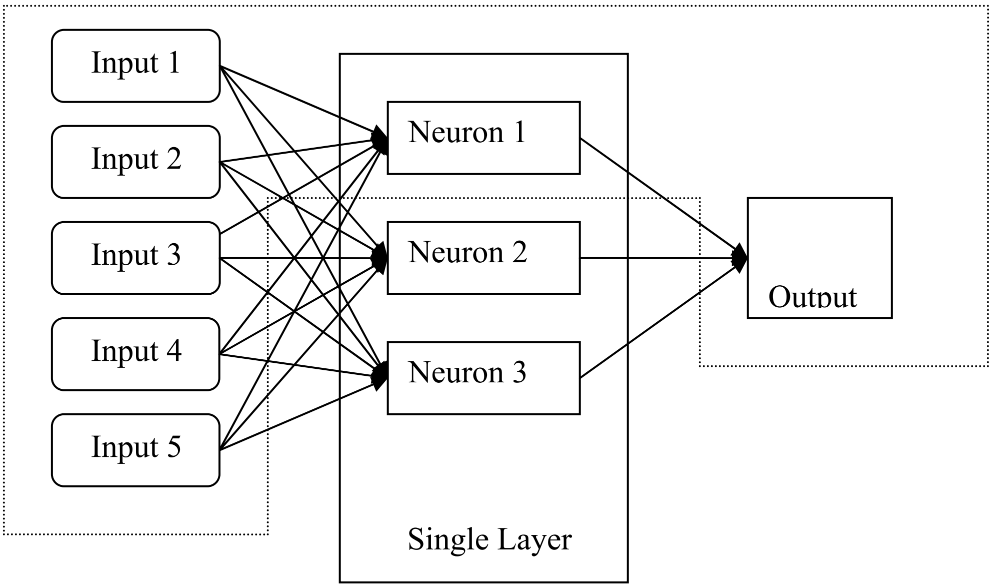

Figure 3 This diagram describes a generic, typical single layer neural network often used in predicting business failures. The inputs are usually financial ratios that are fed into the neurons, which react and process the information to produce the output, classifying firms into financially distressed and non-financially distressed categories. In a typical process there will only be a single winning neuron which will be activated to produce the output. In the context of this example, the output will be produced by either neuron 1, 2, or 3. The dashed region is explained in greater detail in Figure 4 to give a generic picture of the neuron activation process.

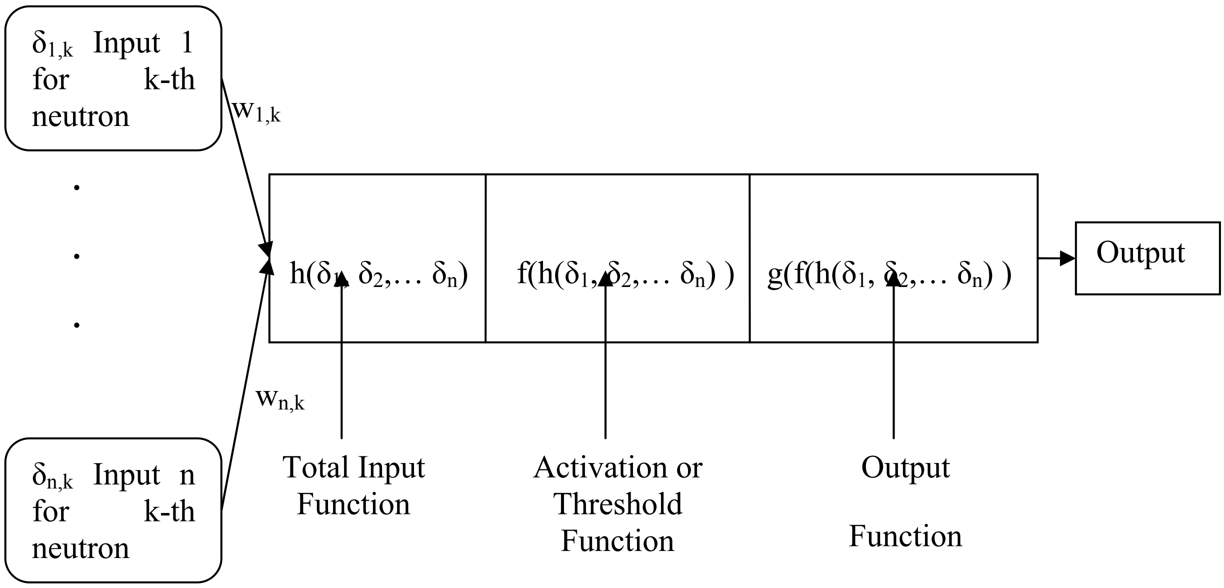

Figure 4 This figure shows a typical single neuron reaction process. An input function summarises the input information through a set of weights denoted by w which can either be random or predefined. These input functions are then activated; this is where computations take place to determine the output of the neural network. Usually the output function is an identity function, i.e. the output of the neuron is same as the activation level. The input, threshold and output functions can vary across different neural network models.

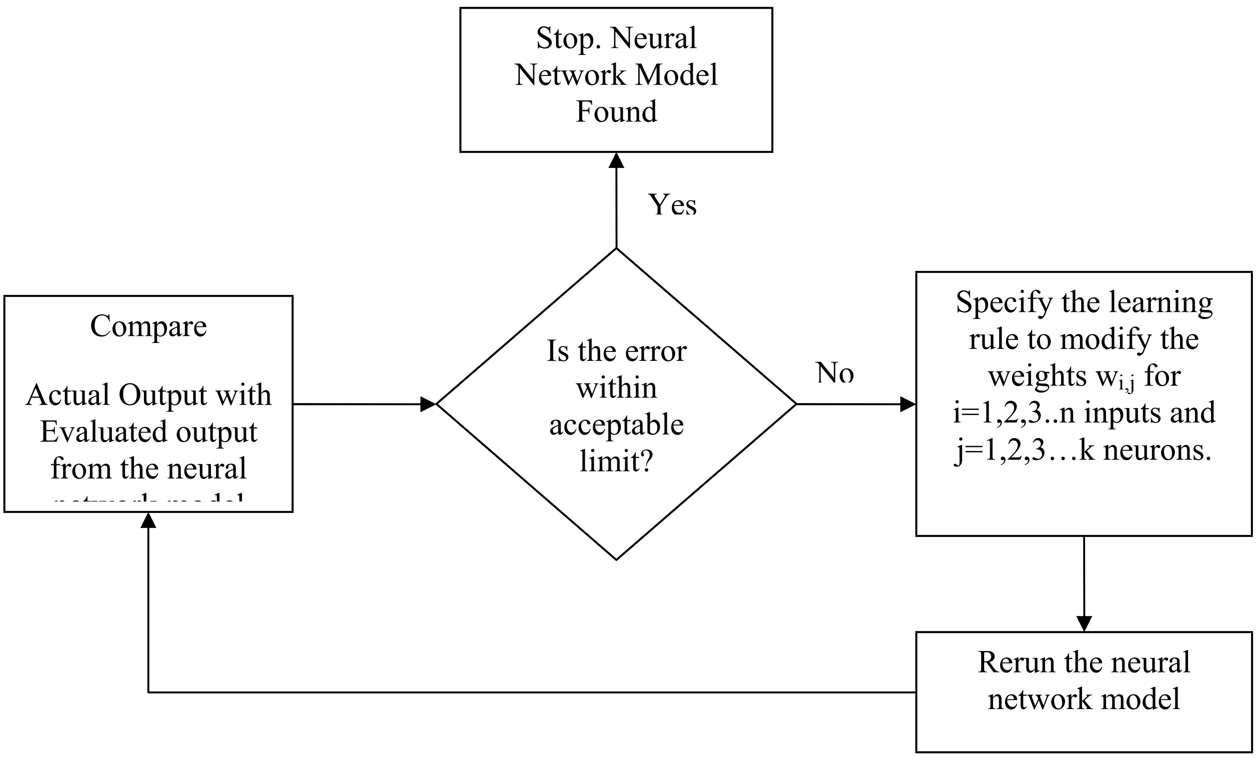

Figure 5 This shows the typical training process of neural network, usually a preset criteria to judge the model performance is set and the input weights assigned to each of the neuron are altered until the errors fall within a reasonable limit. Many variations of different learning rules and model performance criteria exist.

The process described in Figure 3 is a very commonly used neural network model in financial distress modelling (Shah & Murtaza 2000; Charitou, Neophytou & Charalambous 2004). The inputs, usually financial ratios, get feed into the model with different weights and in turn will stimulate different neurons. The winning neuron is usually the one that matches closest to the profile of the input variables, so that the most plausible neuron can be activated to give the output.

Figure 4 shows the process for a single neuron with n inputs or variables. The same process applies to other neurons so there is no loss of generality. In financial distress modelling, the total input function is usually affine, defined by $\sum_{i=1}^{n} w_{ij}\delta_{ij} - a$, with the activation function being the Heaviside or sign function (functions with binary outcomes). This type of model is known as the "threshold device" neural network model. There are also other input functions such as Boolean, linear weighted sums and other activation functions such as sigmoid functions $f(z)=1/(1+\exp(-\beta z))$ (with β being the slope parameter) which can also be used to build different neural network models.

To train the model, a variety of techniques can be used. In most financial distress neural networks, a technique known as back propagation is used. In the training stage, a set of input variables matched with an output is provided. An example of the training set data is provided to the neural network in each trial. The output is calculated from the weights of the current network and these weights may be either randomly generated or defined by the user. Now the total error (sum of squared errors between the neural network output and desired output) is calculated. This error is subsequently back propagated in the neural network and the weights are adjusted. Once the error is below a specified amount, the neural network model is found.

The training phase of the neural network varies between different studies. For example, Shah and Murtaza (2000) adjust weights $w_{ij}(t)= w_{ij}(t-1)+\alpha(x_i- w_{ij}(t-1))$ for the winner neuron that gives output at the correct class. Otherwise a weight of $w_{ij}(t)= w_{ij}(t-1)-\alpha(x_i- w_{ij}(t-1))$ is assigned. Usually, α, the learning rate, is kept to around 0.25. Other studies such as Charitou, Neophytou et al (2004) used conjugate gradient algorithm to minimize the quadratic error between observed and predicted outputs.

Financial distress models using neural networks are fairly popular, and they do appear to give convincing performances. Altman, Marco and Varetto (1994) shows that the accuracy of neural networks are comparable to discriminant analysis models. The acceptance of neural network in the statistics community however is relatively slow. Partly this due to the reason it is difficult to see exactly the set of rules that the computer used to make the decision and this method requires more computational effort than most of the traditional statistical methods.

4. CONCLUSION

In the literature review on business failure in the construction industry a list of major factors to predict such events have been identified. Usually, the underlying model is Altman's *Z*-score approach that can be viewed as a 'standard' against which other methods were assessed. A critique to most of these articles is two-fold: (a) failure is usually considered to be a point process rather than a consequence of a failure process that leads over time to a possible business failure, and (b) the literature review in Section 4 gives a concise overview on the variety of approaches that have been suggested and applied to industry sectors other than the construction and building industry. Both of these comments provide a rich ground for further research.

5. REFERENCES

5.1 References for Section 3

Abidali A. F., and Harris F. (1995), A Methodology for Predicting Company Failure in the Construction Industry, *Construction Management and Economics*, Vol. 13, 189-196.

Altman E. I. (1968), Financial Ratios, Discriminant Analysis, and the Prediction of Corporate Bankruptcy, *Journal of Finance*, Vol. 23 (4), 589-610

Australian Bureau of Statistics (2006), http://www.abs.gov.au/

Clift D. (2006), Financial Requirements for Licensing: Insolvency Research, *BSA working paper,* Queensland Building Services Authority, Australia

Dun and Bradstreet Corporation (2006), http://www.dnb.com/

Freear J. (1980), Business Failure in *The Management of Business Finance*, London: Pitman Publishing

Hall G. (1994), Factors Distinguishing Survivors from Failures Amongst Small Firms in the UK Construction Sector, *Journal of Management Studies*, Vol. 31 (5), 737-760

Jaselskis E. J. and Russell J. S. (1992), Risk Analysis Approach to Selection of Contractor Evaluation Method, *Journal of Construction Engineering and Management*, Vol. 188 (4), 814-821

Kale S. and Arditi D. (1999), Age-Dependent Business Failure in the US Construction Industry, *Construction Management and Economics*, Vol. 17, 493-503

Kangari R. (1987), Business Failure in Construction Industry, *Journal of Construction Engineering and Management*, Vol. 114 (2), 172-190

Koksal A. and Arditi D. (2004), An Input/Output Model for Business Failures in the Construction Industry, *Journal of Construction Research*, Vol. 5 (1), 1-16

Langford D., Iyagba R. and Komba D. M. (1993), Prediction of Solvency in Construction Companies, *Construction Management and Economics*, Vol. 11, 317-325

Lynch F. (2003), Growth-Survival Tradeoffs of Queensland Licensed Homebuilders 1986-1996: An Event History Study, *PhD Thesis, The University of Queensland, Brisbane, Australia*

March J. (1991), Exploration and Exploitation in Organisational Learning, *Organization Science*, Vol. 2 (1), 1-24

Mason R. J. and Harris F. C. (1979), Predicting Company Failure in the Construction Industry, *Proceedings Institution of Civil Engineers*, Vol. 66, 301-307

Russell J. S. (1991), Contractor Failure: Analysis, *Journal of Performance of Constructed Facilities*, Vol. 5 (3), 163-180

Russell J. S. and Jaselskis E. J. (1992), Predicting Construction Contractor Failure Prior to Contract Award, *Journal of Construction Engineering and Management*, Vol. 118 (4), 791-811

Russell J. S. and Zhai H. (1996), Predicting Contractor Failure Using Stochastic Dynamics of Economic and Financial Variables, *Journal of Construction Engineering and Management*, Vol. 122 (2), 183-191

5.2 References for Section 4

Altman E. (1984), The Success of Business Failure Prediction Models, *Journal of Banking and Finance*, Vol. 8, 171-198

Altman E. (2000), Predicting Financial Distress of Companies: Revisiting the Z-Score and Zeta® Models, http://pages.stern.nyu.edu/~ealtman/Zscores.pdf, 2005

Altman E. (2002) Corporate Distress Prediction Models in a Turbulent Economic and Basel II Environment, http://pages.stern.nyu.edu/~ealtman/Corp-Distress.pdf, 2005

Altman E., Haldeman R. and Narayana P. (1977), Zeta™* Analysis, *Journal of Banking and Finance,* Vol. 1, 29-54

Altman E., Marco G. and Varetto F. (1994), Corporate Distress Diagnosis: Comparisons Using Discriminant Analysis and Neural Networks, *Journal of Banking and Finance*, Vol. 18, 505-529

Bajgier S. M. and Hill A. V. (1982), An Experimental Comparison of Statistical and Linear Programming Approaches to the Discriminant Problem, *Decision Sciences,* Vol. 13, 604-618

Balcaen S. and Ooghe H. (2005), 35 Years of Studies on Business Failure: an Overview of the Classic Statistical Methodologies and Their Related Problems, *British Accounting Review,* Vol. 38(1), 1-31

Banks W. and Prakash A. L. (1991), An Efficient Optimal Solution Algorithm for the Classification Problem, *Decision Sciences*, Vol. 22(5), 1008-1023

Beaver W. H. (1966), Financial Ratios as Predictors of Failure, Empirical Research in Accounting: Selected Studies (supplement), *Journal of Accounting Research*, Vol. 5, 179-199

Boyd J. and Melman R. (1980), The Effect of Fuel Economy Standards on the US Automobile Market: an Hedonic Demand Analysis, *Transportation Research*, Vol. 14A, 367-378

Breiman L., Friedman J. H., Olshen R. A. and Stone C. J. (1984), *Classification and Regression Trees*, Monterey, Wadsworth and Brooks/Cole

Burges J. C. (1998), A Tutorial on Support Vector Learning Machines for Pattern Recognition, *Data Mining and Knowledge Discovery,* Vol. 2(2), 121-167

Cardell N. and Dunbar F. (1980), Measuring the Societal Impacts of Automobile Downsizing, *Transportation Research,* Vol. 14A, 423-434

Charalambous C., Charitou A. and Kaourou F. (2000), Comparative Analysis of Artificial Neural Network Models: Application in Bankruptcy Prediction, *Annals of Operations Research,* Vol. 99, 403-425

Charitou A., Neophytou E. and Charalambous C. (2004), Predicting Corporate Failure: Empirical Evidence for the UK, *European Accounting Review*, Vol. 13(3), 465-497

Coen P. G., Gomme E. D. and Kendall M. G. (1969), Lagged Relationships in Economic Forecasting, *Journal of the Royal Statistical Society Series A,* Vol. 132, 133-163

Courtis J. K. (1978), Modelling a Financial Rartios Categoric Framework, *Journal of Business Finance and Accounting,* Vol. 5(4), 371-386

Cox D. R. (1972), Regression Models and Life Tables (with Discussion), *Journal of Royal Statistical Society, Series B,* Vol. 34, 187-220

Cox D. R. and Oakes D. (1984), *Analysis of Survival Data*, London, Chapman and Hall

Dimitras A., Slowinski R., Susmaga R. and Zopounidis C. (1999), Business Failure Prediction Using Rough Sets, *European Journal of Operation Research,* Vol. 114, 263-280

Dimitras A., Zanakis S. H. and Zopounidis C. (1996), A Survey of Business Failures with an Emphasis on Prediction Methods and Industrial Applications, *European Journal of Operation Research*, Vol. 90, 487-513

Dimitras A., Zopounidis C. and Hurson C. (1995), A Multicriteria Decision Aid Method for the Assessment of Business Failure Risk, *Foundations of Computing and Decision Sciences*, Vol. 20(2), 99-112

Doumpos M. and Zopounidis C. (1999), A Multicriteria Discrimination Method for the Prediction of Financial Distress: the Case of Greece, *Multinational Finance Journal*, Vol. 3(2), 71-101

Duchessi P. and Belardo S. (1987), Lending Analysis Support System (LASS): an Application of a Knowledge Based System to Support Commercial Loan Analysis, *IEEE Transaction on Systems, Man and Cybernetics,* SMC-17(4), 608-616

Duchessi P., Shawky H. and Seagle J. P. (1988), A Knowledge Engineered System for Commercial Loan Decisions, *Financial Management,* Vol. 17(3), 57-65

Edmister R. A. (1972), An Empirical Test of Financial Ratio Analysis for Small Business Failure Prediction, *Journal of Financial and Quantitative Analysis,* Vol. 7, 1477-1493

Elmer P. J. and Borowski D. M. (1988), An Expert System Approach to Financial Analysis: the Case of SandL Bankruptcy, *Financial Management,* Vol. 17(3), 66-76

Epanechnikov V. A. (1969), Nonparametric Estimation of a Multivariate Probability Density, *Theory of Probability and Its Applications,* Vol. 14, 153-158

Fisher R. (1936), The Utilization of Multiple Measurement in Taxonomic Problems, *Annals of Eugenics,* Vol. 7, 179-188

Freed N. and Glover F. (1981), A Linear Programming Approach to the Discriminant problem, *Decision Sciences,* Vol. 12, 68-74

Freed N. and Glover F. (1981), Simple but Powerful Goal Programming Models for Discriminant Problems, *European Journal of Operation Research*, Vol. 7, 44-60

Freed N. and Glover F. (1986), Evaluating Alternative Linear Programming Models to Solve the Two-Group Discriminant Problem, *Decision Sciences*, Vol. 17, 151-162

Frydman H., Altman E. and Kao D.-L. (1985), Introducing Recursive Partitioning for Financial Classification: the Case of Financial Distress, *The Journal of Finance*, Vol. XL(1), 269-291

Ganesalinggam S. and Kumar K. (2001), Detection of Financial Distress via Multivariate Statistical Analysis, *Managerial Finance*, Vol. 27(4), 45-55

Gelfland B., Ravishankar C. S. and Delp E. J. (1991), An Interactive Growing and Pruning Algorithm for Classification Tree Design, *IEEE Transaction on Pattern Analysis and Machine Intelligence,* Vol. 13, 163-174

Ghosh A. K. and Chaudhuri P. (2005), On Data Depth and Distribution-Free Discriminant Analysis Using Separating Surfaces, *Bernoulli*, Vol. 11(1), 1-27

Glover F. (1990), Improved Linear Programming Models for Discriminant Analysis, *Decision Sciences*, Vol. 21(4), 771-785

Glover F., Keene S. and Duea B. (1988), A New Class of Models for the Discriminant Problem, *Decision Sciences*, Vol. 19(2), 269-280

Grablowsky B. J. and Talley W. K. (1981), Probit and Discriminant Factors for Classifying Credit Applicants: a Comparison, *Journal of Economics and Business*, Vol. 33, 254-261

Grzymala-Busse J. (1992), LERS - a System for Learning from Examples Based on Rough Sets, *Intelligent Decision Support, Handbook of Applications and Advances of the Rough Set Theory*, R. Slowinski. Dordrecht, Kluwer Academic Publishers, 3-18

Gutpa Y. P., Rao R. P. and Bagvhi P. K. (1990), Linear Goal Programming as an Alternative to Multivariate Discriminant Analysis: a Note, *Journal of Business Finance and Accounting*, Vol. 17(4), 593-598

Hamer M. M. (1983), Failure Prediction: Sensitivity of Classification Accuracy to Alternative Statistical Models and Variable Sets, *Journal of Accounting and Public Policy*, Vol. 2, 289-307

Hendry D. F. (1995), *Dynamic Econometrics*, Oxford, Oxford University Press

Hendry D. F. (2000), Epilogue: the Success of General-to-Specific Model Selection, *Econometrics: Alchemy or Science?,* D. F. Hendry, Oxford, Oxford University Press

Hendry D. F. and Doornik J. A. (2001), *Econometric Modelling using PcGive 10: Volume 1*, London, Timberlake Consultants Press

Hoover K. D. and Perez S. J. (1999), Data Mining Reconsidered: Encompassing and the General to Specific Approach to Specification Search, *Econometrics Journal*, Vol. 2, 167-191

Hotelling H. (1933), Analysis of a Complex Statistical Variables into Principal Components, *Journal of Educational Psychology*, Vol. 24, 471-441,498-520

Izan H. Y. (1984), Corporate Distress in Australia, *Journal of Banking and Finance*, Vol. 8, 303-320

Joachimsthaler E. A. and Stam A. (1988), Four Approaches to the Classification Problem in Discriminant Analysis: an Experimental Study, *Data Mining and Knowledge Discovery*, Vol. 19(2), 322-333

Johnson R. A. and Wichern D. W. (1982), *Applied Multivariate Statistical Analysis*, New Jersey, Prentice-Hall, Englewood Cliffs

Jones S. and Hensher D. (2004), Predicting Firm Financial Distress: a Mixed logit Model, *Accounting Review,* Vol. 79(4), 1011-1038

Kahya E., Ouandlous A. S. and Theodossiou P. (2001), Serial Correlation, Non-Stationarity, and Dynamic Performance of Business Failures Prediction Models, *Managerial Finance*, Vol. 27(8), 1-15

Kahya E. and Theodossiou P. (1999), Predicting Corporate Financial Distress: a Time-Series CUSUM Methodology, *Review of Quantitative Finance and Accounting*, Vol. 13(4), 323-345

Keeney R. L. and Raiffa H. (1993), *Decisions with Multiple Objectives: Preferences and Tradeoffs*, Cambridge, Cambridge University Press

Koehler G. J. and Erenguc S. S. (1990), Minimizing Misclassifications in Linear Discriminant Analysis, *Decision Sciences*, Vol. 21(1), 63-85

Kruskal J. B. (1964), Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis, *Psychometrika*, Vol. 29, 1-27

Kruskal J. B. (1964), Non-metric Multidimensional Scaling: a Numerical Method, *Psychometrika*, Vol. 29, 115-129

Laitinen E. (1991), Financial Ratios and Different Failure Processes, *Journal of Business Finance and Accounting*, Vol. 18(5), 649-673

Laitinen E. (1992), Prediction of Failure of a Newly Founded Firm, *Journal of Business Venturing*, Vol. 7, 323-340

Lam K. F., Choo E. U. and Moy J. W. (1996), Minimizing Deviations from the Group Mean: a New Linear Programming Approach for the Two Group Classification Problem, *European Journal of Operation Research*, Vol. 88, 358-367

Lam K. F. and Moy J. W. (2003), A Piecewise Linear Programming Approach to the Two Group Discrimination Problem - an Adaptation to Fisher's Linear Discriminant Function Model, *European Journal of Operation Research*, Vol. 145, 471-481

Lane W. R., Looney S. W. and Wansley W. (1986), An Application of the Cox Proportional Hazard Model to Bank Failure, *Journal of Banking and Finance*, Vol. 10, 511-531

Lee K. C., Han I. and Kwon Y. (1996), Hybrid Neural Network Models for Bankruptcy Predictions, *Decision Support Systems*, 63-72

Leeuwen P. H. (1985), The Prediction of Business Failure at Radobank, *Journal of Bank Research,* (Summer), 91-98

Lev B. (1971), Financial Failure Information Decomposition Measure, *Accounting in Perspective Contribution to Accounting Thoughts by Other Disciplines*, R. R. Sterling and W. F. Bentz, Southern-West Publishing Co., 102-111

Luconi F., Malone T. and Morton S. (1986), Expert Systems: the Next Challenge for Managers, *Sloan Management Review*, 3-13

Luoma M. and Laitinen E. (1991), Survival Analysis as a Tool for Company Failure Prediction, *Omega*, Vol. 19(6), 673-678

Lussier R. (2005), A Success Versus Failure Prediction Model for the Real Estate Industry, *Mid-American Journal of Business*, Vol. 20(1), 47-53

Mahmood M. A. and Lawrence E. C. (1987), A Performance Analysis of Parametric and Non parametric Discriminant Approaches to Business Decision Making, *Decision Sciences*, Vol. 18, 308-326

Markowski E. P. and Markowski C. A. (1985), Some Difficulties and Improvements in Applying Linear Programming Formulations to the Discriminant Problem, *Decision Sciences*, Vol. 16, 237-247

Martin D. (1977), Early Warning of Bank Failure, *Journal of Banking and Finance*, Vol. 1, 249-276

McFadden D. and Train K. (2000), Mixed MNL Models for Discrete Response, *Journal of Applied Econometrics*, Vol. 15, 447-470

Messier W. F. and Hansen J. V. (1988), Including Rules for Expert System Development: an Example Using Default Bankruptcy Data, *Management Science*, Vol. 34(12), 1403-1415

Micha B. (1984), Analysis of Business Failures in France, *Journal of Banking and Finance*, Vol. 8, 281-291

Michalski R. S. (1983), A Theory and Methodology of Inductive Learning, *Machine Learning: an Artificial Intelligence Approach*, R. S. Michalskis, J. G. Carbonell and T. M. Mitchell. Palo Alto, Tioga Publishing Company

Miller A. J. (1984), Selection of Subsets of Regression Variables (with Discussion), *Journal Royal Statistical Society, Series A*, Vol. 147, 389-425

Molinero C. M. and Ezzamel M. (1991), Multidimensional Scaling Applied to Corporate Failure, *Omega International Journal of Management Science*, Vol. 19(4), 259-274

Neophytou E. and Molinero C. M. (2004), Predicting Corporate Failure in the UK: a Multidimensional Scaling Approach, *Journal of Banking and Finance*, Vol. 31(5and6), 677-710

Oh M. S. and Raftery A. (2001), Bayesian Multidimensional Scaling and Choice of Dimension, *Journal of American Statistical Association*, Vol. 96, 1031-1044

Ooghe H. and Verbaere E. (1982), Determination van Faling: Verklaring en Predicitie, *University of Ghent, Department of Business and Finance Working paper*

Pany K. (1979), Information Analysis of Several Large Failed Banks, *Journal of Bank Research,* (Autumn), 145-151

Patuwo E., Hu M. and Hung M. (1993), Two-Group Classification Using Neural Networks, *Decision Sciences*, Vol. 24(4), 825-845

Pawlak Z. (1982), Rough sets, *International Journal of Information and Computer Sciences*, Vol. 11, 341-356

Pearson K. (1901), On Lines and Planes of Closest Fit to a System of Points in Space, *Philosophical Magazine*, Vol. 2, 557-572

Quilan J. R. (1983), Learning Efficient Classification Procedures and Their Application to Chess End Games, *Machine Learning: an Artificial Intelligence Approach*, R. S. Michalskis, J. G. Carbonell and T. M. Mitchell. Palo Alto, Tioga Publishing Company

Rubin P. A. (1990), A Comparison of Linear Programming and Parametric Approaches to the Two-Group Discriminant Problem, *Decision Sciences*, Vol. 21(2), 373-386

Saaty T. L. (1980), *The Analytical Hierarchy Process*, New York, McGraw-Hill

Saaty T. L. (1992), *Decision Making for Leaders: the Analytic Hierarchy Process For Decisions in a Complex World,* Pittsburgh, RWS Publications

Saaty T. L. (1992), *Multicriteria Decision making: the Analytic Hierarchy Process*, Pittsburgh, RWS Publications

Saaty T. L. and Alexander J. M. (1989), *Conflict Resolution: the Analytic Hierarchy Approach*, New York, Praeger

Saaty T. L. and Vargas L. G. (1982), *The Logic of Priorities: Applications in Business, Energy, Health, and Transportation*, Boston/The Hague/London, Kluwer-Nijhoff Publishing

Salcedo-Sanz S., Fernández-Villacañas J.-L., Segovia-Vargas M. J. and Bousoño-Calzón C. (2005), Genetic Programming For the Prediction of Insolvency in Non-Life Insurance Companies, *Computers and Operations Research*, Vol. 32, 749-765

Schmidt R. (1984), Early Warning of Debt Rescheduling, *Journal of Banking and Finance*, Vol. 8, 357-370

Scholkopf B. and Smola A. (2002), *Learning with Kernels*, Cambridge, MA, MIT Press

Serrango-Cinca C. (1996), Self Organizing Neural Networks for Financial Diagnosis, *Decision Support Systems*, 227-238

Shah J. and Murtaza M. (2000), A Neural Network Based Clustering Procedure for Bankruptcy Prediction, *American Business Review*, Vol. 18(2), 80-86

Shaw M. J. and Gentry J. A. (1988), Using an Expert System with Inductive Learning to Evaluate Business Loans, *Financial Management*, Vol. 17(3), 45-56

Slowinski R. (1993), Rough Set Learning of Preferential Attitude in Multi-Criteria Decision Making, *Methodologies for Intelligent Systems, Lecture Notes in Artificial Intelligence*, J. Komorowski and Z. W. Ras. Berlin, Springer

Slowinski R. and Stefanowski J. (1992), "RoughDAS" and "RoughClass" Software Implementations of the Rough Set Approach, *Intelligent Decision Support, Handbook of Applications and Advances of the Rough Set Theory*, R. Slowinski. Dordbrecht, Kluwer Academic Publishers, 445-456

Slowinski R. and Zopounidis C. (1995), Application of the Rough Set Approach to Evaluation of Bankruptcy Risk, *International Journal of Intelligent Systems in Accounting, Finance and Management*

Srinivasan V. and Kim Y. H. (1987), Credit Scoring: a Comparative Analysis of Classification Procedures, *The Journal of Finance*, Vol. XLII(3), 665-681

Srinivasan V. and Ruparel B. (1990), CGX: An Expert Support System for Credit Granting, *European Journal of Operation Research*, Vol. 45, 293-308

Stefanowski J. and Vanderpooten D. (1994), A General Two-Stage Approach to Inducing Rules from Examples, *Rough Sets, Fuzzy Sets and Knowledge Discovery*, W. Ziarko, London, Springer-Verlag, 317-325

Stein J. H. and Ziegler W. (1984), The Prognosis and Surveillance of Risks from Commercial Credit Borrowers, *Journal of Banking and Finance*, Vol. 8, 249-268

Taffler R. (1984), Empirical Models for the Monitoring of UK Corporations, *Journal of Banking and Finance*, Vol. 8, 199-227

Takahashi K. and Kurokawa Y. (1948), Corporate Bankruptcy Prediction in Japan, *Journal of Banking and Finance*, Vol. 8, 229-247

Tam K. and Kiang M. (1992), Managerial Applications for Neural Networks: the Case of Bank Failure Predictions, *Management Science*, 926-947

Theodossiou P. (1993), Predicting Shifts in the Mean of a Multivariate Time Series Process: an Application in Predicting Business Failures, *Journal of American Statistical Association*, Vol. 88(422), 441-449

Tirapat S. and Nittayagasetwat A. (1999), An Investigation of Thai Listed Firms' Financial Distress Using Macro and Micro Variables, *Multinational Finance Journal*, Vol. 3(2), 103-125

Torgerson W. S. (1952), Multidimensional Scaling, Theory and Method, *Psychometrika*, Vol. 17, 401-419

Wood D. and Piesse J. (1988), The Information Value of Failure Predictions in Credit Assessment, *Journal of Banking and Finance*, Vol. 12, 275-292

Xiao B. (1993), Necessary and Sufficient Conditions of Unacceptable Solutions in LP Discriminant Analysis, *Decision Sciences*, Vol. 24(3), 699-712

Zopounidis C. and Dimitras A. (1998), *Multicriteria Decision Aid Methods for the Prediction of Business Failure*, Boston/Dordrecht/London, Kluwer Academic Publishers

Zopounidis C. and Doumpos M. (1999), Business Failure Prediction Using the UTADIS Multicriteria Analysis Method, *Journal of Operational Research Society*, Vol. 50(11), 1138-1148

Zopounidis C. and Doumpos M. (2000), Building Additive Utilities for Multi-Group Hierarchical Discrimination, the MHDIS Method, *Optimization Methods and Software*, Vol. 14(3), 219-240