Predicting Financial Distress of JSE-Listed Companies using Bayesian Networks

Ziyad Cassim

A dissertation submitted to the Department of Actuarial Science, Faculty of Commerce, at the University of Cape Town, in partial fulfilment of the requirements for the degree of Master of Philosophy.

May 31, 2015

Master of Philosophy specializing in Mathematical Finance,
University of Cape Town,
Cape Town.

Declaration

I declare that this dissertation is my own, unaided work. It is being submitted for the Degree of Master of Philosophy in the University of Cape Town. It has not been submitted before for any degree or examination in any other University.

Ziyad Cassim

May 31, 2015

Abstract

This study aims to test the suitability of using Bayesian probabilistic models to predict bankruptcy of JSE-listed companies. A sample of 132 companies is considered with fourteen years of financial statement information and macroeconomic indicators used as predictor variables. Various permutations of Bayesian models are tested relating to different learning algorithms, intervals of discretisation and scoring metrics. In contrast to previous research, we explore a variety of evaluation measures and it is found that predictive accuracy for bankrupt firms does not exceed 70% in any model augmentation. On comparison to other popular models such as the Altman Z-score and the logit model, it is found that Bayesian networks produce marginally better predictive accuracy. Furthermore, a comparison to previous research on the same subject is carried and reasons for significantly different results are considered. Finally, the reasons for low predictive accuracies is considered with issues relating specifically to South Africa being discussed.

Acknowledgements

The completion of this dissertation brings to a culmination one of the most significant academic endeavours of my career. It would not have been possible without the support and guidance of the following people, to whom I am extremely grateful.

- To Associate Professor Ryan Kruger, for introducing me to this interesting area of research and for his patience and guidance throughout the process of this dissertation.
- To Alex Backwell, for the tireless assistance, support, and encouragement shown to me throughout this process.
- To the Mathematical Finance department at UCT, whose lecturers and tutors, not only lead me through an academically enlightening year but also assisted me and guided me through this task.
- Finally, and most importantly to my parents, whose lifelong encouragement and support have brought me to where I am today.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In today's dynamic economic environment, the number and magnitude of bankruptcy filings have increased significantly and pose a severe risk for investors. Auditors, ordinarily entrusted with making judgements on a firm's position, often fail to make accurate conclusions on their going-concern position as they tend to only take account of static financial statement ratios. Although traditional bankruptcy prediction models of the past have not been entirely satisfactory, there has been strong and continued interest in this field, as accurate instruments would benefit many interested parties (Cybinski, 2001). With the the financial crisis of 2008 still recent in memory, bankruptcy prediction has become even more pertinent due to the knock-on effects the failure of one company can have for others in the market.

In South Africa, company bankruptcy is an essential area that needs focus, with an average of 290 companies[1] going bankrupt every month from 2000 to 2013 (Tradingeconomics.com, 2014). Prior information on such events is invaluable to shareholders, creditors, auditors and employees as well as prospective investors. Current practice for bankruptcy prediction in South Africa has been mainly placed in the hands of auditors who primarily use going-concern methods via the use of financial statement ratios to evaluate a company's distress position by comparative sector analysis. From the investors point of view, the use of these simple methods does not reveal the company's financial distress position early enough to enable accurate decisions to be made.

Techniques to develop bankruptcy prediction have evolved over the past 60 years with many avenues of research attempted from the accounting, statistical and mathematical fields. In recent years researchers have applied a multitude of new dynamic modelling techniques for classification. Often, however, it has been the classical methods such as Altmans Z-Score and the logit model that have in practice been used more, given their ease of implementation.

The introduction of machine-learning techniques has provided a new tool for

[1] This number includes private and public companies.

financial research where substantial value is accrued from forecasting future events. Along with neural networks, Bayesian networks have been recently introduced into financial literature with promising results. A Bayesian network is a graphical model that encodes probabilistic relationships among variables of interest (Heckerman, Geiger and Chickering, 1995) and is an appealing modelling technique given that it is intuitively understandable and mathematically rigorous. Bayesian networks have recently been applied to bankruptcy prediction, but adequate operational guidance has yet to be provided. There is a need for a more comprehensive assessment of their applicability to modelling financial distress and to consider their practical implications. This research attempts to do this.

This study commences with a brief overview of techniques used in identifying financial distress before moving onto a theoretical introduction to Bayesian networks. Given the complexity of this technique, the learning process is explained in detail in Chapter 4 while Chapter 5 outlines the dataset and the challenges encountered with the sample collected. Chapter 6 explains how the network is used for inference and the methodology to be followed while Chapter 7 documents the results. Due to the vast array of models available, a comparison is made to other prediction techniques in Chapter 8. Lastly, Chapter 9 gives a discussion focusing on comparisons with similar research and issues for South African companies and Chapter 10 concludes this study.

Chapter 2

Overview of Techniques used in Financial Distress Research

The assessment of financial distress is a research area which has evolved from as early as the 1950s. This chapter briefly reviews some of the most prominent statistical techniques employed, analysing both the classical and the more recent dynamic models used. Research concerning South Africa is then assessed after which international literature where Bayesian networks have been tested in financial distress prediction is introduced.

2.1 Classical Models and their Limitations

Initial research focused on the use of univariate analysis, and Beaver (1966) used this approach to compare patterns of 29 financial statement ratios for a sample of failed firms. One of the most prominent models which is often still used is the Altman Z-Score (also known as the ZETA model). Altman (1968) developed a formula using business ratios weighted by coefficients determined by multidiscriminant analysis[1] to predict the probability of company bankruptcy within a year using US market data. The model performed successfully at first classifying 94% of the bankrupt companies and 97% of the non-bankrupt companies one year prior to bankruptcy .

The availability of better computing systems allowed for the use of larger datasets, and research challenged Altman's model with more advanced estimation methods. One of the issues with Altman's model and the subsequent versions is that the coefficients are determined using US market data and therefore are not entirely suitable for other markets. The use of linear probability models was shown by Ohlson (1980) who employed a logit model, while Zmijewski (1984) developed a probit model for

[1] A statistical technique used to classify an observation into one of several priori groupings, by reducing the difference between variables, dependent upon the individual characteristics of observations (Altman, Iwanicz-Drozdowska, Laitinen and Suvas, 2014).

predictive use. These regression-based techniques are more tractable as they allow for a model to be built on a particular dataset and can capture the structural characteristics of specific markets.

One of the major problems in bankruptcy research has been the nature of the dependent variable 'failure' which is not as well-defined as it should be for the types of models that have been used. Cybinski (2001) explains that this issue can be visualised as having two samples, one of failed firms and the other of non-failed firms where each sample is normally distributed and using one predictor variable, there lies an area of overlap between the two distributions. This indecisive area is the most difficult to predict but is also that of most interest, as the performance of the model is dependent on its ability to reduce the overlap area, i.e., separate the groups in a multidimensional space.

Another important issue is that the performance of the model is dependent on the firms used in the estimation sample and where they lie on the aforementioned 'success-failure continuum'. Model performance is most successful when data conforms to the expectation that the two groups are already well separated; however, this is often not the case. Gilbert, Menon and Schwartz (1990) provided evidence to support this claim by excluding strong firms in their non-bankrupt sample and concluded that ratio-based models perform poorly in identifying likely bankruptcy in a sample consisting of 'problem' companies. Another provision that previous literature has failed to address, has been to incorporate the external influence of macroeconomic factors which contain substantial information, as bankruptcy rates are highly correlated with economic business cycles (Cybinski, 2000).

2.2 Dynamic Models

In recent years, artificial neural networks have emerged as a technique in the field of corporate bankruptcy prediction and have remained popular despite research showing contrasting results. A neural network fundamentally maps inputs to the outputs (the classifications) using layers and neurons to create a complex learned algorithm (Muller, Steyn-Bruwer and Hamman, 2009). Without being linearly constrained like previous models, neural networks developed a technique where the model can be made to fit the data almost perfectly. In contrast to regression-based models, they can overcome the effect of autocorrelation and have the ability to account for missing values.

Several pieces of literature have proved the superiority of artificial neural networks to multidiscriminant analysis. Udo (1993) compared a neural network set-up to the logit model and concluded that these dynamic models are superior in classification, easier to use, more robust and more responsive than a regression model. There is, however, a concern regarding the 'black box' nature of how neural networks solve a particular problem in the sense that studying its structure does not give any insights on the structure of the function being approximated, as well as there being no output to interpret the causality behind financial distress.

Market based models using the Black and Scholes (1973) contingent claims approach have proved to be an appealing alternative for this problem. These models are advantageous in that they provide a sound theoretical model for firm bankruptcy and their output is not time or sample dependent (Agarwal and Taffler, 2008). There are however, a number of problematic assumptions embedded in these models such as that they do not distinguish between different types of debt as it is assumed that a firm only has a single zero coupon loan. Furthermore, the model requires measures of asset value and volatility which are unobservable. It is therefore not surprising that empirical performance of market-based model has been mixed (Agarwal and Taffler, 2008) and have little forecasting power (Campbell, Hilscher and Szilagyi, 2008).

2.3 Financial Distress Literature in South Africa

Due to the importance of understanding financial distress, previous research has analysed this topic regarding South African companies. Bruwer and Hamman (2006) reviewed previous methodologies applied to South African data from 12 different studies. The same study was used to predict financial distress for industrial JSE companies using recursive partitioning[2]. They further looked at dividing this model for different economic cycles, which assisted in understanding which variables have a specific causal effect on financial distress. Finally, it was concluded that despite not having 'spectacular' success rates, they had managed to overcome many of the deficiencies in previous studies.

A more recent review of South African literature was conducted by Muller, Steyn-Bruwer and Hamman (2009) where they compared the success rates of different models on JSE listed companies. They commented that previous literature had not sufficiently accounted for Type I and Type II errors[3] occurring and hence introduced a measure called the 'Normalised Cost of Failure' to account for these faults. Using this measure, it was found that logit analysis and neural networks produced the best

[2] Recursive partitioning is a dynamic method which creates a decision tree that attempts to correctly classify members of the population based on several independent variables (Wang, Chen, Liu, Zheng, Gu and Xu, 2014)

[3] In hypothesis testing, Type I error can be defined as incorrectly rejecting a true null hypothesis while Type II error is defined as the failure to reject a false null hypothesis.

predictive accuracies when using South Africa data.

2.4 Bayesian networks in Bankruptcy Prediction

While initial applications of Bayesian networks were rooted in the medical field, they have recently found success in financial analysis given their structure in dynamically modelling a multitude of variables to predict specific outcomes. In recent years, Bayesian networks have been used in portfolio analysis for predicting risk and returns as well as providing a modelling technique to better understand operational risk. Despite being a very new method to be applied to financial distress prediction, initial results have been very promising.

Sarkar and Sriram (2001) used naive Bayes classifiers to provide a technique for early warning of bank failures by making use of two different probabilistic models where the independence assumptions differed in each model. Both models were found to be well calibrated with prediction success in excess of 90% but it was the composite attributes[4] augmentation of the naive Bayes model, that performed better. It was concluded that this model variation better reflects the underlying joint distribution across variables of interest compared to the conventional naive Bayes model.

A more thorough analysis was performed on the use of naive Bayes classifiers by Sun and Shenoy (2007) as they attempted to provide operational guidance for bankruptcy prediction. They examined issues regarding whether continuous or discrete distributions provide better model performance, which is commented on later in this research. They suggested one of the future research possibilities was to use variable selection algorithms which have often been omitted in financial applications of Bayesian networks but are an integral component of finding the network structure.

Other studies that have used Bayesian networks for bankruptcy prediction include Aghaie and Saeedi (2009) who found the Bayesian model to have 94% prediction success a year in advance using Iranian data. All three similar studies are likely to have an element of model 'over-fitting' by using a large proportion of data in their training sets (subset of sample used to build the model). This is a crucial issue which is later considered by making a comparison to the aforementioned studies.

[4] A naive Bayes model where the conditional independence assumption among predictive variables was relaxed. The reasons behind these assumptions will be explained in Chapter 4.

Chapter 3

Understanding Bayesian Networks

Bayesian networks combine principles from different academic fields including graph theory, probability theory and computer science. Therefore, there is a fair deal of complexity in constructing the model. This chapter aims to give general introduction of Bayesian networks, before proceeding to the properties related to the modelling problem such as the local probability distributions and the embedded Markovian properties. This lays the foundation to discuss the complexities of constructing the network in the next chapter.

3.1 Defining Bayesian Networks

Bayesian networks belong to the family of probabilistic graphical models and can be used to represent knowledge of an uncertain domain (Ben-Gal, 2007). A Bayesian network can be defined as a directed acyclic graph (DAG) that represents a joint probability distribution of a set of random variables. The structure consists of a set of nodes, each representing a variable, which can each take a number of predefined values referred to as states. Edges (also known as arcs) between nodes denote a conditional dependence between variables. It is important to note that edges signify a relationship in one direction only and there are no paths of edges that start and end at the same node, hence the structure is termed acyclic (Nilsson, 1998). These conditional dependencies in the graph are estimated by using a combination of statistical and computational methods which will be introduced in Chapter 4.

The terminology describing the hierarchy of the graph uses straightforward genealogical terms. A simple graph is shown in Figure 3.1 for explanatory purposes. The initial node A would be denoted the root node and is the parent of node B, which is the child of A as there is directed edge from A to B. Node E would be referred to as a descendant of A and furthermore a node without children is re-

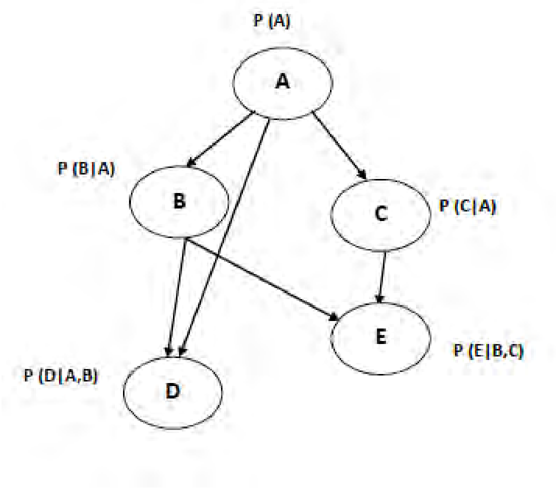

Fig. 3.1: A Simple Network for Illustration.

ferred to as a leaf. Lastly, a node such as C with parents and children is called an intermediate node.

This structure can be used for two purposes. Firstly, *deductive reasoning* which follows the direction of the causal links between variables, and we will use in the learning (construction) process (i.e., given bankruptcy, what is the structure of the descendant nodes). Secondly, *diagnostic reasoning* which goes against the direction of the causal links, which is used for inference (i.e., using the predictor variables to predict bankruptcy).

The topology or structure of the network can be constructed using expert knowledge or some type of learning algorithm. Once this has been specified, the relationships between connected nodes can be quantified using conditional probability tables (CPT). For each node all possible combinations of values of its parent nodes need to be determined. Each such combination is called an instantiation of the parent set. For each distinct instantiation of parent node values, we need to specify the probability that the child will take for each of its values. Root nodes also have an associated CPT, containing only one row representing its prior probabilities (Korb and Nicholson, 2004).

3.2 Local Probability Distributions

The probabilities that can be seen in Figure 3.1, each denote a local probability distribution of that node. The CPT summarises all possible states for that node given the state space for its parents. Hence, a series of CPTs represents the joint probability distribution of all possible states of the network. When the node is discrete, its local probability distributions are modelled using a multinomial distribution[1] parameterised by a set of probability vectors. Where the node is, instead, continuous it is modelled using a Gaussian distribution and the mean state is simply a linear combination of the parent states (Lauritzen, 1996). Where a node does not have a parent, its local probability distribution is said to be unconditional.

In probabilistic terms, the joint probability distribution function (PDF) of the domain can be factorised into smaller, local PDFs each involving a node and its parents only. Therefore, the local PDFs provide the quantitative probabilities that are multiplied together in the fashion prescribed by the qualitative independence relations. These relations are implied by the structure of the network and are sufficient to reconstruct the joint PDF of the domain (Margaritis, 2003).

3.2.1 Mathematical outline

Ben-Gal (2007) shows this can be formally structured as follows:

A Bayesian network is defined by a pair $B = (G, \Theta)$ where G is the DAG with nodes $X_1, X_2..., X_n$ representing random variables. The graph G has the independence assumption such that each variable X_i is independent of its non-descendants given its parents in G. The second component Θ denotes the set of parameters (conditional probability tables) in the network. This set contains the parameter $\theta_{x_i|\pi_i} = P(x_i|\pi_i)$ for each realisation of x_i of X_i conditioned on each π_i which is the set of parents of X_i in G (Ben-Gal, 2007). Therefore, B defines a unique joint probability distribution over the set of random variables and the probability mass function can be denoted as:

$$
\begin{aligned}
P(X_1, X_2..., X_n) &= \prod_{i=1}^{n} P(X_i|\pi_i), && (3.1)\\
&= \prod_{i=1}^{n} \theta_{X_i|\pi_i}, && (3.2)\\
\text{or more simply,} \quad &= \prod_{i=1}^{n} P(X_i|Parents(X_i)). && (3.3)
\end{aligned}
$$

3.3 The Markov Condition

One of the reasons Bayesian networks are described in terms of conditional independence relations is that these 'independencies' simplify the computation of the global joint probability distribution for the network. These independencies are captured by

[1] The multinomial distribution is a generalisation of the binomial distribution. For n independent trials, each of which leads to success for k categories, it gives the probability of any particular number of successes for the various categories (Taylor and Karlin, 1998)

the local semantics of the network and *d-separation* (Pearl, 2011). Local semantics refer to the property that every node in the network is conditionally independent of all its non-descendent nodes given its parent nodes (Pearl, 2011). This property, referred to as the Markov condition of the network, implies that there are no direct dependencies in the system being modelled which are not already explicitly shown via the edges in the structure (Korb and Nicholson, 2004).

3.3.1 d-separation

d-separation, also referred to as the global Markov condition, is a graphical property that can be used to determine whether a set of variables X is independent of another set Y, given a third set Z (Lauritzen, 1996). This concept applies to sets of nodes rather than pairs and captures both the conditional independence and dependence relations that are implied by the Markov condition on the random variables (Ben-Gal, 2007). Nilsson (1998) explains d-separation as follows: Two nodes, X and Y, can be said to be d-separated by a node Z, denoted $< X|Z|Y >_D$, if one of the following hold:

- The path of undirected edges connecting them comprises one edge leading into Z and one leading out of Z (known as a *chain*).
- The path of undirected edges connecting them comprises two edges, both leading out of Z (a *fork*).
- The path of undirected edges connecting them comprises two edges, one of which leads from X to Z and the other from Y to Z (a *collider*).

For a Bayesian network, the Markov condition is informed by all conditional independencies identified by d-separation; however, conditional independencies can exist which are not captured by d-separation (Nilsson, 1998). It is further assumed that Bayesian networks meet the faithfulness condition.

3.3.2 Faithfulness Condition

A DAG G, and a joint probability distribution P, over a set of variables X, are faithful to one another if and only if every one and all independence relations valid in P are also present in G (Margaritis, 2003). This faithfulness condition along with the Markov conditions enables the association of the DAG with the joint probability distribution of the network.

3.3.3 Markov Equivalence

Two DAGs that entail exactly the same set of conditional independence relations in the data (i.e., among the variables) are said to be Markov equivalent. The requirements for Markov equivalence are that the DAGs have the same skeleton and must have the same edges between nodes that are not part of a *collider* (Nilsson, 1998). This property of Bayesian networks is crucial in simplifying the computational complexity of the learning process as a theoretical limit is imposed on learning the structure.

Chapter 4

Building the model

The most challenging task in dealing with Bayesian networks is learning their structure. The learning process requires determining both the structure represented by the DAG and the parameters described by the CPTs, from a dataset. In this section, the primary techniques used in this process, learning algorithms and scoring metrics are presented. Then a simpler augmentation of this model known as the naive Bayes classifier is introduced.

4.1 Learning the Bayesian Network Structure

The construction of an optimal network that accurately models the dependencies in the model is a computationally expensive problem given that the number of possible structures is super exponential. Robinson (1977) shows that the number of possible structures $f(n)$, with n variables can be computed using the recursive relation:

$$f(n) = \sum_{i=1}^{n} (-1)^{i+1} \binom{n}{1} 2^{i(n-i)} f(n-i) \tag{4.1}$$

where $f(1) = 1$.

To put this into perspective with three variables, the number of possible structures is only 25 yet with 10 variables this massively increases to 4.2×10^{18}. This is, however, somewhat simplified by the Markov equivalent condition as the DAGs are required to be statistically distinguishable, reducing the number of relevant structures.

Learning the structure can be conducted via either conditional independence test methods or search and scoring strategies or a hybrid method of both approaches. Conditional independence relations in the data are derived using constraint-based algorithms based on conditional independence tests, and attempt to find a network that best represents these relationships. The more commonly used strategy is search and scoring methods which involve a search strategy to explore the possible structure of the underlying data and a scoring metric on which to select the best structure identified during the search process. Below we outline scoring metrics than can be used before moving onto the learning algorithms (search methods) that identify the best structure and calculate the parameters.

4.2 Choice of a Scoring Metric

A scoring metric is used to evaluate the goodness of fit of the network. Learning a network structure can be considered an optimisation problem where a quality measure of a network structure, given the training data, needs to be maximised. This quality measure can be based on the Bayesian approach, minimum description length or information criterion (Bouckaret, 2004). We consider two metrics in this study namely the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), which determine the relative information lost when using an approximate model to describe the true relationship between a set of variables (Nilsson, 1998). The logic for using these measures is that they attempt to find the model that best explains the data with a minimum number of 'free parameters' and thus discourage overfitting[1].

The AIC has the following form

$$AIC = -2ln(L) + 2k, \tag{4.2}$$

while the BIC has the form,

$$BIC = -2ln(L) + k \cdot ln(n), \tag{4.3}$$

where L represents the maximised value of the likelihood function for the estimated model, k is the number of parameters in the model and n is the number of variables.

Therefore, the first term in both the equations represents the model fit while the second is a penalty for overfitting. The goal in terms of the structure learning is to minimise both AIC and BIC and rank the various different augmentations of the model by their scores with the model with the lowest score accepted as the best.

4.3 Learning Algorithms

The scoring metric works in conjunction with a search method that identifies the DAG for the selected scoring criterion. Various network-search methods exist including genetic, simulated annealing, tabu search and branch and bound algorithms

[1] Overfitting occurs when a model describes the training data very well but explanatory power deteriorates in describing other instances of the same phenomenon making the whole learning process worthless (Nannen, 2003).

(Lin, 2012). Since these learning algorithms involve optimisation problems, certain algorithms cannot always find solutions for the problem presented. For this reason, only two algorithms which consistently found solutions are presented in this study.

4.3.1 The K2 Algorithm

Cooper and Herskovits (1992) define a Bayesian network $B = (B_S, B_P)$, where B_S represents the network structure and B_P, the conditional probabilities (or network parameters). Let D be the dataset in use and Z a set of n discrete variables. We want to calculate $P(B_S|D)$ (the probability of Bayesian network, B_S giving rise to this dataset) which can be reduced to $P(B_S, D)$ by Bayes' theorem. Cooper and Herskovits (1992) equation for calculating $P(B_s, D)$ is based on four assumptions:

1. The database variables, Z, are discrete.
2. Cases occur independently given a Bayesian network model.
3. Variables do not contain any missing values.
4. The density function $f(B_p|B_s)$ is uniform[2] and as a result we are indifferent regarding prior probabilities to place on a network structure B_S (Lin, 2012).

Let X be a variable in Z which has r possible states. The dataset D has m entries, where each entry contains a state for each variable in Z. Each variable X_i in B_S has a set of parents represented by π_i and there are q_i instantiations of π_i, each which is unique. Lastly let N_{ijk} be defined as the number of entries in D in which a variable X_i is in its k^{th} state and π_i is in the j^{th} state. Therefore, N_{ij} can be defined as:

$$N_{ij} = \sum_{k=1}^{r_i} N_{ijk}. \tag{4.5}$$

Cooper and Herskovits (1992) show given assumptions 1 to 4 it follows that:

$$P(B_S, D) = P(B_S) \prod_{i=1}^{n} \prod_{j=1}^{q_i} \frac{(r_i - 1)!}{(N_{ij} + r_i - 1)!} \prod_{k=1}^{r_i} N_{ijk}!. \tag{4.6}$$

Equation (4.6) gives a computable method of comparing the probabilities of a network structures. The optimal structure B_S could be uncovered by iteratively computing $P(B_S, D)$ for all possible network structures given D. This, however is

[2] This represents the density function of $P(B_s, D)$:

$$P(B_S, D) = \int_{B_P} P(D|B_S, B_p) f(B_p|B_s) P(B_S) dB_p. \tag{4.4}$$

not a computationally feasible approach in most instances and therefore a heuristic method is constructed. We further assume that there is an ordering available on all n variables such that if X_i precedes X_j in the ordering then X_i cannot be a parent of X_j if $j > 1$, and that all possible structures B_S are equally likely with a probability c. We can then restate $P(B_S, D)$ as

$$P(B_S, D) = c \cdot \prod_{i=1}^{n} \prod_{j=1}^{q_i} \frac{(r_i - 1)!}{(N_{ij} + r_i - 1)!} \prod_{k=1}^{r_i} N_{ijk}!. \tag{4.7}$$

In order to maximise $P(B_S, D)$ we find the parent set for each variable that maximises the function

$$\underset{B_s}{max}\left[P(B_S, D)\right] = c \cdot \prod_{i=1}^{n} \underset{\pi_i}{max} \left[\prod_{j=1}^{q_i} \frac{(r_i - 1)!}{(N_{ij} + r_i - 1)!} \prod_{k=1}^{r_i} N_{ijk}! \right]. \tag{4.8}$$

Let π^S be the parents of x_i in B_S denoted as $\pi_i^S \rightarrow x_i$, Equation (4.6) can be generalised to

$$P(B_S, D) = \prod_{i=1}^{n} P(\pi_i \rightarrow x_i) \prod_{j=1}^{q_i} \frac{(r_i - 1)!}{(N_{ij} + r_i - 1)!} \prod_{k=1}^{r_i} N_{ijk}!. \tag{4.9}$$

The computational complexity of the problem can then be reduced to polynomial-time by further assuming that there is a limit on the number of parents of any nodes and that $P(\pi_i \rightarrow x_i)$ and $P(\pi_j \rightarrow x_j)$ are independent when $i \neq j$, it then follows that

$$\underset{B_s}{max}\left[P(B_S, D)\right] = \prod_{i=1}^{n} max_{\pi_i} \left[\prod_{j=1}^{q_i} \frac{(r_i - 1)!}{(N_{ij} + r_i - 1)!} \prod_{k=1}^{r_i} N_{ijk}! \right]. \tag{4.10}$$

The K2 algorithm is a greedy heuristic that searches for the network structure B_S that maximises $P(B_S, D)$. The algorithm assumes that a node lacks parents and then incrementally adds that parent whose inclusion increases the probability of the resulting structure according to the function

$$g(i, \pi_i) = \prod_{j=1}^{q_i} \frac{(r_{i-1}!)}{(N_{ij} + r_i - 1)!} \prod_{k=1}^{r_i} N_{ijk}!. \tag{4.11}$$

The algorithm stops adding parents to a node when none of the remaining potential parents are able to improve the probability for the structure (Nilsson, 1998).

4.3.2 The Max-Min Hill Climbing Algorithm

The Max-Min Hill-Climbing (MMHC) algorithm is a hybrid algorithm combining both conditional independence tests methods and search and scoring strategies. The structure of the network is found by using the Max-Min Parents and Children (MMPC) algorithm, while the directionality of the edges between nodes is determined using a hill climbing search. The MMHC algorithm, derived by Tsaardinos, Brown and Aliferis (2006), is briefly outlined below by presenting MMPC algorithm and then explaining how this incorporated into a hill climbing search.

Max-Min Parents Children Algorithm

Tsamardinos, Aliferis, Statnikov and Statnikov (2003) derived the MMPC algorithm where the Max-Min part of the name refers to a heuristic it uses, while the parents-children part refers to its output. The MMPC algorithm runs on a target variable T and provides a way to identify the existence of edges to and from T but without being able to identify the orientation of the edges. By invoking the MMPC with each variable as the target, all the edges can be identified in an unoriented fashion, giving us the skeleton of the Bayesian network (Tsaardinos *et al.*, 2006).

Tsaardinos *et al.* (2006) denote a DAG G, a joint probability distribution P, and set of parents and children of a node (variable) T as PC_T^G. By the faithfulness condition introduced in the previous chapter it can be deduced that for any faithful two Bayesian networks $< G, P >$ and $< G', P >$, $PC_T^G = PC_T^{G'}$ holds. This allows us to denote the parent-child combination for variable T as PC_T for any Bayesian networks faithful to the same distribution.

The algorithm then proceeds by attempting to find a conditioning set Z for which function $I(X, T|Z)$ holds and it can be proven that X does not belong to PC_T^G. As Nilsson (1998) explains, it does this by determining a minimum association[3] between T and all potential parents and children X over all subsets of a feature set Z as follows:

$$MinAssoc(X, T|Z) = min_{S \subseteq Z} Assoc(X, T|S). \tag{4.12}$$

In an iterative process variables then enter a candidate parent-child set, denoted as CPC using a heuristic function. In each iteration the algorithm selects the variable that maximises equation (4.12) with T relative to CPC. This process allows for variables that are highly associated with T to be found even after repeated attempts to make the variable independent of T in the subsequent iteration. The process

[3] The function $Assoc(X; T|Z)$ is an estimate of the strength of association (dependency) of X and T given Z (Tsaardinos *et al.*, 2006)

ceases when the minimum association of all remaining variables with T given some subset of CPC is zero.

If the faithfulness condition holds, the MMPC algorithm will return no false negatives but may uncover false positives. In order to remove any false positives, the algorithm therefore attempts to test whether $I(X, T|S|)$ for some subset $S \subseteq CPC$ and where this condition is true X is removed from CPC.

Hill Climbing Algorithm

While, the MMPC algorithm provides the existence of edges between nodes, it does not provide the direction of the edges; however, this is provided by the hill climbing algorithm. Hill climbing algorithms are particularly popular due to their trade-off between computational demands and the quality of the models learned (Gamez, Mateo and Puerta, 2011). As Nilsson (1998) explains the hill climbing search initialises with an empty, full or randomly-generated structure. The algorithm then iteratively adds, deletes or changes the direction of edges between the nodes to improve the chosen scoring metric until convergence is reached. The search is constrained to only consider adding an edge if it was discovered by the MMPC, explained earlier. This constrained search improves time efficiency as it reduces the possible networks considered by the search procedure.

4.4 Naive Bayes Classifier

In this study, we test both the standard Bayesian network and the naive Bayes classifier. The motivation for using the Bayesian network for this prediction problem is that is a less restricted model in that it allows for relations between predictor variables and furthermore, it has not been previously tested in financial distress literature. The naive Bayes Classifier, which has been the trusted method in previous research, is specifically equipped for classification problems but is more restricted in its construct. In this section we introduce the naive Bayes model with the assumptions embedded in the model, and the classification algorithm that allows for prediction to take place.

4.4.1 Naive Bayes Basics

The naive Bayes model is aptly named due to its simplicity in comparison to the standard Bayesian networks. In a naive Bayes model, the node of interest has to be the root node, which means it has no parent nodes (Sun and Shenoy, 2007). A simple example is given below where node A would be the binary financial distress (bankruptcy) variable and n predictor variables, $B_1 \dots B_n$ are denoted in the tier below.

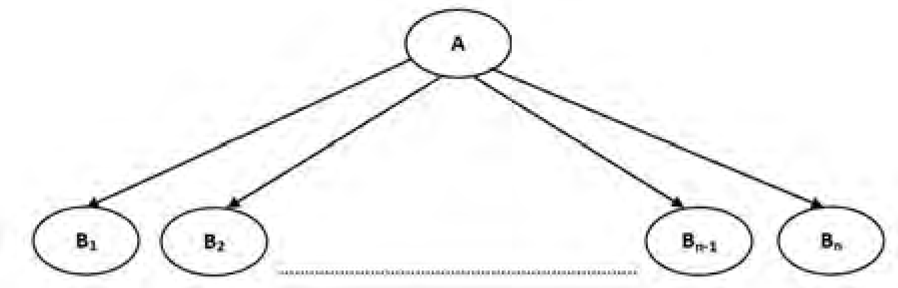

Fig. 4.1: Naive Bayes Structure.

In contrast to a Bayesian network, the naive Bayes assumes conditional independence between predictor variables, given variable A i.e., for $i = 1, 2....n$

$$B_i \perp (B_1, B_2...., B_{i-1}, B_{i+1}, ..., B_n|A). \tag{4.13}$$

There is no structure learning in a naive Bayes model as the layout of the DAG is fixed. This simplifies the computational complexity greatly and leaves only the network parameters to be calculated. Different algorithms exist for parameter learning, in this section we present a classification algorithm as shown by Mitchell (2010).

4.4.2 Naive Bayes Algorithm

Let A be any discrete-random variable and the discrete predictor variables are represented by $B_1 \dots B_n$. Using the definition of conditional independence, if B contains n variables which are conditionally independent of one another given A, it follows that

$$P(B_1 \dots B_n|A) = \prod_{i=1}^{n} P(B_i|A). \tag{4.14}$$

Our goal is to find a classifier that will output the probability distribution over all possible values of A for each new instance of B that is required to be classified. The expression for the probability that A takes on its k^{th} of m possible values according to Bayes' theorem and Equation (4.14) can be written as

$$P(A = a_k|B_1....B_n) = \frac{P(A = a_k) \prod_{i=1}^{n} P(B_i|A = a_k)}{\sum_{j=1}^{m} P(A = a_j) \prod_{i=1}^{n} P(B_i|A = a_j)}. \tag{4.15}$$

Equation (4.15) is the fundamental equation for the Naive Bayes classifier and shows how to calculate the probability that A will take on any given value, given the observed values of B, as well as the distributions $P(A)$ and $P(B_i|A)$, which would be estimated from the training set. Furthermore, if we are interested in the most probable value of A for inference purposes, then we have the classification rule:

$$A \leftarrow \underset{a_k}{argmax} \quad \frac{P(A = a_k) \prod_{i=1}^{n} P(B_i|A = a_k)}{\sum_{j=1}^{m} P(A = a_j) \prod_{i=1}^{n} P(B_i|A = a_j)}. \tag{4.16}$$

Which can be simplified as the denominator does not depend on a_k

$$A \leftarrow \underset{a_k}{argmax} \quad P(A = a_k) \prod_{i=1}^{n} P(B_i|A = a_k). \tag{4.17}$$

Chapter 5

Sample and Data Preprocessing

5.1 Make-up of Dataset

The dataset in use is made up of information on 132 listed companies from 2000 to 2013. These companies can be split into 66 pairs of a company that has gone bankrupt, and a company with similar characteristics that has not gone bankrupt[1]. A full breakdown of the companies in the sample can be found in Appendix A.2.

The definition of bankruptcy or financial distress has varied across previous literature causing conflicting definitions. In this research, it is defined as the situation where a company ceases to exist in its current form and is declared bankrupt. This does not include companies that de-list from the JSE. De-listing does create a problematic influence on our dependent variable as this is often due to a company entering some sort of financial distress, however this cannot necessarily be assumed to be going into bankruptcy as de-listing can occur for contrasting reasons such as management buy-outs. This definition will assist in classifying companies as 'failed' or 'non-failed', and hence we can create a binary dependent variable for prediction.

The dataset contains 67 variables from the financial statements of these companies at their respective financial year-ends as well as information relating to their share price. Variables were sourced from the McGregor BFA database and a list with explanations has been included in Appendix A.1. All variables are expressed as ratios or percentages for them to be comparable across companies and were adjusted for inflation to allow for fair comparisons across time.

To further enhance explanatory power, variables relating to the sector/industry of a company were appended to the dataset. These included two variables, *Industry* which split companies into nine categories while *Industry group* was more specific and had 43 categories. A list with the number of companies in each category has been included in Appendix A.3.

[1] The control group companies were matched using the criteria of sector, market share, market capitalisation and capital structure.

5.2 Missing Data

There is an issue of missing data for certain companies for entire years mainly due to a company de-listing or a company only listing in the course of the 14 year time period. Three bankrupt companies in the original sample were found to have no data available and as a result were dropped.

5.2.1 Fraudulent Companies

When predicting bankruptcy, the model could be negatively influenced by companies who were liquidated due to fraud, as a company could have a good business model and hence characteristics that are more similar to those of non-bankrupt companies. This could bias the model and not allow the 'treatment group' to only contain companies with specific bankruptcy characteristics required for learning. From the sample of 63 bankrupt companies, only three companies were found to have explicit press releases that detailed them as committing fraud and were removed. The final dataset therefore contains 120 companies with 1 136 entries in total (i.e., an average of nine and a half years of information per company).

5.3 Remedies for Missing Variable Information

There is a large number of variables without full information in the dataset: 39 of the 67 company-specific variables have some degree of 'missingness' (the extent of which can be seen in the far right column of Appendix A.4). To overcome this issue, we can proceed in three ways, (1) Ignore the variables with missing information and use a dataset with 28 variables, (2) Use an imputation method to estimate missing values or (3) Allow the Bayesian network to determine missing values. We consider the effectiveness of the latter two strategies in more detail below.

5.3.1 Imputation

Imputation is commonly used to assign values where these are missing and hence complete the dataset. In order to perform imputation we need to make the assumption that the data is missing completely at random (MCAR) or missing at random (MAR). The distinction between these is that MCAR means that the propensity for a data point to be missing is completely random, while MAR means the propensity for a data point to be missing is not related to the missing data, but is related to some observed data. With either of these assumptions, we can proceed with imputation, the only inaccessible mechanism being where data is not missing at random (NMAR) (Donaldson, Graham and Hansen, 1994).

Little (1988) devised a test to determine whether data is MCAR or MAR; however, by definition you cannot determine whether data are NMAR by looking at the observed values. Multiple imputation can give unbiased estimates with NMAR data, but only if the imputation method includes a model of the missingness mechanism (SSCC, 2014). Therefore, before multiple imputation can be carried out, the missing mechanism needs to be identified and furthermore a distributional assumption needs to be made. To determine this, we firstly inspect the distributional properties of the variables.

Variable Distribution

If imputation is to be used, a distributional assumption that data is normally distributed needs to be made. One would expect this to be a fair assumption for financial ratios; however, inspection of the descriptive statistics in Appendix A.4 reveals that variables have a skewed leptokurtic distributional shape. Expected skewness and kurtosis of a normally distributed variable are 0 and 3 respectively; however, it can easily be seen that variables deviate drastically from these expected norms. The non-normality of their distribution is further emphasised by the significantly large differences in their means and medians. The dispersed distribution of the variables is not completely unexpected for the make-up of the companies in the sample. Approximately half of these companies eventually experienced bankruptcy and as a result, you would expect indicators of their performance to be of a highly volatile nature and contain many 'outliers'.

The reason common regression-based imputation methods are often not satisfactory when data is not normally distributed is because the error term is assumed normal and does not take into account the non-normal distribution of the data. Solutions do exist for this such as replacing regression with predictive mean matching[2] which gives a much better fit. However, before the predictive mean matching can be carried out, we need to determine whether data is MCAR or MAR for imputation to be applicable. This is a difficult issue to determine and there is no basis to assume data is MAR as missing data in our case is caused by companies not releasing certain items of information.

[2] Predictive mean matching is similar to regression imputation method except that for each missing value, it imputes a value randomly from a set of observed values whose predicted values are closest to the predicted value for the missing value from the simulated regression model (Schenker and Taylor, 1996).

5.3.2 Bayesian Network Methods for Dealing with Missing Data

Similarly to imputation assumptions, an important distinction concerning missing data in Bayesian Networks is whether the absence of an observation is dependent on the actual states of the variables (Heckerman *et al.*, 1995). Various methods are suggested in the Bayesian network literature for dealing with incomplete data such as Monte-Carlo Methods, Gaussian Approximation and using Expected-Maximisation algorithms. The software in use, *Weka*, also has a preprocessing mechanism for dealing with missing data whereby it replaces all missing values for nominal and numeric attributes in a dataset with the modes and means from the training data (Bouckaret, 2004), however this method has very little mathematical justification.

5.3.3 Conclusion on Missing Data

As stated earlier the predictive mean matching method is suitable for non-normal data but it is not a very robust method as it could lead to various missing entries taking the same value. A rather more important issue is that, we require the assumption of MCAR or MAR and not NMAR. An interesting finding in the data was that the average missingness among variables which had some degree of missing values, was far higher for companies that eventually went bankrupt (27.9%) for those that did not (13.2%). This would support the hypothesis that data is not missing at random and that companies withheld specific information with reason.

Therefore, due to the limitations and lack of mathematical suitability in the methods stated above, it was felt that the dataset of 28 variables was best to use. Furthermore, this being a prediction study, we should be wary of trying to predict the actual predictor variables, as this could greatly influence the final classifications. Another point considered, was that given the dispersed nature of the variables, it would very difficult to arrive at fair approximations for the incomplete data. Lastly, a dataset of 28 company-specific variables is fairly extensive and includes more variables than both Sun and Shenoy (2007) and Aghaie and Saeedi (2009) in their similar studies.

Since, a large proportion of the variables have been dropped, results using a fully imputed dataset by the predictive mean matching method have been included in Appendix B.3 to assess the differences in results.

5.4 Macroeconomic Variable Inclusion

Initial results for the various models were found to have low success rates. One of the reasons for this could be that companies went bankrupt due to exogenous market-based factors impacting their stability. In order to partially account for time dependent structural changes in the economy, three macroeconomic variables were appended to the dataset:

- *Business Confidence Index (BCI)* is generated monthly by the South African Chamber of Commerce and Industry as a measure of the level of business confidence within the South African economy (SACCI, 2015).
- *Consumer Price Index (CPI)* is a primary measure for inflation in South Africa and tracks the rate of change in prices of goods and services purchased by consumers.
- *JSE All Share Index Returns (ALSI)* is a market capitalisation weighted index that tracks the performance of all companies listed on the JSE.

BCI and *CPI* have quarterly releases once a period ends and hence are not known at the time they occur. To account for this, lagged variables were used, meaning all indices were matched to the company's financial year for one quarter in advance, as for fair prediction to take place, only information available at the time should be used. Variables are expressed as returns as well as the change in return.

5.5 Nature of the Dependent Variable

The timing of when to test for bankruptcy is also an essential issue. It would make no sense to test for bankruptcy at the very time it occurs as this would be of no help to any of the interested parties. We are limited, in that information is only available at the time of a company's financial year end; as a result we have constructed three indicator variables that can be tested:

- *Bankrupt year (0y)*: Denoted as distressed at most recent financial year before bankruptcy end: i.e., bankrupt in less than one year.
- *Year before bankruptcy (1y)*: Denoted as distressed at penultimate financial year before bankruptcy: i.e., bankrupt in one to two years.
- *Cumulative Indicator (CI)*: Denoted as distressed in all years of existence before bankruptcy.

5.6 Discretisation

Discretisation is the process of converting a continuous function into a corresponding approximate discrete structure of point or interval form. The Bayesian learning algorithms presented in the previous chapter, both require the underlying data to be discrete in nature. Discretising variables is also advantageous as this allows for non-linear relationships between the variables to be revealed (Muhlenbach and Rakotomalala, 2005).

The numeric variables in use, as can be seen in Appendix A.1, contain financial statement ratios and therefore are all continuous. There are various methods for discretising a continuous variable within this framework, namely the Pearson-Tukey method and Bracket Medians Method (Mihaela-Daciana Craciun and Bala, 2014). These methods, however, require a distributional assumption for a variable. A natural assumption would be that variables are normally distributed; however, as previously explained, distributions of variables have an extreme leptokurtic shape and hence we cannot deduce an assumption of normality.

This somewhat simplifies the process of discretisation, which occurs in two stages. Firstly, a decision needs to be made whether to have intervals of equal frequency or equal length. Given the dispersed nature of the data, it was chosen to have intervals of equal frequency to avoid having intervals containing a single figure number of observations. Secondly, the number of intervals needs to be determined. There is no definitive answer to this question but given the variable distribution of the companies where the variables often contain far outliers, it would be advantageous to have more intervals to reduce the information loss from discretisation.

Previous research on the same topic has used very few intervals. Sarkar and Sriram (2001) divided variables into two intervals. Aghaie and Saeedi (2009) and Sun and Shenoy (2007) both empirically examined how the number of states chosen for discretisation impact a model's predictive power. Comparing results they found predictive accuracy was best when using five states and diminished for intervals higher than five. Consequently, using the idea behind Shenoy's research we try to avoid letting this user-defined decision on the number of intervals, to be a limitation and hence test for different number of intervals, these being 2, 5, 10, 15 and 20.

5.7 Size of Training Set

The training set is a subset of the data used for the learning process to construct the model. The remaining data would be then be used for inference queries to test the model's predictive accuracy. Previous research has used large proportions of the data at their disposal for learning. Sun and Shenoy (2007) used 10-fold analysis, where the entire sample is divided randomly into ten equal sized subsets, nine of these randomly selected to form the training sample, while the remaining subset is used as the test sample to test the models performance. Sarkar and Sriram (2001) used a proportion of 80% for training while the other study on this subject by Aghaie and Saeedi (2009) made no mention of the size used.

One of the most common accusations levelled against Bayesian network models is that of overfitting in that they fail in 'out of sample' testing. Therefore, despite the constraint of a small-scale dataset, the decision taken was to use only 40% (i.e., approximately the first six years) of the data for training purposes[3]. The motivation behind this was to test as rigorously as possible for the applicability of the method and furthermore avoid the model conforming to our dataset but having little use out of sample. This choice is an important theme of classification and hence is emperically examined in Chapter 9.

[3] The training set would therefore contain 24 bankrupt companies using the *bankrupt year* as the dependent variable, 18 using the *year before bankruptcy* variable and 112 cases of bankruptcy using the *cumulative indicator* variable.

Chapter 6

Inference and Evaluation Methodologies

Once the structure has been constructed and parameters calculated, via the processes explained in Chapter 4, a Bayesian network can be used for predictive purposes. In this chapter we further explain how these inference queries are conducted. This allows for the methodology to be formalised, and the different model permutations that are tested to be outlined. Lastly, the different metrics that will be used in performance evaluation are introduced.

6.1 Conducting Inference from Network Structure

The inference process is conducted via a 'flow of information' through the network (Korb and Nicholson, 2004) where information refers to the subset of the sample that was not used in the learning process. Two types of inference are often considered, causal (also referred to as top-down) and diagnostic (referred to as bottom-up) inference. Under causal inference, the state of a node, referred to as the query-node, is inferred from the state of its parent node(s), referred to as the evidence node(s) (i.e, the state of the evidence node(s) is *caused* by the parent node(s)). Diagnostic inference is where we infer the state of a node from its children node(s) (Nilsson, 1998), and is our area of interest. This occurs as the discretised predictor variables are fitted to each of their respective local CPTs and working against the direction of the network edges, CPTs are adjusted all the way to the root node, where a state of financial distress is revealed.

6.2 Methodology

The initial steps in the methodology involve using a discretised time-dependent training set that contained no missing values, to build a direct acyclic graph and calculate the global joint probability distribution of the network. This is performed using the algorithmic techniques detailed in Chapter 4. The remaining portion of the data can then be used to test the model and predict whether a company is financially distressed through diagnostic inference. All models are run using *Weka*, which is a software package for machine learning written in Java and developed by the University of Waikato.

One of the primary aims of this study is to test the applicability of various user-defined decisions in impacting model performance. We therefore want to empirically examine, how different user-defined choices will affect predictive accuracy and determine which of these choices allow us to arrive at an optimal Bayesian network. Overall in excess of 100 permutations are run, and below we seek to outline the areas being investigated:

- **Number of intervals of discretisation:** 2, 5, 10, 15 and 20 as choices for the number of discrete intervals.
- **Learning Algorithms:** For the Bayesian Network model, we test how the K2 and MMHC algorithm impact performance.
- **Bayesian Models:** A naive Bayes classifier and a standard Bayesian network are tested against each other.
- **Output Variables:** Using the *year of bankruptcy*, *one year before bankruptcy* and *cumulative indicator*, we seek to deduce how model performance differs as the dependent variable is changed.

6.3 Performance Evaluation

The *Weka* software gives a large array of output after a model has been tested. One of the weaknesses of previous research in this area has been that only predictive accuracy has been reported. This can be very concerning as this does not account for issues such as the number of 'false positives' predicted, which can be a very important consideration for going concern. Therefore, besides predictive accuracy there is a need for more robust measures, in order to conclusively assess the success of a model. In this section, we introduce the metrics that were considered and reported in this study.

6.3.1 Precision and Recall

Let TP, FP and FN denote true positives, false positives and false negatives respectively. Precision and recall are the simplest measures in classification. Precision is the fraction of retrieved instances that are relevant, while recall is the fraction of relevant instances that are retrieved. Both statistics can be broken down from an overall measure into bankruptcy and non-bankruptcy measures. They are calculated as follows

$$\begin{aligned} Precision &= \frac{TP}{TP+FP}, && (6.1)\\ &= \frac{\text{Correctly classified bankruptcies}}{\text{Total bankruptcies predicted}}. && (6.2)\end{aligned}$$

$$\begin{aligned} Recall &= \frac{TP}{TP+FN}, && (6.3)\\ &= \frac{\text{Correctly classified bankruptcies}}{\text{Total bankruptcies in sample}}. && (6.4)\end{aligned}$$

The sub-metrics for bankruptcy specifically are shown in Equations (6.2) and (6.4), where the non-bankruptcy measure follows similarly. Recall would therefore be the 'predictive accuracy' as normally defined while a low precision would indicate a large degree of false positive predicted. A statistic that combines precision and recall as an overall evaluative indicator is the F-measure (also known as the F1-score) which is the harmonic mean of the two above metrics.

$$F = 2 \cdot \frac{precision \cdot recall}{precision + recall}. \qquad (6.5)$$

The coefficient of the F-measure has no intuitive interpretation other than being a combined metric to measure a test's overall accuracy. Equal weighting is given to both precision and recall and therefore is useful in that it allows for both evaluative factors to be considered in one measure.

6.3.2 Kappa Statistic

The kappa statistic (also known as Cohen's kappa coefficient) is a measure of agreement of prediction with the true classification (observed categorisations), while correcting for agreements that occur by chance, and hence is a more robust measure than the F-score. This can be defined as follows

$$\kappa = \frac{Pr(a) - Pr(e)}{1 - Pr(e)}. \qquad (6.6)$$

$Pr(a)$ is the relative observed agreement (proportion of companies correctly predicted) and $Pr(e)$ is the hypothetical probability of agreement by chance ($Pr(e) = 0.5$ when predicting for binary outcomes). A kappa of 0 would indicate that correct classifications are a result of 'chance agreement' and a kappa of 1 would indicate perfect agreement (Viera and Garrett, 2005).

6.3.3 Area under ROC Curve

Another output statistic that can be used for evaluation in Bayesian Networks is the area under the Receiver Operating Characteristic (ROC) curve. The ROC curve is a graphical plot that illustrates the performance of a binary classifier system as its discrimination threshold is varied. The curve is created by plotting the true positive rate against the false positive rate for every possible classification threshold. Accuracy of the test depends on how well the test separates the two classification possibilities. A value of 0.5 indicates a random model with no predictive ability and a value of 1.0 indicates perfect discrimination (Chava and Jarrow, 2004). An example graph is presented below from Tape (2015). This measure will prove very useful as it allows for comparison across different mathematical models which is used in Chapter 8.

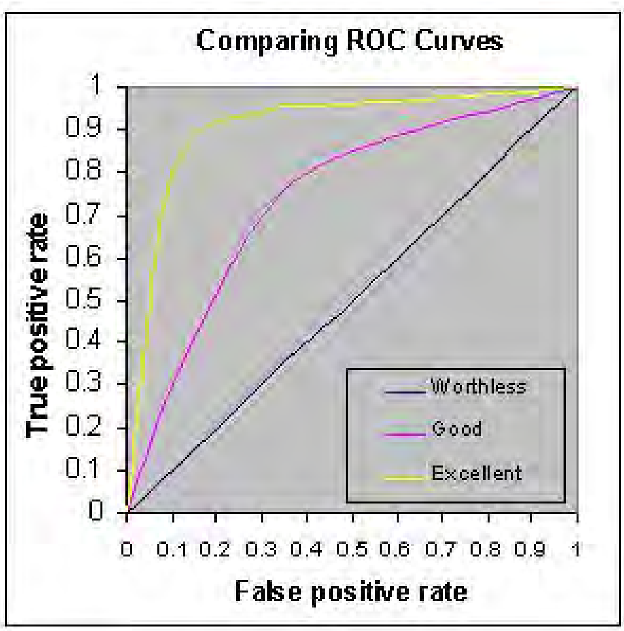

Fig. 6.1: Illustration of a ROC Curve.

Chapter 7

Results

In the previous chapter, we presented the various permutations of Bayesian models that are being tested in this research. Therefore, this chapter aims to isolate the most important findings that have been observed, while complete breakdown of results for each model augmentation has been included in Appendix B.1. We wish to reach a conclusion on the best methods for each augmentation and therefore examine how intervals of discretisation, different Bayesian models, learning algorithms and time-varying dependent variables perform against each other.

7.1 Intervals of Discretisation

The optimal number of intervals of discretisation is an area of interest since there is no conclusive choice. Below we have plotted four metrics, *precision*, *recall*, *F-measure* and *ROC area*, that assess the success of the model against the five predefined choices for the number of intervals. To exclusively focus on discretisation, we do not break down metrics for *bankrupt* and *non-bankrupt* but we do average across predictor variables and learning algorithms, in the case of Bayesian networks.

Given the previous research by Sun and Shenoy (2007) and Aghaie and Saeedi (2009), one would expect some sort of concave shape where accuracy peaks at five and tails off once it passes that level. However the results exhibited in the graph below reveal increasing accuracy for more intervals, peaking at 20, and is a result consistent for both models. It should be noted that the axis for performance accuracy has been adjusted to between 70% and 90% and as a result, the pattern looks slightly accentuated.

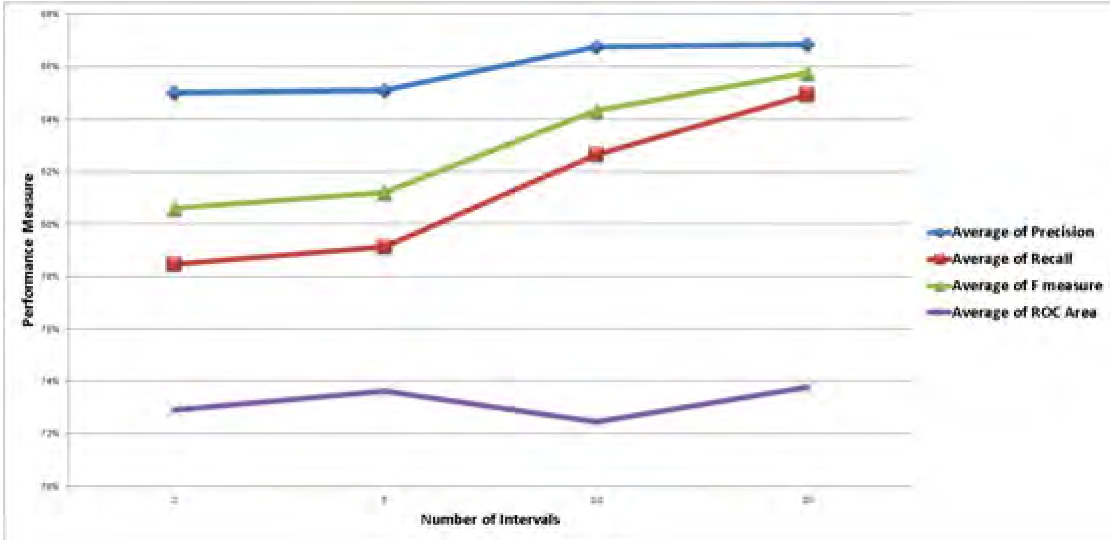

Fig. 7.1: Effect of Different Intervals of Discretisation for Bayesian Networks.

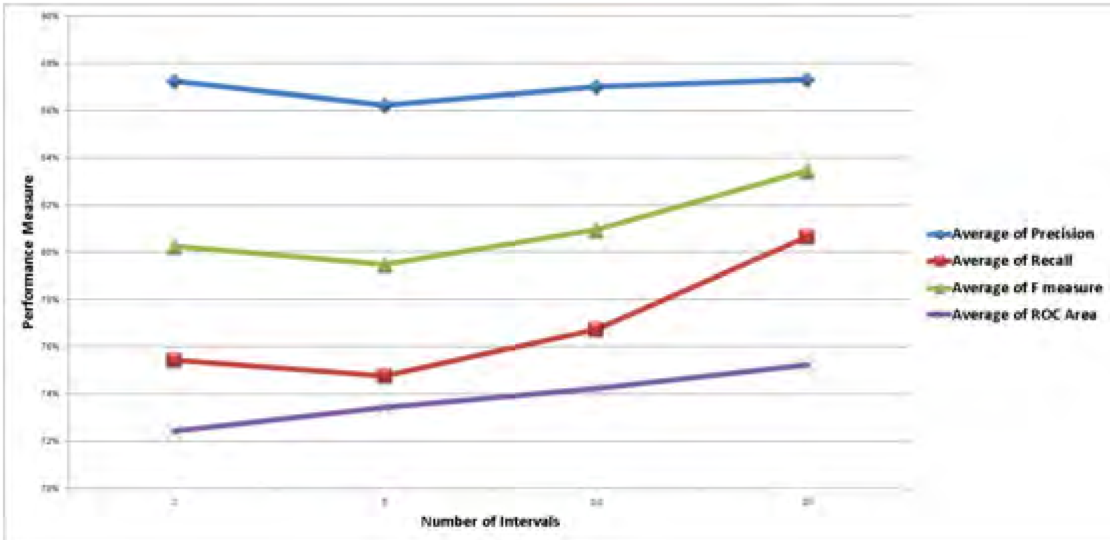

Fig. 7.2: Effect of Different Intervals of Discretisation for Naive Bayesian Classifier.

The results do make some intuitive sense when one considers the nature of the discrete variables. Due to the volatile distributional shape, having fewer intervals can pair values that are vastly different in the same category, resulting in the network losing out on some of the richness in the data. It would make sense, given the upward shape of the lines in both figures, to test for a number higher than 20; however, it was found in many cases that a solution could not be reached for these higher numbers as a result of the learning algorithm not being able to optimise with so many attributes.

7.2 Learning Algorithms for Bayesian Networks

As mentioned earlier, the learning algorithm works in conjunction with a scoring metric and in Chapter 4 we introduced two of the more prominent measures. Throughout the testing process, it was found that AIC and BIC as well as other measures, such as MDL and *entropy*, available in the software package *Weka*, arrived consistently at the same result.

A lot of the complexity of Bayesian networks arises from the understanding of how a learning algorithm arrives at the structure and parameters of the model. The algorithms employed, K2 and MMHC, produced differing results for each permutation they were run on; however, the results were not very dissimilar in the patterns followed.

Since it is the learning algorithms that require discrete variables, it is interesting to note from Figure 7.1 that for both learning algorithms, *recall* in particular was greatly impacted by increasing the number of intervals. However, specifically zoning in on predicting bankruptcy[1] in Table 7.1, a different pattern is observed.

Across the different permutations in Table 7.1, the first noticeable finding is that results are poor with *recall* and *precision* never exceeding 67%. Furthermore, in contrast to the discretisation graphs, it appears that the MMHC algorithm has decreasing success as the number of intervals increases. One of the issues realised is that given the computational complexity of learning algorithms, it is exceedingly difficult to explain why patterns like these occur with the limited size of the dataset being one of the likeliest conclusions.

A number of further deductions can be made. Firstly, across the three dependent variables, it is not surprising that the cumulative indicator (*CI*) is far more successful given that it has a far greater sample of firms that would be denoted bankrupt. Secondly, in terms of the original objective, it does appear that the K2 algorithm dominates the HC algorithm in almost all cases.

It is worth noticing that coupled with the low *recall*, the very low *precision* is a particularly unfavourable characteristic. This means we do not have a type of 'conservative' model which, together with predicting few of the realised bankruptcies, often falsely predicts bankruptcy for stable companies.

Before moving onto the results of the naive Bayes model, since no research has been conducted specifically on Bayesian networks and financial distress before, it is worth considering if there are any ways of further altering the model to improve predictive accuracy. Two alternatives have been identified and are examined below.

[1] The graphs for discretisation, are a weighted average for predicting both non-bankruptcy and bankruptcy and due to being a much larger part of the sample are weighted towards non-bankruptcy predictions.

Tab. 7.1: The Association between Learning Algorithms, Discretisation and Prediction Variables for Bankruptcy Prediction.

		MMHC			K2		
Variable	**Discrete Intervals**	**Precision**	**Recall**	**F-Score**	**Precison**	**Recall**	**F-score**
0y	2	0.062	0.314	0.104	0.074	0.371	0.123
0y	5	0.083	0.286	0.129	0.083	0.343	0.134
0y	10	0.000	0.000	0.000	0.111	0.371	0.171
0y	15	0.000	0.000	0.000	0.116	0.286	0.165
0y	20	0.000	0.000	0.000	0.105	0.286	0.154
0y Average		0.029	0.120	0.046	0.098	0.331	0.149
1y	2	0.047	0.214	0.077	0.074	0.321	0.120
1y	5	0.069	0.174	0.099	0.069	0.321	0.114
1y	10	0.000	0.000	0.000	0.088	0.321	0.138
1y	15	0.000	0.000	0.000	0.073	0.250	0.113
1y	20	0.000	0.000	0.000	0.080	0.214	0.116
1y Average		0.023	0.078	0.035	0.077	0.285	0.120
CI	2	0.405	0.616	0.489	0.406	0.626	0.493
CI	5	0.451	0.610	0.519	0.441	0.616	0.514
CI	10	0.471	0.628	0.538	0.483	0.644	0.552
CI	15	0.524	0.653	0.581	0.519	0.653	0.578
CI	20	0.524	0.659	0.584	0.521	0.667	0.589
CI Average		0.475	0.633	0.542	0.474	0.643	0.545

7.2.1 Markov Blanket

The Markov blanket of a variable X is the smallest set containing all variables carrying information about X that cannot be obtained from any other variable (Pellet and Elisseeff, 2008). A Markov blanket can be incorporated into both the learning algorithms in use. This is put in place at the end of the traversal of the search space, by adding a heuristic which is used to ensure each of the attributes are in the Markov blanket of the classifier node (Bouckaret, 2004).

Table 7.2 produces the same permutations as Table 7.1 with the addition of a Markov blanket heuristic. There is a minor increase in the performance of the first two variables, but all results are still less than 50%. Again, this would be attributable to the small number of bankruptcies in the sample. To put this in perspective, with a training set of 40% and 60 bankruptcies occurring, the learning process would only have approximately 24 bankruptcies to learn from in a sample of 430 entries. The situation is intensified for the $1y$ variable where there are in total only 44 bankruptcies in the entire sample, before learning.

There are, however, promising signs with the cumulative indicator variable exhibiting a definite indication of improvement in recall with success as high as 68.1%.

Increased precision is also noticed and looking at the combined evaluation metric, we can conclude that the Markov blanket heuristic does improve a Bayesian network for financial distress prediction.

Tab. 7.2: The Association between Learning algorithms, Discretisation and Prediction Variables for Bankruptcy Prediction using a Markov Blanket Heuristic.

		MMHC			K2		
Variable	**Discrete Intervals**	**Precision**	**Recall**	**F-Score**	**Precison**	**Recall**	**F-score**
0y	2	0.081	0.200	0.115	0.074	0.371	0.123
0y	5	0.200	0.143	0.167	0.078	0.343	0.127
0y	10	0.000	0.000	0.000	0.108	0.371	0.167
0y	15	0.000	0.000	0.000	0.116	0.286	0.165
0y	20	0.000	0.000	0.000	0.111	0.314	0.164
0y Average		0.056	0.069	0.056	0.097	0.337	0.149
1y	2	0.061	0.107	0.078	0.074	0.321	0.120
1y	5	0.073	0.143	0.097	0.076	0.357	0.125
1y	10	0.000	0.000	0.000	0.087	0.321	0.137
1y	15	0.000	0.000	0.000	0.085	0.286	0.131
1y	20	0.000	0.000	0.000	0.080	0.214	0.116
1y Average		0.027	0.050	0.035	0.080	0.300	0.126
CI	2	0.441	0.651	0.526	0.429	0.634	0.512
CI	5	0.500	0.644	0.563	0.455	0.620	0.525
CI	10	0.526	0.626	0.572	0.491	0.644	0.557
CI	15	0.551	0.626	0.586	0.532	0.656	0.588
CI	20	0.590	0.601	0.595	0.5	0.687	0.599
CI Average		0.522	0.630	0.568	0.481	0.648	0.556

7.2.2 Number of Parents per Node

With the primary objective of each model being to predict financial distress, often a trade-off with computational complexity and causality occurs. The number of parents per node in the previously mentioned results has been restricted to one parent, meaning that a single variable is inferred from another, resulting in a causality restriction. Previous research in Bayesian networks generally restricts the maximum number of parent nodes to at most three, if solutions can be found. Somewhat counter-intuitively, removing this restriction does not improve performance and actually leads it to decline. The results of this finding are reported in Appendix B.4. This may speak to the weakness of using Bayesian networks as despite giving the model more 'freedom', performance has not improved.

7.3 Naive Bayes vs. Standard Bayesian Network

Due to the simplicity of the naive Bayes structure and its success in previous research, we would expect better results than were observed with the Bayesian network. Again, analysing the results for non-bankruptcy prediction can lead us to overestimating the effectiveness of the model and therefore we produce the results only for bankruptcy prediction in Table 7.3.

Tab. 7.3: The Association between Prediction Variables and Discrete Intervals for Bankruptcy Prediction using a Naive Bayes Classifier.

Variable	**Discrete Intervals**	**Precision**	**Recall**	**F-Score**
0y	2	0.066	0.343	0.111
0y	5	0.079	0.343	0.128
0y	10	0.103	0.400	0.164
0y	15	0.105	0.343	0.161
0y	20	0.120	0.429	0.188
0y Average		0.095	0.372	0.150
1y	2	0.091	0.393	0.148
1y	5	0.072	0.357	0.120
1y	10	0.082	0.321	0.131
1y	15	0.075	0.286	0.119
1y	20	0.074	0.250	0.114
1y Average		0.079	0.321	0.126
CI	2	0.409	0.620	0.493
CI	5	0.449	0.626	0.523
CI	10	0.484	0.644	0.553
CI	15	0.517	0.669	0.583
CI	20	0.507	0.693	0.586
CI Average		0.473	0.650	0.547

Conclusions similar to the Bayesian network case are reached, whereby results are very disappointing for the first two prediction variables with very little trust in the model at identifying bankrupt companies. There are, however, much better results using the cumulative indicator with accuracy very close to that computed in Table 7.2, and no conclusive answer can be reached on which Bayesian model better predicts bankruptcy.

Therefore, we attempt a more comprehensive assessment in Table 7.4, where a variety of more robust metrics are displayed to compare the models. Using what has been previously deduced, the results are shown for 20 intervals of discretisation and the K2 algorithm (using a Markov blanket heuristic) in the case of the Bayesian network, and using the cumulative indicator variable for prediction.

Tab. 7.4: Naive Bayes vs. Bayesian Network.

	Bayes			Naive Bayes		
Metric	**Bankrupt**	**Non-Bankrupt**	**Overall**	**Bankrupt**	**Non-Bankrupt**	**Overall**
Precision	0.531	0.887	0.800	0.507	0.887	0.793
Recall	0.687	0.802	0.774	0.693	0.78	0.759
F-measure	0.599	0.843	0.783	0.586	0.83	0.77
Kappa Statistic	-	-	0.4453	-	-	0.4214
ROC Area	-	-	0.827	-	-	0.823

Comparing all the above metrics for the two Bayesian models, it is immediately apparent that there are minor differences between them. The Bayesian network does have slight dominance on all overall metrics, but never by more than 3%. Despite the complexity of the learning algorithms in the construction of the Bayesian network, it would make sense for the results to converge due to their similarities.

The difference between the two models can be explained as the naive Bayes is restricted, in that the financial distress variable is the parent for all nodes and hence there is only one level in the graph for predictor variables. In contrast, the Bayesian network does allow for predictor variables to be parents of other predictor variables and hence there is an unlimited number of levels for these evidence nodes. Looking at the graph for the Bayesian network in Appendix B.2, we can observe that despite giving the Bayesian network the ability to have many levels, all but six variables are children of the root node. This means that it was found that most predictor variables are conditionally independent of one another. Therefore, the graphs for both models are very similar with the Bayesian network marginally improving accuracy as it does not contain a causality restriction.

Chapter 8

A Comparison with Other Models

Due to the interest of many parties in bankruptcy of a company, various methodologies have been used for prediction purposes. The two most commonly used methods in previous literature been have the Altman Z-score and Logit model owing to their simplicity and success. Using the same sample, we test these models for bankruptcy prediction to assess the feasibility of Bayesian models in comparison.

8.1 Altman's Z-score

Altman (1968) introduced the Z-score to predict the probability that a public firm will go bankrupt a year in advance. In spite of the vast research on failure prediction, the original Z-score Model introduced by Altman (1968) has been the dominant model applied all over the world, in research and practice. Since the original model, updated versions have been released including specific models for private companies and manufacturing and non-manufacturing firms. Z-score models consist of a linear combination of variables estimated using multidiscriminant analysis (MDA), which classify observations into one of a number of pre-specified categories or groups (Jackson and Wood, 2013).

Initially, we test against the original Altman model, as it is still regarded as one of the most successful models bankruptcy prediction. The model takes the following form:

$$Z = 1.2A + 1.4B + 3.3C + 0.6D + 1E, \tag{8.1}$$

where

- A=Working Capital/ Total Assets.
- B=Retained Earnings/Total Assets.

- C= EBIT/ Total Assets.
- D=Market Value of Equity/Total Liabilities.
- E=Revenue/Total Assets.

Each company is given a score as an output and specific thresholds to classify scores into three 'zones' are defined, where the 'distress zone' means a company is heading for bankruptcy. Using the Altman (1968) model on the JSE sample of bankrupt companies, we present the results in Table 8.1. Across all three variables, the model performs poorly, identifying less than half of the bankruptcies. Therefore, despite the popularity of the Z-score, it is not useful on our South African dataset and therefore, we seek to use one of the augmentations of this model that may be better suited.

Tab. 8.1: Altman Z-Score Predictions for Bankrupt Companies.

Altman Prediction	**Year of Bankruptcy**	**Year Before Bankruptcy**	**Cumulative Indicator**
Distress Zone	39.22%	37.50%	40.91%
Safe Zone	58.82%	60.00%	57.02%
Grey Zone	1.96%	2.50%	2.07%
Grand Total	100.00%	100.00%	100.00%

Altman (2005) introduced the EMS model for emerging market corporate bonds to better incorporate the particular characteristics of emerging market companies. Despite the model relating to corporate credit, it can easily be related to bankruptcy prediction as the thresholds were reported that classify output scores again into *safe, grey* and *distress* zones. The format of the model can be expressed as follows with the same variables as defined above, except that this model has dropped the E coefficient and included a constant term:

$$\text{EM Score} = 6.56A + 3.26B + 6.72C + 1.05D + 3.25. \tag{8.2}$$

The results of the EMS model are expressed in Table 8.2 but are worse than those of the original 1968 model, predicting just over a third of bankruptcies. This may indicate the difficulty of predicting bankruptcies in South Africa. Furthermore, it is worth noting that despite the success of the Altman model in previous literature, it is a model designed specifically for US companies and there would be structural differences in the performance indicators of South African companies and therefore it would not be a suitable model. With regard to the EMS model, despite its being termed a model for emerging markets, the coefficients in the model were derived using Mexican firms, and again there is no surprise at its lack of suitability to South African firms.

Tab. 8.2: Altman Emerging Market Model Predictions for Bankrupt Companies.

Altman Prediction	Year of Bankruptcy	Year Before Bankruptcy	Cumulative Indicator
Distress Zone	31.37%	37.50%	36.78%
Safe Zone	68.63%	62.50%	62.40%
Grey Zone	0.00%	0.00%	0.83%
Grand Total	100.00%	100.00%	100.00%

8.2 The Logit Model

The logit model is a conditional probability model which uses a non-linear maximum log-likelihood technique to estimate the probability of firm failure under the assumption of a logistic distribution (Jackson and Wood, 2013). Jackson and Wood (2013) explain the logit model is often favoured over the Altman's model as it circumvents some of the important statistical assumptions violated by MDA. Logit models have been favoured in financial distress prediction literature as a benchmark for relative success.

We build the logit model on our current sample using each of the different classification variables as the dependent variable. The same sample of companies is then tested on the built model for inference purposes with the results displayed in Table 8.3. The commonly accepted threshold of greater than 0.5 is used for indicating bankruptcy, as this simply would mean that the probability of financial distress is greater than 50%.

Tab. 8.3: Results from the Logit Model.

Variable	Total Cases	Predicted Correctly	Predictive Accuracy %	R^2	Area under ROC Curve
Cumulative Indicator	280	89	31.79%	0.123	0.73
Year of bankruptcy	59	8	13.56%	0.10	0.72
1 year before bankruptcy	44	10	22.73%	0.19	0.79

The predictive accuracy for bankrupt companies is extremely poor for all three variables and far worse than either of the Bayesian models or the Altman model. Two additional metrics are included in the table to find any justification in the model.

The R^2 statistic[1], which is a measure of how much of the variation in the dependent variable is explained by the predictor variables, demonstrates the lack of explanatory power in the model. The ROC area metric was included as further validation of poor performance, as it gives an indication of success when the threshold is varied, with the realised values far lower than those in Table 7.4.

Focusing simply on predictive ability, it can be deduced that both Bayesian models surpass the Altman and logit models. This result should be taken with caution as, despite their low predictive abilities, there are still advantages in using these classical models. Both models, in contrast to the Bayesian network, give an indication of the degree of bankruptcy as the final output, can be quantified and not only classified (Bayesian classification models give a binary output). This is useful as even if a company does not surpass a predefined threshold, the result can allow for company comparisons.

With regard to the logit model, another advantage is that the detailed output allows for the effect of each predictor variable to be quantified by the realised coefficient. Despite Bayesian networks showing causality relations in their DAG, the relationships cannot be quantified. The coefficients of a logit model would be of great interest in determining what are isolated causal effects of variables on financial distress. Tables showing the logit model output have been included in Appendix C and despite very low levels of significance, interesting outcomes can be deduced. These include that the Debt to Asset ratio has the strongest positive effect on the probability of bankruptcy, while the JSE ALSI return has the largest negative effect. Furthermore, analysing the coefficients of the sectors, we find that consumer goods and consumer services companies are most likely to increase the probability of bankruptcy while industrials have a negative effect on the probability.

8.3 Conclusion on which Model to Use in Practice

Despite predictive accuracy never exceeding 70% for bankruptcy prediction, the Bayesian network results strongly surpassed the success of the Altman Z-score and logit models. Therefore, even if these models are more trusted in practice, it may make sense to use them in conjunction with Bayesian networks. This will allow for the model with the strongest predictive characteristics to be employed and have a score or probability as a validation or indication of the degree of financial distress.

[1] This is a pseudo R^2 as logistic regressions do not produce a conventional R^2 measure because they use maximum likelihood estimation.

Chapter 9

Discussion

The observed success rates for all models in Chapters 7 and 8 have not achieved any 'spectacular' success rates in contrast to previous financial distress studies. This is most likely due to issues relating to our South African dataset. Therefore, this section expands on why this may be the case before moving on to a comparison to other financial distress studies.

9.1 Assessing the Applicability of Bayesian Networks for South African Use

The biggest constraint in this study was the availability of data with very few bankruptcies from which the model can learn. This is, however, out of our control as all bankruptcies occurring over the last 14 years for publicly-listed companies in South Africa were captured in the sample. In total more than 40 000 companies declared bankruptcy in South Africa from 2000 to 2013 (Tradingeconomics.com, 2014); however, only 66 of these were publicly listed companies. The proportion of entries was already very skewed towards non-bankrupt companies and therefore adding more non-bankrupt companies was not deemed to be a feasible solution.

The reason that such a small proportion of all bankrupt companies are listed companies, could be attributable to many of them de-listing prior to bankruptcy. Companies which de-listed from the JSE were not included in our definition of financial distress due to their reasons for leaving the exchange not being available. This however, could have biased our model as de-listed companies may often have performance indicators very close to that of the bankrupt companies, therefore making it difficult for the model to separate companies into two dichotomous groups. It is further worth noting that these groups are not always easily distinguishable on the 'success-failure continuum' as bankruptcy is not declared at some specific point. A large proportion of entries in the sample were lower mid-cap companies and hence their performance indicators are less stable over time, further complicating the learning process.

The long length of the sample (14 years) could also hamper the model's performance. The inclusion of the macroeconomic variables was to explain these exogenous market-based factors and structural changes in the economy; however, not everything could be captured by these variables. To put this in perspective, our training set, which covered the period approximately from 2000 to 2005 would contain the 'tech bubble' of 2002. In contrast, the subset used for inference from 2006 to 2013 would contain the global financial crisis. Therefore, an exogenous shock with specific market characteristics will be embedded in the created model but when used for inference the same characteristics are not prevalent in the market while new extremities cannot not be accounted for.

A remedy for this issue could be to use shorter periods for training and inference that are closer to each other and as a result more structurally alike. However, in the case of South African data, we are constrained in having only 66 bankruptcies and therefore any subset of this would have too few observations of interest. It is important to remember that this sample of bankrupt companies is condensed into the learning and inference samples. The choice of using a conservative 40% proportion for learning was in contrast to previous studies and could be another reason behind low prediction measures, as is detailed in the next section.

9.2 Comparison with Previous Research

The three earlier Bayesian Network studies in this field had highly impressive success rates and were far better than any predictive accuracy realised in this study. In contrast to the previous research, we have attempted to be as mathematically rigorous as possible by applying specific Bayesian learning algorithms to construct the model instead of using expert knowledge and conditional correlation techniques.

A key reason across predictive studies for potential model overfitting is using a large proportion of the data for training and leaving only a small set for inference purposes. This often results in immediate success but has limited predictive power the further out of sample we move. The choice in this study of 40% is far lower than previous research as can be seen in Table 9.1 which presents a summary of their findings.

Tab. 9.1: Previous Research Predicting Financial Distress using a Naive Bayes Classifier.

Research	Sample	Sample Size	Training set size	Discrete Intervals	Non-Bankrupt Predictive Accuracy %	Bankrupt Predictive Accuracy %
Sarkar and Sriram (2001)	US	1139 Banks	80%	2	89.10%	89.90%
Shenoy (2006)	US	7827 firms	10-fold	5	81.85%	81.12%
Aghaie (2009)	Iran	144 firms	Not stated	5	89.00%	90.00%

The obvious advantage of the two US studies is the large sample of data that was used and would lead to a much better basis for model construction. The other striking difference that can be observed is the high proportion of data used in the training set. To determine the impact of the training set proportion on accuracy, the same proportions used by Sarkar and Sriram (2001) and Sun and Shenoy (2007) are applied to our dataset and tested against a naive Bayes model (as this was the model used in their studies).

The results can be noted in Table 9.2, where we see a marginal increase in predictive accuracy and results that are somewhat closer to the previous research. Despite the results for 10-fold analysis improving our success rates, it is worth realising that this methodology makes little practical sense. The 10-fold algorithm picks random subsets from the data sample which have no time dependency, therefore invalidating any attempt to model the time-based relationship between the risk factors and probability of bankruptcy. The 80% subset improves performance but as hypothesised, this could be a result of model overfitting and create a model that has little out-of-sample use.

Tab. 9.2: Predictive Accuracy with Adjusted Training Set.

Traing set Proportion	Non-Bankrupt Predicted Correctly	Bankrupt Predicted Correctly
80 %	82.2 %	71.8 %
10 fold	78.4 %	70.7 %

As a final validation for model success, it would be worthwhile comparing results to South African research. Muller *et al.* (2009) tested a variety of techniques for prediction using a similar-sized JSE dataset. Using four time varying binary variables, they tested for bankruptcy one, two, three and four years in advance. Therefore, results are not comparable to the cumulative indicator variable but to the other two binary variables that have been used. Using MDA, logit analysis, neural networks and recursive partitioning, their results for predicting failed companies one year in advance never exceeded 40%. Noting the results in Appendix tables B.2 and B3, we find that our best predictive accuracies for bankrupt companies in the year of bankruptcy and in the year before bankruptcy 42.3% and 39.2% receptively. We can therefore conclude that despite experiencing low predictive accuracy for bankrupt companies, the results are aligned with that of similar models and further emphasises the difficulty of predicting financial distress for South African companies.

Chapter 10

Conclusion

In this study, the aim was to assess the feasibility of using Bayesian networks for predicting financial distress of JSE-listed companies. A study of this nature has not been conducted on South Africa data despite international research using this increasingly popular methodology. Several methodological issues relating to scoring metrics, learning algorithms and discretisation were examined in order to determine how they impact predictive success. In contrast to previous studies, which only used a naive Bayesian classifier as a prediction model, the use of a standard Bayesian network was analysed due its stronger mathematical foundation and 'freedom' within the model.

An issue of missing data for certain variables was encountered and a decision was made not to impute missing values as it could not be assumed that data was missing at random, as well as due to the highly volatile distribution of the variables. Further, through 'prediction of the predictors', the final output could be greatly affected. This truncated dataset was then split into the first 40% of the sample and used for the learning process, which allows the model to be to constructed using search and scoring methodologies. The remaining portion was then used for inference in order to test the model's predictive accuracy.

Using a variety of robust measures, the success of the different technical elements in affecting predictive accuracy was assessed. It was found that 20 intervals of discretisation gave the best results and this could be due to the volatile distribution of predictor variables requiring many intervals to appropriately categorise the data. Different scoring metrics were found to have no influence, while the K2 algorithm was found to be superior to the MMHC algorithm as a choice for network search methods. Both Bayesian models appeared to perform very similarly, with the Bayesian network performing minutely better; however, this was only after the introduction of a Markov blanket to the learning algorithms for the Bayesian network. The reason behind the Bayesian networks slight dominance could be attributed to it having fewer restrictions concerning the relations between predictor variables in comparison to the naive Bayes classifier.

Overall success rates were found to be less than 70%, at best. The biggest constraint in this was the limited data available for JSE firms, despite the entire population of bankrupt firms being used over 14 years and it was most likely for this reason that results were significantly lower than in previous research. The length of the dataset may also prove problematic as events that affect the entire market could be present in the learning sample but not in the inference sample, resulting in a weakness of the constructed model. A final finding was that of the influence of the proportion used for training, where it was established that increasing the size of the training set improves predictive accuracy. This could, however, be attributed to overfitting and this could in some part explain the very high success rates of previous results.

Bibliography

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

Aghaie, A. and Saeedi, A. (2009). Using bayesian networks for bankruptcy prediction: Empirical evidence from Iranian companies, *Information Management and Engineering, 2009. ICIME'09. International Conference on*, IEEE, pp. 450–455.

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

Altman, E. I. (2005). An emerging market credit scoring system for corporate bonds, *Emerging Markets Review* **6**(4): 311–323.

Altman, E. I., Iwanicz-Drozdowska, M., Laitinen, E. K. and Suvas, A. (2014). Distressed firm and bankruptcy prediction in an international context: A review and empirical analysis of Altman's z-score model, *Available at SSRN 2536340* .

Beaver, W. H. (1966). Financial ratios as predictors of failure, *Journal of Accounting Research* pp. 71–111.

Ben-Gal, I. (2007). *Encyclopedia of Statistics in Quality and Reliability*, Wiley and Sons.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities, *The Journal of Political Economy* pp. 637–654.

Bouckaret, R. (2004). *Bayesian Network Classifiers in Weka*, The University of Waikato.
URL: *http://www.cs.waikato.ac.nz/pubs/wp/2004/uow-cs-wp-2004-14.pdf*

Bruwer, B. and Hamman, W. (2006). Company failure in South Africa: classification and prediction by means of recursive partitioning., *South African Journal of Business Management* **37**(4).

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

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

Cooper, G. F. and Herskovits, E. (1992). A Bayesian method for the induction of probabilistic networks from data, *Machine learning* **9**(4): 309–347.

Cybinski, P. (2001). Description, explanation, prediction–the evolution of bankruptcy studies?, *Managerial Finance* **27**(4): 29–44.

Cybinski, P. J. (2000). The path to failure: Where are bankruptcy studies at now?

Donaldson, S. I., Graham, J. W. and Hansen, W. B. (1994). Testing the generalizability of intervening mechanism theories: Understanding the effects of adolescent drug use prevention interventions, *Journal of Behavioral Medicine* **17**(2): 195–216.

Gamez, J., Mateo, J. and Puerta, J. (2011). Learning b, *Data Mining and Knowledge Discovery* **22**(1-2).

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

Heckerman, D., Geiger, D. and Chickering, D. M. (1995). Learning Bayesian networks: The combination of knowledge and statistical data, *Machine learning* **20**(3): 197–243.

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

Korb, K. B. and Nicholson, A. E. (2004). Bayesian artificial intelligence.

Lauritzen, S. L. (1996). *Graphical models*, Oxford University Press.

Lin, X. (2012). Bayesian network learning and applications in bioinformatics.

Little, R. J. (1988). A test of missing completely at random for multivariate data with missing values, *Journal of the American Statistical Association* **83**(404): 1198–1202.

Margaritis, D. (2003). *Learning Bayesian network model structure from data*, PhD thesis, US Army.

Mihaela-Daciana Craciun, V. C. and Bala, C. (2014). *Methods for discretizing continuous variables within the framework of Bayesian networks*, ACTA Universitatis Apulensis.

Mitchell, T. M. (2010). *Generative and Discriminative Classifiers: Naive Bayes and Logistic Regression*, McGraw Hill, chapter 1.
URL: *www.cs.cmu.edu/tom/mlbook.html*

Muhlenbach, F. and Rakotomalala, R. (2005). Discretization of continuous attributes, *Encyclopedia of Data Warehousing and Mining* **1**: 397–402.

Muller, G., Steyn-Bruwer, B. and Hamman, W. (2009). Predicting financial distress of companies listed on the JSE-A comparison of techniques, *South African Journal of Business Management* **40**(1): 21–32.

Nannen, V. (2003). The paradox of overfitting, *Unpublished masters thesis, Rijksuniversiteit Groningen, the Netherlands* .

Nilsson, N. J. (1998). *Artificial Intelligence: A New Synthesis: A New Synthesis*, Elsevier.

Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy, *Journal of Accounting Research* pp. 109–131.

Pearl, J. (2011). Bayesian networks, *Department of Statistics, UCLA* .

Pellet, J.-P. and Elisseeff, A. (2008). Using markov blankets for causal structure learning, *The Journal of Machine Learning Research* **9**: 1295–1342.

Robinson, R. W. (1977). Counting unlabeled acyclic digraphs, *Combinatorial mathematics V*, Springer, pp. 28–43.

SACCI (2015). Business confidence index.
URL: *http://www.sacci.org.za/index.php?optioncomcontentviewarticleid29Itemid39*

Sarkar, S. and Sriram, R. S. (2001). Bayesian models for early warning of bank failures, *Management Science* **47**(11): 1457–1475.

Schenker, N. and Taylor, J. M. (1996). Partially parametric techniques for multiple imputation, *Computational Statistics & Data Analysis* **22**(4): 425–446.

SSCC (2014). Multiple imputation examples.
URL: *http://www.ssc.wisc.edu*

Sun, L. and Shenoy, P. P. (2007). Using bayesian networks for bankruptcy prediction: Some methodological issues, *European Journal of Operational Research* **180**(2): 738–753.

Tape, T. G. (2015). The area under an roc curve.
URL: *http://gim.unmc.edu/dxtests/roc3.htm*

Taylor, H. M. and Karlin, S. (1998). An introduction to stochastic modeling.

Tradingeconomics.com (2014). South Africa bankruptcies — 1980-2014 — data — chart — calendar — forecast.
URL: *http://www.tradingeconomics.com/south-africa/bankruptcies*

Tsaardinos, I., Brown, L. E. and Aliferis, C. F. (2006). The max-min hill-climbing bayesian network structure learning algorithm, *Machine learning* **65**(1): 31–78.

Tsamardinos, I., Aliferis, C. F., Statnikov, A. R. and Statnikov, E. (2003). Algorithms for large scale Markov blanket discovery., *FLAIRS Conference*, Vol. 2003, pp. 376–381.

Udo, G. (1993). Neural network performance on the bankruptcy classification problem, *Computers & industrial engineering* **25**(1): 377–380.

Viera, A. J. and Garrett, J. M. (2005). Understanding interobserver agreement: the kappa statistic, *Fam Med* **37**(5): 360–363.

Wang, L., Chen, L., Liu, Z., Zheng, M., Gu, Q. and Xu, J. (2014). Predicting mtor inhibitors with a classifier using recursive partitioning and naïve Bayesian approaches, *PloS one* **9**(5): e95221.

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

Appendix A

Sample Information

A.1 Variable Details

Tab. A.1: List of Variables with Calculations.

Variable	**Calculation**
Accounts Rcvb/Tover	(Turnover for 12 months / Months covered by financial statements) / Debtors
Assets / Captal Emp	Total assets / Employment of capital
Book Val / Share (c)	(Ordinary shareholders interest / Number of ordinary shares in issue at year-end) x 100
Cash Flw / Share (c)	(Bottom line earnings + items not representing cash flow) / Number of ordinary shares in issue at year-end) x 100
Cash Flow Div Cover	(Bottom line earnings + Items not representing cash flow) / Number of ordinary shares in issue at year-end) / (Dividends per share / 1 000)
Cash Flow Intr Cover	(Gross income + Items not representing cash flow) / Interest and finance charges
Current Ratio	Current assets / Current liabilities
Debt / Assets	(Long-term liabilities + Current liabilities) / Total assets
Debt / Equity	Long-term liabilities + Current liabilities) / Ordinary shareholders interest + Preference shares + Outside shareholders interest + Directors & shareholder loans Intangible assets)
Dir Rem % Pft BTax	(Directors emoluments / (Investment income + Operating profit + Interest received Interest and finance charges)
Dividend / Share (c)	Dividends per share / 10
Dividend Cover	Headline earnings per share / Dividends per share
Dividend Yield %	(Dividends per share / 1 000) / (Share price at financial yearend / 100) x 100

Variable	**Calculation**
Earnings/ Share (C)	Headline earnings per share / 10
Earnings Yield %	(Headline earnings per share / 1 000) / (Share price at financial year-end / 100) x 100
Inflation Adjusted Profit/ Share (C)	(Profit attributable to ordinary shareholders Inflation adjusted depreciation on fixed assets) / Number of ordinary shares in issue at year-end x 100
Inflation Adjusted Return On Assets %	(Investment income + Operating profit + Interest received + associate companies Inflation adjusted depreciation on fixed assets) / (Total assets + Inflation adjusted other fixed assets)) x 100
Inflation Adjusted Return On Equity %	(Profit attributable to ordinary shareholders Inflation adjusted depreciation on fixed assets) / (Ordinary shareholders interest + Directors & shareholders loans + Inflation adjusted other fixed assets)) x 100
Interest Cover	Gross income / Interest & other finance charges
Leverage Factor	(Profit attributable to ordinary shareholders x Total assets) / ((Ordinary shareholders interest + Directors & shareholders loans) x (Gross income Taxation))
Long-Term Loans % Total Debt	(Long-term liabilities / (Long-term liabilities + Current liabilities)) x 100
N A V / Share (C)	((Total assets Long-term liabilities Current liabilities) / Number of ordinary shares in issue at year-end) x 100
Net Profit Margin %	(Bottom line earnings / Turnover) x 100
Operating Profit /Employee	Operating profit / No of persons employed) x 1 000
Operating Profit Margin %	(Operating profit / Turnover) x 100
Price / Inflation Adjusted Profit	(Share price at financial year-end / 100) / (Investment income + Operating profit + Interest received + Associate companies Inflation adjusted depreciation on fixed assets) / Number of ordinary shares in issue at year-end)
Price / Book Value	(Share price at financial year-end / 100) / (Ordinary shareholders interest / Number of ordinary shares in issue at year-end)
Price / Cash Flow	(Share price at financial year-end / 100) / (Bottom line earnings + Items not representing cash flow) / Number of ordinary shares in issue at year-end
Price / Earnings	(Share price at financial year-end / 100) / (Headline earnings per share / 1 000)

Variable	Calculation
Price / N A V	(Share price at financial year-end / 100) / (Total assets Long-term liabilities Current liabilities) / Number of ordinary shares in issue at year-end
Price / Share (C)	Share price at financial year-end
Quick Ratio	(Current assets Inventory) / Current liabilities
Return On External Investments %	((Investment income + Interest received) / Investments & loans) x 100
Retention Rate	(Retained earnings in current year / Bottom line earnings) x 100
Return On Assets %	((Investment income + Operating profit + Interest received + Associate companies) / Total assets) x 100
Return On Equity %	(Profit attributable to ordinary shareholders / (Ordinary shareholders interest + Directors & shareholders loans)) x 100
Total Assets / Turnover	(Turnover x 12 months / Months covered by financial statements) / Total assets
Total Debt / Cash Flow	(Long-term liabilities + current liabilities) / (Bottom line earnings + Items not representing cash flow)
Turnover / Employee	Turnover / No of persons employed) x 1 000
Return on Capital Employed	Earnings Before Interest and Tax (EBIT) / Capital Employed
Price / EBITDA	Share Price/ Earnings before Interest, Taxes, Depreceation and Amortisation (EBITDA)
Price / EBIT	Share Price/EBIT
Price / Cash	Share Price/(Operating Cash Flow per Share
Return on Average External Investments %	Net Income / Total Average Assets
Return on Average Assets %	((Investment income + Operating profit + Interest received + Associate companies) / Average assets) x 100
Return on Average Equity %	(Profit attributable to ordinary shareholders / (Average Ordinary shareholders interest + Average Directors & shareholders loans)) x 100
Inflation Adj. Return on Average Total Assets %	(Investment income + Operating profit + Interest received + associate companies Inflation adjusted depreciation on fixed assets) / (Average assets + Inflation adjusted other fixed assets)) x 100

Variable	Calculation
Inflation Adjusted Return on Average Equity %	(Profit attributable to ordinary shareholders Inflation adjusted depreciation on fixed assets) / (average Ordinary shareholders interest +average Directors & shareholders loans + average Inflation adjusted other fixed assets)) x 100
Cash Flow Return On Total Net Assets	(Bottom line earnings + Items not representing cash flow)/(Total Assets-Total Liabilities
Cash Flow Return On Total Net Operating Assets	(Bottom line earnings + Items not representing cash flow)/(Net Profit from Operations
Cash Flow To Total Shareholders Equity	(Bottom line earnings + Items not representing cash flow) /Number of Shareholders
Dividend Coverage	(Headline earnings per share-Interest Expense) / Dividends per share
Interest Coverage	(Profit attributable to ordinary shareholders Inflation adjusted depreciation on fixed assets-Interest Expense) / (Ordinary shareholders interest + Directors & shareholders loans + Inflation adjusted other fixed assets)) x 100
Cash Flow (Cata) To Total Debt	(Cash available from total activities after interest and tax)/(Long-term liabilities + Current liabilities)
Cash Flow (Cata) To Current Liabilities	(Cash available from total activities after interest and tax)/Current Liabilities
Cash Flow To Capital	(Bottom line earnings + Items not representing cash flow)/(Total Capital Employed)
Adequacy Ratio	(Total Capital Employed)/(Average Assets)
Reinvestment Rate	1-(Net Profits/Dividends Paid)
Cash Flow (Ncta) To Capital Investments	(Net cash from total activities)/(Total Capital Employed)
Cash Flow (Ncta) To Financial Investments	(Net cash from total activities)/(Financial Investment)
Cash Flow (Ncta) To All Investments	(Net cash from total activities)/(Total Investment)
Cash Flow (Cata) To Turnover (Margin)	(Cash available from total activities after interest and tax)/(Turnover)
Cash Flow (Cata Less Pref. Dividend) Per Share	((Cash available from total activities after interest and tax)-Pref Dividends)/(Total Number of Shares)
Price Per Share To Cash Flow Per Share	(Share Price)/(Cash Flow per Share)
Working Capital To Operating Cash Flow	(Current Assets- Current Liabilities)/(Cashflow from Operations)
Cash Flow (Cata) To Net Earnings After Tax	(Cash available from total activities after interest and tax)/(Net Profit)
Cash Flow Less Interest Paid To Income Before Tax	(Cash Flow Less Interest Paid)/(Income Before Tax)

A.2 Company Information

Tab. A.2: List and Details of Companies in Sample.

Company	Number of years of Data	Industry	Industry Group	Bankruptcy Indicator
ADAPTIT HOLDINGS	14	Computer Services	Computer Services	0
ADRENNA PROPERTY GROUP	13	Real Estate Hold, Dev	Real Estate Hold, Dev	0
ADVTECH	13	Spec.Consumer Service	Spec.Consumer Service	0
AECI	13	Specialty Chemicals	Specialty Chemicals	0
AF & OVR	14	NA	NA	0
AFRIBRAND HOLDINGS L	1	Consumer Finance	Consumer Finance	1
AH	13	Food Products	Food Products	0
ALERT STEEL HDG.	7	Home Improvement Ret.	Home Improvement Ret.	0
ALLIANCE MINING	5	Mobile Telecom.	Mobile Telecom.	0
ALLIED TECHNOLOGIES	14	Electrical Equipment	Electrical Equipment	0
ALTRON	14	Computer Services	Computer Services	0
ALUDIE	4	Electrical Equipment	Electrical Equipment	1
AMBIT PROPERTIES	5	Real Estate Hold, Dev	Real Estate Hold, Dev	1
AMLAC*	1	Consumer Goods	Auto Parts	1
APS TECHNOLOGIES	5	Medical Equipment	Medical Equipment	1
AQUILA GROWTH	5	Internet	Internet	1
ARGENT INDUSTRIAL	14	Divers. Industrials	Divers. Industrials	0
ASPEN PHMCR.HDG.	14	Pharmaceuticals	Pharmaceuticals	0
ASSORE	14	General Mining	General Mining	0
ASTRAPAK	14	Containers & Package	Containers & Package	0
AUSTRO GROUP	7	Industrial Machinery	Industrial Machinery	0
AVI	14	Food Products	Food Products	0
AWETHU BREWERIES	14	Brewers	Brewers	0
BAUBA PLATINUM	14	Plat.& Precious Metal	Plat.& Precious Metal	0
BEGET HOLDINGS	7	Internet	Internet	1
BEST CUT	1	Business Support Svs.	Business Support Svs.	1
BOLTON INDUSTRIAL HO	1	General Mining	General Mining	1
BONATLA PR.	13	Real Estate Hold, Dev	Real Estate Hold, Dev	0
BRIMSTONE INV.	13	NA	NA	0
BRYANT TECHNOLOGY	7	Computer Hardware	Computer Hardware	1
CARGO CARRIERS	14	Trucking	Trucking	0
CCI	3	Software	Software	1
CELCOM GROUP	2	Specialty Retailers	Specialty Retailers	1
CENTURY CARB	1	Ind. & Office REITs	Ind. & Office REITs	1
CITY LODGE HOTELS	14	Hotels	Hotels	0
CMH	14	Specialty Retailers	Specialty Retailers	0
COMAIR	14	Airlines	Airlines	0
COMPU CLEAR.OUTSC.	14	Computer Services	Computer Services	0

Company	Number of years of Data	Industry	Industry Group	Bankruptcy Indicator
CONAFEX HOLDINGS (JSE)	9	Food Products	Food Products	1
CONVERGENET HOLDINGS	13	Computer Services	Computer Services	0
CORE	2	Computer Services	Computer Services	1
COUNTRY FOODS	1	Farming & Fishing	Farming & Fishing	1
CROOKES BROTHERS	14	Farming & Fishing	Farming & Fishing	0
CULLINAN HOTELS & LEIS*	13	Consumer Services	Specialty Retailers	1
DATACENTRIX	14	Computer Services	Computer Services	0
DIALOGUE GROUP HDG.	5	Specialty Retailers	Specialty Retailers	1
DIGICORE	14	Electronic Equipment	Electronic Equipment	0
DNA SUPPLY CHN.INV.	4	Business Support Svs.	Business Support Svs.	1
DON GROUP	14	Hotels	Hotels	0
DORBYL	13	Auto Parts	Auto Parts	0
EC HOLD	3	Computer Services	Computer Services	1
ELEMENTONE	10	Broadcast & Entertain	Broadcast & Entertain	1
ENTER.RS.MAN.	8	Specialty Finance	Specialty Finance	1
EUREKA INDUSTRIAL	9	NA	NA	1
EXCELLERATE HDG.	12	Business Support Svs.	Business Support Svs.	0
FASHION AFRICA	3	Apparel Retailers	Apparel Retailers	1
FRONTRANGE SLTN.	6	Software	Software	1
GENCOR	6	Specialty Finance	Specialty Finance	1
GIJIMA GROUP	14	Computer Services	Computer Services	0
GLOBAL TECHNOLOGY	4	Specialty Finance	Specialty Finance	1
GLOBAL VILLAGE	6	Recreational Services	Recreational Services	1
GLODINA	4	Clothing & Accessory	Clothing & Accessory	1
GRINDROD	13	Marine Transportation	Marine Transportation	0
HARMONY GOLD MNG.	14	Gold Mining	Gold Mining	0
IDION TECH.	6	Software	Software	1
INFRASORS HOLDINGS	6	General Mining	General Mining	0
INTERTRADING	11	Farming & Fishing	Farming & Fishing	1
ISA	13	Computer Services	Computer Services	0
JD GROUP	14	Home Improvement Ret.	Home Improvement Ret.	0
KAGISO MEDIA	14	Broadcast & Entertain	Broadcast & Entertain	0
KAIROS INDL.	12	Industrial Machinery	Industrial Machinery	1
KAP INDUSTRIAL	13	Divers. Industrials	Divers. Industrials	0
KELGRAN	7	General Mining	General Mining	1
KING CONS.	9	Restaurants & Bars	Restaurants & Bars	1
KIRCHMANN	1	Real Estate Hold, Dev	Real Estate Hold, Dev	1
LABAT AFRICA	14	Electronic Equipment	Electronic Equipment	0
LONMIN (JSE)	14	Plat.& Precious Metal	Plat.& Precious Metal	0
MASTERFRIDGE LTD	1	Industrial Machinery	Industrial Machinery	1
MATHOMO GROUP	5	Apparel Retailers	Apparel Retailers	1
MERAFE RESOURCES	13	General Mining	General Mining	0
MESSINA	6	Plat.& Precious Metal	Plat.& Precious Metal	1
METAIR INVESTMENTS	13	Consumer Goods	Auto Parts	0

Company	Number of years of Data	Industry	Industry Group	Bankruptcy Indicator
METOREX	11	Home Improvement Ret.	Home Improvement Ret.	0
MILLIONAIR CHARTER	4	Delivery Services	Delivery Services	1
MUSTEK	14	Computer Hardware	Computer Hardware	0
M-WEB HOLDINGS LTD	1	Internet	Internet	1
NET 1 APPLIED TECH.	4	Computer Hardware	Computer Hardware	1
NEW AF.INVT.	9	Broadcast & Entertain	Broadcast & Entertain	1
NINIAN & LESTER	2	Clothing & Accessory	Clothing & Accessory	1
OAI	1	Software	Software	1
OCTODEC INVESTMENTS	14	Real Estate Hold, Dev	Real Estate Hold, Dev	0
ONELOGIX GROUP	14	Business Support Svs.	Business Support Svs.	0
PACIFIC	6	Hotels	Hotels	1
PALS	8	Clothing & Accessory	Clothing & Accessory	1
PASDEC RESOURCES	7	Electrical Equipment	Electrical Equipment	1
PINNACLE	14	Computer Hardware	Computer Hardware	0
PREMIUM PROPERTIES	14	Real Estate Hold, Dev	Real Estate Hold, Dev	0
PUTPROP	14	Real Estate Hold, Dev	Real Estate Hold, Dev	0
QUEENSGATE HTL.& LEIS.	9	Hotels	Hotels	1
RETAIL APPAREL GP.	2	Apparel Retailers	Apparel Retailers	1
REUNERT	14	Electrical Equipment	Electrical Equipment	0
REX TRUF.CLOTH.	14	Apparel Retailers	Apparel Retailers	0
SABLE	9	Real Estate Hold, Dev	Real Estate Hold, Dev	0
SANTOVA	11	Prop. & Casualty Ins.	Prop. & Casualty Ins.	0
SEARDEL INV.	14	Clothing & Accessory	Clothing & Accessory	0
SECUREDATA HOLDINGS	14	Computer Services	Computer Services	0
SET POINT GROUP	10	Business Support Svs.	Business Support Svs.	1
SILTEK LTD	1	Computer Hardware	Computer Hardware	1
SILVERBRIDGE HDG.	14	Software	Software	0
SOVEREIGN FOOD INVS.	14	Farming & Fishing	Farming & Fishing	0
SPUR	14	Restaurants & Bars	Restaurants & Bars	0
SQUARE ONE SLTN.GP.	9	Computer Services	Computer Services	0
SUPER GROUP*	13	Industrials	Transport Services	1
THE LASER GROUP	2	Trucking	Trucking	1
TIGER WHEELS	7	Auto Parts	Auto Parts	1
TOP-TECH	1	Computer Services	Computer Services	1
UCS GROUP	11	Software	Software	0
UNISP	1	General Mining	General Mining	1
UNITRANS	7	Business Support Svs.	Business Support Svs.	1
UNIVERSAL GROWTH HOL	1	Industrial Machinery	Industrial Machinery	1
VENTER LEIS.COML.TLR.	8	Gold Mining	Gold Mining	1
VERIMARK	14	Broadline Retailers	Broadline Retailers	0
VOLTEX	2	Electrical Equipment	Electrical Equipment	1
WESIZWE	9	Plat.& Precious Metal	Plat.& Precious Metal	0
WINECORP	4	Distillers & Vintners	Distillers & Vintners	1
ZARARA EN.	4	Real Estate Hold, Dev	Real Estate Hold, Dev	1

*Indicates dropped for fraud.

A.3 Sector Information

Tab. A.3: Number of Companies by Sector.

Industry	Number of Companies	Industry Group	Number of Companies
Basic Materials	13	Plat.& Precious Metal	4
		General Mining	6
		Gold Mining	2
		Specialty Chemicals	1
Consumer Goods	17	Brewers	1
		Food Products	3
		Farming & Fishing	4
		Clothing & Accessory	4
		Auto Parts	4
		Distillers & Vintners	1
Consumer Services	24	Travel & Tourism	1
		Apparel Retailers	4
		Home Improvement Ret.	3
		Hotels	4
		Specialty Retailers	3
		Broadcast & Entertain	3
		Broadline Retailers	1
		Restaurants & Bars	2
		Airlines	1
		Spec.Consumer Service	1
		Recreational Services	1
Financials	15	Real Estate Hold, Dev	9
		Prop. & Casualty Ins.	1
		Specialty Finance	3
		Consumer Finance	1
		Ind. & Office REITs	1
Health Care	2	Pharmaceuticals	1
		Medical Equipment	1
Industrials	24	Electronic Equipment	2
		Business Support Svs.	6
		Containers & Package	1
		Electrical Equipment	5
		Divers. Industrials	2
		Industrial Machinery	4
		Trucking	2
		Marine Transportation	1
		Delivery Services	1
NA	4	NA	4
Technology	26	Computer Services	12
		Computer Hardware	5
		Software	6
		Internet	3
Telecommunications	1	Mobile Telecom.	1

A.4 Descriptive Statistics

Tab. A.4: Descriptive Statistics and Percentage of Missing Values for Financial Variables.

Variable	Mean	Standard Deviation	Skewness	Kurtosis	Missing Values %
Accounts Rcvb/Tover	13.24	96.89	22.07	543.68	1.14%
Assets / Captal Emp	1.27	4.94	-21.48	495.43	0.00%
Book Val / Share (c)	1 239.49	3 322.82	6.84	63.46	0.00%
Cash Flw / Share (c)	270.43	867.76	9.34	127.47	0.00%
Cash Flow Div Cover	23.34	82.34	6.79	54.41	26.41%
Cash Flow Intr Cover	201.35	1 894.86	5.23	117.68	2.46%
Current Ratio	2.37	4.88	12.82	231.64	0.00%
Debt / Assets	1.53	32.70	33.68	1 135.08	0.00%
Debt / Equity	1.48	8.78	0.76	337.33	0.00%
Dir Rem % Pft BTax	35.01	467.73	27.08	821.59	1.23%
Dividend / Share (c)	98.79	214.91	6.30	56.54	26.41%
Dividend Cover	6.44	30.52	9.75	106.54	26.41%
Dividend Yield %	211.00	4 575.03	23.34	545.28	26.58%
Earnings/ Share (C)	139.20	655.24	13.89	249.32	0.09%
Earnings Yield %	-1.60	141.51	17.21	502.47	2.73%
Inflation Adjusted Profit/ Share (C)	157.07	769.50	10.36	153.74	0.00%
Inflation Adjusted Return On Assets %	-13.11	517.61	-32.83	1 097.03	0.00%
Inflation Adjusted Return On Equity %	-98.27	2 631.62	-32.36	1 071.86	0.00%
Interest Cover	108.68	2 679.02	-9.20	303.72	2.46%
Leverage Factor	1.65	17.60	-5.26	139.58	0.00%
Long-Term Loans % Total Debt	28.62	25.15	0.89	-0.07	12.68%
N A V / Share (C)	1 395.39	3 763.84	5.61	40.60	0.00%
Net Profit Margin %	-0.19	4 709.29	8.07	560.45	1.14%
Operating Profit /Employee	1.17E+05	1.14E+06	4.36	85.22	18.05%
Operating Profit Margin %	-53.65	6 286.30	-10.79	590.72	1.14%
Price / Inflation Adjusted Profit	16.66	179.18	17.64	359.89	2.64%
Price / Book Value	2.52	16.01	26.09	763.18	2.64%
Price / Cash Flow	11.02	164.67	27.85	857.03	2.64%
Price / Earnings	3.38	375.49	-9.16	402.70	2.73%
Price / N A V	2.60	17.38	18.24	563.98	2.64%
Price / Share (C)	1 629.79	5 131.84	8.69	102.27	2.64%
Quick Ratio	1.96	4.87	13.09	237.56	0.00%
Return On External Investments %	1 421.52	15 167.49	14.96	245.89	13.91%
Retention Rate	117.60	1 178.80	33.41	1 122.44	0.00%
Return On Assets %	-13.17	517.82	-32.79	1 095.24	0.00%
Return On Equity %	-25.83	585.06	-23.00	653.10	0.00%
Total Assets / Turnover	1.58	2.13	18.60	481.76	1.14%
Total Debt / Cash Flow	7.77	101.73	17.54	415.91	13.38%
Turnover / Employee	1.06E+06	1.23E+06	3.20	15.28	18.22%
Return on Capital Employed	1.38	337.06	6.21	222.80	0.00%
Price / EBITDA	2.37	40.31	-13.61	266.45	2.64%
Price / EBIT	-0.62	130.69	-18.57	466.49	2.64%
Price / Cash	320.36	3 452.65	27.32	819.39	3.70%
Return on Average External Investments %	1 725.56	26 029.11	19.44	398.77	12.06%
Return on Average Assets %	7.07	85.80	22.27	677.44	0.00%

Variable	Mean	Standard Deviation	Skewness	Kurtosis	Missing Values %
Return on Average Equity %	-9.42	317.51	-20.62	525.56	0.00%
Inflation Adj. Return on Average Total Assets %	6.37	85.64	22.42	683.03	0.00%
Inflation Adjusted Return on Average Equity %	-71.22	2 149.16	-33.14	1 109.83	0.00%
Cash Flow Return On Total Net Assets	31.57	882.33	17.94	565.05	0.00%
Cash Flow Return On Total Net Operating Assets	-117.41	3 910.09	-33.02	1 103.81	0.00%
Cash Flow To Total Shareholders Equity	8.41	155.60	-2.56	125.42	0.00%
Dividend Coverage	30.34	733.44	16.57	412.46	18.66%
Interest Coverage	22 334.44	2.04E+05	12.24	187.06	2.11%
Cash Flow (Cata) To Total Debt	38.32	323.03	2.10	191.78	0.00%
Cash Flow (Cata) To Current Liabilities	41.86	331.05	1.73	174.77	0.00%
Cash Flow To Capital	-3 545.45	1.09E+05	-32.92	1 090.91	1.85%
Adequacy Ratio	47.39	1 453.71	1.21	93.01	0.26%
Reinvestment Rate	-2.19	701.31	10.39	465.06	0.00%
Cash Flow (Ncta) To Capital Investments	-138.88	8 097.89	10.78	343.18	0.97%
Cash Flow (Ncta) To Financial Investments	-8 631.66	2.36E+05	-23.57	617.75	18.93%
Cash Flow (Ncta) To All Investments	166.63	1 300.62	17.90	352.33	0.09%
Cash Flow (Cata) To Turnover (Margin)	-329.09	10 648.88	-32.91	1 083.34	0.70%
Cash Flow (Cata Less Pref. Dividend) Per Share	2.21	9.02	12.09	199.96	1.32%
Price Per Share To Cash Flow Per Share	609.75	15 897.28	-10.36	247.13	2.38%
Working Capital To Operating Cash Flow	-83.60	1 992.14	-31.50	1 031.15	0.09%
Cash Flow (Cata) To Net Earnings After Tax	164.30	1 001.75	6.99	127.66	0.00%
Cash Flow Less Interest Paid To Income Before Tax	163.70	2 585.01	29.37	949.77	0.00%

Appendix B

Additional Results

B.1 All Results

This section includes results for each different model augmentation. Results are split by the three different output variables. Precision and recall have been decomposed into non-bankrupt (NB) and bankrupt (B). Results are aligned with conclusions reached in Chapter 7, where it was concluded that model performance is best with 20 intervals of discretisation and the standard Bayesian network, using a Markov blanket heuristic, performs marginally better than the naive Bayes classifier.

The only result that may seem contradictory to previous conclusions is related to the learning algorithms. The MMHC algorithm does appear to dominate the K2 algorithm for most of the metrics presented in Table B.1. However, for the metric of most interest, bankruptcy recall (proportion of bankrupt companies were correctly identified by the model), the K2 algorithm performs significantly better. This is most likely the result of a small number of bankrupt companies available. This deduction is emphasised in the K2 algorithms significantly better performance in Tables B.2 and B.3. Therefore, we can conclude that the K2 algorithm performs better than the MMHC algorithm in prediction from a small sample.

Tab. B.1: All Results Predicting *Cumulative Indicator* Variable.

Model	Learning Algorithm	Discrete Intervals	NB Precision	NB Recall	B Precision	B Recall	Precsion Overall	Recall Overall	F-measure	ROC	Kappa
Bayes	K2	2	0.854	0.715	0.416	0.626	0.747	0.693	0.71	0.756	0.291
Bayes	K2	5	0.861	0.752	0.451	0.626	0.76	0.721	0.735	0.788	0.3346
Bayes	K2	10	0.871	0.784	0.493	0.644	0.778	0.75	0.76	0.813	0.3884
Bayes	K2	15	0.879	0.812	0.532	0.656	0.794	0.774	0.781	0.827	0.4346
Bayes	K2	20	0.887	0.802	0.531	0.687	0.8	0.774	0.783	0.827	0.4453
Bayes	MMHC	2	0.873	0.756	0.47	0.663	0.774	0.733	0.747	0.8	0.368
Bayes	MMHC	5	0.873	0.798	0.51	0.644	0.784	0.761	0.769	0.816	0.4064
Bayes	MMHC	10	0.871	0.82	0.531	0.626	0.787	0.773	0.778	0.831	0.4209
Bayes	MMHC	15	0.873	0.834	0.551	0.626	0.794	0.783	0.788	0.836	0.4401
Bayes	MMHC	20	0.865	0.866	0.586	0.583	0.796	0.797	0.796	0.839	0.45
Naive Bayes	-	2	0.851	0.709	0.409	0.62	0.743	0.687	0.71	0.756	0.291
Naive Bayes	-	5	0.86	0.75	0.449	0.626	0.76	0.72	0.733	0.784	0.3323
Naive Bayes	-	10	0.87	0.776	0.484	0.644	0.775	0.744	0.755	0.809	0.3783
Naive Bayes	-	15	0.796	0.796	0.517	0.669	0.791	0.765	0.774	0.822	0.4231
Naive Bayes	-	20	0.887	0.78	0.507	0.693	0.793	0.759	0.77	0.823	0.4214

Tab. B.2: All Results Predicting *Bankrupt year* Variable.

Model	Learning Algorithm	Discrete Intervals	NB Precision	NB Recall	B Precision	B Recall	Precsion Overall	Recall Overall	F-measure	ROC	Kappa
Bayes	K2	2	0.954	0.733	0.072	0.371	0.908	0.714	0.792	0.669	0.0351
Bayes	K2	5	0.956	0.789	0.083	0.343	0.91	0.765	0.826	0.687	0.0529
Bayes	K2	10	0.96	0.835	0.111	0.371	0.915	0.81	0.855	0.681	0.0978
Bayes	K2	15	0.955	0.878	0.105	0.257	0.91	0.845	0.874	0.692	0.0798
Bayes	K2	20	0.956	0.865	0.105	0.286	0.911	0.834	0.868	0.695	0.0832
Bayes	MMHC	2	0.951	0.862	0.074	0.2	0.905	0.827	0.862	0.664	0.0343
Bayes	MMHC	5	0.951	0.965	0.154	0.114	0.909	0.92	0.915	0.719	0.0903
Bayes	MMHC	10	0.947	1	0	0	0.897	0.947	0.922	0.708	0
Bayes	MMHC	15	0.947	1	0	0	0.897	0.947	0.922	0.69	0
Bayes	MMHC	20	0.947	1	0	0	0.897	0.947	0.922	0.675	0
Naive Bayes	-	2	0.952	0.73	0.066	0.343	0.906	0.709	0.789	0.667	0.0243
Naive Bayes	-	5	0.955	0.777	0.079	0.343	0.909	0.755	0.819	0.687	0.0466
Naive Bayes	-	10	0.96	0.806	0.103	0.4	0.915	0.785	0.839	0.683	0.0872
Naive Bayes	-	15	0.958	0.838	0.105	0.343	0.913	0.812	0.855	0.689	0.0875
Naive Bayes	-	20	0.963	0.825	0.12	0.429	0.918	0.804	0.852	0.692	0.1146

Tab. B.3: All Results Predicting *Year Before Bankruptcy* Variable.

Model	Learning Algorithm	Discrete Intervals	NB Precision	NB Recall	B Precision	B Recall	Precsion Overall	Recall Overall	F-measure	ROC	Kappa
Bayes	K2	2	0.965	0.824	0.074	0.321	0.927	0.803	0.856	0.643	0.0562
Bayes	K2	5	0.964	0.808	0.069	0.321	0.927	0.788	0.847	0.665	0.047
Bayes	K2	10	0.966	0.854	0.088	0.321	0.929	0.831	0.874	0.673	0.0774
Bayes	K2	15	0.963	0.86	0.073	0.25	0.925	0.834	0.875	0.662	0.0509
Bayes	K2	20	0.962	0.887	0.077	0.214	0.925	0.858	0.889	0.653	0.0545
Bayes	MMHC	2	0.959	0.928	0.061	0.107	0.921	0.893	0.907	0.648	0.0256
Bayes	MMHC	5	0.959	0.918	0.055	0.107	0.921	0.884	0.902	0.655	0.0174
Bayes	MMHC	10	0.958	0	0	0	0.917	0.956	0.936	0.639	-0.0029
Bayes	MMHC	15	0.958	0	0	0	0.917	0.958	0.937	0.63	0
Bayes	MMHC	20	0.958	1	0	0	0.917	0.958	0.937052	0.622	0
Naive Bayes	-	2	0.969	0.827	0.091	0.393	0.932	0.809	0.861	0.644	0.085
Naive Bayes	-	5	0.966	0.799	0.072	0.357	0.928	0.78	0.843	0.668	0.0542
Naive Bayes	-	10	0.966	0.841	0.082	0.321	0.928	0.819	0.867	0.677	0.0678
Naive Bayes	-	15	0.964	0.844	0.075	0.286	0.927	0.821	0.867	0.659	0.0554
Naive Bayes	-	20	0.963	0.862	0.074	0.25	0.926	0.836	0.876	0.652	0.0521

B.2 Bayesian Network Direct Acyclic Graph

The image below represents the direct acyclic graph for a Bayesian network using 20 intervals of discretisation and the K2 algorithm. The DAGs produced for other model permutations were very similar in nature where it resembled the structure of a naive Bayes classifier with very few variables appearing on the second and third levels.

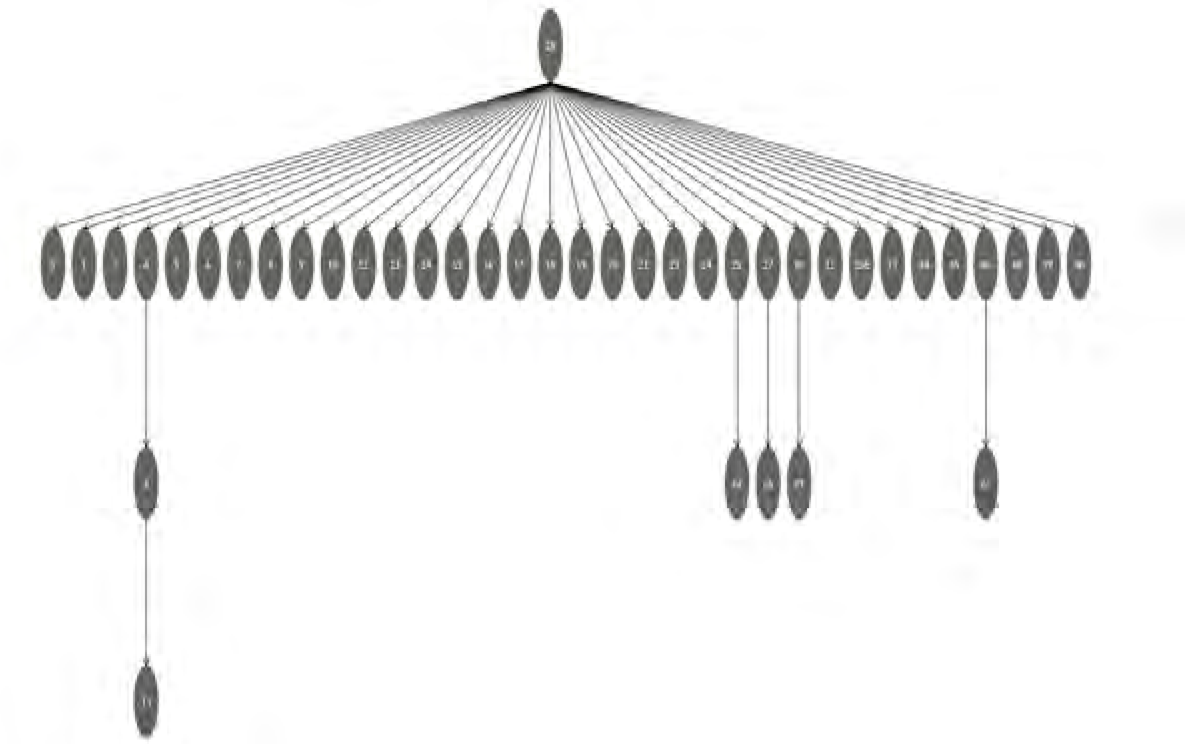

Fig. B.1: Constructed Direct Acyclic Graph for Financial Distress Prediction.

Tab. B.4: Key for Bayesian Network Structure.

0	Assets / Captal Emp	11	Quick Ratio	21	Cash Flow Return On Total Net Operating Assets	31	BCI_Annual
1	Book Val / Share (c)	12	Retention Rate	22	Cash Flow To Total Shareholders Equity	32	JSE
2	Cash Flw / Share (c)	13	Return On Assets %	23	Cash Flow (Cata) To Total Debt	33	CPI_Annual
3	Current Ratio	14	Return On Equity %	24	Cash Flow (Cata) To Current Liabilities	34	BCI_Quartely
4	Debt / Assets	15	Return on Capital Employed	25	Reinvestment Rate	35	BCI_change
5	Debt / Equity	16	Return on Average Assets %	26	Cash Flow (Cata) To Net Earnings After Tax	36	CPI_Quartely
6	Inflation Adjusted Profit/ Share (C)	17	Return on Average Equity %	27	Cash Flow Less Interest Paid To Income Before Tax	37	CPI_Change
7	Inflation Adjusted Return On Assets %	18	Inflation Adj. Return on Average Total Assets %	28	**Financial Distress Variable**	38	lagged_BCI_Quartely
8	Inflation Adjusted Return On Equity %	19	Inflation Adjusted Return on Average Equity %	29	Industry	39	lagged_BCI_change
9	Leverage Factor	20	Cash Flow Return On Total Net Assets	30	Industry Group	40	lagged_CPI_Quartely
10	N A V / Share (C)						

B.3 Results with Predictive Mean Matching Imputed Dataset

Using the Predictive mean matching for imputation we present results for the three output variables in the tables below. The results shown are only those for 20 discrete intervals and using the K2 learning algorithm for the case of the Bayesian network, while the *cumulative indicator* is used as the dependent variable. Taking note of the results, we find that although precision and recall are very high for non-bankrupt companies, the values are lower for bankrupt companies than those realised for the condensed dataset that was not imputed. The inferior performance by the imputed dataset is emphasised by the overall statistics where results worse than those reported in section B. This supports the hypothesis that data was not missing at random and further shows that using imputation did not succeed in constructing a better model.

Tab. B.5: Results using Cumulative Indicator.

Model	NB Precision	NB Recall	B Precision	B Recall	Precsion Overall	Recall Overall	F-measure	ROC Area	Kappa
Nave Bayes	0.852	0.781	0.459	0.578	0.756	0.732	0.741	0.744	0.3303
Bayes	0.849	0.787	0.461	0.566	0.755	0.733	0.742	0.743	0.3276

Tab. B.6: Results using Bankrupt year Indicator.

Model	NB Precision	NB Recall	B Precision	B Recall	Precsion Overall	Recall Overall	F-measure	ROC Area	Kappa
Nave Bayes	0.959	0.802	0.092	0.371	0.915	0.78	0.836	0.677	0.0714
Bayes	0.96	0.855	0.113	0.343	0.917	0.828	0.867	0.672	0.1008

Tab. B.7: Results using year before Bankruptcy indicator.

Model	NB Precision	NB Recall	B Precision	B Recall	Precsion Overall	Recall Overall	F-measure	ROC Area	Kappa
Nave Bayes	0.975	0.82	0.063	0.364	0.945	0.805	0.865	0.669	0.0554
Bayes	0.976	0.874	0.088	0.364	0.948	0.858	0.897	0.679	0.0945

B.4 The Impact of Altering the Number of Parents Restriction for Bayesian Networks

Table B.8 illustrates the decreasing predictive performance as the restriction of the maximum number of parents is altered for a Bayesian network model. Using what we learnt in chapter 7, the results shown are for the K2 learning algorithm and 20 discrete intervals, while the *cumulative indicator* is used as the dependent variable. The differences appear minor and there are certain cases where one parent does not produce the best result. Therefore, using the decomposable metrics, we represent results specifically for bankrupt companies in Table B.2, where a clear dominance can be seen for the one parent column. The precision is lower for one parent, however the recall, which is the metric of most interest, is far higher than those seen in the other two columns.

Tab. B.8: Results when Altering the maximum Number of Parents in the K2 Algorithm.

Number of Parents	**1**	**2**	**3**
Precision	0.8	0.764	0.761
Recall	0.774	0.786	0.782
F-score	0.783	0.756	0.737
Kappa	0.4453	0.2977	0.2388
ROC	0.827	0.824	0.828

Tab. B.9: Results when Altering the maximum Number of Parents in the K2 Algorithm for Bankrupt Companies.

Number of Parents	**1**	**2**	**3**
Precision	0.531	0.636	0.667
Recall	0.687	0.301	0.221
F-score	0.599	0.408	0.332

Appendix C

Logit results

Despite the extremely poor predictive ability of the logit model that was observed in Chapter 8, the detailed output from the model does allow us to get some understanding of the causality behind the dependent variable. The three tables in this section give the coefficients that were realised for each variable in the logit model, for the three prediction variables. It should be noted that the entire dataset was used to construct the model (i.e, a 100% training set). Variables that are in bold font represent cases that were significant at a 95% level.

Tab. C.1: Logit Model Output using *Cumulative Indicator* as Dependent Variable.

Variable	Coef.	Std. Err.	z	P>z
AssetsCaptalEmp	**-0.1505**	**0.0507**	**-2.9700**	**0.0030**
BookValSharec	-0.0000	0.0001	-0.5500	0.5800
CashFlwSharec	0.0004	0.0004	0.8900	0.3760
CurrentRatio	-0.0215	0.1241	-0.1700	0.8620
DebtAssets	**0.9536**	**0.2349**	**4.0600**	**-**
DebtEquity	0.0467	0.0224	2.0800	0.0370
InflationAdjustedProfitShare	-0.0004	0.0004	-0.9400	0.3480
InflationAdjustedReturnOnAss	-0.0052	0.0027	-1.9500	0.0510
InflationAdjustedReturnOnEqu	0.0014	0.0017	0.7800	0.4330
LeverageFactor	-0.0029	0.0062	-0.4700	0.6410
NAVShareC	-0.0000	0.0001	-0.7200	0.4690
QuickRatio	0.0960	0.1249	0.7700	0.4420
RetentionRate	**0.0029**	**0.0010**	**2.9300**	**0.0030**
ReturnOnEquity	0.0014	0.0015	0.9400	0.3480
ReturnonCapitalEmployed	**-0.0026**	**0.0009**	**-2.9000**	**0.0040**
ReturnonAverageAssets	0.0035	0.0792	0.0400	0.9650
ReturnonAverageEquity	-0.0003	0.0004	-0.6100	0.5400
InflationAdjReturnonAverage	-0.0053	0.0793	-0.0700	0.9470
InflationAdjustedReturnonAve	-0.0000	0.0002	-0.2000	0.8390
CashFlowReturnOnTotalNetAs	**0.0005**	**0.0003**	**2.1000**	**0.0360**
CashFlowToTotalShareholders	-0.0011	0.0009	-1.2300	0.2170
CashFlowCataToTotalDebt	**0.0009**	**0.0003**	**2.6400**	**0.0080**
ReinvestmentRate	-0.0006	0.0003	-1.8500	0.0640
CashFlowCataToNetEarnings	-0.0000	0.0001	-0.3700	0.7090
CashFlowLessInterestPaidTo	-0.0000	0.0000	-0.1600	0.8760
BCI_Annual	0.0018	0.0217	0.0900	0.9320
JSE	**-1.6124**	**0.5996**	**-2.6900**	**0.0070**
CPI_Annual	-3.4869	5.7748	-0.6000	0.5460
BCI_change	0.1793	0.5222	0.3400	0.7310
CPI_Change	0.0133	0.0711	0.1900	0.8520
lagged_BCI_Quartely	0.0467	0.0463	1.0100	0.3140
lagged_BCI_change	-0.6511	2.0289	-0.3200	0.7480
lagged_CPI_Quartely	-22.7346	12.9142	-1.7600	0.0780
lagged_CPI_Change	0.1084	0.0732	1.4800	0.1390
Basic_Materials	0.2536	0.3358	0.7600	0.4500
Consumer_Goods	**0.9048**	**0.2734**	**3.3100**	**0.0010**
Consumer_Services	**0.8829**	**0.2515**	**3.5100**	**-**
Financials	0.0674	0.3154	0.2100	0.8310
Health_Care	0.4302	0.5824	0.7400	0.4600
Industrials	**0.5422**	**0.2556**	**2.1200**	**0.0340**
NA	1.0147	0.5260	1.9300	0.0540
_cons	**-2.2894**	**0.5099**	**-4.4900**	**-**

Tab. C.2: Logit Model Output using *Year of Bankruptcy* as Dependent Variable.

Variable	**Coef.**	**Std. Err.**	**z**	**P>z**
AssetsCaptalEmp	-0.1080	0.0685	-1.580	0.115
BookValSharec	-0.0002	0.0004	-0.520	0.606
CashFlwSharec	-0.0002	0.0013	-0.120	0.901
CurrentRatio	-0.0301	0.2398	-0.130	0.900
DebtAssets	0.0015	0.0494	0.030	0.975
DebtEquity	0.0188	0.0322	0.580	0.559
InflationAdjustedProfitShare	-0.0018	0.0010	-1.760	0.079
InflationAdjustedReturnOnAss	0.0002	0.0031	0.050	0.957
InflationAdjustedReturnOnEqu	0.0012	0.0045	0.280	0.782
LeverageFactor	0.0014	0.0110	0.130	0.897
NAVShareC	-0.0001	0.0003	-0.160	0.876
QuickRatio	0.0682	0.2396	0.280	0.776
RetentionRate	-0.0011	0.0018	-0.590	0.556
ReturnOnEquity	0.0004	0.0044	0.080	0.934
ReturnonCapitalEmployed	-0.0019	0.0013	-1.430	0.152
ReturnonAverageAssets	0.0668	0.1511	0.440	0.659
ReturnonAverageEquity	-0.0028	0.0015	-1.860	0.062
InflationAdjReturnonAverage	-0.0716	0.1520	-0.470	0.638
InflationAdjustedReturnonAve	0.0038	0.0022	1.740	0.081
CashFlowReturnOnTotalNetAs	-0.0001	0.0003	-0.170	0.865
CashFlowToTotalShareholders	0.0014	0.0016	0.860	0.387
CashFlowCataToTotalDebt	-0.0001	0.0005	-0.220	0.825
ReinvestmentRate	0.0000	0.0004	0.080	0.937
CashFlowCataToNetEarnings	0.0000	0.0001	0.130	0.897
CashFlowLessInterestPaidTo	0.0000	0.0001	-0.320	0.751
BCI_Annual	0.0067	0.0405	0.160	0.869
JSE	-0.3584	1.0454	-0.340	0.732
CPI_Annual	-6.7464	11.2531	-0.600	0.549
BCI_change	-1.9738	1.0522	-1.880	0.061
CPI_Change	0.2545	0.1383	1.840	0.066
lagged_BCI_Quartely	0.0944	0.0927	1.020	0.309
lagged_BCI_change	-6.7324	4.4504	-1.510	0.130
lagged_CPI_Quartely	-26.0699	22.6970	-1.150	0.251
lagged_CPI_Change	**0.2469**	**0.1203**	**2.050**	**0.040**
Basic_Materials	0.2075	0.6158	0.340	0.736
Consumer_Goods	0.2508	0.5023	0.500	0.618
Consumer_Services	0.2345	0.4509	0.520	0.603
Financials	0.0518	0.5459	0.090	0.924
Health_Care	0.5714	1.1047	0.520	0.605
Industrials	0.2575	0.4489	0.570	0.566
NA	0.8741	1.2543	0.700	0.486
Constant	**-2.4347**	**0.8912**	**-2.730**	**0.006**

Tab. C.3: Logit Model Output using *Year before Bankruptcy* as Dependent Variable.

Variable	**Coef.**	**Std. Err.**	**z**	**P>z**
AssetsCaptalEmp	0.1164	0.1487	0.78	0.434
BookValSharec	0.0000	0.0003	0.14	0.886
CashFlwSharec	-0.0011	0.0021	-0.54	0.591
CurrentRatio	0.0717	0.3619	0.2	0.843
DebtAssets	0.0876	0.0753	1.16	0.245
DebtEquity	-0.0961	0.0801	-1.2	0.23
InflationAdjustedProfitShare	0.0013	0.0020	0.66	0.509
InflationAdjustedReturnOnAss	0.0052	0.0048	1.09	0.278
InflationAdjustedReturnOnEqu	0.0041	0.0027	1.51	0.131
LeverageFactor	-0.0170	0.0156	-1.09	0.275
NAVShareC	-0.0002	0.0003	-0.64	0.521
QuickRatio	-0.0879	0.3628	-0.24	0.809
RetentionRate	0.0003	0.0004	0.71	0.48
ReturnOnEquity	-0.0023	0.0024	-0.94	0.348
ReturnonCapitalEmployed	-0.0027	0.0014	-1.9	0.057
ReturnonAverageAssets	-0.0324	0.1757	-0.18	0.854
ReturnonAverageEquity	0.0017	0.0012	1.5	0.134
InflationAdjReturnonAverage	0.0257	0.1759	0.15	0.884
InflationAdjustedReturnonAve	-0.0021	0.0018	-1.19	0.234
CashFlowReturnOnTotalNetAs	-0.0108	0.0046	-2.36	0.018
CashFlowToTotalShareholders	-0.0028	0.0027	-1.05	0.296
CashFlowCataToTotalDebt	-0.0018	0.0010	-1.73	0.084
ReinvestmentRate	0.0146	0.0062	2.35	0.019
CashFlowCataToNetEarnings	0.0002	0.0001	1.24	0.215
CashFlowLessInterestPaidTo	0.0000	0.0002	-0.23	0.818
BCI_Annual	0.0058	0.0479	0.12	0.903
JSE	-2.6401	1.7293	-1.53	0.127
CPI_Annual	7.6723	13.6878	0.56	0.575
BCI_change	-0.5162	1.2830	-0.4	0.687
CPI_Change	-0.1229	0.1630	-0.75	0.451
lagged_BCI_Quartely	-0.0800	0.1022	-0.78	0.434
lagged_BCI_change	6.8919	4.7122	1.46	0.144
lagged_CPI_Quartely	10.1643	39.4331	0.26	0.797
lagged_CPI_Change	-0.2157	0.3097	-0.7	0.486
Basic_Materials	-0.0649	0.7968	-0.08	0.935
Consumer_Goods	0.4282	0.6072	0.71	0.481
Consumer_Services	-0.0585	0.6055	-0.1	0.923
Financials	0.4965	0.6448	0.77	0.441
Health_Care	0.4543	1.1562	0.39	0.694
Industrials	0.3566	0.5477	0.65	0.515
NA	-0.0844	1.5252	-0.06	0.956
Constant	**-5.9914**	**1.1968**	**-5.01**	**0**