The use of Recursive Partitioning to build a financial distress prediction model for JSE Listed Equities.

Candice Smit

Supervisor: Professor Paul van Rensburg

ABSTRACT

The financial crises of 2008 increased the focus around financial distress and even more so on predicting financially distressed companies prior to the fact. This research paper investigates using recursive partitioning to predict financially distressed companies on the Johannesburg Stock Exchange, taking different business cycle periods into account over the time period 1997-2014. The updated as well as longer time period over which the analysis is conducted distinguishes this research paper from prior research. This paper employs both the CART and CHAID algorithm and obtains financially distressed prediction models which have a higher correct classification rate than chance alone and prior literature in South Africa. This paper also makes use of a matched data sample approach and the manner in which missing data is addressed makes a valuable contribution to financial distress prediction research. Furthermore, support is found for prior literature in that financial variables are statistically significant in predicting financial distress.

Keywords – Financial distress, cost of financial distress, years prediction, recursive partitioning, CHAID.

Table of Contents

1. Introduction

In the aftermath of the financial crisis of 2008, business news headlines have included large public company names, at a stage trading at high stock prices, but now no longer in existence due to financial events (Chen, 2011). Each economy is affected to a certain extent by financial distress and in turn so is every stakeholder. Financial distress is inherent to the science of business management (Pretorius, 2008).

The vast losses experienced by investors in 2008 have created increased caution around the topic of financial distress. Costs of filing for bankruptcy for the company may include the stigma associated with it, filing and associated legal fees, an increased cost of capital for the firm and potentially the loss of workers (Haber, 2005). In South Africa, there is not a vast amount of literature on the indirect costs of financial distress and hence many companies in South Africa may be biased towards the tax benefits debt provides (Tshitangano, 2010). This does not however imply that South African companies are highly leveraged on average, but just that a bias may be introduced as a result. Altman (1984) investigated the direct as well as indirect costs associated with financial distress and found that it came to 4.3% (retail and industrial firms) as well as 4.5% (retailers) and 10.5% (industrial firms) respectively. This finding was based on the market value of the company one year before financial distress. Andrade and Kaplan (1998) also investigated the cost of financial distress and found that it ranged from 10%-20% of the value of the company.

Almeida and Philippon (2008) examined the risk-adjusted costs of financial distress. They highlight that the cost of debt (in the capital structure) is relatively easy to ascertain, but the costs in terms of financial distress is challenging to quantify. Financial distress costs should have a direct impact on the recovery rates which the investors can obtain and hence if the efficient market hypothesis holds then the costs should be exhibited in lower bond prices or equity prices (Almeida & Philippon, 2008). The suggestion is to use risk-neutral probabilities so that the expected distress costs can be discounted at the risk free rate. Financial distress also tends to happen during periods of recession which aligns with investors having a low risk tolerance. Therefore, financial distress costs appear to have a systematic element which should be accounted for. It has been estimated that the cost of financial distress amounts to 3%-5% of the total firm value at the point in time of financial distress (Almeida & Philippon, 2008). Taking into account all the indirect costs, it has been estimated that the total cost could be in the range of 10%-23%, based on the value of the firm prior to financial distress. It due to the cost implications for companies, that there is a vested interest in creating a model to accurately predict financial distress; not only to predict it but to do so earlier rather than later (Deakin, 1972).

Research on financial distress prediction models has been conducted all around the world (Ooghe & Spaenjers, 2010). The focus seems to remain the same, but the statistical techniques vary widely. In recent years, evermore data is being stored electronically in comparison to the past and the volume of electronically stored data is expected to grow at an increasing rate in the future (Hargreaves & Hao, 2013). This provides an ever growing opportunity to use the data to the benefit of investors. Financial distress attracts a vast amount of attention more often than ever before (Pretorius, 2008).

There are many causes behind financial distress, as has been pointed out by numerous authors (Du Jardin, 2009). The major causes can be divided into seven categories. First is accidental causes, which may include a change in the litigation under which the company operates or the passing of a prominent leader. Second is financial distress attributable to market problems which may be as a result of an inadequate product or changing consumer needs which leads to a loss in market share (Du Jardin, 2009). Kale and Arditi (1998) attributed the size of the firm as one of the reasons for potential financial distress. The reason being that a small firm may have a vast amount of initial resources injected which acts as a buffer. Once this time period is over, the probability of financial distress increases significantly as the company no longer has the buffer to protect against it. A large company on the other hand may be more susceptible to financial distress when external market conditions change (Kale & Arditi, 1998). Henderson (1999) provides a different interpretation on the risk of financial distress. Henderson (1999) provides the argument that the risk increases in adolescent companies and that the risk of financial distress is dependent on the long-term strategy the company embarks on. The third cause of financial distress according to du Jardin (2009) may be due to financial threats. This may include aspects such as an increase in the cost of capital, lack of capitalization or defaulting on a loan repayment. Fourth may be as a result of managerial problems in the company such as a lack of a clear organizational structure (Du Jardin, 2009). D'Avenci (1989) supports this as one of the causes of financial distress as managerial issues cause a decline in the company, which then causes managerial problems. D'Avenci (1989) titles this the vicious cycle pattern. It is therefore crucial that to avoid financial distress, management adapts to change in a timely manner. The fifth cause could be macroeconomic factors which may be elements such as higher interest rates, a decline in demand or increased competition which places the company under financial strain. It was confirmed however, that exogenous factors are not considered as important an influence as endogenous variables in the causes of financial distress by small and medium businesses in Singapore (Theng & Boom, 1996). The sixth cause could be attributable to costs such as the high cost of labour or the abrupt loss of a valued supplier (Du Jardin, 2009). The seventh cause may be due to strategy elements which would include the error in entering an unprofitable market or a significant project failing.

In most cases financial distress is not an event that occurs abruptly in the particular company (Fitzpatrick, 1934). Unfavourable conditions may be developing in the company way before the company is actually in financial distress. The company may lack sufficient liquid assets, but might still be sound in terms of overall assets owned. The public, at this stage, would not be aware of any problems. Conditions deteriorate as the company is forced to obtain new financing or else sell some of the assets in order to meet current obligations. Increasingly, total liabilities of the company exceed the total assets. It is only at this point that the public would become aware of the problem. The company will be declared insolvent if no contingency plan is successful. According to Lorange and Nelson (1987), after a time period of strong economic growth there is a tendency for organisational performance to slack off, often due to complacency which is as a result of the good economic times. In the first stages of decline the signals thereof are not always that clear in the company. Financial distress in a company does not generally originate as a result of a single factor, but rather from an accrual of various decisions made within the company (Moss Kanter, 2003). It is due to this that preconditions exists which makes up each unique situation a company in financial distress finds itself in (Pretorius, 2008). Once again, it is clear that it is essential that failure tendencies are picked up earlier rather than later in order to avoid the cost to the economy when a company is declared insolvent (Fitzpatrick, 1934).

Business rescue legislation was introduced on the 1st of May 2011 (Anderson, 2011). This legislation makes allowance for companies which may be in financial distress, but still exhibit economic viability. It allows for these particular companies to have a second opportunity. This would result in tempered statistics on liquidations in South Africa. Companies, previously, either entered compulsory liquidation or voluntary liquidation; both are in most cases due to financial distress. In the case of voluntary liquidation it is a cheaper alternative with the costs amounting to roughly R5000.00 as opposed to compulsory liquidation which could amount to up to R50000.00 (Van der Walt, 2010). The end result of the new Business rescue legislation will be to increase the number of compulsory liquidations in relation to the total number of liquidations (Van der Walt, 2010).

Financial distress prediction is discussed in three branches which are namely: The definition of financial distress, the variables used in the prediction models and finally, the research on the methods used in order to build the financial distress prediction models which will also consider the misclassification rates (Type 1 and Type 2) which can be defined as in Table 1 below.

Table 1: Misclassifications

	Classified as normal company	Classified as financially distressed company
Actual normal company	P_{11}	$1 - P_{11}$ (Type II error)
Actual financially distressed company	$1 - P_{22}$ (Type I error)	P_{22}

The research question this paper sets out to answer is whether a financial distress prediction model for JSE listed equities can be built by means of recursive partitioning techniques using a longer time period. Predicting financial distress has a vast amount of benefits. Firstly, investors in the industry the particular company may be in can act with caution when considering whether to lend to the company when there is information to suggest that they may potentially become financially distressed (Shah, 2014). This may be particularly relevant in South Africa, where the Johannesburg Stock Exchange is predominantly resource based which requires a large amount of capital investment. Companies can also establish long term relationships with other companies which will not fail going forward. This will ensure the prolonged existence and viability of the relationship (Shah, 2014). It is also beneficial for companies themselves, as they become more knowledgeable on their financial position and can take appropriate action based on the predictive model.

Beyond who this paper may be relevant to, this paper delivers a valuable contribution to the literature of financial distress prediction models in South Africa. This research paper investigates using recursive partitioning to predict financially distressed companies on the Johannesburg Stock Exchange, taking different business cycle periods into account over the time period 1997-2014. The updated as well as longer time period over which the analysis is conducted distinguishes this research paper from prior research. This paper also makes use of a matched data sample approach and the manner in which missing data is addressed makes a valuable contribution to financial distress prediction research. This paper aims to develop a statistically significant financial distress prediction model for different business cycles.

The remainder of this paper is organized into six sections. Section 2 of this paper gives a literature review on the subject of predicting financial distress. It investigates the definition of financial distress and then goes on to review the ideas surrounding financial distress as a continuum. The paper then looks at the various variables that could be used in the financial distress prediction models and the categories into which they are classified into, also giving consideration to the debate surrounding financial and non-financial ratios inclusion in the prediction model. Prior literature on financial distress prediction on an international and South African level is then reviewed. The paper then moves on to discuss the number of years prior to failure which is best for predictive accuracy and examines the possible methods which can be used. Classification error is then reviewed in light of financial distress prediction models as well as the ideas around dichotomous variables. Lastly, the data samples used as well as the problems associated with it are examined and a deeper discussion is engaged in regarding recursive partitioning.

Section 3 and 4 will examine the data and methodology respectively which will be employed in this study. Section 5 then moves on to the results and Section 6 draws conclusions based on the findings in Section 5. Lastly, Section 7 makes recommendations for further research in the field of financial distress prediction.

2. Literature Review

There is a vast body of research on financial distress. The pioneers were Beaver (1966) and then later Altman (1968). Since then many researchers have focussed on financial distress prediction and even more so post 2008. The literature review is divided into firstly reviewing possible definitions of financial distress, which is critical for predicting financial distress. Secondly, the literature review gives due consideration to the dependent and independent variables in building a financial distress prediction model. We then move on to considering evidence provided by prior literature on the topic of the time period prior to financial distress and the methods available to build a financial distress prediction model. Literature on the international markets as well as on the South African market is then reviewed with due consideration for the method used in the literature and the accuracy of classification. The liquidity on the Johannesburg Stock Exchange and its' potential implications is discussed, after which the predictive accuracy and cost of misclassification in financial distress prediction models is reviewed.

2.1 Defining financial distress

In building a model to predict financial distress it is imperative that the definition of financial distress be determined. The theoretical challenge of understanding business failure is yet to be met (Pretorius, 2008). The fact that there does not seem to be consensus around the theoretical definition of financial distress may be the reason behind the lack of understanding surrounding this topic (Pretorius, 2008). Altman (1983) has defined financial failure to be when the realised rate of return on the capital which was invested in the company, with an allowance made for the relevant level of risk, is significantly and consistently below that of rates on similar investments. Argenti (1976), states that a company is declared insolvent when it is not able to pay its debt as and when it becomes due or alternatively when the company's net asset value becomes negative. A decision will then be made as to whether the company should continue trading or whether the best decision is for it to be liquidated.

In the context of South Africa, de la Rey (1981) has developed a definition for financial distress also known as corporate failure. De la Rey (1981) defined it to be any company which had equity that became negative; it was placed under judicial management due to the forced discontinuation of its operations; no profit could be reflected in 2/3 years; it could not make preference dividends; it had to default on a loan payment or finally, the share capital nominal value was decreased in order for it to accurately reflect the assets it is suppose to represent.

Haber (2005) states that insolvency could be as a result of a company becoming financially distressed, but just because a company files for bankruptcy does not necessarily mean that they are insolvent. It is for this reason that liquidity ratios may not be the best in differentiating between solvent and financially distressed companies. Court and Radloff (1990) defined financial distress to be when a company delisted and then proceeded to liquidate. Olivier (1992) considered solvent versus insolvent as the definition of financial distress. Le Roux and Olivier (1992) went on to consider the defining factor of financial distress to be a company that delisted due to poor financial performance versus being solvent for 7 years. Van Niekerk (1993) had a definition of financial distress similar to that of Le Roux and Olivier (1992) in terms of poor financial performance. Arron and Sandler (1994) defined financial distress to be when a company is liquidated due to their state of bankruptcy. Bruwer and Hamman (2006) considered the definition of financial distress to be when a company delisted accompanied by a major structural change.

Very few companies apply for bankruptcy in reality, even though they may in fact be financially distressed (Bruwer & Hamman, 2006). Even when a company is in financial distress they may opt for an alternative such as reaching an agreement with a creditor which results in a reduction of the amount of debt. A company may also opt to restructure the debt they have by converting into ordinary shares or else making the decision to merge with another company which may be financially sound (Bruwer & Hamman, 2006). All these options are indicative of the fact that the company could not continue in the form it was.

2.2 Financial distress prediction model: dependent variable

Ohlson (1980) raises the idea that to differentiate financial distress as binary in nature is not realistic. Jones (1987) suggests that financial distress has been used as the dependent variable simply because it is relatively easy to be determined and can be done so objectively. Prior literature suggests that financial distress may not be the ideal dependent variable. Cybinski (2001) identifies the difficulty models have in accurately predicting companies that lie in the continuum between the two extremes. Foster (1986) is in agreement with the suggestions of Cybinski (2001) in that companies can lie on a broad continuum of financial distress and do not always fall into one of two distinct categories.

The norm is mostly a dichotomous classification in financial distress prediction, but the problem is that this makes it less practical for users other than academics (Haber, 2005). If a company, for instance sake, is in the position to file for bankruptcy but then chooses not to- the financial distress prediction model may have predicted it to be bankrupt, but due to the fact that there was no bankruptcy filing recorded it would be counted as an error (Haber, 2005). Stakeholders need more of a continuum in financial distress prediction models, because users of the model will not have prior knowledge of a company being in financial distress or not.

Dichotomous analysis has been conducted in South Africa. Strebel and Andrews (1977) conducted a study on 16 financially distressed and 13 non-financially distressed companies over the time period 1971-1976. One of the findings was that Cash flow/Total debt has predictive power in the model. Daya (1977) on the other hand found that Cash flow/Average current liabilities had the most predictive power in predicting financial distress one year prior. The data used was that of 31 paired financially distressed and non-financially distressed companies. De La Rey (1981) created a financial distress prediction model and obtained an accuracy rate of 96%, one year prior using only financial information on 26 paired financially distressed and non-financially distressed companies.

Lau (1987), on the other hand, did not use a dichotomous dependent variable but instead used an ordinal dependent variable which could take on one of five different values representing an increasing degree of severity of the financially distressed state the company finds itself in. Lau (1987) managed to determine probabilities from the model.

2.3 Financial distress prediction model: Independent variables

Beginning with Altman (1968) through to Agarwal and Taffler (2008), almost all variables that could possibly influence the accuracy of a financial distress prediction model has been analyzed (Du Jardin, 2009). Yet, it appears that not all of the variables have been examined to the same extent. As previously mentioned, financial distress may be as a result of one or many causes which are relatively easy to identify. The challenge is in identifying the variables which would accurately reflect these causes or factors (Du Jardin, 2009). The causes identified prove to be difficult to reduce to a single, measureable parameter. An examination of accounting documents of the relevant companies often shows that the symptoms of the causes are observable. Therefore, financial and accounting documentation is often the basis in financial distress prediction models.

It is important to recognize however, that accounting variables alone cannot provide the most accurate predictive model (Du Jardin, 2009). There are three main categories in which a variable used to predict financial distress can be classified into. The first of these categories is known as the Micro environment which is based on the company itself, including financial variables as well as variables describing characteristics unique to the particular company. There are also characteristics specific to a firm that are non-financial (Court & Radloff, 1993). Ohlson (1980) looked at the fact that the delay in financial statements being made public could introduce a bias into the predictive model. Lawrence (1983) also concluded this and acknowledged that this variable should be taken into consideration in a predictive model. The first to include non-financial variables (which may include political influences, the capital market, GDP growth or short term interest rates), along with financial variables in the financial distress prediction model was Peel, Peel and Pope (1986). They found that by doing this there was an improvement both in terms of the predictive ability and explanatory power. In the South African context, Court (1991) found that non-financial variables (such as the log of Total assets/GDP) displayed greater predictive ability than financial ratios alone in a financial distress prediction model.

The second category is referred to as the Market environment, which navigates around the firm's environment in which it operates. This would include general indicators such as growth in the sector in which it operates for example (Du Jardin, 2009).

The third category is known as the Macro environment which includes factors from the financial markets. Rose, Andrews and Giroux (1982) found that the probability of a company entering financial distress was higher during a period where there was a downturn in the economy. Their findings were based on data from the United States. Another researcher to acknowledge the importance of including macroeconomic variables in a financial distress prediction model was Goudie (1987). In the context of South Africa, Court and Barr (1989) conducted an investigation of financial distress prediction by selecting seven categories of economic variables and then applying factor analysis on these in order to determine the interrelationship.

According to Du Jardin (2009) financial ratio variables were used in 93% of the 190 bankruptcy prediction model studies conducted which were examined. Financial ratios assist in standardizing the data being used in the prediction process. Size may however still be a factor that needs to be accounted for, dependent on the data sample obtained. The primary reason for the popularity of financial ratios in predictive models may be more as a result of the ease of collection thereof than their actual predictive power (Du Jardin, 2009). Financial variables was followed by 28% using statistical variables, 14% using variation variables and 13% using non-financial variables. 6% used market variables and 5% used financial market variables.

A financial distress prediction model created with only financial ratios was found to perform better than a model created with common financial variables (such as debt and assets) by Back, Oosterom, Sere and Van Wezel (1994). Ratio-based models and financial market variable based models were analyzed by Mossman, Bell, Swartz and Turtle (1998) and it was found that the ratio-based models performed slightly better. To verify whether a model that includes non-financial variables either in isolation or with financial variables performs better than a financial distress prediction model built only with ratios, Keasey and Watson (1987) examined three different models. They found that a model which used financial and non-financial variables performed better than the other predictive models built.

Atiya (2001) on the other hand found that a ratio-based financial distress prediction model is more accurate in comparison to a ratio and financial market variable-based model. A model built on financial ratios as well as macroeconomic variables was developed by Tirapat and Nittayagasetwat (1999) and they had an accuracy rate of 70% in their prediction of financial distress. Pompe and Bilderbeek (2005) found that the absolute ratios provide more predictive power than percentage changes in these variables. Court and Radloff (1993) attempted to build a financial distress prediction model by including microeconomic as well as macroeconomic variables. They did this in a two stage model. The first stage was to select the variables that had the most predictive power; these were identified by means of factor analysis. Four macroeconomic factors emerged which indicated monetary policy, general economic activity level, activity in the capital market and politics. Six microeconomic variables were highlighted; these were profitability, liquidity, and solvency, changes in the board of directors, share activity and the publication of audited accounts indicators.

Having seen that financial ratios appears to be most common in financial distress prediction models (perhaps due to the low cost associated in obtaining it), the most important step is then to select a subset of variables from all financial ratios available (Du Jardin, 2009). An important consideration is that the variables need to be as independent as possible and there should be a sufficient number of data points to ensure a model with predictive power. The accuracy a financial distress prediction model has is very much dependent on the data sample used in building it, which would include the quality of the data as well as how readily available it is (Sarlija & Jeger, 2011). There is also the potential limitation that the financial distress prediction model can only be applied to companies with similar characteristics which those in the sample exhibited.

Many financial distress prediction models have however made progress in selecting financial ratios which were found to be significant in multi-ratio prediction models (Sarlija & Jeger, 2011). In the period 1966-1975 many predictive studies were conducted and Chen and Shimerda (1981) reviewed these to ultimately identify 41 financial ratios which were identified as important in these studies. Beaver (1966) found the variable titled Cash flow to total debt, to have the greatest discriminatory power out of all the financial ratios that were examined. Beaver (1968) found that non-liquid asset measures appeared to provide more predictive power than liquid asset measures, which may be attributed to the fact that non-liquid assets are primarily more permanent aspects of a company.

Variable selection in building financial prediction models is often based on a two-step process (Du Jardin, 2009). The first set is usually comprised of many variables chosen often not by statistical means but rather based on their predictive power in prior research or else based on the variables' popularity in other research. The final set of variables to be used in building the financial distress prediction model is selected by an automatic procedure 60% of the time. Du Jardin (2009) examined the criterion that was used to select independent variables to be included in the financial distress prediction models in 190 studies. 40% was based on the popularity or predictive power of the variable in prior literature. It is important to acknowledge that circumstances may have been unique in a particular study and applying the same variable in another model that may have different circumstance could result in that variable not having the same predictive power or significance. Univariate analysis, Stepwise searches, Genetic algorithms, Expert, Non-linear modelling techniques accounted for 17%, 26%, 6%, 4% and 3% respectively. Other criteria such as multiple regression accounted for 4%.

Court and Radloff (1990) used the 15 most popular financial ratios in prior literature and reduced it by means of factor analysis. Court (1991) used motivated non-financial variables as well as 17 financial ratios which were then reduced again by means of factor analysis. A large variety of ratios were used by Olivier (1992), with no specific focus. Le Roux and Olivier (1992) considered ratios in isolation and selected those ratios that on an individual level displayed significant predictive power in predicting financial distress. Cash flow ratios alone were used by Van Niekerk (1993). Court and Radloff (1993) on the other hand applied factor analysis to 15 macroeconomic variables as well as 21 microeconomic variables which had been randomly selected for the study. Arron and Sandler (1994) relied only on 18 financial ratios in their research and Bruwer and Hamman (2006) focused on cash flow variables for the majority of the variables used.

According to Bruwer and Hamman (2006) the defining characteristic underlying financial distress is the fact that the company does not have a sufficient amount of cash on hand in order to meet its obligations. It is for this reason that it can reasonably expected that cash flow ratios have more predictive power in the model, but should be combined with other accrual financial ratios due to the fact that these ratios contain different information. The cumulative cash flow of a company over a time period of more than one year may contain more information, as a company may find it possible to survive through a negative cash flow year. Bruwer and Hamman (2006) accounted for this by including the cumulative three year cash flow ratios.

One concern regarding independent variables across all statistical techniques used to predict financial distress is that most of the techniques are based on the assumption that the variable's distribution does not vary over time (Balcaen & Ooghe, 2006). This would imply an established relationship between the independent variables and the dependent variable. In actuality, the independent variables may change due to various factors. If it is the case that the data is not stable over different economic time periods, this may have an impact on the financial distress prediction model (Balcaen & Ooghe, 2006). A positive relationship between the financial circumstances of a company and their performance during a recession was found by Opler and Titman (1994). Opler and Titman (1994) also found that those companies that are highly leveraged during a recession tend to lose more market share than competitors during that time period. According to Ko, Blocher and Lin (2006) conducting a rank transformation on the independent variables data set assists in making the predictive model less sensitive to non-normality. This approach was taken by Kane, Richardson and Mead (1998) and they recorded an improvement in the accuracy of the financial distress prediction model.

2.4 Financial distress prediction time period

It would appear, from previous studies, that financial distress prediction model's predictive power declines over time (Sarlija & Jeger, 2011). This statement was supported by studies conducted by both Zavgren (1985) and Holmen (1988). Altman (2000) as well as Deakin (1972) both found that a shorter time period prior to financial distress led to higher predictive accuracy of the model. Over a shorter time period, the public may be aware of the situation and there would be more current information available. Andreica (2013) found that the most efficient prediction was obtained one year prior to financial distress. Other researchers such as Beaver et al (2005) however, found that the predictive accuracy decreased as soon as the company approached financial distress and favoured a longer prediction time period.

2.5 Methods available to predict financial distress

Financial distress prediction models can be built using various model categories, of which there are three primary ones (Aziz & Dar, 2006). The first being statistical models, which is concerned with the symptoms of financial distress and uses data mainly from the financial statements of the particular company. The second model category is referred to as artificially intelligent expert system models which are also concerned with the symptoms of financial distress. It also draws much of the information from companies' financial statements, but is very much reliant on technology processing. Lastly, there are theoretical models which pay attention to the qualitative causes of financial distress. It usually uses a statistical technique in order to support the theoretical argument put forward as to why the company may experience financial distress (Aziz & Dar, 2006).

Initially, when financial distress prediction first became topical, the most popular methods used to build the models were Multiple Discriminant Analysis (MDA), Logit Analysis as well as Probit Analysis. Financial distress prediction developed and advanced machine-learning techniques were introduced. Examples hereof would be neural networks and recursive partitioning (Bruwer & Hamman, 2006). Pompe and Feelders (1997) as well as Rees (1995) found that there were no statistically significant differences between the predictive powers of the various methods used in predicting financial distress. Sarlija and Jeger (2011) concluded that, given that economic conditions are stable, the predictive model built should be able to deliver relatively sufficient accuracy over time.

Various methods are used to build a financial distress prediction model. Multivariate discriminant analysis (MDA) which can be used if the dependent variable to be predicted is dichotomous, which in the case of financial distress it is (Rama, 2012). MDA, as a technique, attempts to obtain a linear equation which fits the independent variables the best. This implies that the equation obtained should theoretically minimize the misclassifications. The most important advantage which MDA offers in financial distress prediction is that this technique looks at all the characteristics which the relevant firms have in common, as well as the interactions amongst these properties (Altman, 1968). MDA is comprised of three steps. The first is to approximate the variables coefficients. The second step is then to determine the discriminant score for each observation individually. The last step is then to classify each of the observations on the basis of the cut off score (Jo & Han, 1996). MDA has been determined as the most popular technique used in financial distress prediction by Eidleman (1995). A high chance of financial distress is indicated by a low discriminant score. Deakin (1972) concluded that MDA can be used for the prediction of financial distress, using financial data, three years in advance. MDA has some disadvantages which should be highlighted. Firstly, the independent variable's variance-covariance matrix needs to be the same for the financially distressed as well as the non-financially distressed groups (Rama, 2012). The way in which Altman (1968) applied MDA the significance of each independent variable cannot be determined (Rama, 2012). Multicollinearity may also emerge in MDA, which in its severe form may lead to somewhat deceptive model accuracy (Balcaen & Ooghe, 2006).

Logit Analysis is a relatively recent and advanced technique which is used in order to model distinct outcomes (Rama, 2012). The theory behind this technique is based on the consumer behaviour theory found in microeconomics (Jones & Henser, 2004). Logit analysis builds a model based on the assumption that the observations are drawn from a multinomial distribution and the selection probabilities are based on the actual values of individual qualities as well as their alternatives. A Logit model may be referred to as a causal type model (Rama, 2012). Logit Analysis classifies between financially distressed and non-financially distressed companies by looking at the logit score compared to the cut off score. If the logit score is greater than the cut off score it is likely that the company will become financially distressed. The opposite interpretation holds for non-financially distressed companies. Logit Analysis is based on the assumption that the dependent variable is dichotomous. It also stresses the importance of accounting for the cost of Type 1 and Type 2 misclassification errors when the optimal cut off score is defined (Balcaen & Ooghe, 2006). The arguably most important advantage of Logit Analysis is that it does not require the independent variables to be normally distributed and it is considered robust (Rama, 2012).

Recursive partitioning is known for the fact that it is a non-parametric statistical technique and it is non-linear in nature (Rama, 2012). The classification/decision trees in recursive partitioning are hierarchical and are comprised of a set of rational conditions. The decision tree starts with the entire sample thereafter it is split which is comprised of first determining which independent variable will discriminate best between the observations in the sample followed by determining which variable will best categorize the classes of the node (Rama, 2012). The highest predictive classification will be obtained when the splitting continues until it is no longer possible. Shah (2014) obtained a predictive accuracy rate of 89.5% using Chi-squared automatic interaction detection (CHAID) and 88.4% using the CART algorithm. In using the CHAID algorithm, it was observed that Price/Book value and Net interest coverage were the only two significant variables in terms of distinguishing between financially distressed and non-financially distressed Australian mining companies.

Neural Networks attempt to duplicate what the human neurons conduct (Yoon, Swales & Margavio, 1993). Neural Networks can be used in solving many problems. This technique has the benefit of learning by means of experience and continuous learning. This technique does however present some disadvantages such as the fact that the thought process used to obtain the final result cannot be traced, if there is a problem with the system the user will not be aware thereof and finally, for this technique to be effectively used a large sample of data is required (Rama, 2012).

Univariate Analysis is a technique which is based on the idea that each independent variable is compared to an optimal point at which it can be cut off (Rama, 2012). Each optimal point, for each measure, is compared to the company's value (Balcaen & Ooghe, 2006). The most significant advantage is that this technique does not need any statistical knowledge. The one disadvantage is the fact that this technique assumes that there is a linear relationship between the measures and the financial status of the company (Balcaen & Ooghe, 2006).

Companies inevitably change over time, but many models do no account for this which results in many restrictions (Balcaen & Ooghe, 2006). It may firstly assume that the nature of business these companies engage in does not change, it may not account for economic circumstances changing over time and repetition of applying the financial distress prediction model may result in contradictory predictions (Balcaen & Ooghe, 2006). Most statistical models also assume that the path a financial distressed company follows is the same for all companies, which may not be so in reality.

2.6 International literature

This paper will now examine prior literature at an international level on financial distress prediction models, followed by a look at South African research on this topic. Altman (1968) was the pioneer in financial distress prediction with his infamous Z-score model which aimed to overcome the gap between conventional ratio analysis and the more meticulous statistical techniques (Rama, 2012). The Z-model was able to combine financial as well as non-financial variables (multi-variable model). The multivariate discriminant-function had a success rate of 96% in predicting financially distressed companies, 1 year prior. In another sample, which contained 66 non-failed companies the accurately predicted rate was 79%. Altman (1968) had five independent variables namely:

Net working capital/Total assets, Retained Earnings/Total Assets, Earning before Interest and Tax/Total Assets, Market value of common shares and preferred shares/Book value of debt and Sales/Total Assets.

Altman and Bettina (1977) constructed the underpinnings of the ZETA model which included the following variables: retained earnings to total assets, leverage, the variability of earnings, return on total assets, interest payments coverage, current ratio and the size of the asset base. One year prior, the ZETA model accurately predicted 91%. Five years prior, the model accurately classified 77% of the sample. The variables with the greatest predictive power were retained earnings to assets, closely followed by the stability of earnings.

A similar study to that of the Altman Z-score was conducted by Ko (1982), which was based on data from Japan. The data sample was from the time period spanning from 1960-1980 and included 41 paired financially distressed and non-financially distressed companies. The data was used to complete a variable trend analysis in order to decrease the biases which are relatively common in the reporting practises of Japanese companies. The model which was built delivered an accuracy rate of 90.8%, with the only difference in comparison to the Altman Z-score being that a cut off value of zero was used and only 3 coefficients. The Altman Z-score was adapted by Altman, Baidya and Riberio-Dias (1979) and implemented on the Brazilian economy. The alteration was that one of the coefficients was transformed to Total equity-capital contributed by shareholders/Total assets. From the sample to which the model was applied, the accuracy rate was 88% with a type 1 and type 2 error of 13% and 11.4% respectively.

A financial distress prediction model was also built based on data from Australian companies by Castagna and Matolcsy (2006). Similar to the situation in South Africa, the challenge was to obtain a sufficient number of financially distressed companies in order to conduct reliable discriminant analysis. Data was collected from 1963-1977 on industrial companies. The results obtained from the model were not definitive. Another model which used discriminant analysis was built by Knight (1933). The study examined numerous small companies, which involved interviewing key personnel in the company. The most important findings were that firstly, companies tend to become financially distressed early on and secondly, that most companies become financially distressed due to some aspect of managerial ineptitude. Bilderbeek (1979) also conducted a study in the Netherlands which used discriminant analysis to build an eventual Z-score model which included five independent variables. The data sample was from the time period 1950-1974 and included 38 financially distressed companies and 59 non-financially distressed companies which delivered an accuracy rate of 80%.

Altman and Narayanan (1997) went on to build a financial distress prediction model in Switzerland. The data sample was from the time period from 1960-1971 and included 36 paired financially distress and non-financially distressed companies. The study made use of a technique known as cluster analysis in order to decrease the collinearity. The study found that there were six ratios that were able to discriminate between the two groups.

Norton and Smith (1979) used a model of step by step method linear analysis to predict financial distress 4 years prior. They found that the ratio of cash flow to total asset and cash flow to total debt had the most predictive power in predicting financial distress 3 years prior. Ohlson (1980) used a logarithmic symbol on four factors in order to predict financial distress and found that financial ratios are able to predict financial distress with greater accuracy 1 year prior versus 2 years prior. Altman (1993) observed that financial ratios exhibited a declining trend the closer the company came to financial distress and most importantly, the biggest decline in these financial ratios was observed between the second and third year prior to financial distress.

Salehi and Abedini (2009) investigated financial distress prediction on the Iranian stock exchange. The variables which they used included: working capital to total assets, current assets to current liabilities, profit before interest and tax to total assets, total equity to total assets and sales to total assets. There was a 93% predictive accuracy in the model developed based on these variables. The type 1 error amounted to 7% and the type 2 error amounted to 3%, 1 year prior. It was found that the predictive accuracy of the model declined as the time period was extended to 2 years prior (Salehi & Abedini, 2009).

2.7 South African literature

Court and Radloff (1990) employed the method of MDA and Logistic Regression Analysis (LRA) to predict financial distress using data from the time period 1965-1986. The fact that economic circumstances may vary was not accounted for in the sample of 26 financially distressed and 26 non-financially distressed companies. Olivier (1992) made use of MDA to predict financial distress from a sample of 25 insolvent companies and 54 solvent companies over the time period 1970-1988. Olivier (1992) accounted for varying economic conditions to an extent in that the solvent companies were drawn from a neutral economic time period. Le Roux and Olivier (1992) also made use of MDA on a sample of 39 companies that were delisted in the time period from 1970-1988 along with 60 consistently solvent companies from 1982-1988. There was no consideration given to varying economic circumstances for delisted companies, but solvent companies used were from a year considered to be economically neutral. Court and Radloff (1993) also used MDA in order to build their financial prediction model.

LRA was used by Court (1991) on a data sample from the time period 1965-1986, which also did not account for the fact that economic circumstances may vary. The sample characteristics were the same as that of Court and Radloff (1990). Van Niekerk (1993) also made use of LRA to predict financial distress, using a sample of 18 financially distressed companies (from 1975-1988) and 63 non-financially distressed (during 1990), with no reference made to economic circumstances which may vary. Arron and Sandler (1994) used LRA, MDA as well as neural networks to predict financial distress, using a sample of 28 financially distressed companies along with 40 non-financially distressed companies over the time periods 1966-1976 and 1988-1993 respectively.

Bruwer and Hamman (2006) used recursive partitioning, primarily due to the fact that it is a non-parametric, non-linear method to predict financial distress and can be explained relatively easily to users thereof. The difference between Bruwer and Hamman (2006) and previously mentioned South African literature on financial distress prediction models, is that they did not use a sample but instead used the entire population of interest (namely listed industrial companies). The time period Bruwer and Hamman (2006) considered was from June 1997-May 2002 and separate financial distress prediction models was built for the two different economic circumstances (recessionary and growth).

Bruwer and Hamman (2006) performed a Lilliefors test in order to determine whether the independent variables were normally distributed and concluded that they were not. Therefore, a Kruskal-Wallis test was performed on these variables and it was found that during the defined recessionary time period there were five independent variables which indicated a significant difference and eight during the defined growth phase. In all three of the defined populations (recessionary, growth and recessionary & growth) three of the independent variables could distinguish between financially distressed and non-financially distressed companies. These were Log (Total Assets/GDP), cash flow from operating activities/Sales and the cumulative cash flow from operating activities over the past three years/Sales. Various general characteristics of financially distressed and non-financially distressed companies were found for each of the populations defined. For all three it was true that financially distressed companies are on average smaller and have less cash available. The highest prediction accuracy in the combined model was for that of financially distressed at 66.9%.

2.8 Liquidity considerations on the Johannesburg Stock Exchange

In comparison to international markets, the African market has relatively less liquidity (Oosthuyse, 2014). Liquidity basically refers to whether a large quantity can be sold whilst not having a marked impact on the price (Hattingh, 2014). The lack of liquidity may be attributed to factors such as a lack of standardisation in terms of products, limits which restrict short-selling and a shortage of counters which are listed on the Johannesburg Stock Exchange (JSE) as well as significant transaction costs. Foreign investors may be more inclined to invest in the event that liquidity on these markets is increased. If there is a lack of liquidity in an exchange, it serves to reason that price as a function serves a minimal purpose. Liquidity in a market is important in so far as it has an impact on price discovery (Hattingh, 2014). An impact on the equity price will influence any variables in the financial distress prediction model which includes it as a numerator or denominator. A mere few years ago the JSE had liquidity which amounted to a single digit (Oosthuyse, 2014). Recently, it ranges from 33%-44%, but liquidity is still a concern which should be increased in order to reap the benefits which includes depth in the products offered and more certainty and confidence in the processes in the market.

Hattingh (2014) looked at the equity market on the JSE and observed that equities have various degrees of liquidity. Naturally, if a company exhibits a high probability that the company will default or exhibits poor decisions made by management, it may lead investors to avoid investment in these companies and hence result in illiquidity. A market considered to be liquid, generally exhibits a few common characteristics which includes: the costs involved with completing a transaction is low, the swiftness at which a trade can be completed is relatively fast, the number of orders is large, the volume of orders are high and does not have a significant impact on the price and lastly, an imbalance in orders is quickly corrected (Hattingh, 2014).

2.9 Predictive accuracy and misclassification costs

The predictive accuracy of a particular model has been a popular method to evaluate it (Keasey & Watson, 1991). What this implies is that the cost of misclassified failed firms as well as misclassified non-failed firms has been assumed to be equal. To ensure that the decision making is economically efficient, it is important that the predictive functions accurately reflects the relative costs of both types of misclassification errors to the decision maker (Keasey & Watson, 1991).

Many researchers have attempted to customize the predictive functions in order to account for the costs of both types of misclassification. The cost of the two types of misclassification was directly allowed to differ when Anderson (1958) created an assignment of observations to group functions. Although there has been progress towards accounting for the cost of misclassification in the predictive models, not many prior researchers have accounted for a differential in the cost of misclassification (Keasey & Watson, 1991). Muller, Steyn-Bruwer and Hamman (2009) highlighted the fact that the accurate prediction of the financially distressed companies is somewhat more important in terms of the financial implications for the economy in comparison to the non-financially distressed predictive accuracy.

The success of a financial distress prediction model is defined by whether the model is able to predict better than the rate of chance at 50% (Haber, 2005). According to Haber (2005) a better standard by which to measure the predictive success of a model is to look at how many companies are publically listed and then look at how many companies actually went bankrupt. Say for instance this is 6%, this would mean that randomly selecting a company that will not be financially distressed would be accurate 94% of the time. A financial distress prediction model would be applied on this environment and hence a model should outperform this standard (Haber, 2005).

When financial distress or bankruptcy models are replicated, they are generally not as robust. This means that the results obtained, when replicated, are not as good as in the original study (Haber, 2005). This may be attributed to a variety of reasons, one of which being the variables which are included in the financial distress prediction model. As previously discussed, many variables may be included due to their popularity in prior research studies. Most studies then do not critically evaluate what bankruptcy or financial distress is and then include financial ratios that would reflect this which may be the reason for financial distress prediction models not holding up over various time periods (Haber, 2005).

2.10 Data samples

In a significant amount of prior literature, the data samples used matched financially distressed and non-distressed firms based on size, which evidently does not take into account the proportions of the two different groups in the population selected (Bruwer & Hamman, 2006). Bruwer and Hamman (2006) account for this in their research by taking the total population of their data and not only a sample thereof. This means that they did not only rely on the two extremes in the development of their financial distress prediction model.

The result of using equal samples may be that the classification and prediction accuracy of financially distressed companies are overstated and likewise the classification and prediction accuracy of non-financially distressed companies may be understated (Zavgren, 1983). Many of the statistical techniques suffer from what is known as sampling selectivity (Rama, 2012). The reasoning behind a random sample is so that the results of the study can be expanded to the entire population (Balcaen & Ooghe, 2006). Financial distress prediction however, as has been mentioned, leans towards non-random samples due to the fact that financial statements may be readily available and not many companies fail in a given economic time period and hence paired sample techniques often used is not the most correct procedure due to the low number of financially distressed companies in the economy (Balcaen & Ooghe, 2006).

Another consideration is the fact that most data samples used in prior literature often covers different economic circumstances and does not consider the economic influences (Bruwer & Hamman, 2006). Cybinski (2001) noted that not much attention was paid to the external environment such as where in the business cycle the particular company was. Mensah (1984) also made note of the fact that most data samples are pooled over time and no consideration is given to the varying economic circumstances during the time periods. The economy can either be in a recessionary or growth phase and a financial distress prediction model will vary in its accuracy of prediction, dependent on these economic circumstances (Mensah, 1984). For example during a recessionary period the amount of business activity may decline. When this happens it usually translates into lower sales and hence lower profits for the company. According to Richardson, Kane and Lobingier (1998), the cost associated with obtaining credit increases, whilst the availability of the credit being looked for declines.

3. Data

In the data sample the ideal would be for the control sample obtained to be a random selection of non-financially distressed companies with the data covering the same time period as that of the financially distressed companies (Keasey & Watson, 2001). It is also vital according to Keasey and Watson (2001) that the predictive model is not built on data which would only be available after the occurrence of financial distress.

3.1 Sample

The criterion used in the selection of financially distressed companies is whether the company had delisted from the JSE during 1997-2014 for reasons pertaining to financial distress. This means that the company could not persist in the state is was without financial intervention or restructuring. The sample employed in this study comprises of companies that are not in the financial industry directly, as there are items not on financial companies' balance sheets which make it difficult to build a financial distress prediction model. Another factor hampering financial companies' inclusion in the study is that the financial ratio comparison between these companies and others is a tricky task. Furthermore, the sample will include only companies that are or were listed on the Johannesburg Stock Exchange (JSE), depending on whether they are still listed or whether the company delisted as a result of financial distress. In order to allow for a large as possible sample size, data from 1997-2014 will be used. The liquidity constraint on the JSE was not explicitly address in this study, due to the fact that a variable to measure this was not included nor was the sample altered to account for this in an attempt to allow for as large as possible data sample size. This study is concerned with company listing at any point in time from 1997 to 2014. A sample of 66 financially distressed companies will be selected with each company matched to a corresponding control company, based primarily on similar total asset value. All control companies being fully functional and in operation. The companies used in this study can be found in Table 2.

Table 2: Financially distressed and non-financially distressed companies

Financially distressed		**Non-financially distressed**	
Ticker	**Company**	**Ticker**	**Company**
ALC	Amlac Ltd	**MTA**	**Metair Investments Ltd**
ALD	Aludie Ltd	**DGC**	**Digicore Holdings Ltd**
APL	Net1 Applied Technology	**DCT**	**Data Centrix Holdings Ltd**

DNM	Dynamo Retail Ltd.	**EXL**	**Excellerate Holdings Ltd**
ECH	Ec-Hold Ltd	**CVN**	**Convergenet Holdings Ltd**
FRO	Frontrange Solutions	**ALM**	**Alliance Mining Corporation**
GLL	Global Village Holdings	**AWT**	**Awethu Breweries Ltd**
GMF	Gencor Ltd	**WES**	**Wesizwe Platinum**
HRL	Harwill Investments Ltd	**CUL**	**Culinan Holdins Ltd**
KHO	Kirchman Hurry Properties	**OCT**	**Octodec Investments Ltd**
MWEB	Mweb Holdings Ltd.	**BRT**	**Brimstone Investment Corporation**
MES	Messina Ltd	**LON**	**Lonmin PLC**
MTO	Mathomo Group Ltd	**RTO**	**Rex Trueform Clothing Ltd**
NMB	Nimbus Holdings Ltd	**MST**	**Mustek Ltd**
RAG	Retail Apparel Group	**JDG**	**JD Group Ltd**
STOCHOT	Legacy Hotels & Resorts	**DON**	**Don Group Ltd**
STK	Siltek Ltd	**ASR**	**Assore Ltd**
TOT	Top Info Technology Holdings	**SQE**	**Square One Solutions Group**
UTR	Unitrans Ltd	**CMH**	**Combined Motor Holdings Ltd**
VLX	Voltex Holdings Ltd	**ATN**	**Allied Electronics Corporation Ltd**
WNE	Winecorp Ltd	**OLG**	**Onelogix Group Ltd**
ZRR	Zarara Energy Ltd	**PPR**	**Putco Properties Ltd**
ABR	Afribrand Holdings Ltd	**AVI**	**Anglovaal Industries Ltd**
ABT	Ambit Properties Ltd	**PMM**	**Premium Properties Ltd**
ADT	Advanced Technical Systems Ltd	**ADH**	**AdvTech Ltd**
APE	APS-Technologies Ltd	**APN**	**Aspen Pharmacare Holdings**
AQL	Aquila Growth Ltd	**LAB**	**Labat Africa Ltd**

BLT	Bolton Industrial Holdings	**MTX**	**Metorex Ltd**
BRY	Bryant Technology Ltd	**HAR**	**Harmony Gold Mining Ltd**
CCG	CCI Holdings Ltd	**APK**	**Astrapak Ltd**
CEL	Celcom Group Ltd	**PNC**	**Pinnacle Technology Holdings**
CFO	Country Foods Ltd	**UCS**	**UCS Group Ltd**
CNX	Conafex Holdings	**ALT**	**Allied Technologies Ltd**
CNY	Century Carbon Mining Ltd	**AHL**	**Ah-vest Ltd**
COR	Core Holdings Ltd	**CKS**	**Crookes Brothers Ltd**
DNA	DNA Supply Chain Investments Ltd	**ANA**	**Adrenna Property Group**
ELE	Element1 Holdings Ltd	**SDH**	**Secure Data Holdings Ltd**
ERM	Enterprise Risk Management Ltd	**AET**	**Alert Steel Holdings Ltd**
EUR	Eureka Industrial Ltd	**CLH**	**City Lodge Hotels Ltd**
FGM	MasterFridge Ltd	**KGM**	**Kagiso Media Ltd**
FSH	Fashion Africa Ltd	**GIJ**	**Gijima Group Ltd**
GDA	Glodina Holdings Ltd	**ART**	**Argent Industrial Ltd**
GLT	Global Technology Holdings Ltd	**KAP**	**KAP Industrial Holdings**
IDI	Idion Technology Holdings	**AON**	**African & Overseas Enterprises Ltd**
KLG	Kelgran Ltd	**ISA**	**ISA Holdings Ltd**
KNG	King Consolidated Holdings Ltd	**ADI**	**Adaptit Holdings Ltd**
LSR	Laser Group Ltd	**BAU**	**Bauba Platinum Ltd**
MLL	Millionair Charter Ltd	**SBL**	**Sable Holdings Ltd**
NIN	Ninian & Lester Holdings	**VMK**	**Verimark Holdings Ltd**
OAI	Omega Alpha International IT Holdings	**AFE**	**AECI Ltd**
PAC	Pacific Holdings Ltd	**SUR**	**Spur Corporation Ltd**
PAL	Pals Holdings Ltd	**GND**	**Grindrod Ltd**
PSC	Pasdec Resources Ltd	**COM**	**Comair Ltd**

RDPN	Roadcorp Ltd	**CCL**	**Compu Clearing Outsourcing Ltd**
TIW	Tiger Wheels Ltd	**SER**	**Seardel Investment Corporation Ltd**
UNG	Universal Growth Holdings	**SVB**	**Silverbridge Holdings Ltd**
UNISP	Unispan Holdings	**MRF**	**Merafe Resources**
VTL	Ventel Ltd	**BNT**	**Bonatla Property Holdings**
SPO	Set Point Group Ltd	**RLO**	**Reunert Ltd**
NAI	New Africa Investments Ltd	**IRA**	**Infrasors Holdings Ltd**
BCH	Best Cut Ltd	**ASO**	**Austro Group Ltd**
BEE	Beget Holdings Ltd	**SOV**	**Sovereign food Investment Ltd**
ITR	Intertrading Ltd	**SNV**	**Santova Ltd**
DLG	Dialogue Group Holdings	**DLV**	**Dorbyl Ltd**
KIR	Kairos Industrial Holdings	**SPG**	**Supergroup Ltd**
QHL	**Queensgate Hotels & Leisure Ltd**	**CRG**	**Cargo Carriers Ltd**

The study will include financial variables (which included especially cash-flow variables) for each company as well as select non-financial variables (such as earnings on the All-share Index or short term interest rate) in the economy in the analysis. The situation of financial distress was used as the identifier for financial variable data to then be collected for the specified period of relevance.

The data will be analysed according to the business cycles in existence. In South Africa, it has been noticed that the duration of the business cycles have been getting longer (Provincial treasury, 2012). The expansion phases have exhibited this more in comparison to the contractionary phases. From September 1997-November 2007, South Africa experienced an expansionary period which was the longest recorded. 2007-2009 were the years in which a recession was experienced (Provincial treasury, 2012).

The McGregor BFA database will be used to obtain the data before mentioned. The JSE is relatively small which provides various limitations which are further compounded by the fact that financial companies are not included in this study. Some data is also often missing from the financial statements which further hamper the study.

The problem of missing data is not unique to this study; on the contrary it is not unusual in many quantitative analyses in various fields of study. The way in which missing data is treated could have an impact on the robustness of the model constructed. One option is to neglect the missing data or adjust for it on a case by case basis; this may potentially impact on the parameter estimates as well as the standard errors (Schafer, 1997). If an observation is simply deleted completely, the already small sample size declines even further which has an impact on the statistical techniques applied. This could also potentially introduce bias to the study as the sample is no longer a true representation. A widely used alternative to address the issue of missing data is imputation. This technique takes the average of the observed values for a specific variable under consideration and puts that into where there is missing data. The result is an artificial reduction of the particular variables variance.

In order to address the missing data problem, it is important to identify the reason behind the missing data. If it is completely random and not at all related to the independent variables included in the sample, then it is referred to as missing completely at random (MCAR). In the event that the missing data is actually not random, but in fact related in some way to the values that are observed then the data is referred to as missing at random (MAR). In the event that the data is thought to be related to the unobserved values, then the data is termed not missing at random (NMAR).

One manner in which missing data can be addressed is by means of multiple imputations. This technique imputes values by examining the data which is available at that point in time, any preceding knowledge relevant to the data and lastly any potential relationships between the independent variables. Numerous data sets are then constructed by the model and missing values expected value is the average of the numerous imputations.

The data is in a set represented by the letter D. It is a vector of p variables which includes the dependent (financial distress) as well as the independent variables. D is then broken up into the observed elements as well as the missing elements so that $D=\{D_{obs}, D_{miss}\}$. The matrix named M is used to represent the missing data which is similar in size to that of D. A value of 1 is assigned to observed values; similarly 0 is assigned to missing values. Under the assumption that the data is MCAR, the missing values can be inferred based on the observed values in D in order that $p(M|\ D)=p(M|Dobs)$. Di will represent the vector with p variables which includes both the dependent and independent variables. The assumption is that the data follows a normal distribution with a mean vector μ and a variance matrix Σ. The likelihood function for the data set is then given by:

$$L(\mu, \Sigma|\mathrm{D}) \propto \prod_{i=1}^{n} N(D_i|\mu, \Sigma) \qquad \text{(i)}$$

The assumption is that the data is MAR and, given the prior assumption of normality in the data set, the observed data likelihood is:

$$L(\mu, \Sigma|\mathrm{D}_{obs}) \propto \prod_{i=1}^{n} N(D_{i,obs}|\mu_{i,obs}, \Sigma_{i,obs}) \qquad \text{(ii)}$$

Where Di, obs is representative of the observed elements in row i of D, $\mu_{i,obs}$ is then the subvector of i and $\Sigma_{i,obs}$ is representative of the submatrix of i. Assuming that normality assumption made holds, the missing values can be imputed by means of a linear regression function with random draws made from the posterior which is most appropriate (indicated by ~), where

$$\tilde{D}_{ij} = D_{i,-j}\tilde{\beta} + \tilde{\varepsilon}_i \qquad \text{(iii)}$$

4. Methodology

Decision trees are favoured for many reasons, amongst these being that decision trees give a relatively intuitive graphical depiction of how some target group is determined by means of a few inputs (Neville, 1999). Furthermore, decision trees can also accept various types of variables which includes ordinal, nominal and interval. The most important reason in most cases for favouring decision trees is the fact that they are robust in situations of missing values and the assumptions on the distributions of input variables (Neville, 1999).

Decision trees are most commonly used to create a predictive model, due to the fact that they can overcome some of the hurdles often involved in creating a predictive model such as variable selection, variable importance and missing values (Neville, 1999). Trees can detect relationships by means of observing the interaction of input variables and also disregards redundant input variables. It is also important to know which variables identified are relatively more important- indicating what is the strength of the influence on the dependant variable. Decision trees offer an advantage in terms of missing variables, due to the fact that they split on one input variable at a time. In a regression for example many input variables are combined and in such an instance if an observation is missing it is disregarded (Neville, 1999).

Creating a decision tree is relatively simple. The goal with a decision tree in most cases is to find the best split between input variables (Neville, 1999). Loh and Shih (1997) provide the argument that in the process of splitting each variable, a bias towards nominal input variables with numerous categories is introduced. However, this would not be a concern if the purpose of the model is prediction as opposed to interpretation or variable selection.

The structure of the data can often be better described by splitting the input variable into more than two branches, known as a binary split (Neville, 1999). Kaas (1980) also alluded to the fact that splitting in a binary way may be inefficient. However, a multiple branch split can also be accomplished by numerous binary splits on the same input variable. This way more multi step partitions will be considered and it alleviates the problem of the data often not clearly determining the appropriate number of branches for a multiple branch split (Neville, 1999). Optimally, for a predictive model, researchers would want for the input variable splits to maximize the separation of the target values in the dependent variable (Neville, 1999). When the dependent variable is binary (financial distress) or interval, the best split may be found by not having to look at all the splits. Heuristic algorithms do not guarantee optimality due to the fact that they do not search for the best available split of an input variable, but instead for the best split examined (Neville, 1999).

Recursive partitioning is the basis of a vast amount of algorithms. It concentrates on the optimization of individual input variable splits and does not give much consideration to the quality of the decision tree as a whole. The quality of a decision tree may to a large extent be determined by the size thereof. The data is not described well if the decision tree is too small. Whereas if the decision tree is too large, it may be that some leaves on the decision tree has too little information to make reliable predictions using a different data set (Neville, 1999). One option is that a decision tree can be retrospectively pruned to an appropriate size after the decision tree has been grown to a size which is considered to be too large.

The first Recursive partitioning technique discussed is the CART (Classification and regression tree) algorithm. There are two things that distinguish it from the Chi-squared Automatic Interaction detection (to be discussed below). Firstly, CART makes use of a binary splitting algorithm which means that every parent node is split into only two child nodes and hence each node in the decision tree is either a terminal node or will have two child nodes. Secondly, the CART algorithm makes use of the reduction in diversity criterion which means that the nodes need to be as heterogeneous as possible. The CART algorithm starts at the root node, which includes all the objects together. Hence, there is a certain amount of diversity within the root node in terms of behaviour where some will be financially distressed and others non-financially distressed (Durbach, 2014). The CART algorithm makes use of what is referred to as the diversity index. It is defined by the probability that when you select two objects in the group at random, they will each belong to a different group (Durbach, 2014). In this instance, where the root node is comprised of financially distressed and non-financially distressed, the diversity index would be calculated using the following formula:

Diversity Index (DI) = $1- \rho_1^2-\rho_2^2$

In this formula ρ_1 is the share of objects which belongs to a particular group (financially distressed) and then similarly ρ_2 is the share of objects that would belong to the other group (non-financially distressed). The chance then that two objects selected at random are not the same can be expressed as $1- \rho_1^2-\rho_2^2$. The Diversity Index can never be less than 0, which would occur in the instance where all the objects in a particular node are members of the same group (Durbach, 2014).

The Diversity Index having been determined, the CART algorithm can continue to partition the root node. Each independent variable is taken into consideration in turn and then a diversity index is calculated for each possible partitioning option. The particular partition which would lead to the largest reduction in the diversity within the root node will be selected. Each child node then goes through the same process (Durbach, 2014).

Chi-squared Automatic Interaction Detection (CHAID) is another example of a Recursive partitioning technique (Decision trees). It is a data mining as well as a statistical method (Okwell, 1999). CHAID recursively partitions the data. The input variable split must achieve a certain level of significance in the Chi-squared test of independence, else it is not split. Hence, the Chi-squared test of association is used to determine whether the branch should be split again and if this is the case assists in determining which independent variables should be used to do so (Okwell, 1999). CHAID makes adjustments to mitigate the bias introduced when input variables have many possible values (Neville, 1999). The CHAID algorithm introduces a penalty (adjustment of the relevant p-value) for the input variable splits that may be inclined to introduce bias (Kaas, 1980). This algorithm also adopts the last split examined.

Under the CHAID algorithm each independent variable is evaluated in terms of how many possible splits exist and then a chi-squared test of association is performed on the dependent variable and the various possible partitions of that independent variable (Durbach, 2014). Whichever chi-square test statistic is most significant, determines which independent variable and partition is used at that node. This process described is then duplicated for each child node in the tree until the point is reached where no more significant partitions are available. The formula for the Chi-squared test of association is:

$$\chi^2=\sum \frac{(\text{Observed value}-\text{Expected value})^2}{\text{Expected value}}$$

CHAID, although it has many advantages (amongst which is that it is much easier to identify mistakes and over-fitting), also has weaknesses. Firstly, it requires a large amount of data in order to make sure that the number of observations in the leaf tree nodes is large enough in order to be deemed significant (Okwell, 1999).

Secondly, CHAID requires that continuous independent variables be banded into classes with a categorical quality (Okwell, 1999). Okwell (1999) states that CHAID is the winner in terms of predictive modelling and it is for this as well as previously mentioned reasons that the CHAID algorithm will be used in this study.

Considering the preceding discussion surrounding the decision tree algorithms, it is important to discuss the input variables that will be used in building the financial distress prediction model. The sample described before contains a vast number of variables, which can be seen in Table 3 below. The aim will be to reduce this to a more manageable selection and then apply CART and CHAID to these respectively. A vast majority of prior research used financial variables based on popularity in previous studies. This study will instead use a logit model to obtain a subset of variables that appear to be significant in the financial distress prediction, where the dependent variable is 1(financially distressed) or a 0 (non-financially distressed). Two non-financial variables will be included which are namely the past 6 months returns on the JSE as well as short term interest rates as both should provide a good indication of the state of the economy at certain points in time and can be readily obtained. This subset of variables will then be used in the CART and CHAID algorithms to build the financial distress prediction model for the entire dataset as well as the specified business cycle time period (growth and recession). Thereafter the predictive accuracy will be evaluated.

Table 3: A list of the financial variables used

Factor	**Factor Description**
FAC1	Accounts Receivable/Turnover
FAC2	Assets / Captal Employed
FAC3	Cash Flow to Dividend Cover
FAC4	Cash Flow to Interest Cover
FAC5	Current Ratio

FAC6	Debt / Assets
FAC7	Debt / Equity
FAC8	Earnings Yield %
FAC9	Inflation Adjusted Return On Assets %
FAC10	Inflation Adjusted Return On Equity %
FAC11	Interest Cover
FAC12	Leverage Factor
FAC13	Long-Term Loans % Total Debt
FAC14	Net Profit Margin %
FAC15	Operating Profit Margin %
FAC16	Price / Book Value
FAC17	Price / Cash Flow
FAC18	Price / Earnings
FAC19	Price / N A V
FAC20	Quick Ratio
FAC21	Return On External Investments %
FAC22	Retention Rate
FAC23	Return On Assets %
FAC24	Return On Equity %
FAC25	Total Assets / Turnover
FAC26	Total Debt / Cash Flow
FAC27	Turnover / Employee
FAC28	Return on Capital Employed
FAC29	Price / EBITDA
FAC30	Price / EBIT
FAC31	Price / Cash
FAC32	Return on Average External Investments %

FAC33	Return on Average Assets %
FAC34	Return on Average Equity %
FAC35	Inflation Adj. Return on Average Total Assets %
FAC36	Inflation Adjusted Return on Average Equity %
FAC37	Cash Flow Return On Total Net Assets
FAC38	Cash Flow Return On Total Net Operating Assets
FAC39	Cash Flow To Total Shareholders' Equity
FAC40	Dividend Coverage
FAC41	Interest Coverage
FAC42	Cash Flow (Cata) To Total Debt
FAC43	Cash Flow (Cata) To Current Liabilities
FAC44	Cash Flow To Capital
FAC45	Adequacy Ratio
FAC46	Reinvestment Rate
FAC47	Cash Flow (Ncta) To Capital Investments
FAC48	Cash Flow (Ncta) To Financial Investments
FAC49	Cash Flow (Ncta) To All Investments
FAC50	Cash Flow (Cata) To Turnover (Margin)
FAC51	Cash Flow (Cata Less Pref. Dividend) Per Share
FAC52	Price Per Share To Cash Flow Per Share
FAC53	Working Capital To Operating Cash Flow
FAC54	Cash Flow (Cata) To Net Earnings After Tax
FAC55	Cash Flow Less Interest Paid To Income Before Tax
FAC56	Financial Distress dummy variable

5. Results

In this section the findings will be discussed in two sections. Firstly, when all the imputed data was used in the algorithms run and then when the data was divided into different time periods to determine whether there were any significant differences. Each consideration will then also look at both the CART and CHAID recursive partitioning algorithms to determine whether there are any significant differences between these in terms of predictive accuracy.

5.1 Analysis for significant variables

Upon conducting the analysis, as set out before, to determine which variables are significant and should be included when running the recursive partitioning algorithm, the following was found:

Table 4: Significant variables on all imputed data

Variable/factor name	P-value	Average of the variable
Cash flow to Interest coverage	0.0026	243.06
Debt/Assets	0.0247	1.11
Earnings yield %	0.0201	-4.34
Interest coverage	0.0000	-33.15
Net Profit Margin %	0.0005	-24.50
Return on External Investments %	0.0029	219.07
Return on Assets %	0.0444	-16.28
Total Assets/Turnover	0.0002	1.65

Turnover/Employee	0.0387	117.52
Price/EBIT	0.0005	-5.62
Price/Cash	0.0117	504.82
Return on Average External Investments %	0.0002	-192.48
Return on Average Assets %	0.0443	-0.18
Dividend Coverage	0.0002	69.71
Adequacy Ratio	0.0000	-32.38

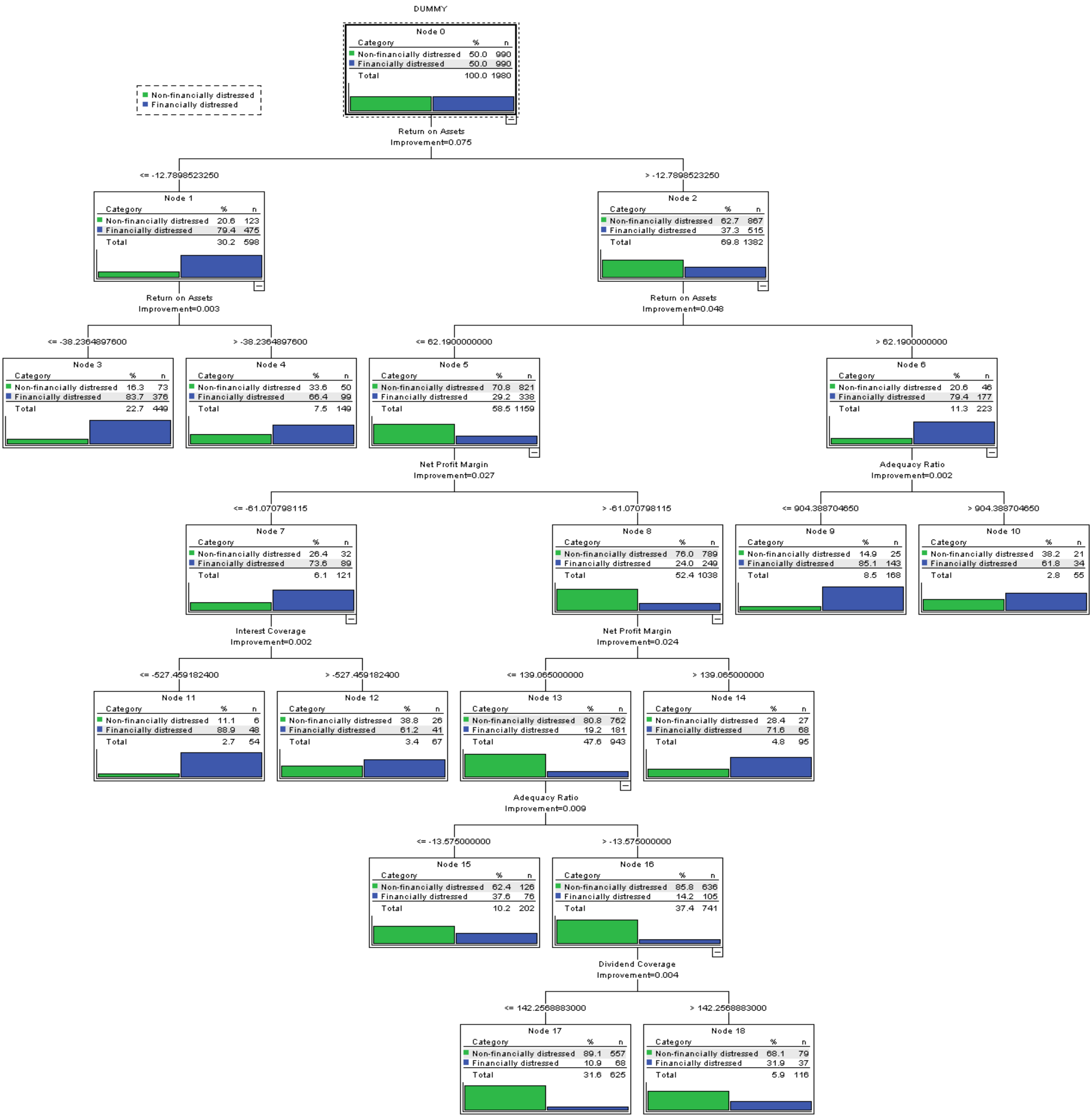
All the variables in Table 4 above are statistically significant at the 5% level of significance which would be indicated by a p-value of less than 0.05.

5.2 Entire time period: CART algorithm

All classification trees start with what is known as the root node which consists of all objects placed together with no independent variables having been taken into consideration yet (Durbach, 2014). A node is what the classification tree creates through the use of the independent variables which get grouped into what we hope to be exhaustive and mutually exclusive (Durbach, 2014). Nodes where the group cannot be split any further are referred to as terminal nodes.

The CART algorithm was run on the significant variables in Table 4. The root node is split into node 1 and node 2 by Return on Assets. If Return on Assets is less than or equal to 12.78 then node 1 is reached which predicts financially distressed, whereas if it is greater than 12.78 then node 2 is reached which predicts non-financially distressed. Node 1 can then be split using independent variable Return on Assets again, where if the variable value is less than or equal to -38.23 then node 3 is reached which predicts financially distressed. Similarly node 4 predicts financially distressed and is reached when Return on Assets is greater than -38.23. Node 2 can then be split using Return on Assets. If the variable value is less than or equal to 62.19 then non-financially distressed is predicted at node 5. If Return on Assets is greater than 62.19 then node 6 is reached which predicts financially distressed. Node 5 can be further split by the Net profit margin variable; if the value is less than or equal to -61.07 then financially distressed is predicted and if the value is greater than -61.07 then node 8 is reached which predicts non-financially distressed. Node 11 and 12 is reached by means of Interest coverage being used at Node 7 to split. Node 11 is reached if Interest coverage is less than or equal to -527.45 and similarly node 12 is reached if the variable value is greater than -527.45. Regardless of the value, node 11 and 12 predicts financially distressed. Node 8 can then be split even further by the Net profit margin variable. If the value is less than or equal to 139.06 node 13 is reached and similarly, if the value is greater than 139.06 then node 14 is reached. Non-financially distressed and financially distressed is predicted by nodes 13 and 14 respectively. Node 13 can be split into node 15 and 16 if the Adequacy ratio is less than or equal to -13.57 or greater than -13.57 respectively. Both nodes 15 and 16 predict non-financially distressed. Node 17 and 18 is reached from node 16, which is split using independent variable Dividend coverage. If Dividend coverage is less than or equal to 142.25 then node 17 is reached and if Dividend coverage value is greater than 142.25 then node 18 is reached. Both node 17 and 18 predict non-financially distressed. A visual representation of the decision tree described above can be seen in Figure 1.

Figure 1: Decision Tree (CART algorithm) on entire data period



As discussed in the literature review, due to the relative cost differentials involved with non-financially distressed and financially distressed, it would be important to consider the error in classification in the predictive model. Researchers would hope to observe that the error in predicting financially distressed is lower than that of non-financially distressed, but this would also be dependent on what context the model is being used for.

Table 5: Correct classification table for CART on full data period

Classification

Observed	Predicted		
	Non-financially distressed	Financially distressed	Percent Correct
Non-financially distressed	762	228	77.0%
Financially distressed	181	809	81.7%
Overall Percentage	47.6%	52.4%	79.3%

Growing Method: CRT

Dependent Variable: DUMMY

This means that each of these is better than chance alone, which is relatively good in comparison to prior literature as well. Using a matched sample, Haber (2005) defined the success of a financial distress prediction model by whether it is able to predict better than the rate of chance of 50%. By this standard the financial distress prediction model built using the CART algorithm on the entire data period in this paper is better. The overall model can also be evaluated in terms of the correct classification rate which could be compared to the goodness-of-fit of the model. For this model in particular the correct classification rate is 79.3%. It is important to compare this relative to another similar study. Bruwer and Humman (2006) used a similar variable selection method and analysed across different economic periods and hence would be appropriate to use. Bruwer and Humman (2006) had a correct classification rate of 65.9%. Splitting the total correct classification rate into financially distressed and non-financially distressed, the correct classification rate of this paper is 81.7% and 77% respectively, which is higher than that of 66.9% and 65.3% found by Bruwer and Hamman (2006). This paper aligns with the objective to correctly classify financially distressed companies in this model better than that of non-financially distressed companies.

5.3 Growth and Recessionary periods: CART algorithm

As mentioned before, not a vast amount of prior literature takes different time periods into consideration in predicting financial distress. This is something this paper aims to take into account in the South African environment.

5.3.1 Growth period

As per the previous discussion on the growth period under consideration, the CART algorithm was once again run on the significant variables in Table 4. The root node can be split using Return on Average External Investments If the variable value is less than or equal to -483.51 then node 1 is reached whereas if the variable value is greater than -438.51 then node 2 is reached. Node 1 and node 2 predict financially distressed and non-financially distressed respectively. At node 1, Price/Cash can be used to further split the node into node 3 and 4. Node 3 is reached if Price/Cash is less than or equal to 187.80 which predicts financially distressed whereas if the value is greater than 187.80 then node 4 is reached which predicts financially distressed as well. Node 3 can then be split using Price/Cash independent variable once again. If Price/Cash at node 3 is less than or equal to -20.25 then node 7 is reached which predicts financially distressed and if the variable value is greater than -20.25 then node 8 is reached which predicts non-financially distressed. Node 2 can then be split into node 5 and node 6 using Debt/Assets. If Debt/Asset is less than or equal to -0.16 then financially distressed is predicted at node 5. If the variable value is greater than -0.16 then node 6 is reached which predicts non-financially distressed. Node 6 can then also be split by using Debt/Assets. If the variable value at node 6 is less than or equal to 1.27 then non-financially distressed is predicted at node 9 and if the variable value is greater than 1.37 then financially distressed is predicted at node 10. A visual representation of the decision tree described above can be seen in Figure 2.

Table 6: Correct classification table for CART on growth data

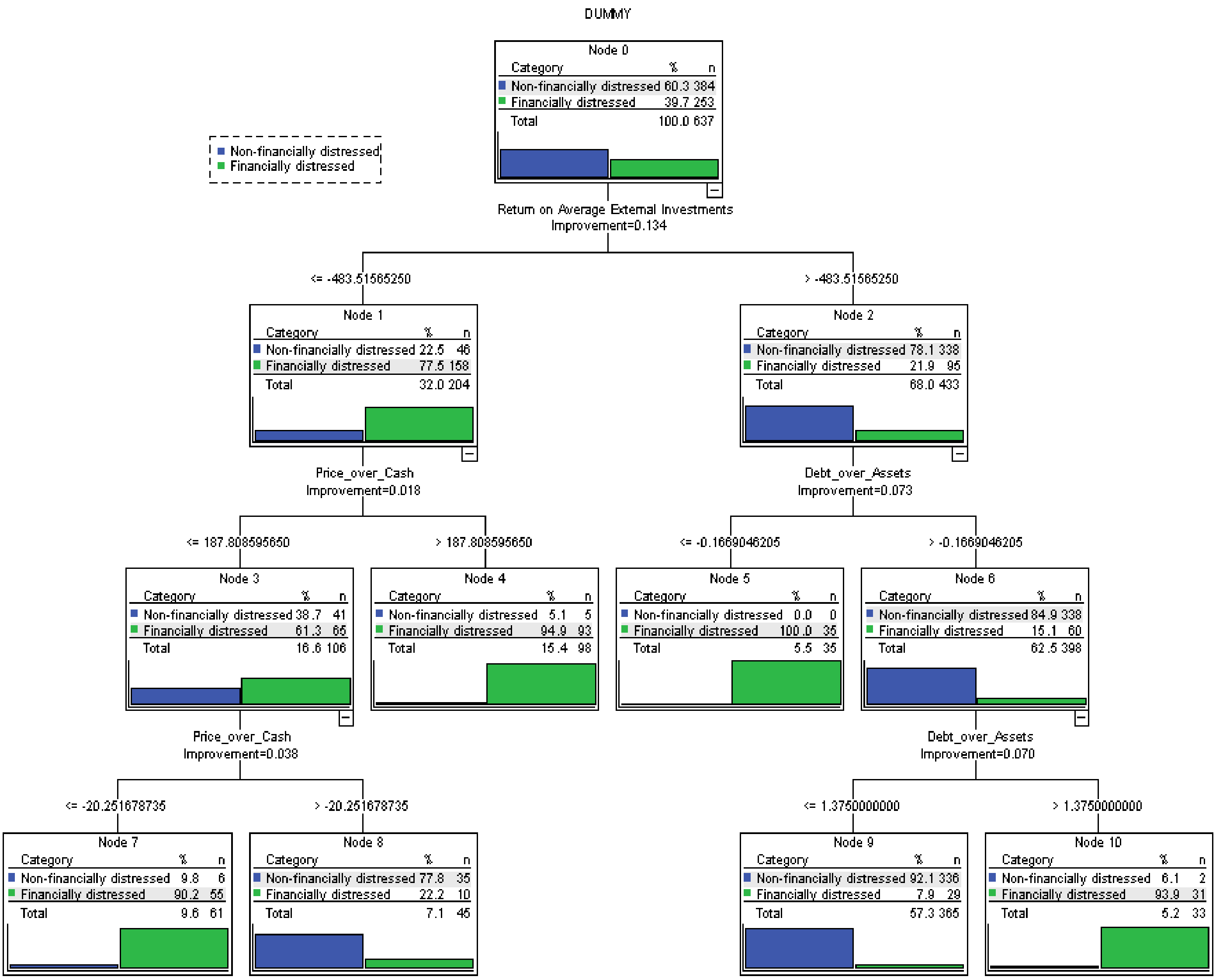
Classification

Observed	Predicted		
	Non-financially distressed	Financially distressed	Percent Correct
Non-financially distressed	371	13	96.6%
Financially distressed	39	214	84.6%
Overall Percentage	64.4%	35.6%	91.8%

Growing Method: CRT
Dependent Variable: DUMMY

Once again it is relatively important to consider the errors in classification. The overall model can then be evaluated on the correct classification rate, which for this particular model is 91.8%. which is substantially higher than that of 67.1% obtained by Bruwer and Hamman (2006). Splitting this then into financially distressed and non-financially distressed correct classification rates of this paper is 84.6% and 96.6% respectively which is considerably better than chance alone, but beyond that is higher than comparable literature. Bruwer and Hamman (2006) obtained 57.1% and 73.1% respectively over the growth period.

Figure 2: Decision Tree (CART algorithm) on Growth period data



5.3.2 Recessionary period

Following from the previous discussion on the recessionary period under consideration, the CART algorithm was run on the significant variables outlined in Table 4. The root node can be split using Return on External Investments. If this variable value is less than or equal to -2628.87 then node 1 is reached which predicts financially distressed, whereas if the variable value is greater than -2628.87 then node 2 is reached where non-financially distressed is predicted. Node 1 can be further split using Price/EBIT. If

Price/EBIT is less than or equal to -2.27 then node 3 is reached which predicts financially distressed. Similarly, if Price/EBIT is greater than -2.27 then node 4 is reached which also predicts financially distressed. Node 5 and node 6 is reached if Return on Average Assets is less than or equal to -3.57 or greater than -3.57 respectively. Node 5 predicts financially distressed and node 6 predicts non-financially distressed. A visual representation of the decision tree detailed above can be seen in Figure 3.

Figure 3: Decision Tree (CART algorithm) on Recessionary period data

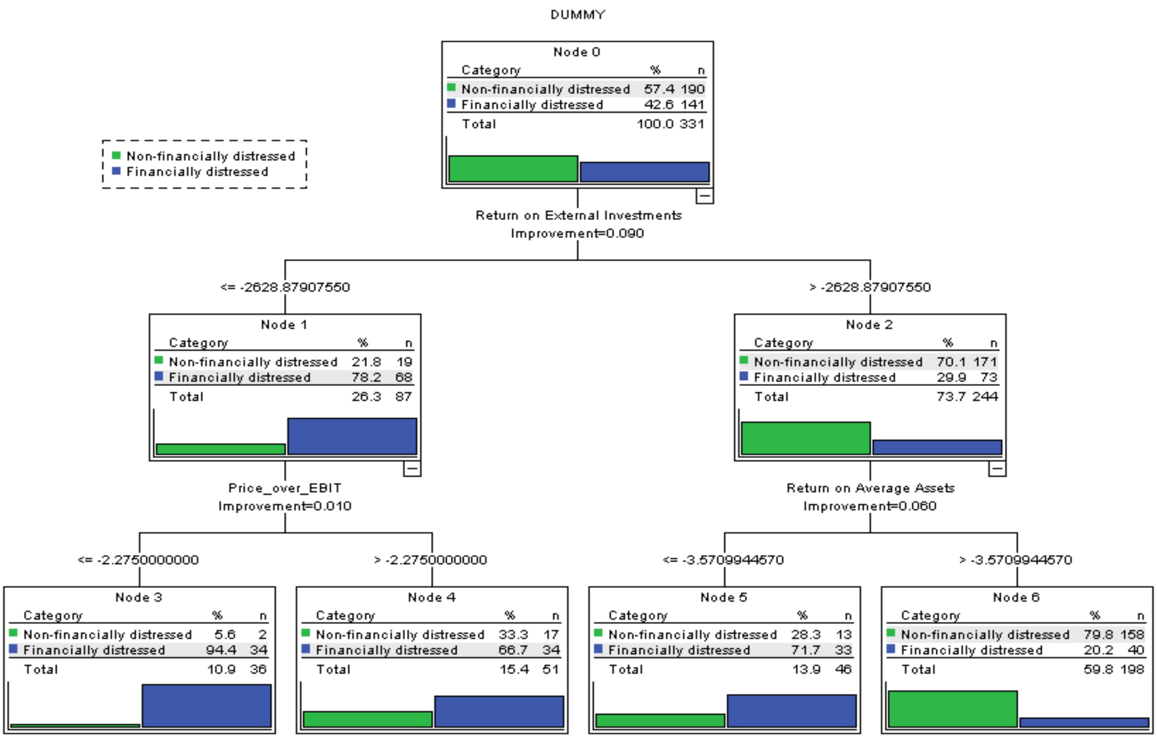


Table 7: Correct classification table for CART on recessionary period

Classification

Observed	Predicted: Non-financially distressed	Predicted: Financially distressed	Percent Correct
Non-financially distressed	158	32	83.2%
Financially distressed	40	101	71.6%
Overall Percentage	59.8%	40.2%	78.2%

Growing Method: CRT

Dependent Variable: DUMMY

For this recessionary time period it is also relatively important to consider the error in classification. The overall model in terms of the correct classification rate comes to 78.2%, which is higher than the 71.2% correct classification rate obtained by Bruwer and Hamman (2006). Upon splitting this into financially distressed and non-financially distressed this paper obtains a correct classification rates of 71.6% and 83.2% respectively which is better than chance alone as well as the 60.8% and 78.6% obtained by Bruwer and Hamman (2006).

5.3.3 Summary of findings on the CART algorithm

It is interesting to note that the variables that could most significantly predict financially distressed companies were financial variables and the market variables did not even enter the model. This finding coincides with past literature where only financial variables were included in the model and macroeconomic variables added no predictive power to the financial distress prediction model such as the study conducted by Mossman, Bell, Swartz and Turtle (1998). It is also interesting to note the classification rules that come about from each node. If Return on Assets is low, financially distressed is predicted. This may be indicative of the fact that the assets are not being used effectively to generate income for the company. If Return on Assets is high, as well as the Adequacy ratio then financially distressed is predicted again. This may indicate that the assets are being sweated too hard to generate the returns and not enough leverage is being used to maximize the returns the company is indeed earning. If Return on assets is on average, Net profit margin is on the upper extreme in this prediction model then financially distressed is also predicted. The fact that net profit margin is on the upper extreme, may be indicative that once again the company is not growing at a sustainable rate. In essence, it is growing too fast.

It appears that non-financially distressed have variable values on average, whereas the financially distressed companies lie to the extremes in terms of these variable values they possess. This makes intuitive sense as it means that the financially distressed company will either be at the extreme lower level and already in trouble or alternatively at the extreme upper end which may not be sustainable and may be indicative of the fact that the company will soon slow down and then there may be inherent financial problems. The accuracy of prediction is also relatively high which may align with prior literature which suggests that a longer predictive time period (years prior to financial distress) leads to higher predictive accuracy of financially distressed companies as they may show signs of problems in existence long before they actually become financially distressed. An example would be the study conducted by Beaver (2005) which favoured a longer prediction period.

When considering the growth period, it is interesting to note that slightly less variables are significant in predicting financial distress in the CART algorithm. There was also a higher correct classification rate for the growth period in comparison to the entire time period as well as the recessionary period, which may be indicative of the fact that a slightly shorter time period, although still long in comparison to prior literature, leads to higher predictive accuracy in terms of financially distressed firms. As supported by De la Rey (1981) who obtained a correct classification rate of 96% over a shorter time period. It is also interesting to note that in what is classified as a growth period, the predictive accuracy of a non-financially distressed company was higher than that of a financially distressed company. This may be attributed to the fact that generally, in growth periods financial ratios are improving for all companies regardless of their situation prior which may mean that even companies that may be in financial difficulty can make ends meet which filters through to the financial ratios which these models are built upon. It is interesting to take note of some of the classification findings that come about in the growth period. Only in the full data period, is the financially distressed predictive accuracy higher than that of non-financially distressed companies. As previously discussed, the economy always experiences business cycles and a growth period is always followed at some point by a contraction/recession period. If the Return on Average External Investments is low and Price/Cash is at the lower end then financially distressed is predicted. This may be indicative of the fact that the company in question may make poor investment decisions and hence returns are not enhanced above the earnings from operation. Furthermore, a low Price/Cash ratio may be indicating that there is a high cash reserve which is not being used effectively to enhance returns or that there are no viable research and development opportunities for growth opportunities. It could also be that the market sentiment on the company is low.

5.4 Entire time period: CHAID algorithm

The decision tree constructed on the entire data period found that the most significant variable in splitting the root node is Net profit margin. If the Net profit margin is less than or equal to -889.92 then node 1 is reached where financially distressed is predicted. If the Net profit margin is greater than -889.92, but less than -245.72 then node 2 is reached where financially distressed is also predicted. Therefore, if the Net profit margin variable value is less than -245.72, the company is predicted as financially distressed. If the Net profit margin is greater than -245.72, but less than -14.66, then node 3 is reached which is then split by the most significant variable, which is Total Assets/Turnover. If Total Assets/Turnover variable value is less or equal to 0.57 then Non-financially distressed is predicted at node 9. If the variable value is greater than 0.57 then node 10 is reached, where financially distressed company is predicted. At node 10, it can be further split by significant variable Debt/Assets. For Debt/Assets value either less than or equal to or greater than 0.77, financially distressed is the status predicted. At nodes 4 and 5, if Net profit margin values is greater than -14.66 but less than -3.84 then non-financially distressed is predicted. Node 6 is reached if the Net profit margin is greater than -3.84, but less than or equal to 18.85 where non-financially distressed is predicted. Node 11 and 12 both predict financially distressed, regardless of the value of Return on Average External Investments. Node 12 is further split by Return on Assets, where node 20 and 21 is reached. At both these nodes, regardless of the value of Return on Assets, non-financially distressed is predicted. Node 21 can then further be split, based on the significant variable of Return on Average External Investments where node 24 and 25 is then reached, both predicting non-financially distressed. At node 7, Net profit margin being greater than 18.85, but less than or equal to 132.48 then non-financially distressed is predicted. Node 13 and node 14 are then reached by splitting by Earnings Yield. If the value of Earnings yield is less than or equal to 4.16 then financially distressed it predicted, whereas if the value is greater than 4.16 non-financially distressed is predicted. At node 8, Net profit margin is greater than 132.48. This node is then further split by Debt/Assets leading to nodes 15, 16 and 17; regardless of the value of Debt/Assets at each of these nodes financially distressed is predicted. Node 22 and 23 is reached when Debt/Assets at node 15 is less than or equal to 0.11. At both nodes 22 and 23, regardless of the value of Return on Average Assets, financially distressed is predicted. A graphical depiction of this information can be seen in Figure 4.

Figure 4: Decision tree (CHAID algorithm) on data from the entire period

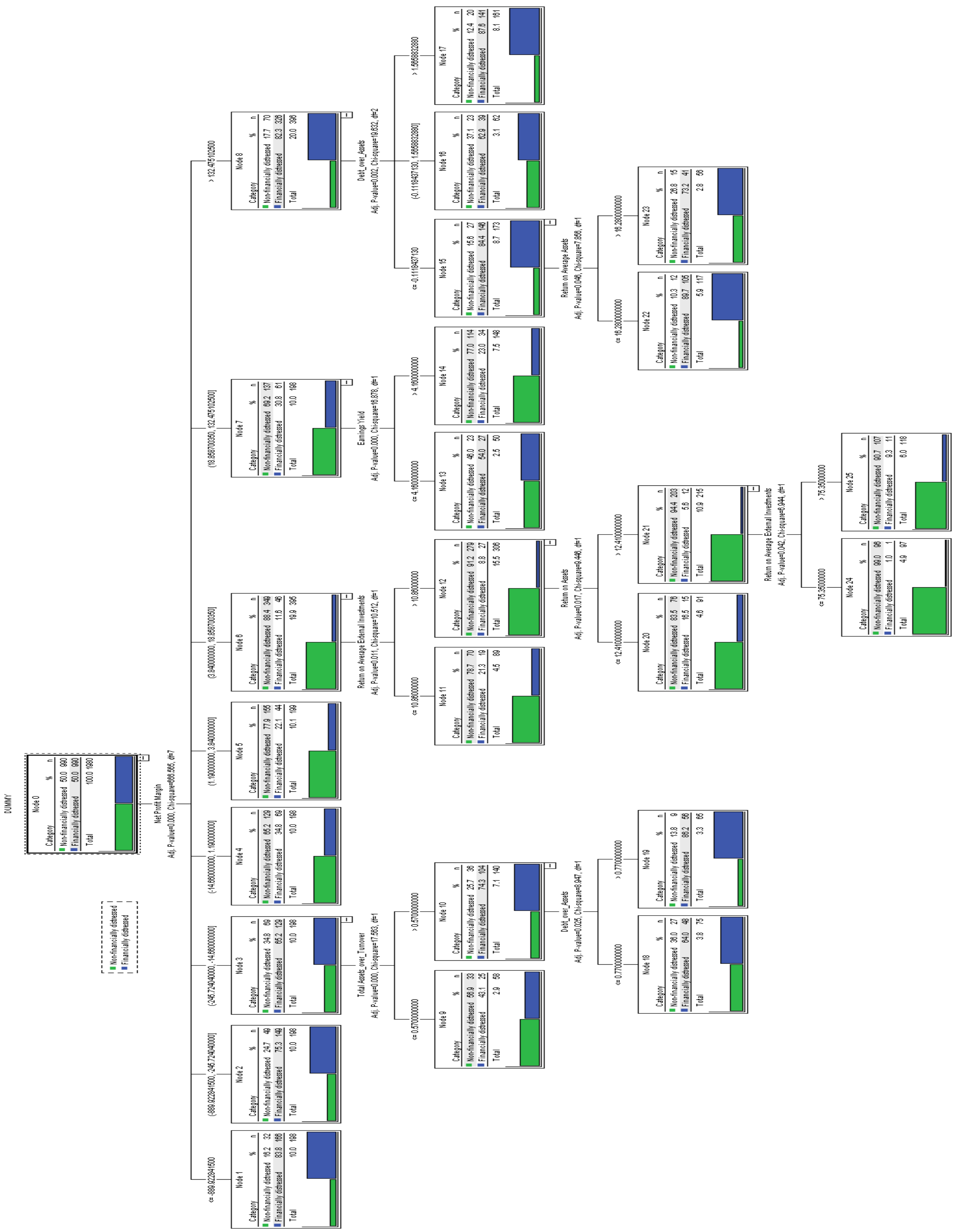


Table 8: Correct Classification table for CHAID on the entire data period

Classification

Observed	Predicted		
	Non-financially distressed	Financially distressed	Percent Correct
Non-financially distressed	780	210	78.8%
Financially distressed	218	772	78.0%
Overall Percentage	50.4%	49.6%	78.4%

Growing Method: CHAID
Dependent Variable: DUMMY

As discussed in the CART algorithm section, it is once again important to evaluate the correct classification rate of the prediction model. For the CHAID algorithm, on the entire time period, the correct classification rate obtained is 78.4%. In comparison to prior literature such as Bruwer and Hamman (2006) who obtained a correct classification rate of 65.9%. The prediction accuracy of the non-financially distressed is slightly higher than that of the financially distressed at 78.8% and 78.0% respectively, but both are substantially better than chance alone and both higher than that found by Bruwer and Hamman (2006) of 65.3% and 66.9%. It is interesting that this paper, under the entire time period and full time period results in a higher correct classification rate of non-financially distress companies.

There are some interesting findings on the entire time period, using the CHAID algorithm. For instance, if Net profit margin is greater than or equal to -245.72, but less than -14.66, Total Assets/Turnover is greater than 0.57 and for any Debt/Assets value, financially distressed is predicted. If Net profit margin is greater than 132.48 then regardless of the value of Debt/Assets or Return on Average Assets, financially distressed is predicted. If Net profit margin is greater than or equal to 18.55, but less than 132.48 and Earnings yield is less than or equal to 4.16, then financially distressed is also predicted. It is interesting to note than if the Net profit margin is at the extremes, either very high or very low it seems to be indicative of financial distress. Either the company is growing too fast or not at all and not being able to meet obligations.

5.5 Growth and recessionary periods: CHAID algorithm

5.5.1 Growth period

The decision tree constructed, using the growth period data, found that Debt/Assets independent variable is the most significant variable in splitting the root node. If Debt/Assets is less than or equal to 2.94 then node 1 is reached and if debt/Assets is greater than 2.94, but less than or equal to 0.10 then node 2 is reached. Both node 1 and node 2 predict financially distressed. If Debt/Assets is greater than 0.10, but less than or equal to 0.97, then node 3 is reached which predicts non-financially distressed. Node 3 can then be further split using Net profit margin. If Net profit margin is less than or equal to 0.48 then node 6 is reached, whereas if it is greater than 0.48 node 7 is reached. Both these nodes predict non-financially distressed. Node 7 can then be further split using Return on External Investments, which reaches node 8 and 9, where both predict non-financially distressed. Node 4 is reached, when Debt/Assets is greater than 0.97, but less than or equal to 5.01 and this node predicts financially distressed. Similarly, node 5 predicts financially distressed when Debt/Assets is greater than 5.01. It is interesting to take note of the fact that once again, at the extremes of the Debt/Assets variable value financially distressed is predicted. This is indicative of either a large amount of leverage in the company which can be the result of downfall or the fact that there is a large amount of assets which is not being fully utilized.

Figure 5: Decision tree (CHAID algorithm) on data from the Growth period

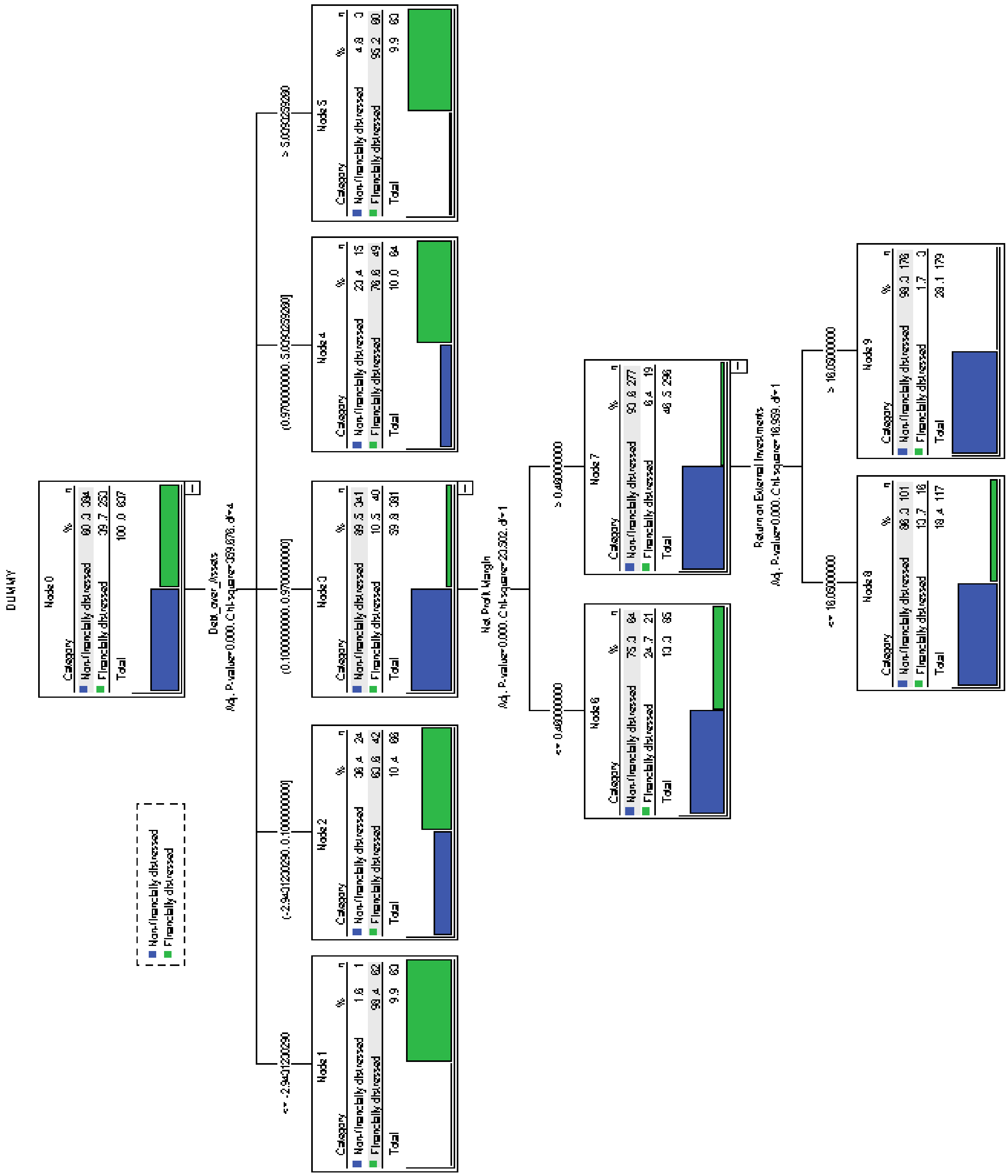


Table 9: Correct Classification table for CHAID on growth period

Classification

Observed	Predicted		
	Non-financially distressed	Financially distressed	Percent Correct
Non-financially distressed	341	43	88.8%
Financially distressed	40	213	84.2%
Overall Percentage	59.8%	40.2%	87.0%

Growing Method: CHAID

Dependent Variable: DUMMY

It is once again important to evaluate the correct classification rate of the predictive model. For the CHAID algorithm, on the growth time period, the correct classification rate for the entire time period is 87.0% which is substantially higher than that obtained by Bruwer and Hamman (2006) of 67.1%. The prediction accuracy of the non-financially distressed is again slightly higher than that of the financially distressed at 88.8% and 84.2% respectively, but both are substantially better than chance alone and significantly higher than that obtained over the entire time period as well as the 73.1% and 57.1% obtained by Bruwer and Hamman (2006).

5.5.2 Recessionary period

The CHAID algorithm run on the recessionary period finds that the most significant variable in splitting the root node is Price/EBIT. If Price/EBIT is less than or equal to 0.57 then node 1 is reached which predicts financially distressed. If Price/EBIT is greater than 0.57 but less than or equal to 13.51, then node 2 is reached, which predicts non-financially distressed. At node 2, the CHAID algorithm finds Return on Assets to be the most significant variable to split the node on. Node 4 is reached when the value of Return on Assets is less than or equal to 14.57 and Node 5 is reached when it is greater than 14.57. Both node 4 and node 5 predicts non-financially distressed. If Price/EBIT is greater than 13.51 then node 3 is reached and this node predicts financially distressed. Once again, it is interesting to take note of the fact that at the extremes, financially distressed is predicted. If Price/EBIT is low it may be that EBIT is unsustainably high and assets are being burnt too fast to produce those high earnings. At the higher extreme, it may be that EBIT is very low and hence the company may not be able to meet its' obligations going forward or the price is low due to market sentiment on the company. A graphical depiction of the decision tree over the recessionary period can be seen in Figure 6 below.

Figure 6: Decision tree (CHAID algorithm) on data from the Recessionary period

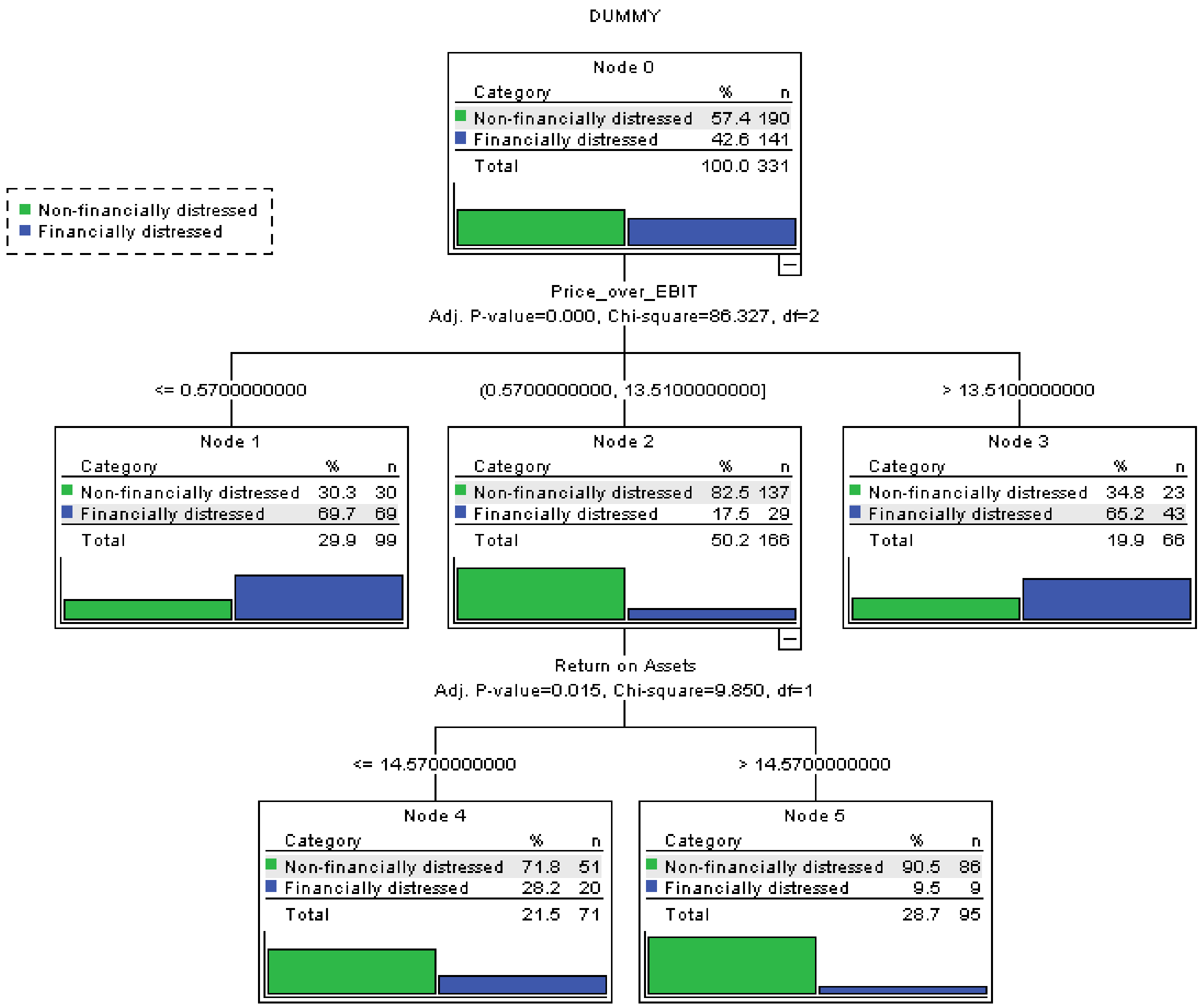


Table 10: Correct classification table for CHAID on the recessionary data period

Classification

Observed	Predicted: Non-financially distressed	Predicted: Financially distressed	Predicted: Percent Correct
Non-financially distressed	137	53	72.1%
Financially distressed	29	112	79.4%
Overall Percentage	50.2%	49.8%	75.2%

Growing Method: CHAID
Dependent Variable: DUMMY

For the CHAID algorithm, on the recessionary time period, the correct classification rate for this time period is 75.2% in comparison to the 71.2% obtained by Bruwer and Hamman (2006). The prediction accuracy of the financially distressed is higher than that of the non-financially distressed at 79.4% 72.1% respectively, but both are substantially better than chance alone. Similarly, this paper obtains correct classification rates higher than that obtained by that of Bruwer and Hamman (2006) of 60.8% for financially distressed and 78.6% for non-financially distressed.

5.5.3 Summary of findings on the CHAID algorithm

It is interesting to note that in the entire data period, 7 independent variables were identified as significant in splitting a particular node in the decision tree. It also important to consider the event that a company may be doing well, but if Net profit margin is low and declining it may be indicative of financial distress due to the fact that margins being squeezed means the reserves will need to be called on in the near future and cannot be sustained indefinitely. It is also important to note, as previously mentioned, that leverage can aid in good times, but can also be to a company's detriment and it is important that management look at what is a sustainable level of debt given their situation. If on the other hand, Net profit margin is at the higher extreme it may be indicative of financial distress in that the assets may be sweating too hard to produce the earnings and cannot be sustained either. Net profit margin for the entire data period appears to be the most important determinant. The reason being that for instance for any value of Debt/Assets or Return on Average Assets, financially distressed is predicted for low values of Net profit margin. The entire data period has a relatively high correct classification rate, with non-financially distressed having a slightly higher correct classification rate.

Then considering the growth period, it is interesting that far fewer independent variables were picked up as significant in splitting nodes in the growth period model in comparison to the entire time period model. Debt/Assets came up as a significant variable again with the extreme variable values being indicative of financial distress. At the lower end of the variable value, it may mean that there is a large asset base, but which is not being utilized effectively to enhance returns or alternatively not enough debt is being used to leverage the returns. On the higher extreme of the Debt/Assets variable value, it may mean that there is too much leverage in the company and could lead to financial distress going forward as their obligations cannot be met and similarly the asset base may be small, so in the event that there are bad circumstances there is not a large asset base to sell to meet obligations. Net profit margin and Return on External Investments are also found to be significant in splitting a certain node in the model, but all values for these variables lead to non-financially distressed being predicted. The correct classification rate for the growth period model at 87% is significantly higher than that of the entire time period model. The non-financially distressed correct classification is once again higher than that of the financially distressed. Both classification rates are however still higher than chance alone.

Lastly, considering the recessionary time period, only 2 significant independent variables were used in the model. The first being Price/EBIT; once again financially distressed companies are predicted as those at the extreme values of Price/EBIT. At the lower value of Price/EBIT, it may be indicative of unsustainably high earnings being produced which may mean assets are sweating too hard. At the higher extreme of the variable value it may be indicative of very low earnings values, which may mean that the company is not producing earnings in line with what is needed to meet all obligations. Similarly, the price in the market may be either too low based on poor market sentiment during a recessionary period when some problems come to the surface or the price is inflated due to misperceptions of the value in the company. The correct classification rate in the recessionary period is the lowest in comparison to the other two time periods considered. This may be as a result of many financial variables being influenced during a recessionary period in the market. Furthermore, using the CHAID algorithm, the only period in which the predictive accuracy of the financially distressed companies is higher is in the recessionary period. In the full time period and growth time period the predictive accuracy of non-financially distressed companies is higher than that of financially distressed.

6 Conclusions

When considering the recessionary period under the CART algorithm, it is interesting to note that even less variables are significant in predicting financial distress in comparison to that of the entire time period and growth period previously discussed. In a recessionary period an increased amount of volatility may filter through to the financial ratios that were used in the analysis. If the Return on External Investments is relatively low, then regardless of the Price/EBIT variable value the company is predicted to be financially distressed. The reasoning under the growth period stands. If then the Return on External Investments is relatively high and Return on Average Assets is relatively low, financially distressed is predicted. It may be indicating that the assets are not being optimally utilized to produce sustainable earnings. It may be outdated or the likes.

It is also important to consider the fact that for this particular paper with this particular data set using the CHAID algorithm over the recessionary period, the correct classification rate is somewhat lower than that of the entire data period and growth period considered for the CHAID algorithm. What this may imply is that financially distressed companies can be predicted in the time period prior to the recessionary period with more accuracy than within a recessionary period. As per Muller, Steyn-Bruwer and Hamman (2009), the accurate prediction of financial distress is somewhat more important than predicting non-financially distressed. This again, may provide support for prior literature, which proposed a somewhat longer time period in order to predict financially distressed companies as supported by Bruwer and Hamman (2006) who also separated according to economic periods. The concept of a matched sample and longer time period is also supported by Balcaen and Ooghe (2006). It is important for investors to get out of a company before it is actually financially distressed and hence when making decisions in a growth period, it may be useful to more accurately predict non-financially distressed than financially distressed companies as has been found in this paper.

It is interesting that during growth time period, the non-financially distressed has a higher correct classification rate than the financially distressed companies. Using the CHAID algorithm, the predictive accuracy is higher for financially distressed than non-financially distressed which is important in that financially distressed has more cost implications as supported by Rose, Andrews and Giroux (1982). The importance of taking different business cycles into account is highlighted in the findings discussed and is crucial for stakeholders in the particular company they are invested or considering investing in. The fact that financial ratios can be obtained with relative ease provides support for the importance of this research in assisting investors in a cost effective manner. Furthermore, the contribution of this paper to literature on longer data samples period and the manner in which missing variables was addressed will be useful for further study in the field of financial distress prediction on the JSE. This paper provides evidence to support the research question on whether recursive partitioning can be used to build a statistically significant financial distress prediction model on the JSE.

7. Recommendations for future research

Predicting financial distress may be a valuable area of research for many years to come as economic circumstances change and developments come about. Recommendations for future research would firstly include conducting research over similar time periods abroad in other developing economies similar in characteristics to that of South Africa. This will allow assessing whether some of the financial variables identified in this study are unique to South Africa in certain business cycles or whether they are statistically significant in other countries as well during specified economic conditions. It may also be insightful to make a comparison between developing and developed economies. If the same financial variables are statistically significant in predicting financial distress that is useful to know, but so is knowing whether other financial variables are significant and if so, how investors could potentially benefit from this investment knowledge abroad.

Another future research recommendation would be to incorporate a comparison of other methodologies for predicting financial distress and assessing the statistically significant variables over the same time period. It may also prove useful to include more macroeconomic variables in the model to evaluate whether any other macroeconomic variable may be significant in predicting financial distress across various methodologies. In terms of variables, it is also important to take note of the fact that management decisions play a large role in the success of a company. Future research may include a variable to take account of the confidence in management and how they are running the business by means of their decision making processes.

Future researchers predicting financial distress on the Johannesburg Stock exchange or any other exchange may want to take the composition of the exchange into account and make relevant adjustments for this. For example the Johannesburg Stock exchange is heavily resource based which may introduce a bias to the results. It may prove a viable idea to separate the sectors on the exchange and attempt to build a financial distress prediction model for particular industries of interest. This may not be viable on the Johannesburg Stock Exchange due to pure size constraints which constrains the data sample size, but it may be possible on somewhat larger exchanges or those that have been in existence for a longer period of time.

Lastly, obtaining data for analysis on smaller exchanges is relatively difficult. The fact that some data had to be imputed in this research paper may lead to potential biases being introduced. It is however important to acknowledge the data constraint and attempt to account for it as accurately as possible. Furthermore, the size of the companies in the data sample was not accounted for in that it may have been relatively large market capitalisation companies as well as companies with a relatively small market capitalisation. This may potentially once again influence the results of the recursive partitioning in the splitting values. Further research should attempt to take the relative size of the company into account to ensure ease of comparison and more specific implementation.

References:

Almeida, H. and Philippon, T. 2008. Estimating risk-adjusted costs of financial distress. *Journal of applied corporate finance.* 20(4): 110-114.

Altman, E. I. 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The journal of finance.* 1(1): 589-609.

Altman, E. I., Baidya, T. and Riberio-Dias, L. M. 1979. Assessing potential financial problems of firms in Brazil. *Journal of international business studies.* 10(3): 102-105.

Altman, E. I. and Bettina, L. 1977. A financial early warning system for over-the-counter broker-dealers. *The journal of finance.* 1(1): 1201-1217.

Altman, E. I. 1983. Corporate financial distress. *A complete guide to prediction, avoiding and dealing with bankruptcy.* Wiley, New York. First edition.

Altman, E. I. 2000. Predicting Financial Distress of Companies: Revisiting the Z-score and Zeta model. Journal of banking and finance. 25(1): 589-609.

Anderson, A. 2011. *Liquidations plummet to levels seen before 2007.* BusinessDay. [Online]. Available: www.bdlive.co.za [2016, May 21].

Anderson, T. W. 1958. An introduction to multivariate statistical analysis. Wiley, New York. Third edition.

Andrade, G. and Kaplan, S. N. 1998. How costly is financial (not economic) distress? Evidence from highly leveraged transactions that become distressed. *Journal of Finance.* 53(1): 1443-1493.

Andreica, M. E. 2013. Early warning models of financial distress: Case study of the Romanian firms listed on RASDAQ. *Theoretical and applied economics.* 5(582): 7-14.

Argenti, J. 1976. Corporate collapse: the causes and symptoms. *McGraw- Hill, London.* First edition.

Arron, J. and Sandler, M. 1994. The use of neural networks in predicting company failure. *De Ratione.* 8(2): 57-78.

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

Aziz, M. A. And Dar, H. 2006. Predicting corporate bankruptcy: where we stand? *Corporate governance.* 6(1): 18-33.

Back, B. , Oosterom, G. , Sere, K. and Van Wezel, M. 1994. A comparative study of neural networks in bankruptcy prediction. *Finnish Artificial Intelligence Society*. (1): 140-148.

Balcaen, S. and Ooghe, H. 2006. 35 Years of studies on business failure: An overview of statistical methodologies and their related problems. *The British accounting review.* 38(1): 63-93.

Beaver, W. H. 1966. Financial ratios as predictors of failure. *Journal of accounting research.* 1(1): 71-111.

Beaver, W. H. 1968. Alternative accounting measures as predictors of failure. *The accounting review*. 1(1): 113-122.

Beaver, W. H. 2005. Have Financial Statements become less informative? Evidence from the ability of financial ratios to predict bankruptcy. *Review of Accounting Studies*. 10(1): 93-122.

Bilderbeek, J. 1979. The continuity factor of evaluation instruments of a company. *Journal of accounting and finance.* 4(3): 58-61.

Bruwer, B. W. and Hamman, W. D. 2006. Company failure in South Africa: classification and prediction by means of recursive partitioning. *South African journal of business management.* 37(4): 7-18.

Castagna, A. D. and Matolcsy, Z. P. 2006. The Market Characteristics of Failed Companies: Extensions and further evidence. *Journal of business finance and accounting.* 8(4): 467-483.

Court, P. W. 1991. An investigation into certain firm-specific non-financial variables as predictors of corporate failure in South Africa. *De ratione.* 5(2): 3-15.

Court, P. W. and Radloff, S. E. 1990. A comparison of multivariate discriminant and logistic analysis in the prediction of corporate failure in South Africa. *De Ratione*. 4(2): 11-15.

Court, P. W. and Radloff, S. E. 1993. A two-stage model for the prediction of corporate failure in South Africa. *Investment analysis journal.* 1(38): 9-19.

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

D'Avenci, R. A. 1989. The aftermath of organisational decline: A longitudinal study of the strategic and managerial characteristics of declining firms. *Academy of Management Journal.* 32(4): 577-608.

Daya, K. 1977. Financial ratios as predictors of corporate failure in South Africa. *South African Journal of Business Management.* 45(2): 93-111.

Deakin, E. B. 1972. A discriminant analysis of predictors of business failure. *Journal of accounting research.* 10(1): 167-179.

De la Rey, J. H. 1981. Finansiele verhoudingsgetalle en die voorspelling van finansiele mislukking by nywerheidsondernemings in die republiek van Suid Afrika. *Buro van Finansiele Analise, University of Pretoria.* 8(1): 36-43.

Du Jardin, P. 2009. Bankruptcy prediction models: how to choose the most relevant variables? *MPRA.* (98): 39-46.

Durbach, I. 2014. Applied multivariate data analysis: PART ||- Predictive techniques. Course notes for STA3020F: research and survey statistics. 169-194.

Eidleman, G. J. 1995. Z-scores- A guide to failure prediction. *CPA journal.* 65(2): 52-54.

Fitzpatrick, P.J. 1934. Transitional Stages of Business failure. *The accounting review.* 9(4): 337-340.

Foster, G. 1986. Financial statement analysis. Second edition. Englewood Cliffs: Prentice-Hall.

Goudie, A. W. 1987. Forecasting corporate failure: The use of Discriminant Analysis within a disaggregated model of corporate sector. *Journal of the royal statistical society.* 1(1): 69-81.

Haber, J. R. 2005. Assessing how bankruptcy prediction models are evaluated. *Journal of business and economics research.* 1(3): 87-92.

Hargreaves, C. and Hao, Y. 2013. Prediction of stock performance using analytical techniques. *Journal of emerging technologies in web intelligence.* 5 (2): 136-142.

Hattingh, A. 2014. JSE Equity market liquidity: The impact of technology and resultant microstructure changes. University of the Witwatersrand, Johannesburg. 1-23.

Henderson, A. D. 1999. Firm strategy and age dependence: A contingent view of the liabilities of newness, adolescence and obsolescence. *Administrative science quarterly.* 44(1): 281-314.

Holmen, J. S. 1988. Using financial ratios to predict bankruptcy: An evaluation of classic models using recent evidence. *Akron business and economic review*. 19(1): 52-63.

Jo, H. and Han, I. 1996. Integration of case-based forecasting, neural networks and discriminant analysis for bankruptcy prediction. *Expert systems with applications*. 11(4): 415-422.

Kane, G. D., Richardson, F. M. and Mead, N. L. 1998. Rank transformations and the prediction of corporate failure. *Contemporary accounting research*. 1(1): 145-166.

Keasey, K. and Watson, R. 2001. Financial distress prediction models: A review of their usefulness. *British journal of management.* 2(1): 89-102.

Knight, F. 1933. Risk, uncertainty and profit. London school of economics and political science, London.

Ko, C. J. 1982. A delineation of corporate appraisal models and classification of bankruptcy firms in Japan. New York University, New York.

Ko, L., Bocher, E. J. and Lin, P. P. 2006. Prediction of corporate financial distress: An application of the composite rule induction system. *The international journal of digital accounting research.* 1(1): 69-85.

Jones, F. L. 1987. Current techniques in bankruptcy prediction. *Journal of accounting literature.* 1(6): 131-164.

Jones, S. and Henser, D. 2004. Predicting firm financial distress: A mixed logit model. *The accounting review.* 79(4):1011-1038.

Kaas, G. V. 1980. An exploratory technique for investigating large quantities of categorical data. *Applied Statistics.* (29): 119-127.

Keasey, K. and Watson, R. 1987. Non-financial symptoms and the prediction of small company failure: A test of Argenti's hypothesis. *Journal of business finance and accounting*. 14 (3): 335-354.

Keasey, K. and Watson, R. 1991. Financial distress prediction models: A review of their usefulness. *British journal of management.* 2 (1): 89-102.

Lau, A. H. L. 1987. A five-state financial distress prediction model. *Journal of accounting research.* 4(1): 127-138.

Lawrence, E. C. 1983. Reporting delays for failed firms. *Journal of accounting research.* 21(2): 287-296.

Le Roux, N. J. and Olivier, P. 1992. Diskriminantanalise en trosanalise as beskrywende tegnieke van finansiele mislukking. *Bestuursdinamika*. 1(1): 55-66.

Loh, W. And Shih, Y. 1997. Split selection methods for classification trees. *Statistica.* (7): 815-840.

Lorange, P. and Nelson, R. T. 1987. How to recognize and avoid organisational decline. *Sloan management review.* 28(3): 41-46.

Mensah, Y. M. 1984. An examination of the stationarity of multivariate bankruptcy prediction models: A methodological study. *Journal of accounting research*. 22(1): 380-395.

Moss Kanter, R. 2003. Leadership and the psychology of turnarounds. *Harvard business review.* (1): 58-67.

Mossman, C. E. , Bell, G. G. , Swartz, L. M. and Turtle, H. 1998. An empirical comparison of bankruptcy models. *Financial review.* 33(2): 35-53.

Muller, G. H., Steyn-Bruwer, B. W. and Hamman, W. D. 2009. Predicting financial distress of companies listed on the JSE-A comparison of techniques. *South African journal of business management.* 40(1): 21-31.

Neville, P. G. 1999. Decision trees for predictive modelling. SAS Institute Inc.

Norton, C. L. and Smith, R. E. (1979). A comparison of general price level and historical cost financial statement in the prediction of bankruptcy. *The accounting review.* 54(1): 72-87.

Okwell, J. 1999. Neural Networks versus CHAID: A white paper from smartFOCUS. (2): 1-14.

Ohlson, J. A. 1980. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of accounting research.* 18(1): 109-131.

Olivier, P. 1992. Mislukkingvoorspelling vir ondernemings met inagneming van tydsdimensies. *South African journal of business management*. 23(1): 21-26.

Ooghe, H. and spaenjers, C. 2010. A note on performance measures for business failure prediction models. *Applied Economics.* 17(1): 67-70.

Oosthuyse, D. 2014. Increasing market liquidity- identifying barriers to liquidity, enabling cross-membership, pooling liquidity across markets. *Building African financial markets 2014.*

Opler, T. C. and Titman, S. 1994. Financial distress and corporate performance. *The journal of finance.* 49(3): 1015-1040.

Pompe, P. P. M. and Feelders, A. J. 1997. Using machine learning, neural networks and statistics to predict corporate bankruptcy. *Microcomputers in Civil engineering*. 12(1): 267-276.

Pompe, P.P.M, and Bilderbeek, J. 2005. The prediction of bankruptcy of small and medium sized industrial firms. *Journal of business venturing.* 20(6): 847-868.

Pretorius, M. 2008. Critical variables of business failure: A review and classification framework. *SAJEMS.* 4 (1): 408-430.

Provincial Treasury. 2012. Business cycles and their impact on the South African economy. *Gauteng Province: National Treasury, Republic of South Africa.* 1-33.

Rees, B. 1995. Financial analysis. Englewood Cliffs: Prentice Hill.

Richardson, F. M., Kane, G. D. and Lobingier, P. 1998. The impact of recession on the prediction of corporate failure. *Journal of business finance and accounting*. 25(1): 167-186.

Rose, P. S., Andrews, W. T. and Giroux, G. A. 1982. Predicting business failure: A macroeconomic perspective. *Journal of accounting and finance.* 1(6): 20-31.

Sarlija, N. And Jeger, M. 2011. Comparing financial distress prediction models before and during recession. *Croatian operational research review.* 2(1): 133-142.

Salehi, M. and Abedini, B. 2009. Financial distress prediction in emerging market: Empirical evidences from Iran. *Business intelligence journal.* 2(2): 398-409.

Schafer, J.L. 1997. Analysis of Incomplete Multivariate Data. Chapman & Hall, London. Second edition. New York: Wiley.

Shah, N. 2014. Developing financial distress prediction models using cutting edge recursive partitioning techniques: A study of Australian mining performance. *Review of integrative business and economics research.* 3(2): 103-143.

Strebel, P. J. and Andrews, G. S. 1977. A simple predictor of company failure: Fact and opinion papers. *Graduate school of business administration, University of the Witwatersrand.* 2(1): 1-9.

Tirapat, S. and Nittayagasetwat, A. 1999. An investigation of Thai listed firms' financial distress using macro and micro variables. *Multinational finance journal*. 3(2): 103-125.

Theng, L. G. and Boon, L. W. 1996. An exploratory study of factors affecting the failure of local small and medium enterprises. *Asia pacific journal of management*. 13(2): 47-61.

Tshitangano, F. 2010. Cost of financial distress model for JSE listed companies: A case of South Africa. University of Pretoria. 1-103.

Van der Walt, J. 2010. Business rescue legislation in South Africa. CRS Business Rescue.

Van Niekerk, M. C. 1993. Die voorspellingswaarde van kontantvloei-inligting. *Meditari.* 1(1): 89-111.

Yoon, Y., Swales, G. and Margavio, T. 1993. A comparison of discriminant analysis versus artificial neural networks. *The journal of the operational research society*. 44(1): 51-60.

Zavgren, C. V. 1983. The prediction of corporate failure: The state of the art. *Journal of accounting literature.* 1(2): 1-38.

Zavgren, C. V. 1985. Assessing the vulnerability of failure of American industrial firms: A logistic analysis. *Journal of business and finance accounting.* 12(1): 19-45.