AN ARTIFICIAL INTELLIGENCE MODEL TO PREDICT FINANCIAL DISTRESS IN COMPANIES LISTED ON THE JSE

by

Francois van der Colff

Submitted in fulfilment of the requirements for the degree

Doctor of Philosophy

in Financial Management Sciences

in the
Faculty of Economic and Management Sciences
at the
University of Pretoria
Pretoria
South Africa

Study leader: Professor F.N.S. Vermaak

Co-study leader: Professor L.M. Brümmer

August 2017

DECLARATION

I, Francois van der Colff, declare that the thesis, *An artificial intelligence model to predict financial distress in companies listed on the JSE,* which I hereby submit for the degree Doctor of Philosophy (Financial Management Sciences) at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution, and that all sources that I have used or quoted herein have been indicated and acknowledged by means of a complete reference system.

Francois van der Colff	August 2017
Signature	Date

ETHICS STATEMENT

I, Francois van der Colff, have obtained, for the research described in this work, the applicable research ethics approval. I declare that I have observed the ethical standards required in terms of the University of Pretoria's Code of ethics for researchers and the Policy guidelines for responsible research.

ACKNOWLEDGEMENTS

I wish to express my sincere gratitude to the following people contributing to the completion of this thesis:

- Prof Frans Vermaak and Prof Leon Brümmer, supervisor and co-supervisor of the thesis, for the professional guidance, encouragement and support they gave me to complete the study;
- Lidia de Wet, for her assistance in the language editing of the thesis; and
- my wife, Hazel, for her inspiration to embark on the study and her continued support and encouragement to complete the study.

Francois van der Colff
August 2017

ABSTRACT

As a prerequisite for an informed decision, a company's financial results are undoubtedly one of the most important aspects to be considered in a financial distress prediction model. To rely purely on financial results for prediction is a risk. The dilemma is that financial variables are backward-looking and point-in-time measures of a company's financial results. Ever-changing quantitative non-financial variables could enhance the decision-making process and should therefore be taken into consideration.

For this research, an artificial intelligence model based on a unique combination of financial, market and quantitative non-financial variables was developed and tested against internationally and South African-developed financial distress prediction models in order to determine its prediction accuracy. Various levels of the artificial intelligence model were separately tested against the two statistical financial distress prediction models. Empirical results of the study proved that a financial distress prediction model enhanced with market and quantitative non-financial variables yielded more accurate results than a model based purely on financial variables.

A two-pronged overview of the theoretical development of financial distress prediction models was given to establish a foundation for the development of a financial distress prediction model for the study.

The reliability, popularity and further development of a statistically based financial distress prediction model were constrained. Constraints such as reliance on outdated financial information in a highly dynamic operating environment and the advent of computer technology and artificial intelligence contributed to a new era in financial distress prediction.

Despite its purported success, neural networks were also subject to various limitations. In an effort to overcome the critical limitations and constraints

experienced in the application of neural network models, researchers have developed derivative financial distress prediction models.

Most of these models are still at the stage of static modelling and are built with sample data, which is collected over an extended period of time. However, variables in the economic and company environment change over time and if the financial distress prediction model is not aligned or adjusted to these changes, the financial distress prediction model could lead to financial distress concept drift. This important criticism against the financial distress prediction models formed the foundation of the study.

In an attempt to deal with the constraints experienced with neural network models, the study applied support vector machines to the financial distress prediction problem. The main difference between neural networks and support vector machines is the principle of risk minimisation. While neural networks implement empirical risk minimisation to minimise the error on the training data, support vector machines implement the principle of structural risk minimisation to minimise the generalisation error by constructing an optimal separating hyperplane in the hidden feature space, using quadratic programming in order to find an optimal solution.

The primary objective of the study was to develop an artificial intelligence-based financial distress prediction model, which incorporated a unique combination of financial and quantitative non-financial variables from a South African perspective. The intention with the proposed financial distress prediction model was to provide a more accurate and timeous company financial health and distress prediction on a financial distress continuum compared with a statistical financial distress prediction model.

A phased approach was followed, first by identifying the variables most often applied to financial distress prediction studies. A principal component analysis was conducted in the final selection of financial and market variables and the model development. The leading, coincident and lagging business cycle indicators as published by the South African Reserve Bank were selected as proxy for quantitative non-financial variables.

A financial distress prediction model was developed based on machine learning principles, enhanced with market and quantitative non-financial variables and compared with existing financial prediction models. The empirical results demonstrated that different combinations of financial, market and quantitative non-financial variables enhanced the accuracy of financial distress prediction models.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND

Company financial distress or bankruptcy can be attributed to a multitude of factors. It can be the result of a single catastrophic event such as non-payment by a major debtor or loss of a contract for some or other reason. Alternatively, bankruptcy can be a terminal point in an extended period of financial distress (Brabazon & O'Neill, 2004:364).

However, *company financial distress*[1], in the context of this study is not synonymous with *bankruptcy*. Financial distress signals that the company is experiencing a crisis, which could be the result of poor management decision-making or its inability to pro-actively identify and remedy potential issues that may affect the company's financial health. Should timeous appropriate remedial action not be taken, the company could potentially descend into bankruptcy.

An alternative view is that financial distress can be attributed to continued high interest rates, recession-squeezed revenue and profit and a heavy debt burden.

Whether financial distress is the result of internal or external factors or a combination of these factors, financial distress prediction has become a preoccupation for researchers and practitioners and has been extensively studied since the early sixties. The consequence of financial distress and potential bankruptcy can be detrimental to stakeholders and the economy at large. In addition, the collective number of failing companies can be regarded as an important indicator of the financial health and robustness of a country's economy (Ahn, Cho & Kim 2000:65).

[1]Hereafter, the term *financial distress* is used.

Financial distress is broadly a situation where a company's cash flow is insufficient to cover its immediate financial obligations such as trade creditors, interest expense and tax. Failure to take pro-active or remedial action to deal with the causes of financial distress could in time force the company into bankruptcy and liquidation. Therefore, timeous and accurate information concerning impending financial distress is of critical importance to company management.

Because the economic cost of bankruptcy is high, there is a need for financial distress prediction models providing timeous and accurate predictions. These financial distress prediction models can be grouped into two broad categories, namely statistical and artificial intelligence models. The following two sections provide an overview of statistical and artificial intelligence or machine-learning models[2]. A third section provides an overview of the application of quantitative non-financial variables to both the statistical and artificial intelligence models.

1.1.1 Statistical financial distress prediction models

Various financial distress prediction models based on the pioneering work of Beaver (1966) on univariate analysis and Altman (1968) on multivariate analysis have evolved and resulted in a large body of research over a number of years. *Univariate analysis* takes into account the relationship between individual financial ratios and financial distress. *Multivariate analysis*, on the other hand, uses multiple financial ratios and weightings to determine a prediction function of financial distress.

Several constraints and disadvantages in the application of the Beaver (1966) and Altman (1968) models compelled researchers and practitioners to explore for alternatives or enhancements. According to Bioch and Popova (2003:1), statistical models are based on several assumptions that are not always present in real-life data.

[2] Machine-learning is a branch of artificial intelligence and is according to Mayer-Schönberger and Cukier (2013:12) misleading because it is not about trying to "teach" a computer to "think" like humans. It is instead about applying mathematics to a large quantity of data in order to infer probabilities: whether a company is financially healthy or distressed.

Hereafter, the terms *artificial intelligence* and *machine-learning* models are used interchangeably.

Other constraints highlighted by Van Gestel, Baesens, Suykens, Espinoza, Baestaens, Vanthienen and De Moor (2003:1) and Kumar and Tan (2004:2) are the reliance of these models on linear separability and normality assumptions among input variables. This is the linear separation between financially distressed and healthy companies and where ratios are treated as independent variables (Neves & Vieira 2004:2).

As a result of these constraints, more sophisticated statistical models such as the Fisher discriminant analysis and the popular logistic regression (logit)[3] models were developed (Chen, 2011b:11262). These models were used both for classification and estimation of financial distress. The reliability of these models has, however similarly been questioned when non-linearity and complexity are present in datasets (Yang 2003:47).

Neves *et al.* (2004:2) and Becerra, Galvão, and Abou-Seada (2005:36) highlighted several additional limitations. Firstly, the choice of regression is a strong bias that restricts the outcome. Secondly, these statistical models are sensitive to exceptions, which are common in financial distress prediction with atypical companies that could compromise the predictions. Thirdly, the patterns need to be linearly separable. Fourthly, samples are assumed to follow a multivariate normal distribution. Fifthly, it is hypothesised that the groups being identified have identical co-variances. Lastly, most of the conclusions, such as the confidence interval, have an implicit Gaussian distribution, which does not hold for many cases.

Although these models may achieve low errors on training data, it may perform poorly on generalisation. In this context, Kim and Yoo (2006:1) describe *generalisation* as the capacity of the model to respond to unknown or unseen inputs that differ from training samples. In order to develop an accurate and reliable financial distress prediction model stable generalisation is required.

[3] Hereafter, the terms *logistic regression model* and *logistic regression* are used interchangeably.

Constraints experienced with statistical models as indicated above and the advancement in both computer technology and artificial intelligence, encouraged and provided the impetus for researchers to develop alternative financial distress prediction models based on artificial intelligence principles.

1.1.2 Artificial intelligence financial distress prediction models

According to Wu, Lee and Tan (2006:328), artificial intelligence models are a class of non-parametric computational intelligence techniques, which relax the functional forms assumed by various statistical models and the assumptions of data distribution. Artificial neural network models[4], decision trees, genetic algorithms, rough sets, and support vector machines are a few examples of artificial intelligence models.

Although neural networks have application in many disciplines such as medicine and engineering, they found application in financial distress prediction with improved results over statistical models.

Neural networks, as non-linear architectures in particular, have to some extent gained initial popularity due to their powerful modelling capability for pattern recognition, object classification and future prediction without many unrealistic a-priori assumptions about the specific model structure and data generating process (Zhang 2007:3).

Other advantages of neural networks highlighted by Chen and Hsiao (2008:1146) are their parallel processing capabilities with great volume, high error tolerance and noise filtering, together with the convenience of not requiring a statistical hypothesis to establish models.

Despite these advantages, neural networks have some weaknesses such as their requirement for a high volume of controllable parameters with the risk of obtaining locally optimal solutions and overfitting. Probably the most important drawback is the fact that their internal functional structure remains unknown once it has been trained,

[4] Hereafter, the terms *neural network models* and *neural networks* are used interchangeably.

referred to as the black-box phenomenon. Although useful results may be produced by the neural network, the actual decision-making process may be difficult or impossible to explain.

This major obstacle resulted in academic interest in neural network modelling with application to financial distress prediction waning to an extent in recent years. A more powerful approach had to be found with better explanation ability as well as high prediction performance.

Because of their relatively high discriminatory power, both the statistical and artificial intelligence distress prediction models are still widely accepted. However, according to Grunert, Norden and Weber (2005:511), they show some disadvantages. One particular aspect evident from the above is that both these categories are predominantly based on financial results of subject companies. Factors such as solvency, profitability and liquidity are primary considerations, which are mostly backward-looking and point-in-time measures. This places the usefulness and reliability of a ratio or financial-based financial distress prediction model in question (Lussier, 1995:8).

1.1.3 <u>Non-financial distress prediction models</u>

It is evident from the above and a more detailed discussion in the following chapters that extensive research has been produced and has proved the suitability of financial variables in the prediction of financial distress. Although consideration of non-financial variables[5] such as management quality and industry perspectives is beyond controversy, there is a lack of quantitative research on this issue, according to Grunert, *et al.* (2005:510).

---

[5] Qualitative and quantitative non-financial variables must be differentiated. For the purposes of the study, *qualitative non-financial variables* refer to variables such as management quality and reputational risk. *Quantitative non-financial variables* refer to macroeconomic indicators such as gross domestic product (GDP) and consumer price index (CPI). See Sections 1.7.6 and 1.7.7.

There appears to be limited consensus on which non-financial variables to include in a financial distress prediction model. Hol (2007) evaluated the co-movement between financial variables and the business cycle in a financial distress prediction. Zong-jun, Hong-xia and Xiao-lan (2006) based their research on a combination of various non-financial indices (ownership concentration coefficient, affiliated debt, pledge and affiliated exchange) with financial indices.

Cybinski (2001), Dunis and Triantafyllidis (2003), Van Gestel *et al.* (2003), Kumar and Tan (2004), Argyrou (2006), Masekesa (2010) and Zhou, Lai and Yen (2010) identified the following macroeconomic indicators as quantitative non-financial variables: the level of activity or demand or growth factor, the cost of capital borrowing factor, the labour market tightness factor, the construction factor and the expenditure (private, public, business) factor.

Furthermore, Becchetti and Sierra (2003:2104) contributed to the body of literature on financial distress prediction by broadening the test on the significance of non-financial variables, such as market share, customer concentration, strength of local competitors, sub-contracting status export status and presence of large competitors in the same region.

Contrary to the above, Lam (2004:568) found that a combination of financial and macroeconomic variables cannot generate significantly higher returns than the average index.

Wang and Li (2007:104) focused on the effect of financial and non-financial ratios on the probability of financial distress, but applied market variables such as: growth ratio per share of equity, net return on assets, earnings per share, interest coverage, net profit margin, pledge, retained earnings ratio and total assets turnover.

In another variation of the application of financial and non-financial variables Sun and Li (2009) focused on the value and importance of experts' experiential knowledge and non-financial information in financial distress prediction. A group decision-making approach was proposed. Yazdipour and Constand (2010:92) agree with the approach of Sun and Li (2009) and argue that the human/managerial/decision-making side of a company cannot be ignored in the area of financial distress prediction.

Scarlat and Delcea (2010:4) based their argument on the view of Prieto and Revilla (2006) that even if non-financial performance indicators have no intrinsic value for company management, they can be applied as a leading indicator of financial performance, especially for future financial performance that is not yet contained in financial results. This argument concurs with the view proposed in the current study, namely that the historical variables as a basis, combined with current or predictive non-financial variables could be useful in determining a company's financial health.

1.2 RESEARCH PROBLEM STATEMENT

1.2.1 Background

In a dynamic and globalised operating environment it is an essential requirement for company stakeholders, whether consisting of management, an investor or a representative labour union, to formulate critical decisions based on internal and external variables. The consequence of these decisions could have far-reaching implications and could potentially affect a company's financial health over a period of time.

As a prerequisite for an informed decision in a dynamic and globalised environment a number of variables, financial and quantitative non-financial, need to be taken into consideration. An array of mechanisms such as artificial intelligence-based financial distress prediction models capturing a combination of these variables is available to assist stakeholders in this decision-making process.

A company's financial results are undoubtedly one of the most important aspects to be considered in a financial distress prediction model. To rely purely on financial results is questionable. As indicated earlier, the dilemma is that financial variables are backward-looking and point-in-time measures of a company's financial results. Ever-changing quantitative non-financial variables could enhance the decision-making process and should therefore be taken into consideration.

Informed decision-making in a dynamic operating environment is considered important in maintaining company financial health and pro-actively avoiding financial distress. However, most financial distress prediction models still rely on static historical financial information and do not take cognisance of both financial and quantitative non-financial variables.

1.3 RESEARCH OBJECTIVES

The study is guided by the specific research objectives set out below.

1.3.1 Primary research objective

The primary objective of this study is to develop an artificial intelligence-based financial distress prediction model that incorporates a unique combination of financial and quantitative non-financial variables from a South African perspective. The intention with the proposed financial distress prediction model is to provide a more accurate and timeous company financial health and distress prediction on a financial distress continuum compared with a statistical financial distress prediction model.

1.3.2 Secondary research objectives

In order to achieve and support the primary research objective relating to the application of an artificial intelligence-based financial distress prediction model from a South African perspective, the following secondary research objectives are emphasised:

- to identify and select a sample representative of South African-listed companies. Predetermined criteria will be applied to identify the sample from a population group of industrial companies listed on the JSE through the INET BFA database, covering a 10-year test period from 2005 to 2014;

- to identify and select financial and quantitative non-financial variables based on predetermined criteria and review of applicable literature;
- to evaluate and validate the De la Rey K-score and Altman Z-score models representative of statistical financial distress prediction models and based on South African data to determine their predictive accuracy;
- to test the null-hypotheses by establishing whether the combination of financial and quantitative non-financial variables in an artificial intelligence-based financial distress prediction model outperform the De la Rey K-score and Altman Z-score models.

Finally, recommendations for future research based on key findings are provided.

1.4 RESEARCH HYPOTHESES

In order to achieve the research objectives and solve the research problem, a detailed comparison of the prediction accuracy of the financial distress prediction models is required. In order to compare the prediction accuracy of the statistical and proposed artificial intelligence financial distress prediction models, Type I and Type II errors are used, which are well-known performance measures in financial distress prediction.

Prediction accuracy is determined as the percentage of subject companies that are correctly classified as either healthy or financially distressed. Zanganeh, Rabiee and Zarei (2011:18) describe Type I and Type II errors as follows:

- Type I error occurs when a healthy company is classified incorrectly as a company in financial distress, and
- Type II error occurs when a financially distressed company is classified incorrectly as a healthy company.

Because a Type I error is more critical than a Type II error in a financial distress prediction problem, models with higher prediction accuracy (least Type I error) are preferred.

Against this background, the study's empirical research (H_0 – null hypotheses) expects to conclude that financial variables in conjunction with quantitative non-financial variables will improve the ability of the proposed artificial intelligence model to predict company financial health more accurately on a financial distress continuum than a statistical financial distress prediction model will do. The study investigates the following research hypotheses:

Hypothesis 1 (H_1): The Altman Z-score statistical financial distress prediction model has higher Type I and Type II errors than the South African-based De la Rey K-score statistical financial distress prediction mode.

Hypothesis 2 (H_2): The artificial intelligence model based on financial variables only has higher Type I and Type II errors than those of the proposed artificial intelligence model based on a combination of financial and quantitative non-financial variables.

Hypothesis 3 (H_3): The statistical financial distress prediction models have higher Type I and Type II errors than those of the proposed artificial intelligence model based on a combination of financial and quantitative non-financial variables.

The research hypotheses are examined by testing its antithesis, the null hypothesis (H_0). By rejecting the null hypothesis, the results of the study will corroborate the research hypotheses (H_1 to H_3). Conversely, should the study fail to reject the null hypothesis (H_0), it can be concluded that the study did not succeed in demonstrating the existence of the investigated relationship.

1.5 IMPORTANCE AND BENEFITS OF THE STUDY

1.5.1 <u>Background</u>

Financial results are usually published a number of months following the company's financial year-end, and by the time an evaluation is done and corrective action instituted, it may be too late. The application of a financial distress prediction model based on historical financial variables only may limit the remedial action that could be taken to prevent eventual default and bankruptcy.

Combining financial and quantitative non-financial variables in a financial distress prediction model may enhance the ability of a particular stakeholder to identify financial distress early, and where applicable, take the appropriate remedial action to avoid default, and ultimately, bankruptcy. The earlier the financial distress is detected, the better the likelihood of avoiding bankruptcy.

There are a number of stakeholders who can benefit from using a dynamic financial distress model based on a combination of financial and quantitative non-financial variables. The following section highlights the areas where the proposed financial distress prediction model is expected to have a meaningful impact.

1.5.2 Determine an appropriate investment strategy

Investors can determine whether financial distress is of a temporary or permanent nature affecting the company's share price negatively. A temporary drop in the share price may be an opportunity to invest in the company in anticipation of an appreciation in the share price once the effect of remedial action has materialised. Should the potential investor determine that the financial distress is more of a permanent nature, investment in this particular company can be avoided. Alternatively, in the case of an existing investment, the investor can decide to divest from the company entirely.

1.5.3 Establish an optimal risk-return lending policy

Funders can determine whether or not to provide new or additional funding to a company. The outcome of the proposed financial distress prediction model should provide an indication of the possibility of restructuring existing funding. On the negative end of a financial distress continuum[6], where default and/or bankruptcy is

[6] See Section 1.7.4.

inevitable, funders can respond timeously to maximise their loan recovery by exercising their legal rights to their security and/or proceed with legal action.

1.5.4 <u>Establish reasonable suppliers' credit terms</u>

Suppliers can utilise the proposed financial distress prediction model in negotiating payment terms with the subject company. The payment terms will depend on where the company is positioned on the financial distress continuum. Further to the right on the continuum, where financial distress becomes more permanent, the supplier may call in payment terms and even request the company to pay cash on delivery of goods and/or services.

1.5.5 <u>Establish a fair monetary and fiscal policy</u>

Public policy makers such as the South African Reserve Bank (SARB) and South African Revenue Services (SARS) could use an aggregated result of the proposed financial distress model in a particular or all industries and formulate monetary and fiscal policy accordingly. The proposed model could, for example, be used as a guideline by these policy-makers to ease interest rates and/or provide tax incentives to stimulate economic growth should there be a general distress situation, or limited to a particular industry, as a result of recessionary conditions.

1.5.6 <u>Establish fair and reasonable wage negotiations</u>

Labour unions could use the proposed financial distress prediction model in their wage negotiations. These negotiations can be fine-tuned and wage increase demands made more realistic in line with the company's financial results.

1.5.7 <u>Improve auditors' going-concern opinion</u>

According to Sun (2007:56), in a dynamic economic environment, auditors with a good knowledge of a company's internal situation often fail to provide an accurate judgement on a company's going-concern conditions. A well-developed financial distress prediction model could serve as an aid to improve an auditor's going concern judgement.

1.5.8 Refine the ability to determine company financial health

Chapter 6, Section 129 (1)(a) of the Companies Act, 71 of 2008, provides for the company board to resolve that the company voluntarily commences with business rescue proceedings and to place the company under supervision, provided that the board has reasonable grounds to believe that the company experiences financial distress.

The proposed financial distress prediction model should allow company management as well as other stakeholders to determine when a company is in fact in financial distress, or where on the financial distress continuum it is positioned.

This will prevent any "affected person or party"[7] to apply through the courts to have the company placed under supervision without reasonable grounds.

The flexibility of the proposed financial distress prediction model will allow each stakeholder to adapt the model according to his or her unique requirements. All stakeholders will utilise the same financial results, and the addition of a non-financial module will allow a particular stakeholder flexibility to tailor the model to its unique requirements.

1.5.9 Value add to the existing knowledge base

An overview of the historical evolvement of financial distress prediction models, from the basic statistical models to the more sophisticated artificial intelligence models, indicates that there is no unified approach in financial distress prediction. In addition,

[7] In terms of Clause 128 (a) of Chapter 6 of the Companies Act No. 71, 2008 an affected person, in relation to a company, means the following:

i. a shareholder or creditor of the company;
ii. any registered trade union representing employees of the company; and
iii. if any of the employees of the company is not represented by a registered trade union, each of those employees or their respective representatives.

it is evident that financial distress prediction is still in its infancy in South Africa, both in academic and practical application.

The study adds and expands on the existing knowledge base in the academic community by comparing the financial distress predictive power of the De la Rey K-score model with that of the Altman Z-score model and also both these models with the results of an artificial intelligence financial prediction model (referred to as the F-score model). Furthermore, the study introduces the benefits derived from combining financial, market and various macroeconomic variables in the models.

The study shows that there is unlimited scope for the successful application of an artificial intelligence financial distress prediction model and its derivatives in South Africa. The study, therefore, introduces a modular financial distress prediction model, which any stakeholder can adapt to his or her unique requirements and circumstances, by adding different combinations of financial, market and quantitative non-financial variables into one financial distress prediction model.

1.6 LIMITATIONS OF THE STUDY

1.6.1 Background

The study has five key limitations, which relates to the context, concepts and theoretical perspectives. These limitations are discussed in the section below.

1.6.2 Limited geographical coverage

The study is limited to companies based primarily within the South African context. Although some of the sample or subject companies trade internationally, their strategy is affected by unique local variables, for example fiscal and monetary policy, legal system, labour demands, and political regime.

1.6.3 <u>Limited to listed companies</u>

Unlisted or private companies are not considered for inclusion in the study sample because access to information is a problem due to confidentiality and public availability of the information.

1.6.4 <u>Limiting sampling criteria</u>

The study is limited to companies listed on the JSE. The study is furthermore limited to companies listed within the Industrials, Consumer Goods, Health Care, Consumer Services, Telecommunications and Technology sectors of the JSE. In addition, companies listed on Development Capital and Venture Capital Boards and Alternative Exchange are considered for inclusion in the study.

All mining and mining-related companies and financial service providers such as bank and long- and short-term insurance companies are excluded from the proposed sample. This criterion is used because of the differences in accounting systems and financial reporting formats, analysis and interpretation methodology, which may be materially different from those applied by companies in the study sample.

1.6.5 <u>No unified approach in financial distress prediction</u>

The literature review covers the general evolvement of the various financial distress prediction models, from the univariate model developed by Beaver (1966) to statistical models and more recently sophisticated artificial intelligence financial distress prediction models. In addition to the existence of a large number of financial distress prediction models, each claiming superior performance over one or the other, no unified approach can be identified in both financial and quantitative non-financial variable selection.

1.6.6 Reliance on pre-programmed software

Sophisticated and advanced software programming ability is required to develop artificial intelligence software. The researcher had to rely on pre-programmed software because he had no ability to develop the required software. XLSTAT® version 2016.3, a statistical data analysis solution for Excel® developed by Addinsoft, was used for model development and all statistical data analysis.

1.7 DEFINITION OF KEY TERMS

1.7.1 Background

The study covers a number of key concepts, namely *financial distress*, *financial health*, *financial default*, *bankruptcy*, *financial distress continuum*, *financial variables*, and *quantitative non-financial variables*. Interchangeable reference is made to the terms *default*, *bankruptcy* and *financial distress* throughout literature reviewed on financial distress prediction. Hereafter, the term *financial distress prediction* will be used.

The following description emphasises the actual distinction to be made between these concepts.

1.7.2 Financial distress

Financial distress can be described as a situation where the company experiences a cash flow constraint for one or another reason. This constraint can be of a temporary nature provided that management has the capability and ability to implement corrective action.

For the purposes of the study, the definition of *financial distress* as defined by the Section 128(f)(i-ii) of the Companies Act 71 of 2008, is relied on:

> "…in reference to a particular company at any particular time, mean that –
>
> (i) It appears to be reasonably unlikely that the company will be able to pay all of its debts as they may become due and payable within the immediate ensuing six months; or
>
> (ii) It appears to be reasonably likely that the company will become insolvent within the immediate ensuing six months."

1.7.3 Financial health

A company's financial results can be either positive or negative, indicating its financial health. In this study, a company is broadly regarded as financially healthy if it reflects a positive free cash flow and optimal gearing for at least two consecutive financial periods.

Should a company reflect negative results it is regarded as financially unhealthy or distressed, described in Section 1.7.2. A company's health indicator can be positioned anywhere on the financial distress continuum, as described in Section 1.7.5.

1.7.4 Default and bankruptcy

As financial distress becomes more of a long-term or permanent event over time or a company defaults regularly on scheduled loan repayments, financial failure and ultimately, bankruptcy may become inevitable. Once a company has reached the point where it is unable to honour its immediate debt obligations, implying commercial insolvency, it can ultimately result in the company becoming factually insolvent where its total liabilities exceed its total assets.

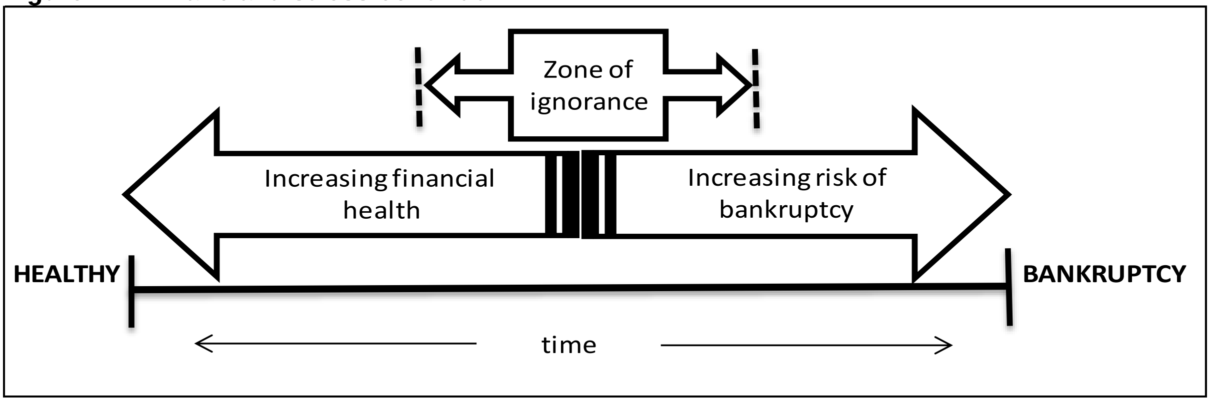
Factual insolvency can result in the company's affairs being wound up, whereby its assets are sold in execution and the net proceeds, if any, distributed amongst creditors by way of a liquidation dividend.

1.7.5 Financial distress continuum

The process where a company experiences financial distress over the short term and progresses over time into a situation of imminent failure and bankruptcy can best be demonstrated on a financial distress continuum (see Figure 1.1).

On a financial distress continuum, as described by Cybinski (2001:29), financial distress can be of a temporary nature at the one end, or over time can become more of a permanent nature at the other end. With movement further to the right-hand side of the continuum where the company's health deteriorates further, any potential remedy or effort to return the company back to financial health, diminishes.

Figure 1.1: Financial distress continuum

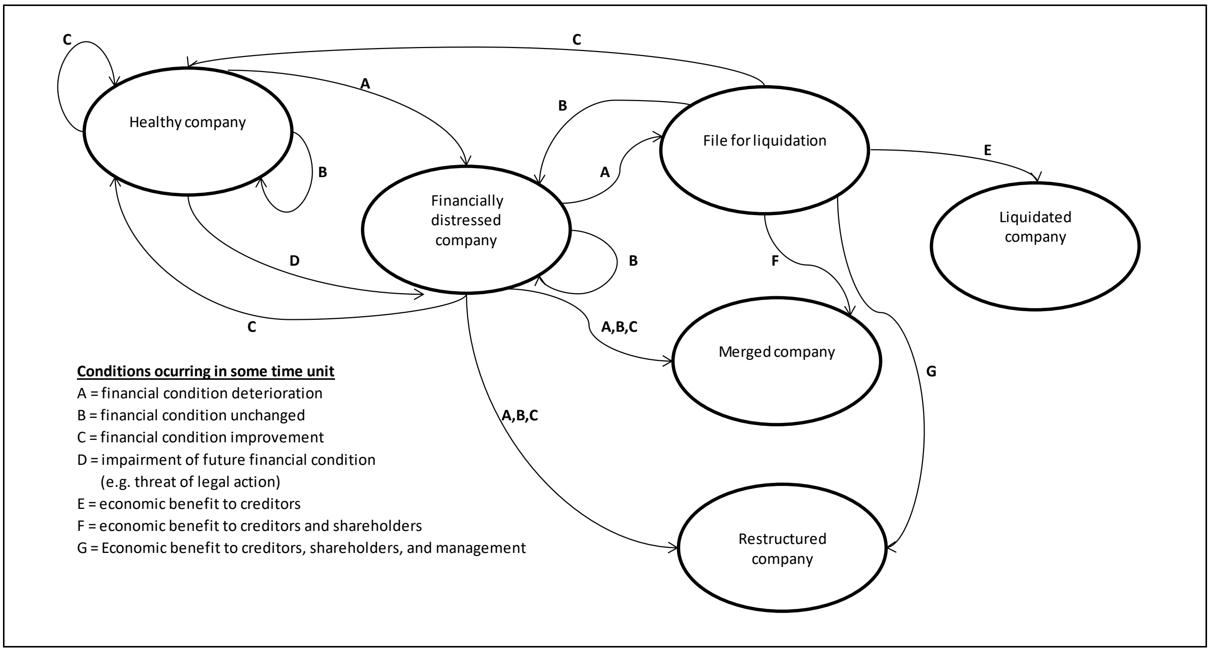


Source: Adapted from De la Rey (1981) and Cybinski (2001:30).

The zone of ignorance represents the highlighted area or period where the company cannot be described as either healthy or financially distressed (De la Rey, 1981). Any extraordinary event, positive or negative, may result in the company becoming either healthier or financially distressed. Early or pro-active detection of potential financial distress, as far as possible to the left-hand side of the continuum, is therefore crucial.

In Figure 1.2, Lensberg, Eilifsen and McKee (2006:680-681) used basic event graph relationships to illustrate a company's continuity/discontinuity relationships. An event is illustrated by a circle and a scheduling edge is illustrated by the connecting line between the circles or event nodes. Over a period of time an event remains the same, represented by the line curving directly back to the same event. In the case where a condition occurs, it leads to another event or event node.

Figure1.2: Event graphs of possible company continuity / discontinuity relationships



Source: Adapted from Lensberg, Eilifsen and McKee (2006:681).

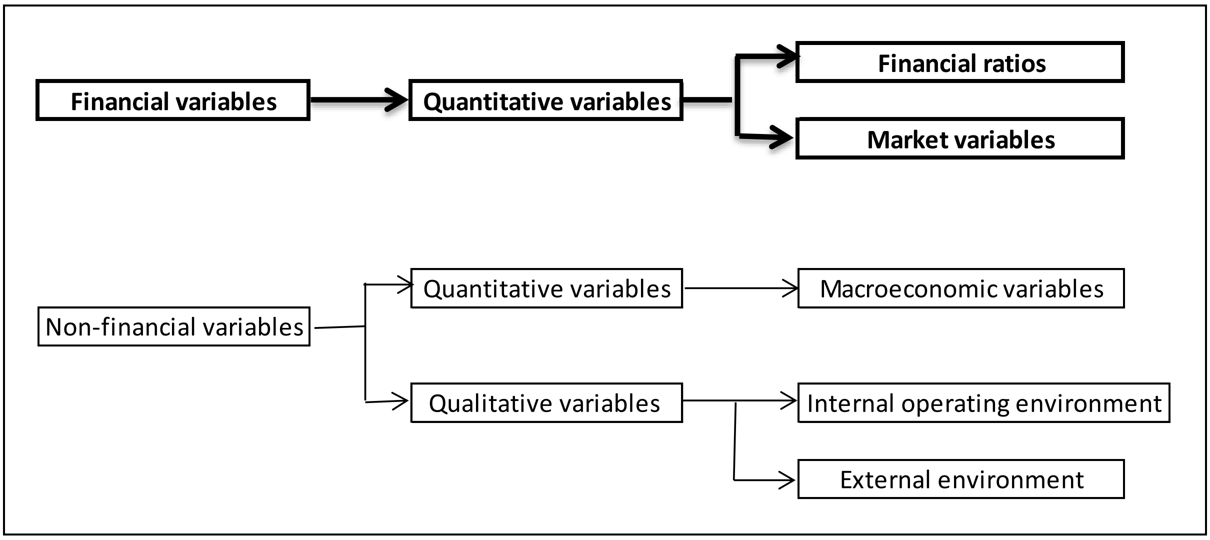
In support of the financial distress continuum illustrated in Figure 1.1 and for the purposes of the study, a direct relationship can be drawn to Lensberg's event graph.

As indicated previously, the financial health of a company can change over time for one or another reason and can either become financially distressed or healthy again. Depending on the severity of financial distress, company management may follow one or a combination of basic options – financial and/or operational restructure, merge with another company, or in a worst-case scenario, file for liquidation. Any one or a combination of these options may result in the company returning to financial health.

1.7.6 Financial variables

Financial variables incorporate two broad categories, which can be described as the quantitative outcome of an analysis and interpretation of a company's financial statements and share price movement. These variables are depicted in the highlighted section of Figure 1.3.

Figure 1.3: Primary categories for financial variables



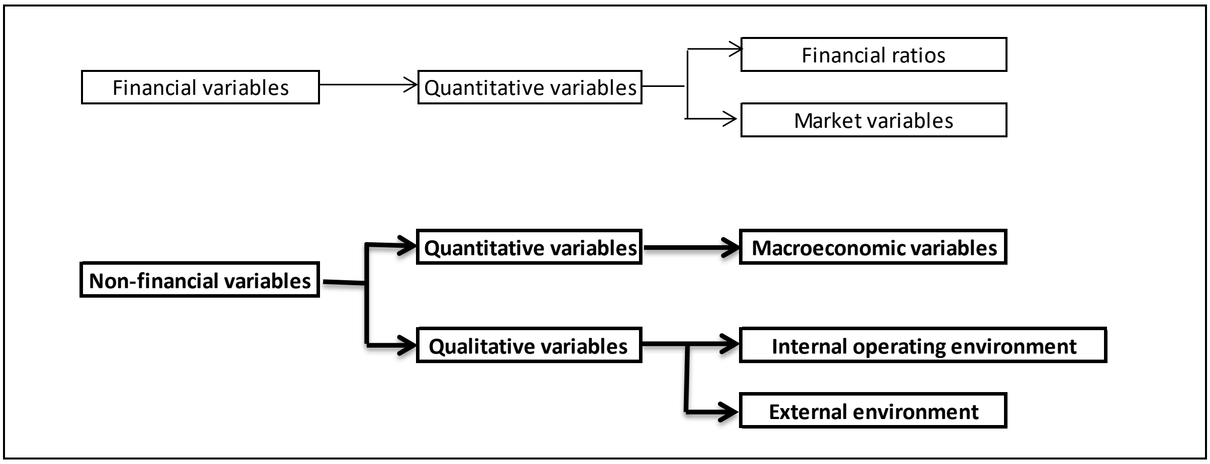
Source: Own compilation.

Financial ratios in particular are derived from an analysis and interpretation of the statement of financial position, statement of comprehensive income and retained earnings and statement of cash flows, and in addition, the supporting notes. On the other hand, market variables are based on an interpretation of the effect of a company's share price movement on its financial results, and *vice versa.*

1.7.7 Non-financial variables

Non-financial variables are those variables not related to a company's financial statements and can be divided into two primary categories, namely quantitative and qualitative variables. These variables are illustrated in the highlighted section of Figure 1.4.

Figure 1.4: Primary categories for non-financial variables



Source: Own compilation.

A secondary division of the quantitative category in Figure 1.4 consists of macroeconomic variables such as the gross domestic product (GDP) and the consumer price index (CPI) or prime lending rate.

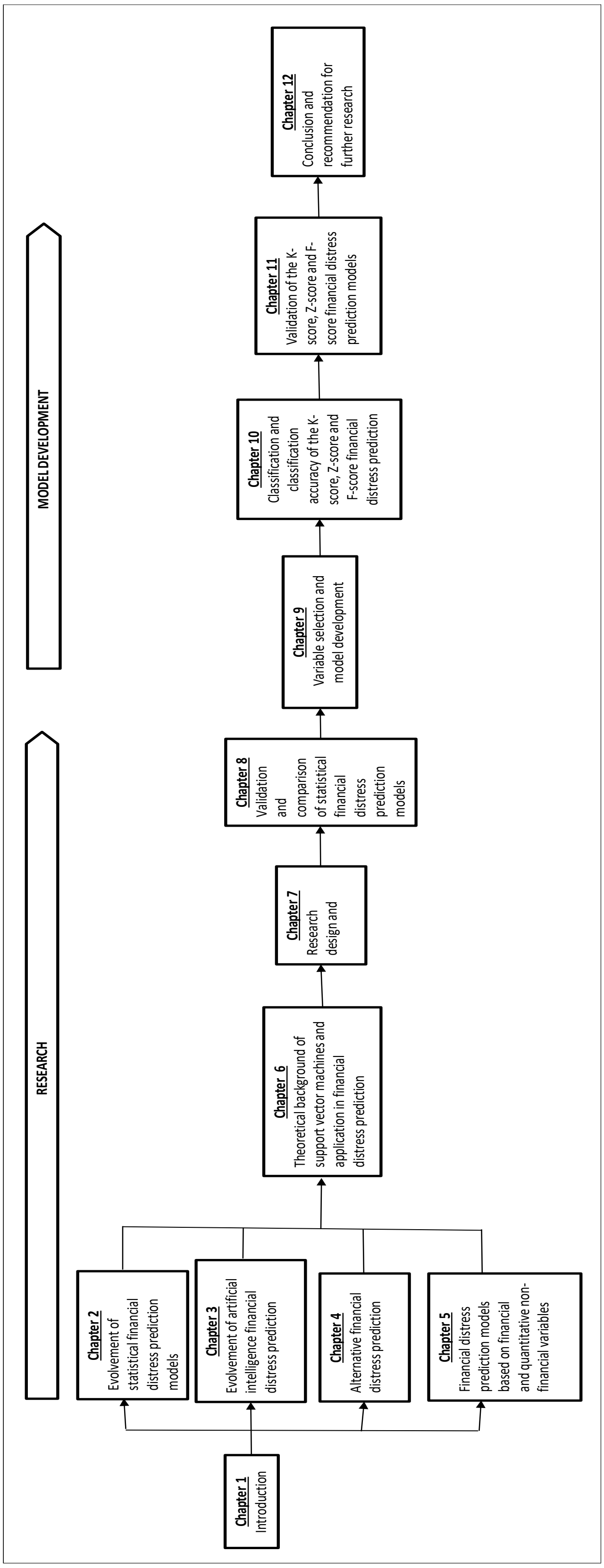
The qualitative category consists of variables from both the internal and external environment in which the company operates, such as record-keeping and financial control, industry experience, management experience, planning, professional advisors, education, staffing, product/service timing and economic timing.

1.8 STRUCTURE OF THE THESIS

1.8.1 Background

Figure 1.5 is a graphic depiction of the study process followed.

Figure 1.5: Research process flow



Source: Own compilation.

Following the introductory chapter, the process consists of two key components. Firstly, Chapters 2 to 8 form the body of the research component. A literature review establishes the basis of the research design and methodology in Chapter 7. Secondly, an artificial intelligence financial distress prediction model is developed and tested in Chapters 9 to 11, based on the research component established in Chapters 2 to 8.

A more detailed overview of the chapters is set out below.

1.8.2 Chapter 1: Introduction

Chapter 1 forms the point of departure of the study. This chapter provides an overview of the evolvement of financial distress prediction models since the mid-sixties and highlights the fragmentation and the fact that there is no unified financial distress prediction model. More importantly, the research problem, objectives and hypothesis are discussed. This is followed by the identification of the limitations of the study and definition of key terms.

1.8.3 Chapter 2: Evolvement of statistical financial distress prediction models

In Chapter 2, the introductory section provides an overview of the historical development of financial distress prediction. Because a substantial volume of research has already been conducted on statistical financial distress prediction models over a number of years, the discussion is limited to an overview of the pioneering research done within this category.

1.8.4 Chapter 3: Evolvement of artificial intelligence financial distress prediction models

Chapter 3 makes up the main body of the literature review section. This chapter focuses on the historical development of artificial intelligence financial distress prediction models. This section of the literature review covers various branches of artificial intelligence models, such as neural network, decision trees, recursive partition, rough sets, and more recently, genetic algorithms and support vector machine models.

1.8.5 <u>Chapter 4: Alternative financial distress prediction models</u>

Chapter 4 focuses on the grouping of artificial intelligence financial distress prediction models into a further three broad sub-categories, i.e. those inspired by the workings of biological neurons (neural networks), those inspired by an evolutionary metaphor (genetic algorithm, genetic programming and grammatical evolution), and finally, those inspired by studies of social interaction (particle swarm and ant colony models). Of these, neural network models have received the most attention, and to a lesser extent, genetic algorithms and programming and grammatical evolution models.

1.8.6 <u>Chapter 5: Financial distress prediction models based on financial and quantitative non-financial variables</u>

This chapter reviews literature on financial distress prediction models incorporating quantitative non-financial variables and its application to financial distress prediction, and covers the final section of the literature review. Because only limited quantitative research has been conducted in this area, it forms the basis for the research problem of the study.

1.8.7 <u>Chapter 6: Theoretical background of support vector machines and application to financial distress prediction</u>

Chapter 6 provides an overview of the theoretical background of support vector machines with its application to financial distress prediction. The first section of the chapter provides an overview of the development of support vector machines, which evolved from constraints experienced with neural network models. The second section focuses on the support vector machine architecture and its proposed application in the study.

1.8.8 Chapter 7: Research design and methodology

Chapter 7 presents the research design and methodology and describes the selection of data and the selection of a sample to be used in the proposed artificial intelligence distress prediction model. The chapter concludes with a discussion on the quality and rigour of the proposed research design and research ethics.

1.8.9 Chapter 8: Validation and comparison of the K-score and Z-score financial distress prediction models

The purpose of this chapter is to validate the K-score and Z-score model and to determine each model's financial distress predictive ability over a number of lead periods based on the Mann-Whitney U, Spearman's rho and weighted efficiency tests.

1.8.10 Chapter 9: Variable selection and model development

Financial, market and quantitative non-financial variables will be identified and selected through a variable reduction process, which will include a principle component analysis. The selected variables will form the basis for the proposed financial distress prediction model.

1.8.11 Chapter 10: Classification and classification accuracy of the SVM-K-score, SVM-Z-score and F-score financial distress prediction models

In this chapter, XLSTAT® will be used to classify the SVM-K-score, SVM-Z-score and F-score models into financially distressed and non-distressed categories. The classification accuracy of each model will be analysed in the second section of the chapter.

1.8.12 Chapter 11: Validation of the SVM-K-score, SVM-Z-score and F-score financial distress prediction models[8]

The XLSTAT® will be used to establish how well the classifier for each of the financial distress prediction models performed.

1.8.13 Chapter 12: Conclusion and recommendation for further research

The final chapter concludes the research with a brief overview of the research objectives and hypothesis and discusses its contribution. Lastly, a proposal for future research based on the research findings is made.

1.9 CONCLUSION

Chapter 1 formed the foundation of the study. An overview of the historical development of statistical and artificial intelligence-based financial distress prediction models was provided. This chapter concludes with a research process flow and outline of the study.

The following chapter provides a detailed review of the historical development of financial distress prediction models from the early sixties to the present. The main focus of Chapter 2 is a review of statistical financial distress prediction models, which established the foundation of most financial distress prediction models currently in use.

[8] The original K-score and Z-score models have been renamed as SVM-K-score and SVM-Z-score (see Section 10.1).

CHAPTER 2

EVOLVEMENT OF STATISTICAL FINANCIAL DISTRESS PREDICTION MODELS

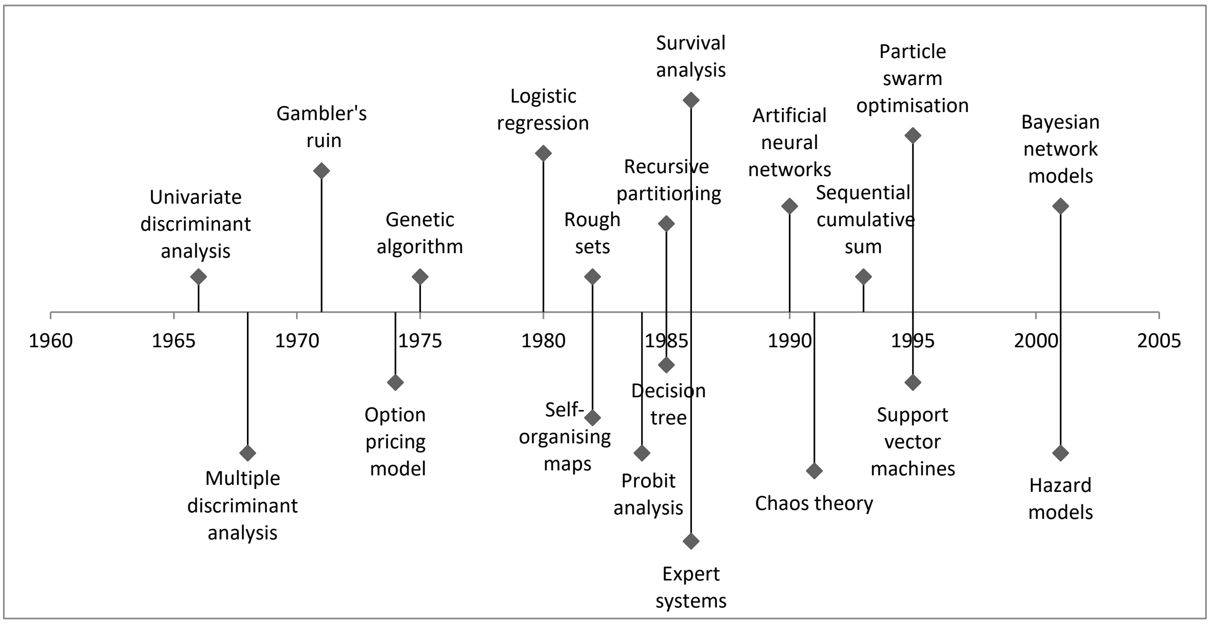
2.1 INTRODUCTION

Companies function in a dynamic environment where financial results are ultimately affected by a combination of internal and external factors. These very same factors can potentially contribute to financial distress, and ultimately, a company's demise. The increasingly volatile economic conditions, increasingly complex socio- and political environment and changes in regulatory policy are examples of factors that could potentially affect a company's financial health. This, in turn, necessitates a more complex decision-making process.

The ability of stakeholders to anticipate financial distress and implement remedial measures timeously is essential, and depends on the identification and application of appropriate analytical tools (Bunyaminu & Issah, 2012:7). As companies and their operating environment become more dynamic and complex, the demand for more sophisticated financial distress prediction models becomes imperative.

Since the mid-1960s, a number of financial distress prediction models have been developed. For the purposes of the study, a survey was conducted of available literature. Figure 2.1 depicts a timeline of the development of the more important financial distress prediction models.

Figure 2.1: Historical development of financial distress prediction models



Source: Own compilation.

In this timeline, there is no clear differentiation between the changeover from the period of statistical model development to the period of neural network model development, and more recently, those models inspired by an evolutionary metaphor.

Based on a review of available literature, there appears to have been a gradual changeover between the various stages. The interest in statistical model development diminished gradually as new research was conducted on the application of neural network models in financial distress prediction.

Aziz and Dar (2006:19) categorised financial distress prediction models into three broad categories: statistical models, artificial intelligent models and theoretical models. The study analysed of 46 articles reporting on 89 empirical studies relating to financial distress prediction.

Aziz and Dar (2006:29) conclude that the statistical models, specifically the multiple discriminant analysis and logistic regression models are the most frequently used in predicting financial distress. The artificial intelligence and theoretical models are less frequently used because they are relatively new and complex.

The more important models in the three categories are set out in Table 2.1.

Table 2.1: Three categories of financial distress prediction models used in studies

Statistical Models	Artificial Intelligence Models	Theoretical Models
• Univariate discriminant analysis (UDA)	• Recursively partitioned decision trees	• Balance sheet decomposition measures (BSDM)
• Multiple discriminant analysis model (MDA)	• Case-based reasoning models (CBR)	• Gambler's ruin theory
• Linear probability model (LPM)	• Artificial neural network models (ANNM)	• Cash management theory
• Logistic regression models	• Genetic algorithms (GA)	• Credit risk theories
• Probit models	• Rough set model (RS)	
• Cumulative sums procedures (CUSUM)	• Support vector machines (SVM)	
• Partial adjustment processes		

Source: Adapted from Aziz & Dar (2006:19-22).

The predictive accuracy of all the statistical models is found to be generally good. Although the artificial intelligence and theoretical models reflect a slightly better predictive ability than the statistical models, the assertion is based on a smaller number of studies. This is compared with the consistently higher accuracy of the multiple discriminant analysis and logistic regression models, which was achieved through a larger number of studies with smaller adjusted standard deviations. This suggests that the multiple discriminant analysis models may provide the most reliable methods of financial distress prediction. On the other hand, not much attention has been given to theoretical models from an empirical point of view.

Kim and Kang (2012:9308) narrow these financial distress prediction models into two broad categories, namely statistical and artificial intelligence financial distress prediction models. Multiple regression, discriminant analysis and logistic regression models are some of the more well-known statistical models used in financial distress prediction. Examples of artificial intelligence financial distress prediction models are decision trees, neural networks, support vector machines and genetic algorithms.

In order to establish the nature of financial distress prediction model development, a survey of methodologies applied was conducted for the purposes of this study. The survey of 219 articles published in academic journals covered a 13-year period from 2000 to 2012. These methodologies were categorised as either an evaluation of an existing model, the development of a combined model, a comparison between various models, the development of a model derived from an existing model, or uncategorised where none of the categories were applicable. The result is illustrated in Table 2.2.

Table 2.2: Categories of methodologies applied in studies from 2000 to 2012

	Number of studies surveyed	Model evaluation	Hybrid of one or more models	Comparison between models	Derivative of one or more models	New development	Not categorised	TOTAL *
2000	10	-	50.0	10.0	40.0	-	-	**100.0**
2001	5	-	-	-	80.0	-	20.0	**100.0**
2002	5	-	-	20.0	80.0	-	-	**100.0**
2003	11	-	-	27.3	72.7	-		**100.0**
2004	13	7.7	15.4	53.8	15.4	7.7		**100.0**
2005	18	22.2	-	50.0	22.2	-	5.6	**100.0**
2006	21	38.1	4.8	28.6	28.6			**100.1**
2007	23	43.5	4.3	30.4	8.7	-	13.0	**99.9**
2008	17	29.4	23.5	41.2	5.9	-	-	**100.0**
2009	14	14.3	35.7	7.1	28.6	7.1	7.1	**99.9**
2010	21	38.1	28.6	4.8	19.0	4.8	4.8	**100.1**
2011	28	42.9	21.4	28.6	3.6	3.6	-	**100.1**
2012	33	39.4	12.1	30.3	18.2	-	-	**100.0**
TOTAL	**219**	**28.8**	**15.5**	**27.9**	**22.8**	**3.0**	**1.4**	**99.4**

* Total may not add up to 100% due to rounding errors

Source: Own compilation.

From Table 2.2, it can be concluded that the largest percentage or 28.8% of methodologies applied relates to an evaluation of existing models, with the next largest category at 27.9% relating to a comparison of two or more models. The following category at 22.8% relates to models derived from existing models. This category can technically not be regarded as a new development. Although, still not regarded as a primary or new development, the hybrid category can be viewed as a movement in the direction of a new development. This category consists of 15.5% of the total number of methodologies surveyed. A further category, falling into the new development category, makes up only 3% of the total number of models surveyed. What could be established was that models in this category were not ground-breaking as they appear not to have attracted further meaningful academic attention after the date of publication.

Based on the above survey, it can be expected that future developments will either emanate from a hybrid model based on the combination of two or more existing models or a derivative of an existing model. These developments should originate from the three broad categories devised by Aziz and Dar (2006:19), namely the statistical models, artificial intelligence and theoretical models.

Due to the magnitude of literature available on financial distress prediction and for ease of reference and simplicity, the discussion in the following section is based on three broad categories. Emphasis will be placed on univariate analysis, multiple discriminant analysis (representing statistical models) and artificial intelligence financial distress prediction models, which will form the basis for the literature overview in the study.

The aim of the following section is to provide a review of the evolvement of statistical financial distress prediction models over the years, with the focus on the univariate analysis and multiple discriminant analysis models. The chapter concludes with a review of a South African-based multiple discriminant analysis model.

2.2 UNIVARIATE ANALYSIS

The initial research employed to examine the use of financial ratios as an indicator of financial distress focused on a single ratio identification that provided predictive utility. The emphasis was therefore on univariate failure analysis.

According to Hall (2002:37), when research is based on a singular characteristic of a set of objects regardless of other variables and characteristics exhibited by an object, the focus is then on univariate analysis or the characteristic of one variable. Univariate analysis explores each variable in a data set separately. Subsequent analysis evaluates a range of values, as well as the central tendency of the values. Univariate analysis describes the pattern of response to the variable. Further, it describes each variable in isolation and represents a statistical procedure founded on the employment of one dependent measure.

Beaver (1966) produced the pioneering research in financial distress prediction based on a single financial ratio and was the first researcher to apply a univariate discriminant analysis on 30 financial ratios of a paired sample of financially distressed and healthy companies in order to predict financial distress. These ratios were grouped into six groups: cash flow ratios, debt-to-total assets ratio, liquid assets to total assets ratio, liquid assets to current debt ratio, turnover ratio, and net income ratio.

In selecting the financial ratios to be included in his univariate analysis model, Beaver (1966) use a dichotomous classification test in order to identify those ratios that were the best in classifying the companies as financially distressed or healthy.

According to Hanson (2002:43), Beaver's (1966) univariate analysis model is based on four concepts, everything else being equal. These concepts are as follows:

- Firstly, the more liquid a company's assets are, the smaller the possibility of financial distress.
- Secondly, the larger the net cash flow from operations, the smaller the possibility of financial distress.
- Thirdly, the larger the amount of liquid assets required to fund operating expenditure, the greater the possibility of financial distress.
- Lastly, the larger the amount of liquid asset necessary to fund operating expenditure, the greater the probability of financial distress.

The results of Beaver's (1966) study indicate that the ratios for financially distressed companies are different from those of healthy companies. To establish the extent of the difference, the relative frequency distribution of each ratio was determined, followed by the identification of the ratio value at which point the possibility of being classified into the appropriate distressed *versus* healthy group, is high or low for each ratio.

The Beaver (1966) study further established that the predictive ability of specific financial ratios, especially the cash flow to total debt ratio, provided useful information in assessing the possibility of financial distress. The results indicate that financially distressed companies have lower cash flow and smaller amounts of liquid assets than healthy companies do.

In addition, it was established that ratios tended to be more successful in predicting companies that were not susceptible to financial distress, without any uniformity in their level of accuracy. Ratios in general did not predict equally. This result was expected by Beaver (1966); the ratio for net income was found to be the second-best predictor because correlation with the best ratio was much higher than other ratios.

The Beaver (1966) study concludes that financial ratios can clearly indicate the difference between financially distressed and healthy companies at a 78% accuracy level prior to financial distress. However, according to Thevnin (2003:31), financial ratios do not explain the significance of the difference.

According to Thevnin (2003:8), four key empirical tests can be derived from Beaver's (1966) research: firstly, dichotomous classification ("yes" or "no" or "1" or "0"); secondly, comparison of means; thirdly, industry effects; and fourthly, analysis of the likelihood that financial ratios or accounting data can predict companies that are susceptible to financial distress for at least five years prior to financial distress.

Other studies following on from Beaver's (1966) study argue that ratio analysis, although a good predictor, lacks accuracy. Most particularly, the operational cash flow ratio lacks accuracy in turbulent economic times. The argument against the cash flow ratio is that it does not have the incremental predictive ability of the accrual-based ratios (Thevnin, 2003:31).

Beaver's (1966) study is highly credited since he initiated ratio analysis in the study of financial distress prediction. Using cash flow concepts as a theoretical framework, it was found that some financial ratios could be used to discriminate between financially distressed and healthy companies. According to He (2002:16), the results of the univariate analysis indicate a certain success of predictive accuracy up to five years before financial distress. However, this is of limited use when the company's financial statements are used after or shortly before financial distress (Van der Colff, 2012:14-16).

Balcaen and Ooghe (2004:6) highlight the important advantages and disadvantages of univariate analysis. The main advantage related to univariate analysis is its simplicity; it does not require any statistical knowledge. For each ratio, the ratio value for a particular company is compared with a predetermined cut-off point, and a classification is then decided on accordingly.

The view of Balcaen and Ooghe. (2004:6-7) is extended to include the following additional disadvantages: firstly, a company can be classified into only one ratio at a time, which may cause inconsistent classification, should multiple ratios be considered for a single company, also referred to as an 'inconsistency problem'; secondly, because most variables are correlated to one another, the use of ratios in a univariate analysis makes it problematic to assess the importance of a single ratio.

In the same context, the univariate analysis model contradicts reality in that the financial status of a company is a complex multi-dimensional concept, which cannot be assessed by a single ratio. In addition, the optimal cut-off point for a specific variable is chosen by trial and error and on an *ex post facto* basis, which implies that the actual financial distress status of the companies in the sample is known. The consequence is that the cut-off points may be sample-specific and it is possible that the classification accuracy of the univariate analysis model is lower when used in a predictive context.

Gepp (2005:6) supports the view of Balcaen and Ooghe. (2004), namely that Beaver's (1966) univariate approach did not contain an overall measure of financial distress, which led to the problem that different ratios may provide conflicting prediction results for a particular company. Gepp (2005:6) emphasises the fact that a single ratio cannot encompass the complexity of financial distress. The error in Beaver's (1966) model was estimated at 22% Type I error and 5% Type II error for one year ahead of financial distress. Further, although the Type I error remained constant for longer-term financial distress prediction, the Type I error increased as the financial distress prediction period lengthened.

Another disadvantage emphasised by Argyrou (2006:11-12) is that univariate analysis fails to take the relative cost of Type I and Type II errors into account. Beaver (1966) assumed that these costs were equal.

The fact remains that Beaver's (1966) innovative research was the primary impetus for statistical financial distress prediction as it is known today.

2.3 MULTIPLE DISCRIMINANT ANALYSIS MODELS

In an effort to solve the question of Type I and Type II error classification accuracy in the predictive context, Altman (1968) proceeded to convert from univariate analysis to multiple discriminant analysis models. The *multiple discriminant analysis model* can be described as a statistical technique used to classify an observation into one of several *a-priori* groupings dependent on the observation's individual characteristics and attempts to derive a linear combination of the characteristics, which best discriminates between the groups.

Altman's (1968) Z-score model and subsequent derivatives such as his zeta model have become the basis or reference point for a number of comparative studies.

The strength of the multiple discriminant analysis model lies in its ability to measure a company's financial attributes by analysing several ratios simultaneously as well as the interaction between these ratios. A composite multiple discriminant analysis score is compared with a single score in the univariate analysis model to differentiate between a financially distressed and a healthy company.

An example of a multiple discriminant analysis model is found in Altman's (1968) Z-score model in the equation below:

$$Z = 0.012X_1 + 0.014X_2 + 0.033X_3 + 0.006X_4 + 0.999X_5 \quad (1)$$

where: Z = overall index

X_1 = working capital / total assets

X_2 = retained earnings / total assets

X_3 = earnings before interest and taxes / total assets

X_4 = market value of equity / total liabilities

X_5 = sales / total assets

In this equation, Altman (1968) used data from 1946 to 1965 on 33 pairs of financially distressed and healthy companies matched by industry and size. Altman (1968) states in his study that all companies with a score greater than 2.99 fall into the healthy category, while those with a Z-score below 1.81 are categorised as financially distressed companies. The area between 1.81 and 2.99 is regarded as a grey area because of its susceptibility to classification error. The midpoint of the interval is 2.675, which is chosen as the Z-value that discriminates best between financially distressed and healthy companies.

Altman (1968) states that if the multiple discriminant analysis model is used correctly and periodically, it has the ability to predict financial distress early enough to enable management to realise its extent in time and consider corrective action to avoid financial distress. He admitted to some limitations of his study, such as that the predictive ability of the multiple discriminant analysis model decays rapidly when the prediction horizon is extended beyond one year. Altman (1968) was unable to discriminate as accurately when the horizon was greater than two years prior to financial distress, whereas the Beaver (1966) univariate analysis model could show some predictive power up to five years before financial distress (Van der Colff, 2012:18).

There were various subsequent attempts to improve on the Altman (1968) research (Thevnin, 2003:35). One such example is the Deakin (1972) study where the Beaver (1966) and Altman (1968) research was combined into a new model. In his study, Deakin (1972) sought to determine whether there was a linear combination for companies in financial distress. He wanted to predict beyond a specific time when a company was susceptible to financial distress, and he established that in times of financial distress, companies tended to change their behaviour with respect to their capital structure.

Some of the variables in the Deakin (1972) study were modifications to a previous study that he completed. He attempted to revert to the original ratios tested in the Beaver (1966) univariate model and incorporate a random, rather than matched, sample of healthy companies. The resulting discriminant equation outperformed the classificatory accuracy Altman (1968) had achieved and was able to discriminate effectively up to three years in advance of financial distress. However, when tested against a validation sample, Rees (1995:307) indicated some inconsistency suggesting that there was considerable instability in the estimated model.

The overall conclusion in Deakin's (1972) study was that discriminant analysis could be used with a high degree of accuracy in financial distress prediction and that some ratios contributed more than others in financial distress prediction. It was also concluded that some ratios provided better predictability than others close to the point of financial distress. The study indicated that the multiple discriminant analysis model could be used with a high degree of accuracy to predict financial distress three years in advance, which was an improvement on the Altman (1968) study where financial distress with an accuracy of two years in advance was predicted.

Libby (1975) modified Deakin's (1972) study to demonstrate that financial ratios could have better predictive values in conjunction with multivariate techniques. The primary purpose of Libby's (1975) study was to determine whether financial ratios could provide useful information *ex ante* with respect to financial distress. The study confirmed that financial ratios, although lacking certain abilities *ex ante*, could provide a prognosis of the financial affairs of a company with respect to potential financial distress. Libby's (1975) evaluation of the predictive power of financial ratios shows that they provide company management with the ability to predictively evaluate financial distress.

The Blum (1974) study notes that financial ratios provide relatively accurate financial distress predictions, but concludes that the predictability decreases if the forecast extends beyond two years. The most conclusive findings indicate that traditional financial ratios alone can provide accurate information when companies are susceptible to financial distress for a limited period. However, it can predict financial distress beyond a two-year period without substantial variability across industries.

Thevnin (2003:37) questions the accuracy of Blum's (1974) assessment of financial distress because the study lends itself to interpretations that are confusing and faulty. One of the faulty interpretations is that liquidity, which is a normal trend to the point of being above average, can lead one to construe that the company is financially healthy, when it might not be the case. Furthermore, the liquidity ratios might not be as good in assessing companies that could be susceptible to financial distress because the emphasis tends to be very specifically on liquidity to such an extent that other warning signs might be left undetected.

Irrespective of whether the univariate analysis or multiple discriminant analysis models are used in financial distress prediction research, previous research results indicate that ratios can provide an accurate measure on a short-term basis. Beyond five years, deterioration and decrease in the level of accuracy of these ratios become noticeably apparent (Thevnin, 2003:37).

Various studies focused on resolving the weakness of financial ratios beyond the extended time horizons (Altman 1968). The Beaver (1966) study provided the same conclusion and established the foundation of ratio analysis, which the Altman study expanded upon using multiple discriminant analysis (Thevnin, 2003:38).

2.4 A MULTIPLE DISCRIMINANT ANALYSIS MODEL IN THE SOUTH AFRICAN CONTEXT

In the South African context, De la Rey (1981) developed a financial distress prediction model based on the multiple discriminant analysis model, namely the K-score model. The objective was to distinguish between financially healthy and financially distressed companies.

The model was developed by paying attention to various combinations of financial ratios. The following techniques were applied in determining the most appropriate combination of ratios:

A step-by-step procedure of discriminatory analysis was used to test the ratios for the different combinations.

Standard deviations were used to point out ratios which were showing a distinct difference between healthy and financially distressed groups. The ratios were then arranged so that ratios determining the liquidity position of a company were in one group, profitability ratios in another group and leverage ratios in the next group. Standard deviations were thus applied to determine the best ratio per group and these were then used as combinations.

A third technique used in the choice of ratio combinations was factor analysis because it had the advantage that ratios with a high correlation or those which showed a certain relationship were grouped together. This reduced the number of ratios that were available as variables in a model. With the selection of different ratios, an attempt was made to include at least one of the following groups of ratios: liquidity ratios, profitability ratios, cash flow funds ratios and other liquidity ratios.

The fourth technique was used to test the combinations recommended by other researchers.

Lastly, combinations that were developed intuitively, according to a trial and error method, were tested. In this manner, De la Rey (1981) tested a total of 194 combinations.

The K-score model used financial information of 26 pairs of healthy and financially distressed South African companies over a period from 1972 to 1997. The K-score model bears the following notation:

$$K = -0.01662a + 0.0111b + 0.0529c + 0.086d + 0.0174e + 0.01071f - 0.06881 \tag{2}$$

where: K = overall index

a = (total outside financing / total assets) x 100

b = (income before interest and tax) / average total assets) x 100

c = total current assets and listed investments / total current liabilities

d = (income after tax / average total assets) x 100

e = (net cash flow / average total assets) x 100

f = (stock / inflation-adjusted total assets) x 100

The function of the "-0.06881" at the end of the equation is to return the point of separation between healthy and financially distressed companies to zero. Had this not been done, the point of separation would have been 0.06881 rather than zero, which, according to Steyn, Warren and Jonker (2000:104), is not acceptable. The model's zone of ignorance stretches from –0.19 to +0.20. Any company with a K-score below –0.19 is certain to fail unless positive corrective steps are taken, while a K-score above +0.20 is regarded as relatively safe.

The K-score model successfully scored 94.5% of the healthy companies and 98.6% of the financially distressed out of a sample of 138 financially distressed and 255 healthy companies. The average success rate was 96.6%.

2.5 CONCLUSION

Research by Beaver (1966) and Altman (1968) provided the impetus for all subsequent research and application of financial distress prediction models. The main criticism against the Beaver univariate analysis model is based on the fact that the financial status of a company as a complex multi-dimensional concept cannot be assessed by a single ratio.

To overcome this constraint, Altman (1968) developed a multiple discriminant analysis model, which strength lies in its ability to measure a company's financial attributes by analysing several ratios simultaneously as well as the interaction between these ratios. The main constraint associated with the multiple discriminant analysis model is related to its predictive ability, which decays rapidly when the prediction horizon is extended beyond one year.

Several subsequent attempts by Deakin (1972), Blum (1974), Libby (1975) and Thevnin (2003) followed, but none with notable improvement in resolving the predictive ability of the model beyond a certain point. However, in a recent study Altman, Iwanicz-Drozdowska, Laitinen and Suvas (2017:132) state the positive is that the discriminant analysis model is non-complex and can still be used with some level of accuracy.

In the South African context, De la Rey (1981) developed the K-score model, based on the multiple discriminant analysis model. This model is also non-complex. However, it is subject to similar constraints to any other international multiple discriminant analysis models.

The K-score model and some other popular international rating models such as the Moody's, Standard & Poor's and KMV models are all subject to a critical constraint in that they tend to be reactive rather than predictive. These models, including the K-score model, predominantly rely on historical and point-in-time financial information. The information might be completely outdated by date of publication, and to some extent, not reliable in a highly dynamic operating environment.

These constraints and the advent of computer technology and artificial intelligence contributed to a new era in financial distress prediction, discussed in the following chapter.

CHAPTER 3

EVOLVEMENT OF ARTIFICIAL INTELLIGENCE FINANCIAL DISTRESS PREDICTION MODELS

3.1 BACKGROUND

In an effort to overcome the restrictions of multiple discriminant analysis and other statistical financial distress prediction models, researchers applied artificial neural network analysis to a variety of problems with a special emphasis on financial distress prediction.

Building on Altman's (1968) multiple discriminant analysis model, the Odam and Sharda (1990) study was one of the first to apply an artificial neural network model to financial distress prediction. The authors applied the financial ratios used in the Altman (1968) multiple discriminant analysis model to a neural network model for the purpose of comparison.

The neural network model architecture applied by Odam and Sharda (1990) consisted of a three-layer neural network with five hidden nodes. Considering 128 companies, approximately 191 000 iterations were run in a 24-hour period, correctly classifying all sample entities as either financially distressed or healthy compared with an 86% accurate classification rate of a benchmark multiple discriminant analysis model.

According to the Odam and Sharda (1990) and Atiya (2001:930) tests, the neural network achieved a Type I correct classification accuracy in a range between 77.8% and 81.5% and Type II accuracy in a range between 78.6% and 85.7%. The corresponding accuracy for the multiple discriminant analysis model was in an accuracy range of between 59.3% and 70.4% for Type I accuracy and for Type II accuracy in a range between 78.6% and 85.7%

Coats and Fant (1993:143) cite a number of studies indicating that neural networks are at least as successful as the multiple discriminant analysis model in terms of overall accuracy in financial distress prediction.

Charalambous, Charitou and Kaourou (2000:403) state that neural network models can effectively capture and represent complex relationships in areas where statistical models do not perform well   Considerable research has been conducted comparing financial distress predictive accuracy of neural networks with other traditional models.

Neural network models fall within the broader category of artificial intelligence models and have become a popular research subject in various disciplines such as medicine, politics, technology and business (Charalambous *et al.,* 2000).

Application of neural network models to the domain of financial distress prediction is evident in the high volume of related research, which indicates that neural network models can effectively capture and express complex relationships where statistical models are limited.

The following sections provide an overview of a number of contributions to the neural network architecture and the types of neural networks. This is followed by a discussion of constraints experienced in the application of neural networks, which led to the transition to derivative models where various other artificial intelligence financial distress prediction models were developed.

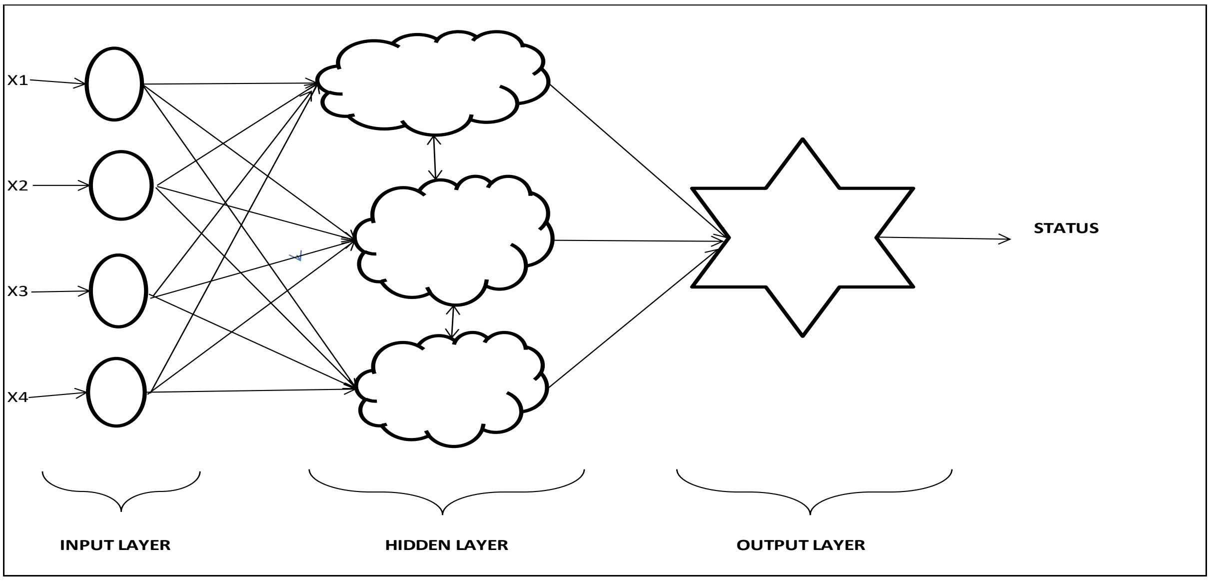
3.2 TYPOLOGY OF A NEURAL NETWORK

Coats and Fant (1993:143) define a *neural network* as a calculation model made up of a number of simple, highly interconnected processing elements which process information by their dynamic state responses to external inputs. This definition highlights two key elements in a neural network, namely processing elements and interconnections.

Coats and Fant (1993:143-144) expand as follows on the key elements: each processing element receives and combines input signals and transforms them into a single output signal. Each output signal, in turn, is sent (from its processing element) as an input to many other processing elements. Signals are passed around the network *via* weighted interconnections or links between processing elements. Network knowledge is stored both in the way the processing elements connect in order to transfer signals and in the nature and strength of interconnections.

Neural network models broadly emulate the data processing function of the human brain. Data are observed through various human sensors such as visual and audio. The brain interprets the data by way of a complex learning process to develop information on which the human formulates a decision or action. The architecture of a typical neural network is shown in Figure 3.1.

Figure 3.1: Architecture of a typical neural network model



Source: Adapted from Olson, Delen and Meng (2012:466).

Koh and Low (2004:464) similarly describe *neural networks* as being modelled after the human brain, which can be perceived as a highly connected network of neurons (called nodes in neural network terminology). Each node, in a layer of nodes, receives inputs from at least one node in a previous layer and combines the inputs and generates an output to at least one node in the next layer.

One or more hidden layers may exist between the input and output layers or nodes. A common neural network consists of one input, one hidden and one output layer each. In combining inputs and generating outputs, each node does a calculation (to combine the inputs) and a transformation (to generate an output).

Each connection between nodes has a weighting that determines how the input from a prior node is to be combined with other inputs to be received by the next node.

Muller, Steyn-Bruwer and Hamman (2009:24-26) provide more details and describe a *neural network* as having multiple layers: the input layer, weighting, the hidden layer (of which there may be several), and the output layer.

These layers have the following functions:

- **Input layer:** The function of the input layer of neurons is to feed the input variables into the network without processing any of the input data. Each input corresponds to a single attribute. For example, in financial distress prediction, an attribute can be one of a multitude of financial ratio values.
- **Weight:** This is a key element in a neural network. The weight expresses the relative strength of the initial entering data or the various connections that transfer data from layer to layer. As highlighted by Trippi and Turban (1993:8), weight expresses the relative importance of each input to a processing element. Weight fulfils a crucial role; it is through repeated adjustments of weights that the neural network learns.
- **Hidden layer:** The function of each hidden layer of neurons is to process the input variables. This is completed by weighting the connection of each input, summing the total of all the inputs, checking whether the total meets the threshold value and applying the activation function. It is the weights between the input and hidden units that determine when each neuron in the hidden layer is active, and by modifying these weights, a hidden layer may fire or not.
- **Output layer:** The behaviour of the output layer of neurons is similar to that of the hidden layer, where each input is processed as a hidden layer. The hidden layer and output layer neurons are connected to all of the neurons in the

preceding layer. The output is the solution to the problem and is expressed as a numerical value, for example, "1" for yes and "0" for no. The purpose of the neural network is to calculate the value of the output.

Another perspective is that of Abid and Zouari (2002:605), who describe a neural network model approach as a mathematical algorithm for creating a perfect mapping between the input and output values for a set of training data. The neural network's training process incrementally captures knowledge about the relationship between the input and output pattern in order to categorise correctly the training situation.

Generally, the independent variable consists of the input layer and the dependent variable consist of the output layer.

3.3 SIMPLISTIC MATHEMATICAL EXPRESSION OF A NEURAL NETWORK

Trippi and Turban (1993:8-10) provide a simplistic expression of the processing elements of a neural network, as follows:

Summation function

This function finds the weighted average of all the input elements to each processing element. The summation function multiplies the input values (X_i) with the weights (W_{ij}) and totals it for a weighted sum (Y_j). For N inputs i into one processing element j, the following equation:

$$Y_i = \sum_i^n X_i W_{ij} \qquad (3)$$

Transformation function

The summation function calculates the internal stimulation of the neuron. Based on this level, the neuron may or may not produce an output. The relationship between the internal and external activation level and the output may be linear or non-linear.

Such relationships are expressed by a transformation function. Although there are several different types, a popular non-linear function is called a sigmoid function:

$$\gamma_T = \frac{1}{1+e^{-y}} \tag{4}$$

where: γ_T = the transformed or normalised value of γ.

The purpose of this transformation is to modify the output levels to a reasonable value, for example, between 0 and 1. This transformation is done before the output reaches the next level. Without such transformation, the value of the output may be very large, especially when multiple layers are involved.

Learning or training

A neural network learns from its mistakes. The usual process of training consist of the following three tasks:

- Calculate outputs.
- Compare outputs with desired answers.
- Adjust the weights and repeat the process.

The training process usually starts by setting the weights randomly. The difference between the actual output (γ or γ_T) and the desired output (Z) is called Δ. The objective is to minimise Δ (or better, reduce it to zero). The reduction of Δ is done by incrementally changing the weights.

During the training stages, the interconnection weights change in response to training data presented to the system. In a training network, the training data set is divided into two categories, namely test cases and training cases.

3.4 TYPES OF NEURAL NETWORK MODELS

Muller *et al.* (2009:24-26) identify two basic types of neural network topologies, as follows:

- **Feed-forward neural networks:** The connections between neurons only occur in the direction from input to output. There are no feedback paths and as a result, the speed of the feed-forward neural network is usually very fast. Feed-forward neural networks tend to be straightforward networks that associate inputs and are extensively used in pattern recognition.
- **Feedback or back propagation neural networks:** Signals can travel in both directions by introducing loops in the network. Typically, the error signal at the output of the neural networks is passed back into the neural network to correct the connecting weights. As such, feedback neural networks are powerful as well as complex and consume large amounts of processing power.

Among the various types of neural network models, the back propagation neural network model is probably the most commonly used and is also the most mature algorithm, according to Maditinos and Chatzoglou (2004:10) and Xiangguang and Xiaozhong (2010:1). Although the feed-forward and back propagation neural networks are applied together quite often, with the back propagation network as a training algorithm for the feed-forward model, it is not a requirement. Both models can be applied individually to generate outputs.

Xiangguang and Xiaozhong (2010:1) favour the back propagation neural network in view of its high prediction accuracy and because it does not require assumptions of corporate data samples and the tests and analysis of the exact sample data parameter distribution. It also has strong fault tolerance, learning ability and error correction capability.

The back propagation neural network poses one important constraint because a large number of iterations are required in the learning process, resulting in an excessive training time, according to Yi and Prybutok (2001:17). More details relating to general constraints in the application of neural networks are provided in the following section.

3.5 CONSTRAINTS IN THE APPLICATION OF NEURAL NETWORK MODELS

One of the initial studies to highlighted constraints in building and using neural networks is that of Shin and Lee (2002). Firstly, it is an art to identify the most appropriate neural network model that reflects the problem characteristics due to the multitude of neural network model architectures, learning methods and parameters available. Secondly, and probably the most important constraint highlighted, is that the final rules of neural networks cannot be readily comprehended, often referred to as the black-box phenomenon.

Further to the Shin and Lee (2002) study, Baek and Cho (2003) focus on data imbalance. There are more healthy companies than financially distressed companies. Neural networks trained with an imbalanced data set of this nature tend to produce a "no signal" output, resulting in an increased false negative error rate. Although not associated with neural network modelling only, data imbalance can potentially affect statistical models as well.

Min and Lee (2004) identify several additional constraints in the application of neural networks. The main constraint relates to the difficulty in the interpretation of neural network models, which directs most studies to focus on prediction accuracy. Neural networks suffer from difficulties with generalisation because of overfitting, and fully rely on a researcher's experience or knowledge for pre-processing of selecting a large number of control parameters, which include input variables, hidden layer size, learning rate and momentum.

Kim and Yoo (2006:1) support the view expressed by Min and Lee (2004) and indicate that although a large number of financial distress prediction models have achieved high accuracy levels over a number of years, most of these models tend to show poor generalisation, where Kim and Yoo (2006:1) describe *generalisation* as the capacity of a model to respond to unknown or unseen inputs, which differ from training samples.

According to Kim and Yoo (2006:1), overfitting caused by poor generalisation is an inherent problem of the non-parametric model approach such as neural networks, which does not specify any assumption about the underlying probabilistic distribution because it relies heavily on available sample data.

To resolve this particular limitation, Kim and Yoo (2006) employed a semi-parametric approach where a parametric model and non-parametric neural network were combined. The proposed model was then compared with pure parametric models such as multiple discriminant analysis and logistic regression models and a pure non-parametric neural network model. The results indicated that the proposed semi-parametric model shows superior performance in terms of model stability and financial distress prediction accuracy. However, a significant limitation of this result is that a company is classified as either distressed or healthy rather than a condition that changes over time.

Neural networks have become popular in the search for an accurate financial distress prediction tool (Zhang, 2007). The popularity of neural network models is largely due to its powerful modelling capability for pattern recognition, object classification and future prediction without many unrealistic *a priori* assumptions about the specific model structure and data generating process.

Despite the growing popularity of neural network models, Zhang (2007:3) states that many problems, pitfalls and misuse emerge in neural network research and application. Some of the main shortcomings and pitfalls are as follows:

- **The black-box phenomenon:** This phenomenon is a lack of explanatory capability in terms of the "incapacity to identify the relevance of independent variables and to generate a set of rules to express the operation of the model".
- **Overfitting and underfitting:** Overfitting occurs when an overly large neural network is built and/or the in-sample data used to train a neural network are too small. Underfitting occurs when a neural network model is underspecified or not well-trained.

- **Data-related problems:** Neural network models are data-driven methods, and rely on quality input data. According to Zhang (2007:6), quality characteristics of secondary or primary data are rarely considered by neural network researchers, and on many occasions data are used as if free of any errors and are representative of the true underlying process. The reliability of a neural network model depends to a great extent on data quality.
- **Model building:** Building a predictive neural network model is a complex matter. Contributing to the complexity is the multitude of issues and parameters that needs to be considered, for example, the lack of process standards and the large number of controversial rules of thumb and guidelines in the literature, and problems relating to data preparation, input variable selection and network architecture parameters.
- **Software uses:** A large number of neural network software programs are available, ranging from standalone freeware software to high-end commercial packages. These packages, according to Zhang (2007:10), vary in features, options, training algorithms, programming capability, and user interface. Although some of these packages are powerful and easy to use, they are also exposed to risk of misuse and errors.
- **Model evaluation and comparison:** Zhang (2007:11) names three problem areas that requires consideration with regard to evaluation and comparison and which may affect validity of neural network research: firstly, comparison with established methods; secondly, use of true out-of-sample for testing; and lastly, use of a reasonable sample size. The study emphasises that statistical evaluation should be a minimum requirement for evaluations.
- **Publication bias:** There is a tendency to publish positive or mixed rather than negative results, according to Zhang (2007:12). Data snooping can be encouraged by repeatedly tuning the model architecture and other parameters if initial results on the holdout sample are not satisfactory. Another problem is that details of many aspects of the modelling process such as data, data processing, test design, model selection and parameter settings are not disclosed in published articles.

In a subsequent and contrasting study, Zhou and Elhag (2007b) emphasise the importance and superiority of neural network models in financial distress prediction compared with statistical models, such as the logistic regression model. Test results indicated a training accuracy of 97.19%, and testing accuracy of 97.53%, compared with the best classification accuracy of 96% achieved by statistical models.

However, Zhou and Elhag (2007b:52) point out additional limitations of neural network models. Firstly, there is the complexity of building a neural network model. The highly interconnected neural network is sensitive to its structure, for example, size of each layer, the number of nodes in the model and data splitting set. Secondly, there is weakness of overfitting and difficulty of choosing parameters and lastly, imitations related to the selection of input variables.

Chen and Hsiao. (2008:1146) added a number of additional weaknesses in neural network models: firstly, they require a great volume of controllable parameters with the risks of obtaining locally optimal solutions and overfitting; secondly, neural network models have been criticised for a long time and many techniques are required revising the great number of neuron linkages when establishing the classifying models.

Peat and Jones (2012:91) express the view that inconclusive evidence of the performance of neural network models may have resulted in a decline in academic interest in neural network models for financial distress prediction in recent years. Among other issues already raised, neural network models are generally more costly and complex to construct and suffer from interpretability issues associated with its black-box phenomenon.

The Peat and Jones (2012) study added to the debate by investigating the performance of the neural network in the context of forecast combination. A neural network framework was used to combine forecasts from a number of predictive models. It was noted that the introduction of market price information significantly improved the accuracy of neural networks, particularly when combined with certain financial variables. The neural network model was tested using a sophisticated equity-based measure of default, notably distance to default measures. Furthermore, the performance of the neural network model was tested with the most widely used discrete choice model, namely logistic regression.

Peat and Jones (2012:100) state that a combined information model, which includes sophisticated equity-based measures of default, significantly improves the predictive performance of neural network models relative to simpler approaches such as logistic regression.

A theoretical analysis by Shi (2009:1) covered the following aspects: firstly, the neural network model is a non-linear mapping from input to output and does not rely on any mathematical model. Secondly, neural network models can only deal with explicit data classification and are not suitable for the expression of rule-based knowledge. However, unascertained systems can handle abnormal, incomplete and uncertain data. Finally, the neural network model's greatest strength is memory, learning and inductive functions; unascertained systems do not have the learning function.

In theory, Shi (2009:60) established that combining unascertained systems with feed-forward neural network models can obtain more reasonable and more advantageous non-linear mapping, which can handle more complete and comprehensive types of data.

Most of the commentators are of the view that neural network models produce accurate and superior financial distress prediction results compared with the traditional statistical models. However, constraints experienced with neural network models necessitated researchers and practitioners to explore alternative avenues to develop accurate financial distress prediction models. A number of these models are reviewed in the following section.

3.6 TRANSITION TO ALTERNATIVE MODELS

The first obvious option to develop alternative financial distress prediction models was to improve on existing neural network models by combining it with another model, such as rough sets, logistic regression or recursive partitioning.

Ahn, Cho and Kim (2000:65) highlight the importance of an early-warning system that can aid company stakeholders to formulate and execute decisions pre-emptively. A derivative intelligent system was proposed combining rough-set theory and a neural network model. The authors reasoned that their proposed derivative model was expected to deal with the constraints experienced separately in the two models. The effectiveness of the proposed derivative model was verified by comparing the traditional multiple discriminant analysis and neural network models with the proposed derivative model.

Charalambous *et al.* (2000:403) hold the view that the increasing popularity of neural network application to financial distress prediction was based on its effectiveness in capturing and representing complex relationships in areas where traditional statistical models have limited capacity. The predictive capability of three contemporary neural network models was compared with the capability of the logistic regression model and a simple feed-forward neural network model using the back-propagation algorithm. Empirical findings indicated that the contemporary neural network model achieved better results than the feed-forward and logistic regression in all three periods prior to financial distress.

McKee and Greenstein (2000:20) criticised the application of statistical models (logistic regression and multiple discriminant analysis models), neural network models and inductive inference algorithms. The essence of the criticism is that many prior studies have not used true prediction models, but rather classification methods. In addition, most of these studies have used unrealistically proportioned data sets (50% distressed against 50% non-distressed entities). Further, the statistical and neural network models must typically be re-estimated for each time period examined. In an effort to resolve these constraints, the predictive accuracy of an easily implemented financial distress assessment model was examined using the recursive partitioning algorithm known as the interactive dichotomiser 3 (ID3). The results were then compared with both logistic regression and neural network models. The overall accuracy of the model ranged from 95% to 97%, which outperformed both the logistic regression model at 93% to 94% and the neural network model at 86% to 91%.

Charitou, Neophytou and Charalambous (2004:466) re-confirmed the need for reliable empirical models that provided timeous and accurate financial distress prediction, which would enable stakeholders to take either preventative or corrective action. The focus of their study was the incremental information content of operating cash flow in predicting and explaining financial distress, which built on the earlier work of Casey and Bartczak (1984). Logistic regression and neural network models were used to test several cash flow-based ratios. The Charitou *et al.* (2004:492) study showed that neural network models achieved a higher average classification rate of 78% compared with the 76% of the logistic regression model, three years prior to financial distress. The limitation of this study was, however, the lack of a sound theoretical framework for the selection of variables with the best potential to predict financial distress.

Cheng, Chen and Fu (2006) developed a financial distress prediction model that combined the approaches of neural network learning and logistic regression analysis in a radial basis functional network. This derivative approach retains the advantages and avoids the disadvantages of the neural network and logistic regression models.

The radial basis functional network approach proposed by Cheng *et al.* (2006:587) provides two advantages. Firstly, neural network models usually outperform statistical models when the explanatory variables are a mix of nominal and non-nominal variables. Secondly, the embedded logistic regression model in the radial basis functional networks can produce prediction that is expressed as the probability of financial distress and hence make the prediction more interpretable.

Cheng *et al.* (2006) demonstrated that the proposed derivative radial basis functional network outperformed both the logistic regression model and back-propagation neural networks individually in the predictive accuracy for unseen data.

Tsakonas, Dounias, Doumpos and Zopounidis (2006) attempted to introduce a new derivative intelligent approach for financial distress prediction. This has led to the development of efficient classification schemes formed as simple logic-based decision rules. The derivative intelligent model proposed is an evolutionary neural logic network by which evolutionary programming techniques are used for obtaining the best possible topology of a neural logic network.

The genetic programming process is guided using a context-free grammar and indirect encoding of the neural logic networks into the genetic programming individuals. According to Tsakonas *et al.* (2006:459), the resulting discrimination rules provide satisfactory results in financial distress prediction. Their proposed model outperforms both intelligent and statistical techniques in terms of prediction accuracy, and is characterised by its simplicity.

Although a number of statistical and neural network models have been developed over the years and have been applied to a specific data set, there was no result indicating that one model consistently outperformed any other, according to Hung, Chen and Wermter (2007:1). The study state that different models have different advantages on different data sets and different feature selection approaches.

Hung *et al.* (2007) hold the view that an ensemble of three popular models, namely decision trees, back-propagation neural networks and support vector machines, drawing advantages and avoiding disadvantages of different classifiers, outperform other stacking ensembles based on a weighting or voting strategy.

Sai, Zhong and Qu (2007:1) highlight the progressive increase in the volume of information available in considering and applying financial distress prediction models, and in addition, the need for an approach to effectively and efficiently utilise the information.

Over the years researchers have compared the performance or prediction accuracy of neural network models and other models such as multiple discriminant analysis and logistic regression. According to Sai *et al.* (2007:1), the results were generally better with neural networks than with traditional statistical models, but there were some disadvantages to neural networks such as long training times and the possibility of converging to a local situation.

The Sai *et al.* (2007) study proposed a derivative genetic algorithm-back propagation model. First, a back-propagation algorithm was used to design a near-optimal network architecture. The genetic algorithm was then used to train the interconnecting weights and thresholds of neural networks. The result indicated that the time to convergence could be shortened, and at the same time, the proposed model could simultaneously search in multiple directions, thus greatly increasing the probability of finding a global optimum.

Lee (2008:151) acknowledges the difficulty experienced with neural networks. Neural networks may produce useful results, but the black-box phenomenon limits the ability to interpret the decision-making process.

In view of the limitation experienced with neural networks, decision trees as an alternative approach was investigated. The advantages of decision trees are that they are simple to understand and interpret, require little data preparation, are able to handle nominal and categorical data, perform well with large data in a short time and the explanation for the condition is easily explained by Boolean logic. Lee (2008:151) further highlights their limitations, as follows: they ignore the relationship between attributes, output attribute must be categorical, algorithms are unstable, and lastly, trees created from numeric data sets can be complex.

In order to overcome the difficulties experienced with the decision trees approach, Lee (2008) considered a hybrid approach, namely genetic programming-decision tree, which evolved the decision tree through genetic programming. It was established that the genetic programming-decision tree model yielded the best classification accuracy though the approximate decision rules inferred were less intuitive and humanly understandable.

Much reliance is placed on information obtained from financial statements to predict financial distress. According to Chen and Vieira (2009:1), due to the complexity of financial statements, too few variables are insufficient for financial distress prediction, while too many, on the other hand, may lead to complications with dimensionality, i.e. the amount of training data needed increases exponentially with the number of variables in order to cover the decision space. Furthermore, data usually contain irrelevant, redundant and correlated variables, not only decreasing classification precision, but are also consuming calculation time and space unnecessarily.

The Chen and Vieira (2009) study proposed a derivative learning vector quantisation algorithm to solve the financial distress prediction problem. An independent component analysis was used as a pre-processing tool to eliminate the dimensionality problem.

The results demonstrated that the proposed algorithm was of higher stability and generalisation power than plain learning vector quantisation without independent component analysis. The proposed derivative learning vector quantisation model performed well compared with a number of other classification models.

Hossari (2009:333) questions the appropriateness of the classification system traditionally used that adopts a ratio-based multivariate approach signalling financial distress. The key criticism raised is that by using one statistical model at a time, the accuracy of a prediction model is examined by benchmarking a particular model against another. The author postulates that such an approach can be problematic in that it might compromise the accuracy of classifying distressed and healthy companies into their corresponding categories. The effect is that the two independent statistical approaches "…work against each other, rather than with each other".

In order to circumvent the problem, the Hossari (2009) study proposes a dual classification scheme for signalling financial distress, namely multiple discriminant analysis and multi-level modelling.

The results indicated that the dual-classification scheme generated 80% overall classification accuracy, which compared favourably with each of the schemes, where the one using only a multiple discriminant analysis model produced 72% overall accuracy and the multi-level modelling 78% accuracy.

Many subsequent financial distress prediction studies report that neural networks produce superior prediction accuracy compared with that of traditional statistical models. According to Jing-rong and Jun (2009:274) a key advantage of neural networks over traditional statistical models is that they do not require pre-specification of a functional form, or the adoption of restrictive assumptions concerning the distributions of model variables and errors. However, the learning problems of poor convergence or undesirable local minimum of neural networks present obstacles to apply them to financial distress prediction.

The Jing-rong and Jun (2009) study proposed an alternative solution by using an evolutionary-based wavelet network technique to discriminate between financially distressed and healthy companies. The advantage of the proposed approach was its global search abilities, its multi-resolution as well as localisation of the wavelets.

The results were encouraging, compared with multiple discriminant analysis, logistic regression models and neural networks. Furthermore, comparisons of the average absolute error also verified that the proposed approach was superior to that of the multiple discriminant analysis, logistic regression and neural network models.

Masten and Masten (2009) investigated the merits of semi-parametric and non-parametric methods used in financial distress prediction. The logistic regression model was used as benchmark for comparison with two alternative models: firstly, the semi-parametric estimator of binary choice model; and a second method based on the classification and regression tree. The first model was chosen because of its superior theoretical properties among the available semi-parametric estimators; and the second model was chosen because of its simplicity, clarity and interpretation, and most importantly, because it did not suffer from the black-box phenomenon. In addition, classification and regression trees can be used in the variables selection phase.

Masten and Masten (2009:4) found that choice-based sampling had a significant effect on prediction accuracy.

Shi, Xi, Ma and Hu (2009) proposed the bagging ensemble to improve the prediction performance of neural networks. Bagging is an ensemble learning technique used in building ensembles of unstable classifiers.

A bootstrap sampling technique was used to generate multiple training sets from the original data set. The bootstrap sampling was obtained by uniformly sampling with replacement patterns from the training set. The size of each sample was equal to the size of the original set, and so each pattern could appear repeatedly or not at all in any particular sample. Every training set was used to learn an individual classifier. Then, through the bagging process or algorithm, the predictions of all individual classifiers were combined *via* majority voting.

The prediction accuracy of bagging of the neural networks was 75.6%, which was approximately 1.6% higher than that of neural networks and 5.3% higher than that of support vector machines.

Behbood, lu and Zhang (2010) developed a financial distress prediction model, which integrated a fuzzy logic-based adaptive inference system with the learning ability of a neural network to generate knowledge in the form of a fuzzy rule base. The proposed financial distress prediction model contained three main phases: firstly, a pre-processing technique called synthetic minority oversampling technique to deal with imbalanced data set problems in distress prediction; secondly, a clustering technique and specifying the network structure and rule formulation algorithm to dynamically calculate the input fuzzy clusters and fuzzy rules from numerical training data; and thirdly, an adaptive inference system with a parametric t-norm operator in the learning algorithm to reduce prediction error.

Behbood *et al.* (2010:37) were confident that this model could significantly improve the financial distress prediction accuracy. Along with supplying a valuable and comprehensive financial knowledge base, this model was expected to have superior performance. The novelty of the proposed model not only organised the appropriate phases together as a framework for establishing a financial distress prediction model, but also presented an efficient neural network structure, rule generation and learning algorithms to gain better results.

The most important characteristic of financial distress prediction is its two forms of accuracy, namely classification and prediction, according to Gepp, Kumar and Bhattacharya (2010:538). A model's classification accuracy is obtained by assessing its accuracy on the data set from which it was developed. Following that, the more important prediction accuracy of the model is assessed by its application to brand a new set of data, which reflects how well the model will perform on future predictions. The authors provide empirical evidence to support the claim that less complex, more parsimonious models are better predictors than more complex models. Further evidence was provided showing that decision tree models are superior classifiers and predictors of financial distress. The classification and regression trees and recursive partitioning analysis and decision tree models produced very similar results and were the best overall predictors. In particular, recursive partitioning analysis was preferred because it had slightly better predictive ability.

The process that underlies such procedure suffers from two major deficiencies, according to Hossari (2010:99). Firstly, it cannot determine which ratios in which industries require alteration; and secondly, when modification of the ratio is required, the traditional model does not allow for variations in the adjustment schemes between the observed sectors.

In order to deal with these shortcomings, Hossari (2010) proposed multi-level modelling as a more parsimonious approach for modelling financial distress. To ascertain the robustness of multi-level modelling, the traditional two-step procedure was utilised as a benchmark for comparison.

Accordingly, the initial step prior to model derivation involved subtracting the industry average from the raw values of each of the useful ratios for each company in the data sample. For consistency, the same industry classification adopted in generating the multi-level modelling-based model, was also applied for the purpose of adjusting the financial ratios using the traditional two-step procedure. The next step involved generating the financial distress prediction model based on the adjusted financial ratios using logistic regression.

The results of the Hossari (2010:105) study indicated that the multi-level modelling-based model correctly predicted 64% of financially healthy companies and 83% of the financially distressed ones, thereby generating an overall prediction accuracy of 74%. Similarly, the model derived using the traditional two-step procedure correctly predicted 65% of financially healthy companies and 73% of distressed companies, which when combined, translated to an overall prediction accuracy of 69%. Therefore, not only did the alternative multi-level modelling approach enhanced procedural efficiency, it did so without compromising the accuracy of signalling financial distress vis-à-vis the traditional two-step procedure.

Jian-guang, Xiao-feng and Jie (2010:148) express the view that statistical models are in general outperformed by artificial intelligence models. There is one common denominator in both these groups: they only compare the validity of different classifiers when applied to financial distress prediction, however, ignore the benefit of combining multiple classifiers. The multiple classifier system could take advantage of the strengths of single classifiers, avoid their weaknesses, and improve prediction accuracy.

Jian-guang *et al.* (2010:148) quote a number of studies where two or more classifiers are combined or hybridised. The general consensus drawn from these studies is that the prediction accuracy of the ensemble systems outperforms all other financial distress prediction models.

Against this background, the main criticism in the Jian-guang *et al.* (2010) study is that financial distress prediction models are still at the stage of static modelling. The model is built with sample data, which is collected over a special extended period of time. Over time, variables in the economic and company environment change and if the financial distress prediction model is not aligned or adjusted to these changes, it is called *financial distress concept drift*. The necessity for the change of the current model due to the change of data distribution is called *virtual concept drift*.

The traditional statistic models cannot satisfy the predicting need of the company in a dynamic economic environment. The purpose of the Jian-guang *et al.* (2010:148) study was to propose a framework for building a dynamic model to adapt the financial distress concept drift. Through this method, the proper window size can be selected to pursue the time window, which selects different classifiers towards different testing samples.

From the test results it can be established that the single model based on the selected time window is better than the static models, and the combined model improves the predictive performance further.

In spite of neural network models being a popular technique for financial distress prediction, its main disadvantage is its inability to explain its internal decision-making steps, namely the black-box phenomenon raised earlier. Fuzzy set theory, on the other hand, has also received attention since its introduction in 1965 (Zanganeh *et al.,* 2011:15). The purpose of fuzzy logic is to map one space (input) to another (output) with relative precision (applying if-then rules). Fuzzy systems do not have much learning capability and it is difficult for humans to tune the fuzzy rules and membership functions from the training data set.

According to Zanganeh *et al.* (2011:15), there are three different approaches to combining neural network and fuzzy systems; firstly, the concurrent neural-fuzzy model. In this model, the fuzzy system is used either before or after neural networks. The neural network does not change any parameters in the fuzzy system. Secondly, there is the co-operative neuro-fuzzy model. In this model, a neural network is used to learn certain parameters of fuzzy sets, fuzzy rules or weights of fuzzy systems. The result is a pure fuzzy system. Lastly, there is the derivative neuro-fuzzy inference system. In this model, the neural network and fuzzy system are combined.

Zanganeh *et al.* (2011) propose the combination of fuzzy modelling with a neural network model (adaptive-network-based fuzzy inference system) because it is recognised as a powerful tool that can facilitate the development of financial distress prediction models. The rule-based nature of fuzzy models allows better use of information expressed in the form of natural language statement and consequently, makes the model interpretation easier.

The contribution by Zanganeh *et al.* (2011:15) is two-fold: firstly, the adaptive-network-based fuzzy inference system structure is applied to a data set and the performance is compared with the logistic regression model. Secondly, the adaptive-network-based fuzzy inference system and logistic regression are combined with the following two alternative feature selection procedure methods and the results based on each of them are compared: selection of the most frequent variables in the former financial distress prediction literature and selection of variables by T-statistic method.

The analysis of empirical results indicated that, firstly, the adaptive-network-based fuzzy inference system model outperformed the logistic regression model; and secondly, the subset of frequent variables in the former literature yielded better prediction models rather than variables, which were selected based on the T-statistic feature selection method.

Olson *et al.* (2012:464) acknowledge the popularity of neural networks compared with statistical models and reiterate a critical constraint limiting its applicability. Neural networks are black boxes lacking transparency (seeing what the model is doing, or comprehensibility) and transportability (being able to easily deploy the model into a decision support system for new cases). Olson *et al.* (2012:464) argue in favour of decision trees, which can be as accurate, and provide transparency and transportability that neural networks are often criticised for.

3.7 CONCLUSION

The advent of computer technology and artificial intelligence provided the impetus for the transition from traditional statistical financial distress prediction models to more accurate neural networks, with the ability to do complex calculations. These complex relationships were captured and represented in areas where the performance of statistical models was unacceptable.

Neural networks generally outperformed statistical models, as expressed by Zhou and Elhag (2007b). Test results indicated a training accuracy of 97.19%, and testing accuracy of 97.53%, compared with the best classification accuracy of 96% achieved by statistical models.

Similar sentiments were expressed in a number of studies; however, constraints experienced with neural network models (Zhang, 2007:3) of which the black-box phenomenon was certainly the single most critical constraint, inhibited further development. These constraints necessitated researchers and practitioners to explore alternative avenues to develop accurate financial distress prediction models.

In an effort to overcome the critical limitations and constraints experienced in the application of neural network models, a number of derivative financial distress prediction models were developed. In most of these studies models were developed where two or more classifiers were combined or hybridised. The general consensus drawn from these studies was that the prediction accuracy of the ensemble systems outperformed all other financial distress prediction models (Jian-guang *et al.,* 2010:148).

Most of these models are still at the stage of static modelling and are built with sample data, which is collected over a special extended period of time. Over time, variables in the economic and company environment change and if the financial distress prediction model is not aligned or adjusted to these changes, the financial distress prediction model may lead to financial distress concept drift. This important criticism, raised by Jian-guang *et al.* (2010) against the financial distress prediction models, forms the foundation of the current study.

The next chapter reviews a number of important developments, still in the artificial intelligence category, but independent of the neural networks category.

CHAPTER 4

ALTERNATIVE FINANCIAL DISTRESS PREDICTION MODELS

4.1 INTRODUCTION

Artificial intelligence financial distress prediction models can further be grouped into three broad sub-categories, namely those inspired by the workings of biological neurons (neural networks), those inspired by an evolutionary metaphor (genetic algorithm, genetic programming and grammatical evolution), and finally, those inspired by studies of social interaction (particle swarm and ant colony models). Of these, neural network models have received the most attention, and to a lesser extent, genetic algorithms, ant colony and other models.

The first sub-category based on biological neurons was discussed in Chapter 3. The second and third sub-categories based on evolutionary metaphors and social interaction-based models respectively are reviewed in the following section.

4.2 EVOLUTIONARY METAPHOR-BASED MODELS

4.2.1 Genetic algorithms and genetic programming

In a dynamic environment, more precise financial distress prediction becomes critical. Identifying critical or useful ratios from financial statements can affect the accuracy of financial distress prediction. According to Ko and Lin (2006:84), factor analysis and stepwise models are traditional statistical models that have been used to extract the critical financial ratios. However, depending on the extracting sequence, the extracted financial ratio set is different because of the linear searching characteristics. In a dynamic macroeconomic environment, critical financial ratios can change.

Ko and Lin (2006:85) introduced a general evolutionary architecture with modularised evaluation functions, which extract minimal critical financial ratios based on any evolutionary algorithm such as genetic algorithm (GA) and particle swarm optimisation (PSO). This allows for the integration of various evaluation modules by assigning distinct weights. Based on the test results it appears that by using the proposed approach better forecasting accuracy is achieved with a minimum critical consideration of financial ratios than using conventional statistical models, such as stepwise regression.

The model developed by Ko and Lin (2006:85) acknowledges the fact that a more accurate financial distress model is required in an increasing dynamic environment. Although this approach allows for better accuracy, the mechanism is primarily based on financial ratios and fails to incorporate quantitative non-financial variables such as macroeconomic variables.

Case-based reasoning (CBR) is one of the most popular data-driven approaches in building effective financial distress prediction models, according to Ahn and Kim (2009). The case-based reasoning approach is easy to apply, has no possibility of overfitting and provides good explanation for the output. However, the main constraint is its limited prediction accuracy, which is usually much lower than the accuracy of a neural network model (Ahn & Kim, 2009:599).

In order to enhance the prediction performance of the case-based reasoning approach, Ahn and Kim (2009:600) proposed the simultaneous optimisation of feature weighting and the instance selection for case-based reasoning by using genetic algorithms. The model proposed to improve the prediction performance by referencing more relevant cases and eliminating noises.

According to Martin, Gayathri, Saranya, Gayathri and Venkatesan (2011:13), most financial distress prediction models developed in the past indicated key cash flow and debt ratios to be important predictors. These models have quite heterogeneous characteristics but most of them share the common feature of relying on multivariate techniques with financial descriptors as the main input variables. Alternative models were based on market information and group decision analysis.

Based on the view that alternative models incorporating market information and group decision analysis do not significantly outperform multivariate models, researchers have focused their attention on non-parametrical and derivative models. Non-parametrical models are suitable for financial distress prediction tasks due to specific features of financial information (i.e. non-normality and heteroscedasticity). Derivative models combine classification methods and achieve greater accuracy than that of individual models.

Against this background, Martin *et al.* (2011:13) propose a strategy for the construction of a derivative model which is based on the sequential application of genetic algorithm, fuzzy c-means clustering and multivariate adaptive regression spines. The reason for the choice of these models is, firstly, that the use of clustering techniques is motivated by the existence of different failing procedures, because financially distressed companies may have dissimilar financial features. Secondly, a multivariate adaptive regression spines model is a flexible procedure, which models relationships that are nearly additive or involve interactions with fewer variables.

Kim and Kang (2012:9308) highlight the evolvement of the two primary research fields in financial distress prediction, namely statistical and artificial intelligence models. A major drawback of statistical models is that it should be based on strict assumptions. These assumptions include linearity, normality, independence among predictor variables and pre-existing functional forms relating to the criterion variables and predictor variables.

Artificial intelligence models such as decision trees, neural networks and support vector machines were subsequently developed to overcome the restrictive assumptions of statistical techniques. A more recent model applied to financial distress prediction is ensemble learning. Ensemble learning is an artificial intelligence model for improving the performance of individual classifiers and predictors, according to Kim and Kang (2012:9308). Ensemble learning constructs a highly accurate classifier (a single strong classifier) on the training set by combining an ensemble of weak classifiers, each of which needs only to be moderately accurate on the training set.

Kim and Kang (2012:9309) proposed a genetic algorithm-based coverage optimisation system for ensemble learning. The optimal (or near-optimal) classifier subset was selected based on prediction accuracy and diversity measurement represented as statistical value of the variance influence factor. The proposed coverage optimisation was applied to a financial distress prediction task to validate the effect on the performance improvement. Test results indicated that the proposed genetic algorithm-based coverage optimisation could help to design a diverse and highly accurate classification system.

Advances in the field of financial distress prediction have led to the development of genetic programming, which is a developing subarea of evolutionary algorithms.

A number of researchers expanded on genetic algorithms to genetic programming, which are often considered as a separate but related field in evolutionary metaphor-based models. According to Woodward (n.d.:1), genetic algorithms use fixed length linear separation, whereas genetic programming uses a variable size tree representation. Woodward argues that these differences are not critical and that the actual important difference lies in the interpretation of the representation; whether it is a one to one mapping between the description and the object being represented, or a many to one mapping between the description and the object being represented.

The accuracy of traditional financial distress prediction models with respect to theory building is negatively affected by either modelling assumptions or model complexity, according to Lensberg *et al.* (2006).

Lensberg *et al.* (2006:677) applied genetic programming (GP) that minimised the amount of an *a priori* structure that was associated with traditional functional forms and statistical selection procedures, but still produced easily understandable and implementable models. Using the same variables, it was concluded that the genetic programming model was more accurate than the logistic regression model.

An analysis of the interaction of variables in the genetic programming model reveals the following results: firstly, liquidity improves the financial distress status of the company, irrespective of the value of other variables such as profitability and size, and that unprofitable companies can maintain high levels of liquidity to offset low profitability. Secondly, financial distress risk decreases with increased size, except in the case of a financial loss. Thirdly, an unfavourable audit report has more negative financial distress status impact for large companies than for smaller companies. According to Lensberg *et al.* (2006:695), this might be interpreted that the model indicates that accounting information (including the audit opinion) is more important for larger companies than smaller ones. Lastly, it is also suggested that liquidity and non-financial information can be the most important information.

Divsalar, Roodsaz, Vahdatinia, Norouzzadeh and Behrooz (2012:505) describe *genetic programming* as a supervised artificial intelligence model that searches a program space instead of a data space. Recently, a new variant of genetic programming was employed, called linear genetic programming to classify financially distressed and healthy companies. Gene expression programming (GEP) is a more recent extension to genetic programming. Gene expression programming evolves computer programs of different sizes and shapes encoded in linear chromosomes of fixed length. Multi-expression programming is another new variant of genetic programming with a linear representation of chromosomes. Based on numerical test, gene expression programming and multi-expression programming approaches can be utilised as efficient alternatives to the traditional genetic programming.

The main purpose of the Divsalar *et al.* (2012) study was to derive a new model for classifying financially distressed and healthy companies using the gene expression programming and multi-expression programming models. Four effective predictive financial ratios, identified through extensive literature reviews, were used as input variables. The generalised regression neural network, logistic regression and least-squares regression based models were also developed to benchmark the gene expression programming and multi-expression programming (MEP) models.

Findings by Divsalar *et al.* (2012:520) were as follows:

- The gene expression programming and multi-expression programming models gave reliable estimates of the financial distress classification. The gene expression programming and multi-expression programming models provided superior performance to the generalised regression neural network, logistic regression and least-squares regression models.
- Unlike classical statistical models, gene expression programming and multi-expression programming were capable of modelling the financial distress without any need to pre-define equations.
- According to the frequency values, financial distress prediction was more sensitive to the quick assets to total assets ratio and the total liability to total assets ratio compared with other variables.
- The gene expression programming and multi-expression programming models give the user an insight into the relationship between input and output data. An interesting feature of these approaches is the possibility of getting more than one prediction model by selecting various parameters and function sets involved in their algorithms.
- Another feature of the gene expression programming and multi-expression programming models is the high level of interactivity between the user and the methodology. User insight can be used to make propositions on the elements and structures of the evolved functions.

4.2.2 Grammatical evolution

Brabazon and O'Neill (2004:363) focussed their attention on this lesser-known model. The grammatical evolution model has been found to have general function for rule-induction applications and is able to automatically evolve quality classifiers for financial distress prediction from primary financial data. In performing this task the grammatical evolution model was required to evolve its own financial ratio representation of the financial data rather than modeller-defined ratios as is generally the case with financial distress prediction.

The Babazon and O'Neill (2004) grammatical evolution model is novel regarding the application of model-generated ratios rather than modeller-defined ratios. The main criticism was that non-financial ratios had not been considered. This aspect was raised as a caveat in their study, and they indicated that the inclusion of non-financial variables could further improve classification accuracy.

4.2.3 Self-organising maps

Self-organising maps (SOM) is a specific type of artificial neural network that learn through its own unsupervised competitive learning, and attempts to map its weights to conform to the given input data. According to Guthikonda (2005), self-organising maps can be applied to resolve difficult interpretable and complex data sets. The advantage is that self-organising maps can be simple enough to code complex data sets in a relatively few number of lines, utilising a limited number of equations.

Simić, Kovačević and Simić (2011:2) state that economic and financial theories on financial distress prediction do not provide a rigorous basis for selecting particular ratios. As a result, empirical studies rely on standard sets of financial ratios that are important in explaining a company's financial health.

Simić *et al.* (2011:2) proposed a model by applying a combination of the multiple discriminant analysis model and a specific neural network topology – self-organising maps to the assessment of company financial health.

The derivative multiple discriminant analysis self-organising maps model yielded good results and showed correct estimates of financial health in 95% of all companies in the sample.

Although this hybrid model is not overly superior to other techniques, it involves significant advantages, according to Simić *et al.* (2011:10-11). The first advantage relates to the suitability of the research for analysis of financial ratios. Company financial ratios have significant leverage over the use of a company's independent financial values. Secondly, multiple discriminant analysis models can be used for the calculation of Altman's (1968) Z-score. This was regarded as very important because financial ratios and Z-score results represent input values for the self-organising map. Furthermore, the implementation of the self-organising map in classification problems is a contemporary technique that represents the outcome of long-term neural network testing and implementation. It is possible to optimise parameters by considering a small number of parameters based on another algorithm. In addition to reducing the number of parameters for optimisation, these methods have the potential to control several other properties such as weight distribution and connection topology. All of the above-mentioned characteristics represent the main technical and technological advantages of the model.

Most prior financial distress prediction studies were based on a static snapshot of a company's financial situation. Chen, Ribeiro, Vieira and Chen (2012:385) based their study on the premise that financial distress trajectory, which characterised the dynamic changes of a financial situation received limited attention.

Chen *et al.* (2012) studied the changes or trajectory patterns of a company's financial situation through a two-step clustering process. A self-organising map clustering approach was proposed to analyse and visualise the effect of temporal evolution of some financial variables in order to assess and establish eventual scenarios of financial distress.

Initially, Chen *et al.* (2012:385) constructed a fuzzy self-organising map to characterise a company's financial distress risk. Afterwards, the instantaneous observations of temporal sequence were successfully projected on the map and the positions were concatenated to a trajectory vector. The trajectory patterns were then learnt by a trajectory self-organising map and shown through an appropriate visual presentation.

The test results demonstrate the promising functionality of a self-organising map for financial distress trajectory clustering and visualisation. Taking the perspective of decision support, the described method might give experts insight into patterns of financial distress or healthy company development.

Financial ratios have historically been the focus of financial distress prediction research (Gunnersen, Smith-Miles & Lee (2012:1). More recently, there has been a movement from the use of financial ratios towards the consideration of market information, though it is acknowledged that financial information such as financial ratios continue to capture critical information in assessing financial distress. The authors hold the view that previous research that utilised financial ratios has often done so without considering the effect that data selection may have. The example was presented where a neural network model was used to classify financial distress, and used the same ratios as the Z-score. In most cases, this approach was used to demonstrate the superiority of a new technique, but had the effect of failing to identify potentially better financial ratios than which, in turn, could be used in future research or used to better understand the causes of financial distress.

A number of research efforts have utilised limited datasets based on predetermined selection criteria. While this selection criteria is useful in ensuring that the results of the predictive model are not constrained by exceptional cases or poor data quality, such selection is not truly representative of the predictive accuracy of the model had it been applied to an entire data set or population, according to Gunnersen *et al.* (2012:1). On the one hand, the use of larger data sets comes with the potential limitation of not building multiple predictive models for each smaller sub-group within the data set. On the other hand, the smaller groups may be limited by not being exposed to as many cases in the training algorithm. Multi-level self-organising maps was used to cluster data sets with the goal of objectively reducing the data presented to different classification models to increase classificatory accuracy.

Financial distress prediction is typically conducted with a short-term perspective, such as financial distress in the next six months, and it is generally believed that prediction rates will be poorer for longer periods. Any focus on feature selection or test case selection should consider a longer timeframe to ensure that features that are useful over the medium term are not necessarily eliminated.

The study of Gunnersen *et al.* (2012:1) was aimed at two aspects of data selection: firstly, the feature selection methodology; and secondly, the case selection methodology that is approached as a clustering problem.

According to Gunnersen *et al.* (2012), it is clear that conducting a non-deterministic, random generation, accuracy-based initial feature selection immediately improves accuracy on the out-of-sample data set, and performing a further heuristic, forward-generation, accuracy-based feature selection almost always improves results of the out-of-sample set, and in the few cases where it did not, the differences were minimal.

Furthermore, while restricting a financial distress prediction data set to only companies in a particular industry can improve accuracy, better accuracy gains can be achieved by conducting objective clustering such as Gunnersen *et al.* (2012) demonstrated with the Deboeck-Kohonen multi-level self-organising maps in combination with SpexVCMV.

It was established that genetic programming generally yielded better results on the out-of-sample sets, possibly due to overfitting occurring in the cascade-correlation neural network utilising the Lachenbruch jack-knife method (Bellovary, Giacomino & Akers, 2007:7).

4.3 SOCIAL INTERACTION MODELS

4.3.1 Ant colony algorithm

Although a number of studies and tests confirm the usefulness of neural network models, there are limitations in building and using the models. Wang, Wu, Zhang and Zhou (2009:137) reiterate the following limitations: firstly, it is an art finding an appropriate neural network model, which can reflect problem characteristics because there are numerous network architectures, learning methods and parameters.

Secondly, the user cannot readily comprehend the final rules that the neural network model acquires. This characteristic was previously referred to as the black-box phenomenon.

Wang *et al.* (2009) propose an unconventional technique in order to solve the above limitations. The ant colony algorithm mimics the techniques employed by real ants to rapidly establish the shortest route from a food source to their nest.

First, the t-test method was used to select five features from the 55 original features. Secondly, the rule encoding was constructed. Thirdly, the ant colony algorithm was utilised to find the optimal rule. The results show that rule extraction via the ant colony algorithm for financial distress prediction is acceptable and it is proposed that future research focuses on improving the prediction accuracy and fasten the algorithm.

4.3.2 Particle swarm optimisation model

Particle swarm optimisation is a biologically inspired computational search and optimisation model developed in 1995 by Eberhart and Kennedy based on the social behaviour of birds flocking or fish schooling (Rini, Shamsuddin and Yuhaniz, 2011).

Rini *et al.* (2011:25) explain that the process of particle swarm optimisation in finding values, follows the work of an animal society which has no leader. Particle swarm optimisation consists of a swarm of particles, where a particle represents a potential solution or better condition. The particle will move through a multi-dimensional space to find the best position in that space.

Particle swarm optimisation has many similarities with genetic algorithms discussed in Section 4.2.1. The model was initialised with a population of random solutions and searches for optima by updating generations. However, compared with genetic algorithms, the difference is that particle swarm optimisation has no evolution operators such as crossover and mutation. In particle swarm optimisation, the potential solutions, called particles, float through the problem space by following the current particles (Hu, 2008).

Rui (2010:557) also acknowledges the popularity of neural network as a financial distress prediction model due to its excellent performance in treating non-linear data with self-learning capability. However, key limitations of the artificial neural network are that it suffers from slow convergence and the black-box phenomenon.

On the other hand, fuzzy logic as a rule-based development in artificial intelligence cannot only tolerate imprecise information, it also makes a framework or approximate reasoning, avoiding the black-box phenomenon. The main disadvantage of fuzzy logic is its lack of effective learning capability.

In order to overcome these limitations, Rui (2010) developed an improved particle swarm optimisation and combined the model with fuzzy logic and neural networks, which he named the particle swarm optimisation-fuzzy neural network. The results indicated higher predictive accuracies than those obtained from neural networks.

4.4 OTHER ARTIFICIAL INTELLIGENCE MODELS

Haber (2005:87) expressed the view that financial distress prediction has largely remained unchanged since the mid-1960s. A matched-pair sample design compared with a dichotomous classification test has been the standard. This has been useful from an academic perspective but failed in a practical application.

In a suggestion for further research on survival analysis models, Haber (2005) proposed that additional models within survival analysis approach could be added. For example, an accelerated failure time model or survival analysis model estimated by a neural network could be added, which have had little application to financial distress prediction thus far. Furthermore, time-dependent explanatory variables could be added to the survival analysis model to observe whether it improved the empirical performance of the model.

As indicated previously, neural network models have become a popular alternative financial distress prediction model to traditional statistical models such as discriminant analysis and logistic regression. Neural networks' popularity stems from its associated memory characteristics and generalisation capability. This is in contrast to its disadvantages such as its long training process in designing the optimal network's topology, the difficulty in identifying the relative importance of potential input variables and certain interpretive aspects.

In order to overcome these difficulties, Lee and Chen (2005) explored the performance of a two-stage derivative modelling procedure using neural networks and multivariate adaptive regression splines.

Sprengers (2005) tested a relatively new distress prediction model, classification and regression tree and benchmarked it against the Z-score multiple discriminant analysis model introduced by Altman (1968).

The classification and regression tree model, at 79.17% prediction accuracy, reported poorer results than the Z-score model at 85.42% prediction accuracy. According to Sprengers (2005:54), this is evidence of the Z-score model's popularity, and is understandable because it is relatively easy to construct and comprehend.

Baixauli and Módica-Milo (2010:60) proposed a financial health indicator to define the degree of a company's financial health or distress. The binomial logistic regression model was used to examine the likelihood of a company becoming financially distressed.

Results were obtained under four different models. Baixauli and Módica-Milo (2010:76) showed in-sample and out-of-sample financial distress prediction. Focusing on healthy companies, the financial health indicator allowed companies to be classified before the estimation process. This procedure permitted the heterogeneity of the companies to be reduced as well as identifying a strong company sample to estimate the financial distress probability accurately. The in-sample and out-of-sample evaluation based on the cumulative accuracy profile and receiver operating activity variables led to the conclusion that the models estimated under the strong company sample were much more accurate.

In order to develop a more accurate and generally applicable prediction approach, data mining and machine-learning techniques including genetic algorithm, support vector machine and radial basis function neural network have been proposed in the past (Fanping & Shiwei, 2010:1).

Fanping and Shiwei (2010) proposed a rough set-based principle component analysis-radial basis functional neural network model for predicting financial distress. To improve radial basis function neural network performance, the rough sets and principle component analysis method was employed to reduce the dimension of the input vector space. In the proposed model, the rough set model was applied to reduce the input data sample at first, and then the principle component analysis was used to select the feature variables as the inputs.

Against this background, Fanping and Shiwei (2010:6) claimed that the rough set-based principle component analysis-radial basis functional neural network model performed significantly better than the conventional radial basis functional neural network when applied to financial distress prediction.

Ranking-order case-based reasoning is a newly developed method for financial distress prediction, which has the capability of predicting financial distress accurately in an easily understandable manner. Based on this, Li and Sun (2011) proposed the combination of forward feature selection with ranking-order case-based reasoning to generate a new predictive financial distress prediction model named forward ranking-order case-based reasoning.

Neves and Vieira (2006:254-255) applied neural networks to financial distress prediction by introducing a new method called a hidden layer learning vector quantisation (HLVQ-C) to improve the prediction of multilayer neural networks. The main advantages of this algorithm were, firstly, it could use a larger set of variables without compromising generalisation. Secondly, that it was capable of improving the network predictions for difficult cases and outliers; and lastly, it gave an easy estimate of the prediction accuracy.

Neves and Vieira (2006:268) introduced a measure of classification efficiency to evaluate the method's performance, and concluded that hidden layer learning vector quantisation clearly outperformed the linear discriminant analysis, Z-score model and traditional neural networks. In addition, it was concluded that it was desirable to use balanced data sets, containing the same number of financially distressed and non-distressed companies, thereby constraining Type I errors.

4.5 CONCLUSION

This chapter provided an overview of a number of lesser-known, but novel financial distress prediction models, broadly grouped into three categories: models inspired by the workings of biological neurons, those inspired by an evolutionary metaphor, and finally, those inspired by studies of social interaction.

The main criticism against these studies was that, although it was directed at improving financial distress prediction accuracy in a dynamic operating environment, none of the studies reviewed, considered non-financial variables. In addition, it appears that none of these evolutionary financial distress prediction models has attracted noticeable academic interest.

The following chapter reviews a number of studies incorporating qualitative and quantitative non-financial variables in a financial distress prediction model.

CHAPTER 5

FINANCIAL DISTRESS PREDICTION MODELS BASED ON FINANCIAL AND QUANTITATIVE NON-FINANCIAL VARIABLES

5.1 INTRODUCTION

Companies operate in an ever-increasing dynamic and complex environment, requiring equally dynamic management decision-making. Relying purely on historical financial information would result in reactive decision-making because financial results are a point-in-time view of the company's operational achievements.

In order to make pro-active decisions, company management is required to augment the financial variables such as company financial results and market variables with quantitative non-financial variables. Financial variables include financial ratios and market variables. Non-financial variables include macroeconomic and qualitative variables such as customer relations, employee skills, innovations and knowledge-intensive services. Qualitative variables are more complex to interpret and to include in a financial distress prediction model. It is more likely to be a subjective interpretation of an aspect (Van der Colff, 2012).

The following section focusses on a review of the various financial and non-financial variables and is divided into two sub-sections. The first sub-section reviews studies based on financial variables, namely financial ratios and market variables, and the second sub-section reviews studies based on non-financial variables, namely macroeconomic and qualitative variables.

5.2 FINANCIAL VARIABLES

5.2.1 Financial ratios

Financial ratios as a sub-section of financial variables provide an objective evaluation or measure of a company's financial health, based on publically available financial information (Balcaen and Ooghe 2006:82).

The pioneering work of Beaver (1966) and Altman (1968) established the foundation of all subsequent financial distress prediction models based on financial ratios. Most of these subsequent studies involved determining which financial ratios best predict financial distress, primarily employing statistical models such as multiple discriminant analysis and logistic regression models. A more detailed review of the historical evolvement of statistical financial distress prediction models is provided in Chapter 2.

Alfaro, García, Gámez and Elizondo (2008:114) and Zhou, Bai, Zhang and Tian (2008:3) highlight a number of commonly used criteria for selecting financial ratios applied to financial distress prediction models:

- The ratios should be commonly used in financial distress prediction literature.
- The information required to calculate these ratios should be readily available.
- The researcher's own decision should be based on their experience in previous studies or based on preliminary trails.

Based on the above criteria, financial ratios selected for financial distress prediction models broadly include liquidity, profitability, efficiency, growth and cash flow ratios. Financial ratios are simplistic and allow for a basic and general understanding of a company's financial strength and profile.

Contrary to these advantages, financial ratios are subject to a number of constraints. The outcome of a financial ratio calculation, whether in a standalone or index format (for example, Z-score or K-score) is a function of a company's financial statements.

Agarwal and Taffler (2008:1542) expressed the view that the very nature of financial statements on which financial distress models are based casts doubt on its validity.

They highlight the following:

- Financial statements present past performance of a company and may or may not be informative in predicting the future.
- Conservatism and historical cost accounting imply that the true asset values may be very different from the recorded book values.
- Accounting numbers are subject to manipulation by management.
- Because financial statements are prepared on a going-concern basis, they are by design of limited use in financial distress prediction.

These constraints are possibly valid for smaller and unlisted companies. However, with listed companies, strict stock exchange requirements and scrutiny by an informed investor's community, for example, could largely eliminate some of these constraints.

5.2.2 **Market variables**

In addition to financial ratios, market variables also fall within the category of quantifiable financial variables and can be considered for inclusion in a financial distress prediction model. Market variables include aspects such as market capitalisation, earnings per share (EPS), earnings yield (EY), and price-earnings (PE) ratio. The following section reviews studies that combined market and financial variables

In an attempt to improve prediction ability, a more dynamic financial distress prediction model was developed by Shumway (2001:51), combining financial and market-driven variables. Five selected variables, which were found to be statistically significant in the empirical test from previous work, evaluated the main aspects of financial position and market reaction within a company's solvency situation. This model could predict more accurately than alternative models using financial ratios only.

He (2002) utilised the financial ratios and market measures in Shumway's (2001) model to develop a financial distress prediction model and concluded that the financial and market predictors exhibited dramatic differences of performance between failed and non-failed companies several years prior to financial distress. When combined into a financial distress prediction model, the predictive variable contributed to improving the discriminatory accuracy in the classification and prediction tests.

Berg (2007) evaluated several accounting-based models for financial distress prediction. In the study, Berg (2007:129-130) acknowledged that accounting and market-based or structural models should be differentiated. A comparison between these two types of models indicated the superiority of market-based models.

Market-based models are based on the market value of the company set by the share price as a proxy on the stock exchange. Berg's study was constrained by the limited number of listed companies on the Norway stock exchange, and had to focus on financial variables only. The generalised additive model was introduced as a flexible non-parametric alternative for financial distress prediction. The generalised additive model was a generalisation of the linear regression model and replaced the usual linear function of a covariate with a sum of unspecified smooth functions. This assisted in discovering potential non-linear shapes of covariate effects and it was concluded that the generalised additive model performed significantly better than discriminant analysis, linear and neural network models.

A key aspect raised by Berg (2007:130), which could have an implication for the current study, was that in order to consider non-financial or market variables, a company had to be listed on a stock exchange.

The primary objective of the Kim and Partington (2008) study was to use dynamic rather than static variables in estimating a financial distress prediction model. A time-dependent Cox regression model was applied, which allowed for dynamic changes of a company's risk levels and its corresponding survival probabilities through time.

The results indicate that companies with higher book leverage, less cash flow generating ability and less market value relative to debt are more likely to fail. Furthermore, the results indicate an improvement in the accuracy of the financial distress probabilities as the time horizon lengthens. However, the predictive power of the model was modest and there was scope for improvement. Based on this result, Kim and Partington (2008:20) suggest that the model should be extended to incorporate more timely market information such as stock return and volatility, and to capture macroeconomic changes over time.

5.3 NON-FINANCIAL VARIABLES

5.3.1 Macroeconomic variables

Macroeconomic variables are quantifiable variables, but sub-categorised as non-financial variables, or variables not related to a company's financial statements. Macroeconomics, as a sub-section of economics, focusses on the movement and effect of certain key variables on the economy as a whole. Key variables include for example the level of employment/unemployment, gross domestic product (GDP), balance of payments (BOP), inflation (CPI), money supply (M) and prime lending rate (P).

These macroeconomic events may affect a particular company's health directly or indirectly. It requires company management to be vigilant and at all times acutely aware of the effect of the risks of these variables on the company's current and future financial health and to adapt strategy pro-actively.

Macroeconomic variables and other non-financial variables should be considered in conjunction with historical financial results to determine a company's health. A number of studies combining macroeconomic variables and financial variables are reviewed below.

Cybinski (2001:30) supports the view that in addition to the use of financial ratios to measure a company's internal health, macroeconomic variables should be included to properly model the company's external environment. Cybinski (2001:34) identifies the following five macroeconomic variables; the level of activity/demand or growth factor, the cost of capital borrowing factor, the labour market tightness factor, the construction factor, and the expenditure (private, public, business) factor.

The Dunis and Triantafyllides (2003) study reiterated the conclusion of Becchetti and Sierra (2003:2117) and applied both financial and economic conditions to a neural network regression model. This would allow stakeholders to make a better assessment of the likelihood of future financial distress.

Because companies have a higher propensity to become financially distressed in times of economic recession or downturn than in times of economic prosperity, Dunis and Triantafyllides (2003:2) state that the inclusion of macroeconomic variables in a financial distress prediction model should prove helpful. The data set includes the following macroeconomic variables: real gross domestic product, real money supply, rate of unemployment and the output gap as measures of the business cycle, consumer price index as a measure of inflation, the FTSE 100 stock index, the three-month Treasury Bill as the short-term interest rate the 10-year Government Bond as the long-term interest rate, the real effective exchange rate, the terms of trade and the number of insolvencies as endogenous variables.

Kumar and Tan (2004) developed a hybrid financial distress prediction model based on a review of various techniques for financial distress prediction. An important aspect raised by them is that it may be beneficial to include macroeconomic variables as additional variables in a financial distress prediction model. However, it may provide a better differentiation between financially distressed and healthy companies, but they are of the opinion it may not provide the most accurate number for distress probability.

Argyrou (2006) examined whether macroeconomic variables, in relation to financial ratios, could enhance the ability of neural network models to predict financial distress. Two multi-layer programming-based neural network models, trained by back-propagation algorithm, were constructed. The one model received as input 12 financial ratios, whereas the second model received as input the aforementioned ratios supplemented by five macroeconomic variables over a five-year period. These variables included: terms of trade, gross domestic product, 12-month interest rate, total household disposable income, and cost of living and price index. The two models were then applied to 16 data sets comprising varying proportions to financially distressed and healthy companies.

The results of the study indicate that macroeconomic variables can enhance the ability of a multi-layer programming-based model to classify healthy companies. There is, however, no difference in classifying financially distressed companies. The result further indicate that macroeconomic variables do not enhance the ability of multi-layer programming-based models to predict either financially distressed or healthy companies.

Hol (2007) evaluated a financial distress prediction function on the basis of both financial statement analysis and movement in the business cycle. The combination was found to improve the financial distress prediction compared with financial statements alone. The gross domestic product gap, a production index and the money supply in combination with some financial health variables for individual companies were found to be significant predictors during both recovery and expansion.

Zhang (2006) investigated the effect of incorporating macroeconomic variables such as economic growth, monetary conditions, inflation and stock market performance into a financial distress prediction model. The study focused on prediction accuracy and parameter stability. The results show that the stability of parameters in the prediction model is improved with macroeconomic variables added. In terms of prediction accuracy, the model augmented with macroeconomic variables performed better in a Lauchenbruch jack-knife prediction, but not in out-of-sample predictions.

Because accurate financial distress prediction is of importance to all stakeholders it is important not only to consider company financial statements but also macroeconomic variables. Zhou *et al.* (2010) support this view and explored the effect of macroeconomic variables on improving the financial distress prediction accuracy with neural network models.

The macroeconomic variables included with a set of financial ratios compiled by Zhou *et al.* (2010:82-83) comprised the gross domestic product, personal income index, consumer price index, and money supply index, reflecting the amount of money supply in the economy. Zhou *et al.* (2010:83) concluded that based on the test results the performance of neural network models improved moderately when macroeconomic variables were included.

A number of studies, however, are not in favour of augmenting financial variables with macroeconomic variables. Van Gestel *et al.* (2003:1) express the view that the common assumption underlying financial distress prediction is that macroeconomic variables such as inflation and interest rate together with company-specific variables such as market share and management quality are discounted in a company's financial statements. Data obtained from financial statements are then utilised, in some instances, in sophisticated models such as artificial neural networks, to predict future financial distress prediction. Least-squares support vector machines apply ridge regression in the high dimensional kernel-induced feature space, while practical expressions for model training and evaluation are obtained in terms of the kernel function.

Lam (2004) added to the Van Gestel *et al*. (2003) opinion and investigated the ability of the back-propagation neural network to integrate fundamental and technical analysis for financial performance prediction. The predictor attributes included 16 financial and 11 macroeconomic variables. Lam (2004) compared the neural network performance with the average return from the top one-third returns in the market (maximum benchmark), which approximated the return from perfect information as well as with the overall market average return (minimum benchmark), which approximates the return from highly diversified portfolios. The test results show that a combination of financial and macroeconomic variables cannot generate significantly higher returns than the average index.

In support of the above and more recently, Charalambakis and Garrett (2016:25) established that market variables combined with accounting information or financial variables do not impact on the probability of financial distress in an emerging market.

5.3.2 <u>Qualitative variables</u>

Qualitative variables can potentially contribute to financial distress but pose a problem when considered for inclusion in a quantitative financial distress prediction model. This constraint, however, does not make it a less important aspect to consider.

Keasey and Watson (1987) criticised financial distress prediction models based solely on financial ratios. Their study examined whether it was possible to achieve financial distress predictions from publically available non-financial information, alone or in conjunction with financial ratios. Sources of information such as reporting lags, audit qualifications, the number of directors and the existence of loans secured on company assets could aid financial distress prediction. The study was based on the 1976 Argenti model where several non-financial variables were tested empirically. Although this model lacked empirical evidence, it was nevertheless a first attempt in the field of a hybrid financial distress prediction model.

In another earlier study, a non-financial model was developed and tested by Lussier (1995). The study included the following qualitative non-financial variables: record-keeping and financial control, industry experience, management experience, planning, professional advisors, education, staffing, product/service timing and economic timing. The company success, or health, *versus* a financial distress prediction model, reliably outperformed the random classification of a group of companies as financially distressed or healthy over 99% of the time.

Various factors can lead to financial distress and may vary from company to company. Some proponents attribute this to continuing high interest rates, recession squeezed profits and heavy debt burdens. Furthermore, industry-specific factors, in the context of qualitative variables that may lead to financial distress, include government regulation and a company's nature of business (Neophytou, Charitou & Charalambous, 2000:3)

Becchetti and Sierra (2003) criticise empirical results based on statistical financial distress prediction models. This view is based on the fact that the scope of statistical models restricts researchers to analysis of financial statement variables only. Becchetti and Sierra (2003:2117) concluded that empirical results should not be generalised because the significance of relevant variables tend to be sample specific. In solving this problem, they suggest that non-financial statement and qualitative variables such as customer concentration, subcontracting status, export status and presence of large competitors in the same region, should be included to improve the explanatory powers of models predicting financial distress.

Lee (2004) focuses primarily on the incorporation of financial ratios and specifically, intellectual capital by integrating artificial neural networks with the multivariate adaptive regression spines approach. Intellectual capital, as an intangible asset is often a major determinant of a company's competitiveness and continuous growth, especially where the company derives its income from non-traditional or intangible assets such as customer relations, employee skills, innovations and knowledge-intensive services.

Zong-jun *et al.* (2006) based their research on a combination of various non-financial indexes (ownership concentration coefficient, affiliated debt, pledge and affiliated exchange) with financial indices. It was concluded that among all the non-financial indices adopted, ownership concentration coefficient had the strongest prediction ability, while the others only had little prediction ability. The result was ascribed to the fact that these non-financial indices served as dummy variables, decreasing its prediction ability. In order to advance the prediction ability of a model, it was not advisable to set indices up as dummy variables. Prediction rules combining both financial and non-financial indices had higher prediction accuracy according to single financial indices. It indicates that when establishing a financial distress prediction model, every aspect of the company should be considered, not only financial variables.

The lack of a unified theory on financial distress results in most studies on financial distress prediction focusing on model accuracy rather than model interpretation (Cortés, Martínez & Rubio, 2007:29). A two-fold process was followed by Cortés *et al.* (2007) in an effort to solve this problem. Firstly, a discerning measure was introduced to rank independent variables in a generic classification task. Secondly, a boosting technique was applied to improve the accuracy of a classification tree.

Cortés *et al.* (2007) applied both financial variables, and qualitative non-financial variables such as company size, activity and legal structure. The results showed that the approach followed decreased the generalisation error by about 30% with respect to the error produced with a classification tree.

Sun (2007:56) holds the view that in a highly dynamic environment, statistical models used for comparison with auditors' opinion do not keep pace with the development of financial distress prediction modelling research.

Sun (2007) proposed a statistical model enhanced with non-traditional variables such as a composite measure of financial distress, industry failure rate, abnormal share returns and market capitalisation. Secondly, a hazard model was employed incorporating both financial and non-financial variables. The study results showed that the prediction ability of the hazard model with incorporation of non-financial ratio variables was superior to that of auditors' going concern opinions.

Wang and Li (2007) applied the rough sets model to test the effect of financial and non-financial or qualitative variables on the probability of financial distress. Financial variables used were as follows: growth ratio per share of equity, net return on assets, earnings per share, interest coverage, net profit margin, pledge, retained earnings ratio, and total assets turnover. Qualitative non-financial variables used were as follows: ownership concentration coefficient, affiliated debt pledge and affiliated exchange.

Wang and Li (2007:106) concluded that prediction rules combining financial and qualitative non-financial variables outperformed those rules containing financial ratios only. A key characteristic raised was that the result of the study implied the necessity of considering every aspect of a company and not limiting the analysis to financial characteristics when constructing a financial distress prediction model.

Wu (2007) added to the importance of adding non-financial variables to financial variables in construction of a financial distress prediction model. The study focused on corporate governance as a function of qualitative non-financial variables because corporate governance had been proved that it had a substantial impact on a company's performance.

Wu (2007:28) adopted 10 governance variables and employed the binary logistic regression model to establish a financial distress model. The results indicated that seven variables had a significant impact on the financial distress predictive probability. These variables were as follows: the percentage shares held by institutional shareholders, the extent of concentration, cash flow rights, the ratio of cash flow to control rights, the ratio of board seats held by outside directors and supervisors, management participation and stock pledge ratio.

Zhou and Elhag (2007a:301) acknowledge that bankruptcy is not an abrupt event. Signs of financial distress leading to actual bankruptcy are evident well in advance. They developed a four-variable logistic regression model to predict bankruptcy. A prediction accuracy of 81% with a cut-off point of 0.7 was achieved, while Type I error was 92% and Type II error was 70%.

The overall performance of a logistic regression model indicated that the predictors, standing for a company's profitability, operational efficiency and human resources management, could distinguish the healthy and financially distressed companies. Zhou and Elhag (2007a:306) concluded that although their study used financial ratios as predictors only, the inclusion of qualitative non-financial information such as company size, maturity, research and development expenses and country risk measures could improve the prediction ability of a logistic regression model.

Altman, Sabato and Wilson (2008:5) explored the value added by qualitative information such as the number of employees, the legal structure of the company, the region where the main business was carried out, and industry type in a financial distress prediction model. It was found that the information, when available, was likely to significantly improve the prediction accuracy of the model by up to 13%. The authors qualified their conclusion that this result was more important for small and medium-sized enterprises considering that for a large part of them financial information was often quite limited.

Huang, Tsai, Yen and Cheng (2008) presented a hybrid model composed of static and trend analysis models such as financial structure, credit standing, operating standing, profitability and short-term credit standing. The test results of the proposed hybrid model using a back-propagation neural network produced good prediction accuracy and outperformed other models including the multiple discriminant analysis model, decision trees and the back-propagation model alone. The strongest advantage was that the proposed hybrid model could predict risk by comparison with other companies, and thus it could adapt the changes such as time, economic, environment and other factors, according to Huang *et al.* (2008:1040).

Wen-tsao and Wei-yuan (2008) applied the probabilistic neural network to construct a financial distress prediction model. A probabilistic neural network was characterised as simple, fast and having a high calculation capability and flexibility. The results obtained indicated that the classification model constructed through fine adjustment of the smoothing parameter of probabilistic neural network by the genetic algorithm had better classification capability than those of back-propagation neural network and logistic regression models.

Wen-tsao and Wei-yuan (2008:138) considered financial variables only. Qualitative non-financial variables such as the entire environment and industry characteristics were not taken into account in the study. Based on the results it was proposed that qualitative non-financial variables, such as: human resources, management strategy and the audit opinion, all related to the business operation performance and the potential root cause of financial distress should also be considered for inclusion in a financial distress prediction model.

Sun and Li (2009) focused on the value and importance of experts' experiential knowledge and non-financial information in financial distress prediction. A group decision-making approach was proposed. A qualitative attribute system including seven first-level attributes was constructed for financial distress diagnosis. Because it was common for different experts to have different opinions and even conflict in group decision-making, a multi-expert negotiation mechanism was designed to weigh qualitative attributes and the new concept of expert's expected negotiation factors was put forward. A method integrating linguistic label and interval value was adopted for experts to express preference of attributes, and the experts were divided into several groups which had at least some kind of common preference.

The results of the Sun and Li (2009:904) study established that experts' experiential knowledge and financial and non-financial information were fully utilised to financial distress prediction.

The purpose of the Lin, Liang and Chu (2010) study was two-fold. Firstly, it not only explored the role of financial variables but also that of non-financial variables in financial distress prediction. For this purpose, the study empirically examined whether the combined consideration or hybrid of both financial and non-financial variables led to more accurate financial distress prediction than an exclusive examination of either variable. Secondly, the support vector machine was adopted to predict financial distress based on both financial and non-financial variables.

Lin *et al.* (2010) integrated the qualitative non-financial variables based on the concept of *corporate governance* to diagnose the financial health of a company. In order to enhance the model's performance, feature selection was undertaken by employing stepwise regression to identify the critical features as the input variables.

The empirical results indicated that examining the selected qualitative non-financial features in addition to traditional financial variables provided a promising solution for assessing the risk of financial distress. The proposed hybrid model achieved an overall predictive accuracy rate of 94.4%, superior to those of the model based exclusively on financial variables and the model considering qualitative non-financial variables only.

Scarlat and Delcea (2010:1) highlight the increasing dynamic and complex environment in which a company has to operate, and further, the increasing possibility of financial distress.

Against this background, Scarlat and Delcea (2010) changed the approach in their study and proposed a model assessing a company's financial health, instead of financial distress, in order to maintain or improve it, based on each company's goals.

The proposed model combined the advantages of three well-known theories: grey systems theory, q-fuzzy subset theory and artificial neural networks. Both financial and non-financial variables were taken into consideration in order to achieve better financial distress prediction results.

The consideration of non-financial variables in financial distress prediction was regarded by Scarlat and Delcea (2010) as important because it was often regarded as a leading indicator of financial performance that had not yet been contained in the accounting measures. The following qualitative non-financial variables were considered: customer satisfaction, company reputation, employee satisfaction/morale, employee efficiency, long-term relation with suppliers and customers, quality of products and services, on-time delivery, and growth in number of customers.

Because feature selection was an important step to select more representative data from a driven dataset to improve financial distress prediction performance, Scarlat and Delcea (2010:7) concluded that the prediction accuracy in their proposed Case III model outperformed the other two cases. In addition, it was concluded that even if non-financial variables were not considered, feature selection made through the synthetic degree of grey incidence made the neural networks pattern recognition perform better than in the case whereby both financial and non-financial variables were included.

As indicated earlier, in assessing company health, management should analyse both quantitative and qualitative data. Quantitative data are said to be objective, and based on accounting details obtainable from the company's financial statements. Qualitative non-financial data are regarded as subjective and based on personal knowledge. For subjective knowledge, input from an expert is required. An expert in a domain can be a person with good knowledge or experience in a particular domain. Qualitative variables do not have any measurement, only the rating of the risk factors can be done by the experts based on the corresponding domain (Martin, Aswathy & Venkatesan, 2012:27).

Financial distress prediction is an analysis to ensure the stability or health of a company, according to Martin *et al.* (2012:27). Many companies enter into financial distress due to inappropriate analysis of its operations. The success of a company mainly depends on the timeous and appropriate decisions that are taken by the management at an appropriate time.

Martin *et al.* (2012) proposed a qualitative financial distress prediction model for generating decision-making rules. An ant colony algorithm was used to generate qualitative financial distress prediction rules. The generated rules using the ant colony algorithm were then clustered based on various characteristics by using the associated rule mining algorithm and the best rules among this were extracted using the particle swarm optimisation technique. By using this, the redundancy in rules and false rules in the prediction process was avoided.

In the exploratory study by Van der Colff (2012), a number of qualitative non-financial variables based on a company's strategic capabilities were evaluated. Five primary variables were identified, namely vulnerability, flexibility, effectiveness, resources and capabilities. Each of these variables was expanded to include one or more generic questions to appraise a company's strategic capability and the effect on its financial results. The directors' report was used as the primary source to evaluate the questions on a 1 to 5 scale over a 10-year observation period. From a global perspective, the result was insufficient to prove an outright positive relationship between qualitative non-financial variables and financial distress.

The study of Tinoco and Wilson (2013) can be singled out as an important and one of a limited number of studies that empirically evaluated whether the inclusion of market variables, among financial and macroeconomic (or quantitative non-financial variables) improved the prediction accuracy of financial distress prediction models. Their study clearly indicates that these three variable categories are not mutually exclusive and do in fact act as complement in a financial distress prediction model.

In contradiction to studies supporting the inclusion of non-financial variables in a financial distress prediction model, certain studies express the view that aspects such as poor management, autocratic leadership and difficulties in operating successfully in the market are reflected in a company's financial results (Van Gestel, Baesens, Suykens, Van den Poel, Baestaens & Willekens, 2006:980). This is similar to the view expressed by Van Gestel *et al*. (2003) in Section 5.3.1, namely that macroeconomic variables have been consolidated in a company's financial results.

5.4 OTHER RELATED STUDIES

A number of studies excluded non-financial variables in the initial testing phase, but concluded that in order to enhance the results, it would be a requirement to include non-financial variables in the test.

Shin and Lee (2002) proposed a genetic algorithm approach for financial distress prediction based on nine financial variables. The study achieved promising results; however, it was suggested that in order to improve results further, qualitative variables should be combined with quantitative variables in extracting prediction rules.

Quah and Srinivasan (2005) applied a feed-forward back-propagation neural network model to predict financial distress and concluded that year-1 neural networks predicted at a 94.44% level of accuracy. Their study demonstrated the usefulness of neural network models in financial distress prediction.

Quah and Srinivasan (2005) highlight that neural networks can be applied to a dynamic environment as they are able to learn from historical data for future prediction. In order to adapt neural networks to a dynamic environment, it is necessary to update the training sample continuously and carry out periodic retraining. This will ensure that the neural network is captured with the latest information about financial ratios and the company's financial health.

The fact that Quah and Srinivasan (2005:1) acknowledged that a neural network can be adapted to function in a dynamic or changing environment is of key importance. However, their study can be criticised to the extent that reliance is placed on financial variables only, which is re-active in nature. The model is only updated or retrained once new or actual financial data become available. Historical financial data can be regarded as the end-result of management decisions in response to or interpretation of other current and future internal and external variables in a dynamic environment.

Quah and Srinivasan (2005:4) acknowledged the limitations of their study and proposed that qualitative non-financial variables should be incorporated as well.

Santos, Cortez, Pereira and Quintela (2006:349) differentiated between two types of financial distress prediction models, namely accounting-based and market-based models. Their study focused on the first approach. In this context, 16 distinct models were evaluated by comparing different algorithms (for example, neural networks and decision trees), training strategies (for example balanced training sets), and feature selection (for example the use of data for one or all three years).

The 16 models were separated into two group sets: the first group set, corresponding to a multi-year approach (three consecutive years), and the second group set, based on a one-year approach (the final year). Accuracies of between 86% and 99% were achieved indicating that the followed approach enabled the use of a data mining model to predict financial distress.

However, what is lacking in the Santos *et al.* (2006:357) study is the consideration of macroeconomic (for example, tight monetary policy, investors' expectations about economic conditions, and the general state of the economy) and other qualitative non-financial variables (for example, whether a budgetary control system is in place, and whether the skills of board members are unbalanced). The study recommend this to be a subject for future research.

Although linear financial distress prediction models are simplistic, they may require unrealistic statistical assumptions. Neural networks, on the other hand may be overly complex in design and interpretation. Zheng and Yanhui (2007:1) propose the use of decision trees, which not only have non-linear architecture, but are able to discriminate between patterns that are not linearly separable and allow data to follow any specific probability distribution. In addition, they are easier to interpret.

Results obtained by Zheng and Yanhui (2007:4) support the use of the decision tree model as a financial distress prediction model, but acknowledged that financial distress is not only detected by financial variables. The inclusion of non-financial variables in a 'multi-method' should provide more valuable results.

5.5 CONCLUSION

The purpose of this chapter was to review studies based on financial and non-financial variables with the objective of identifying the most commonly used financial and non-financial variables in financial distress prediction modelling. The chapter was divided into three broad sub-sections, namely studies based on financial variables such as financial ratios and market variables, studies based on non-financial variables such as macroeconomic and studies based on qualitative variables. The final section reviewed a number of studies that included financial variables, but concluded that the inclusion of non-financial variables would contribute to more timeous and accurate financial distress prediction.

Differentiation was made between quantitative and qualitative variables. Quantitative variables consist of easy identifiable market and macroeconomic variables. Qualitative variables are more problematic to identify and categorise because they cover a diverse range of potential variables. Those studies that included qualitative variables relied predominantly on expert opinion, which was subjective and questionable from an empirical research point of view. This constraint does, however not render qualitative non-financial variables less important and can be considered a subject for future research.

In addition to the above, it is evident from the review that no unified approach was followed in financial distress prediction modelling and non-financial variables selection.

CHAPTER 6

THEORETICAL BACKGROUND OF SUPPORT VECTOR MACHINES AND APPLICATION IN FINANCIAL DISTRESS PREDICTION

6.1 BACKGROUND

Early financial distress prediction studies based on statistical models such as univariate analysis, multiple discriminant analysis and logistic regression were subject to various restrictive assumptions. For example, the reliability of models such as multiple discriminant analysis and logistic regression models was questioned when complexity and non-linearity were present in the data set (Yang, 2003:7).

These restrictions were overcome with the advent of computer technology and the development of artificial intelligence modelling, which contributed to a new era in financial distress prediction. Van Gestel *et al.* (2003:1) suggest that neural networks may be an acceptable replacement for multiple discriminant analysis models because of their universal approximation property and non-linear modelling capacities. The study report on various empirical studies that prove the superiority of neural networks over multiple discriminant analysis and logistic regression models.

Despite its purported success, neural networks are subject to various problems such as the non-convex training problem with multiple local minima, difficulties with generalisation due to overfitting, its dependence on the researcher's experience and knowledge for pre-processing of selecting a large number of control parameters that include relevant input variables, hidden layer size, learning rate and momentum, and in addition, the black-box phenomenon, according to Van Gestel *et al.* (2003) and Min and Lee (2004).

The foundation for the support vector machine model was established by Cortes and Vapnik (1995). The model is one of the models within the broad category of artificial intelligence expert systems and machine learning models. In early 2000, the support vector machine model was suggested as an attempt to provide a financial distress prediction model with better explanatory power and stability and has since attracted increasing attention from both the research and practitioners' community (Min, Lee & Han, 2006).

The main difference between neural networks and support vector machines is the principle of risk minimisation, according to Min *et al*. (2006:652). While neural networks implement empirical risk minimisation to minimise the error on the training data, support vector machines implement the principle of structural risk minimisation to minimise the generalisation error by constructing an optimal separating hyper plane in the hidden feature space, using quadratic programming to find an optimal solution.

The following sub-section reviews various studies relating to the performance of support vector machines compared with statistical and artificial neural network models. The second sub-section provides an overview of constraints experienced with and enhancements to support vector machines.

6.2 DEVELOPMENT OF SUPPORT VECTOR MACHINES

6.2.1 Performance of the support vector machine compared with that of statistical and artificial neural network models

This section highlights a number of studies which contributed to the development of support vector machines. Each of the studies highlights the constraints experienced with both statistical or artificial neural networks and the improved contribution of the support vector machines.

Yang (2003:47) criticises the use of statistical models as reliable financial distress predictors when complexity or non-linearity is present in a data set. Artificial neural network models such as back-propagation neural networks, self-organising maps and probabilistic neural networks have since the early nineties been used to deal with shortcomings of the statistical financial distress prediction models.

The Yang (2003) study supports Van Gestel *et al*.'s (2003) supposition, namely that the support vector machine-learning model outperforms most other models, such as the Fisher discriminant analysis, logistic regression, back-propagation neural networks and probabilistic neural networks. It is comparable with the heteroscedastic probabilistic neural network, which produces the best prediction accuracy prior to the application of support vector machines.

According to Min and Lee (2004:1), a limited number of studies applied artificial neural networks to provide a better understanding of the financial distress prediction process. The support vector machine was applied to solve financial distress prediction problems in an attempt to provide a model with improved explanatory powers. In order to evaluate the support vector machine's prediction accuracy, the study compared its performance with three-layer fully connected back-propagation neural networks. The results of the empirical analysis showed that the support vector machines outperformed the back-propagation neural networks. The result could be attributed to the fact that support vector machines implemented the structural risk minimisation principle, leading to better generalisation than conventional artificial neural networks.

The predictive power of artificial neural networks has empirically been proved to perform better than statistical financial distress prediction models. However, according to Min and Lee (2005), it is commonly reported that artificial neural networks require a large amount of training data to estimate the distribution of input patterns, with difficulty of generalising the results because of their overfitting nature. In addition, it was highlighted that artificial neural networks fully depend on researchers' experience or knowledge of pre-processing data in order to select control parameters including relevant input variables, hidden layer size and momentum.

In an attempt to deal with these constraints, Min and Lee (2005) applied support vector machines to the financial distress prediction problem. Mapping input vectors into high-dimensional feature space, support vector machines transformed complex problems (with complex decision surfaces) into simpler problems that could use linear discriminant functions. The study showed support vector machines' attractive prediction power compared with other existing artificial neural networks.

Shin, Lee and Kim (2005:127) list the following limitations of the back-propagation neural network model:

- Firstly, it is difficult to identify an appropriate artificial neural network which can reflect problem characteristics because of the large numbers of controlling parameters and processing elements in the layer.
- Secondly, the gradient descent search process to calculate the synaptic weights may converge to a local minimum solution that is a good fit for training examples.
- Finally, the empirical risk-minimising principal that seeks to minimise the training error does not guarantee good generalisation performance.

Shin *et al.* (2005) applied the support vector machine model in order to deal with these limitations. The results of their study demonstrated that the support vector machine model had a higher level of accuracy and better generalisation performance than the back-propagation neural network because the training set size was getting smaller sets.

In analysing the determinants of financial distress, Henchiri, Benammou and Hamza (2009) applied the support vector machine model in an effort to improve financial distress predictability. Because the optimal parameter search of the support vector machine played a crucial role in building a financial distress prediction model with high accuracy and stability, a five-fold cross-validation and grid-search technique was applied in order to identify the correct value parameter of the kernel function of the support vector machine.

A comparison of the support vector machines with the artificial neural networks, multiple discriminant analysis and logistic regression model indicated that the support vector machine and artificial neural network approach slightly outperformed the logistic regression and multiple discriminant analysis models in terms of prediction performance of the test data. In terms of training data, the support vector machine showed superior performance in relation to the artificial neural network and statistical models.

Gorgani, Moradi and Yazdi (2010) tested a support vector machine in company going-concern or financial health prediction. Two different data sets were used to assess support vector machines. Different training and testing proportions of each data set were used to train and test support vector machines. In addition, the support vector machine classifier was trained by different kernel functions in order to compare it with the benchmark of the neural network model. Using different kernel functions and the determination of optimal values of the parameters to train support vector machines led to different results. These issues affected the ability to make the final conclusion reliable.

Őnder (2010) applied the support vector machine to predict financial distress based on financial ratios and compared the prediction results with those of the logistic regression model. In order to achieve the best prediction power of the support vector machine model, a grid-search technique was used to select the optimal values of radius used in kernel function and capacity.

In the best case, overall testing accuracy ratio obtained through applying the support vector machine was found to be 75%. On the other hand, the ratio reached the peak at 71.8% with the logistic regression model. Through the Lorenz curves, Őnder (2010:41) determined that the support vector machine model posed a stronger financial distress prediction performance than the logistic regression model. However, in predicting solvent companies, the logistic regression model demonstrated better performance.

Őnder (2010:41) highlights a number of factors on which the performance of the support vector machine model depends. These factors are as follows:

- the size of training and test sets;
- the ratio of training set to test set;
- the utilisation of the proper kernel function;
- the selection of financial ratios as input variables.

By changing these factors, it is possible to improve the classification power of the support vector machine model.

One limitation of the support vector machine model was isolated, namely the size of the training set. The larger the training set, the worse the support vector machine performs.

Financial ratios are used as input variables in certain financial distress prediction models. A key limitation in the utilisation of financial ratios is that it can affect the effectiveness of a particular financial distress prediction model if applied across industries. This has prompted the development of procedures for adjusting the financial ratios so that the same prediction models can be used across a range of industries.

The primary aim of a study by Lin, Liang and Chen (2011) was to select financial variables which might have been ignored in previous studies but could be useful to obtain better financial distress prediction results. A set of 10 financial ratios served as potential candidates for the construction of prediction models. In the second phase, support vector machine models based on the selected features consisting of 10 financial ratios were constructed.

Further analysis by Lin *et al.* (2011) indicated that a support vector machine model built with a feature set of five ratios yielded the best performance. This model was also compared with other models based on the feature sets recommended by prior studies. Results indicated that their model outperformed other models in prediction accuracy.

Cleofas-Sánchez, García, Marqués and Sánchez (2016:144) concur with the above studies in support of the use of popular computational intelligence tools, such as support vector machines. The models are capable of extracting meaningful information from imprecise data and detecting trends that are too complex to be discovered by either human or conventional systems. The authors have established that the accuracy and generalisation performance of a support vector machine is usually better that that of statistical and other soft computing techniques.

6.2.2 <u>Constraints and enhancements to support vector machines</u>

In an early South African study, Kornik (2004:91; 228) acknowledged the applicability of support vector machines as a machine learning financial distress prediction model. The author expressed the opinion that he supported the vector machine, which was almost identical in performance to the kernel ridge regression (KRR) model, and that there was little to choose between these two models in terms of generalisation accuracy.

A number of practical differences were identified by Kornik (2004:228), such as the use of the hinge loss function in the support vector machine, which tended to result in a more sparse solution, with many of the α-values becoming zero. Such a sparse model resulted in a faster run-time classification and a lower data storage requirement. The kernel ridge regression model was viewed as more appropriate for moderately sized data sets.

Based on this view, Kornik (2004) preferred the kernel ridge regression model and proceeded to run tests against the k-nearest neighbour (K-NN) model. Results of the tests indicated that the kernel ridge regression model outperformed the k-nearest neighbour model with large subsets. Conversely, the k-nearest neighbour outperformed the kernel ridge regression model with smaller feature subsets.

Min *et al.* (2006:652) emphasise the principle of risk minimisation as the main difference between artificial neural networks and support vector machines. While, as indicated earlier, artificial neural networks implement empirical risk minimisation to minimise the error on training data, support vector machines implement the principle of structural risk minimisation by constructing an optimal separating hyper-plane in the hidden feature space by using quadratic programming to find a unique solution.

Min *et al.* (2006) improved on the support vector machine model by integrating it with a genetic algorithm. The integrated genetic algorithm-support vector machine (GA-SVM) model was effective in finding the optimal feature subset and parameters of support vector machine, and improving the financial distress prediction accuracy.

Zhou and Tian (2007) proposed a classifier hybridised rough set and wavelet support vector machine. This hybrid model originated from limitations identified in the practical application of the support vector machine.

Although the support vector machine was successfully applied to financial distress prediction, it was found that irrelevant variables in the sample data could spoil the classification of the support vector machine classifier, resulting in the increasing unwanted calculations, and decreasing the real-time capacity of financial distress prediction (Zhou *et al.* 2007:602).

In order to solve the problem and extract the required features Zhou and Tian (2007) suggested the use of dimensionality-reducing methods such as the principle component analysis. These dimensionality-reducing methods usually lose useful information while discarding some redundant variables. Especially, when there are many correlated variables originally, the number of principal components gained by principle component analysis is still large if enough fault messages are to be retained. Too many principal components will equally bring in many irrelevant messages and consequently reduce the efficiency of financial distress prediction. Better and more efficient methods are required.

On the other hand, kernel function selection is another problem in forming an efficient support vector machine classifier. If the kernel function is not selected correctly, generalisation ability will be poor. The performance of the classifier will therefore be influenced.

Although the support vector machine outperformed other financial distress prediction models, such as multiple discriminant analysis models and artificial neural networks, it employs the structural risk minimisation principle; the risk of misclassification may be high at a point closer to the optimisation hyperplane. In addition, support vector machines may be sensitive to outliers or noises due to an overfitting problem.

In using a support vector machine to establish the diagnosis model for financial distress prediction, some factors must be considered, according to Chen and Hsiao (2008:1154):

- the choice of data base to sample data sets;
- the variables in the training data set;
- the selection of significant features;
- the optimal parameter combinations to establish the model through training; and
- the way to improve model diagnosis capabilities.

Chen and Hsiao (2008) integrated the genetic algorithm and the support vector machine to establish the diagnosis model for financial distress prediction by using their traits in parameter evolution as well as data training and classification. The genetic algorithm-support vector machine model achieved an average testing accuracy of 95.56% by using only five financial features and one intellectual capital feature.

The heterogeneous nature of companies and its financial status formed the basis for a study by Ribeiro, Silva, Vieira, Gaspar-Cunha and Das Neves (2010). The asymmetry of information made available by companies for use by stakeholders added to the complexity in determining financial distress.

Against this background, Ribeiro *et al.* (2010) proposed the development of a comprehensive method incorporating a holistic perspective. The support vector machine plus (SVM+) model was proposed, which provided a formal way to incorporate privileged information in order to improve generalisation.

According to Ribeiro *et al.* (2010:2), the support vector machine plus is a new paradigm, while upholding the main principles of the support vector machine, it extends its concept by incorporating the essence of 'untold' information often not handled in a learning problem.

Test results in the setting of a heterogeneous data set demonstrated that the model proposed by Ribeiro *et al.* (2010) showed superior performance in terms of financial distress prediction accuracy.

Yanqing, Shiwei, Junfeng and Lei (2010) highlight some restrictive assumptions of the univariate analysis, multiple discriminant analysis and logistic regression models, such as linearity, independence amongst predictors and pre-existing functional form. Many subsequent studies have shown that an artificial neural network is less vulnerable to these restrictive assumptions. Yanqing *et al.* (2010: 373) support the view that the main difference between an artificial neural network and the support vector machine is the principle of risk minimisation.

The principal component analysis (PCA) is used to select relevant variables and reduce the complexity of the support vector machine. The proposed principle component analysis-support vector machine approach has two distinct advantages. One is that the computation complexity of the principle component analysis-based support vector machine is reduced by the decrease of model inputs and running speed will be accelerated. Another advantage is that the principle component analysis-based support vector machine can avoid some defects of artificial neural networks, such as local minima and overfitting.

Test results demonstrated that the model proposed by Yanqing *et al.* (2010) outperformed a statistical, back-propagation neural network and individual support vector machine. Especially compared with the back-propagation neural network, their proposed model revealed higher accuracy and better performance. In the modelling process, Yanqing *et al.* (2010) employed the principal component analysis method in simplifying the input vector of the support vector machine and the test results reported its efficiency.

Chaudhuri and De (2011:2472) state that signs of potential financial distress are evident long before bankruptcy occurs. The causes leading to financial distress can be divided into economic, financial neglect, fraud disaster and others. Further economic factors include industry weakness and poor location, and financial factors include excessive debt and insufficient capital. Financial difficulties are the result of managerial error and misjudgement. When errors and misjudgement proliferate, it can be a sign of managerial neglect.

The support vector machine is the most widely used non-parametric machine-learning model in the broad artificial intelligence model category and is deemed to be most accurate. It has a flexible structure and produces better classification results than parametric models. The support vector machine has attractive properties and provides a single solution characterised by a global minimum of optimised functional and multiple solutions associated with local minima. It does not rely on heuristics and thus is an arbitrary choice to model various problems.

Support vector machines are based on very few restrictive assumptions and can reveal effects overlooked by other methods. Support vector machines have been able to produce accurate classification results in other areas. However, according to Chaudhuri *et al.* (2011:2485), real-life company data have an inherent degree of uncertainty and impreciseness; it is obvious that unpredictable results may emerge.

Against this background, Chaudhuri *et al.* (2011) investigated the effectiveness of support vector machines with fuzzy membership functions embedded in them leading to the development of the fuzzy support vector machine to financial distress prediction. The fuzzy support vector machine is effective in finding optimal feature subset and parameters thus improving financial distress prediction. There are several arguments supporting the observed high accuracy of the support vector machine by choosing the appropriate value of parameters, which plays an important role in the performance of the fuzzy support vector machine model.

The performance of the fuzzy support vector machine was illustrated by test results, which showed that it was better capable of extracting useful information from company data than traditional financial distress and bankruptcy prediction models; though extensive data sets were required in order to fully utilise their classification power.

The results of the Chaudhuri *et al.* (2011) study demonstrated that the fuzzy support vector machine was effective in finding optimal feature subset and parameters. This improved financial distress prediction.

Chen (2011c:2) reached the following conclusions:

- The artificial neural network is a suitable model for development of financial prediction models.
- Genetic algorithm is a good solution for tuning parameters for artificial neural network, and the integrated optimisation technique is also an important issue in artificial neural networks.
- The hybrid approach outperforms any single approach in terms of prediction performance.

Against this background, Chen (2011c:2) purports that a limited number of studies have been conducted on swarm-inspired optimisation techniques for financial distress prediction. Because most real-world problems are multi-criteria problems, it was regarded appropriate to use multi-objective algorithms in searching for solutions.

Chen (2011c:10) proposed a hybrid evolution approach to particle swarm optimisation with the support vector machine model for financial distress prediction. The main objectives were as follows:

- to use financial and non-financial ratios and a macroeconomic index to improve accuracy of the financial distress prediction model;
- to adopt swarm-inspired optimisation techniques to construct a financial distress prediction model;

- to compare the accuracy of particle swarm optimisation-support vector machine and another artificial neural network approach;
- to expand the proposed model so that it would work within a financial distress prediction system as a type of early-warning system.

Chen (2011c:1) provided four critical contributions, as follows:

- It was found that when a third-iteration principle component analysis was applied to all variables, a 94.41% total explained variance was obtained. All non-financial ratios and macroeconomic indices were eliminated by the first-iteration principal component analysis. This showed that financial prediction performance was mainly influenced by financial ratios, as opposed to non-financial ratios or macroeconomic indices.
- Using principal component analysis, the study selected 12 critical variables, approximately 70% fewer variables as input than other methods, but the model was still able to provide highly accurate financial distress forecasts.
- The particle swarm optimisation-support vector machine model yielded higher classification accuracy than the grid-support vector machine. The Wilcoxon statistics also clearly showed that the particle swarm optimisation-support vector machine significantly outperformed grid-support vector machine.
- The particle swarm optimisation-support vector machine model generally produced better prediction accuracy than the grid-support vector machine, genetic algorithm, simple support vector machine and self-organising map.

Lin, Yeh and Lee (2011:95) believed that financial distress symptoms can be observed prior to financial difficulty or crisis. Accurate financial distress prediction models are therefore of critical importance in terms of decision-making of various stakeholders, as the models provide them with timely warnings of the company’s actual situation.

According to Lin *et al.* (2011:95), prior studies demonstrated that the support vector machine outperformed the artificial neural network, multiple discriminant analysis and logistic regression models in financial distress prediction. In order to achieve a better classification performance in a support vector machine for financial distress prediction, the data inputs for the classifier required special treatment during preparation to guarantee a good performance in the classifier.

Against this background, Lin *et al.* (2011:95-96) proposed a hybrid model of manifold learning approach which combined both the isometric feature mapping algorithm and support vector machine. By using the isometric feature mapping algorithm to conduct dimension reduction, which was then used as a pre-processor in order to improve the financial distress prediction capability of the support vector machine. The effectiveness of the proposed hybrid model was verified by tests that combined compared principal component analysis with support vector machines.

The results indicated that there was no significant difference between the combined isometric feature mapping and support vector machine and the combined principle component analysis and support vector machine in terms of prediction accuracy. However, when the average prediction results were considered, the combined isometric feature mapping and support vector machine proved the better model. On the other hand, by examining the Type I and II errors of these models, the hybrid model proposed by Lin *et al.* (2011) produced fewer Type II errors.

The support vector machine, has become increasingly popular. The formulation of the support vector machine simultaneously embodies the structural risk (a maximum margin classifier) and empirical risk minimisation principles. Consequently, support vector machines combine excellent generalisation properties with a sparse model representation (Huang, Tang, Lee & Chang, 2012:3855).

Although support vector machines have demonstrated superior performance in numerous areas of pattern recognition, the traditional support vector machine does not make efficient use of both labelled training data and unlabelled testing data. Moreover, high-dimensional and non-linear distributed data generally degrade the performance of a classifier due to the curse of dimensionality in financial distress prediction.

To solve these problems Huang *et al.* (2012:3855) proposed a hybrid classifier which integrated kernel local Fisher discriminant analysis with a manifold-regularised support vector machine. Kernel local Fisher discriminant analysis was employed to find an optimal projection which maximised the margin between data points from different classes at each local area of data manifold, while manifold-regularised support vector machine data-dependently wrapped the structure of feature space to reflect the underlying geometry of the data manifold.

Empirical results of the study showed that the proposed system was more accurate and robust than pure support vector machine classifiers, and also outperformed conventional techniques when applied to financial distress prediction.

By examining the kernel local Fisher discriminant analysis, Huang *et al.* (2012) found a good low-dimensional projection, which respected the discriminant structure inferred from the labelled data points, as well as the local geometrical structure inferred from both labelled and unlabelled data points. Traditional linear discriminant analysis only preserved the global discriminant structure, while fully ignoring the local geometrical structure. Kernel local Fisher discriminant analysis maximised the margin between data points from different classes at each local area of data manifold. Consequently, its performance was better than that of linear discriminant analysis. Integrating kernel local Fisher discriminant analysis into a classifier could reduce its computational loading and simultaneously enhance its performance. In the second stage, a manifold-generalised semi-supervised support vector machine was used for classification. The manifold-regularised support vector machine used the data-dependent norm on reproduced kernel Hilbert spaces to alter the structure of the reproduced kernel Hilbert spaces to reflect the underlying geometry of the data. The success of the proposed hybrid classifier mainly enhanced the combination of two techniques.

Ribeiro, Silva, Chen, Vieira and Das Neves (2012:10140) are of the view that in spite of many advanced financial distress prediction models that have been proposed, no comprehensive model incorporating a holistic perspective has hitherto been considered. Thus, the existing models for financial distress prediction lack the whole coverage of contextual knowledge which may prevent the decision-maker to make the right decision. The study proposed the support vector machine plus model, which provided a formal way to incorporate additional information or non-financial variables (not only training data) into the learning models, thereby improving generalisation.

The study showed that, firstly, the support vector machine plus not only yielded a better prediction model than the baseline support vector machine, but also a better model compared with a similar approach of multi-task learning, and secondly, the most salient data parameters per group both in the kernel decision space and in the kernel correcting space were optimised, whereby the parameters and parameter ranges that shaped the various company profiles were exposed. The classification results demonstrated the prediction and robustness of the proposed method.

Li, Sun, Li and Yan (2013) investigated the use of a two-stage ensemble of multiple discriminant analysis and logistic regression model for financial distress prediction. A constructive algorithm of the implementation of the forecasting model was employed in order to obtain a parsimonious ensemble classifier. The technique of using various data representations as input of the same model (i.e. multiple discriminant analysis or logistic regression models) was adopted to generate different classifiers. Principal component information was extracted from four optimal feature subsets representing samples by using principle component analysis. The concept of *majority voting* was used to integrate prediction of the eight principal component statistical models in the classifier level of multiple discriminant analysis or logistic regression models and the second stage of ensemble. For classifier weighting, ranking order information on the base of the model's total accuracy rate was used.

For comparison, Li *et al.* (2013:2) employed the best classical and intelligent models. The statistical models or multiple discriminant analysis and logistic regression models in their best stand-alone modes were employed to compare them with the proposed two-stage ensemble model. Although artificial neural networks have often been used in financial distress prediction, no consensus on their superiority with respect to multiple discriminant analysis and logistic regression models could be established. Furthermore, the black-box phenomenon also prevents artificial neural networks from being successfully applied to management science, and specifically, the financial distress prediction problem.

The support vector machines consistently compare favourably with statistical models and artificial neural networks. Support vector machines can also be viewed as a specific implementation of an artificial intelligence or machine learning model. Against this background, Li *et al.* (2013) compared the support vector machine with the two-stage ensemble model. The resulting ensemble model compared favourably with the multiple discriminant analysis, logistic regression and support vector machine models and also with all its component models.

Based on previous research where performances of different single classifiers were compared, Sun and Li (2012:2264) found the support vector machine to have performed better. However, the ensemble model for financial distress prediction has gained popularity in recent years and warranted further research.

Sun and Li (2012) proposed a new support vector machine ensemble model for financial distress prediction, in which the criteria for selecting base support vector machine classifiers' predictive ability and their diversity degree were considered as the criteria for selecting base support vector machine classifiers from candidate ones. This assisted in building an effective support vector machine ensemble for financial distress prediction, and avoided added complexity to the support vector machine-based financial distress prediction system without meaningful performance improvement. Weighted majority voting was applied as combination mechanism according to base classifiers' cross-validation accuracy on the training data set. The candidate support vector machine classifiers for the ensemble were constructed through different support vector machine kernels and different features subset. The applied support vector machine kernel included linear kernel, polynomial kernel, radial-basis function kernel and sigmoid kernel, and the feature selection/extraction models included stepwise multiple discriminant analysis, stepwise logistic regression and principal component analysis.

Test results indicated that the performance of the support vector machine ensemble proposed by Sun and Li (2012) for financial distress prediction was significantly better than that of individual support vector machine classifiers when the number of base support vector machine classifiers was properly set. Empirically, a support vector machine ensemble with more than nine base classifiers tended to result in acceptable prediction performance in the test, and at least more than three base classifiers were needed to avoid an invalid support vector machine ensemble. If a single support vector machine classifier was used for financial distress prediction, a radial-basis function-support vector machine with features selected by a stepwise multiple discriminant analysis model was an acceptable choice.

The support vector machine is regarded as a state-of-the-art classification method and is one of the most promising among recently developed financial distress prediction models, according to Härdle, Prastyo and Hafner (2012:2). The support vector machine was applied to financial distress prediction and typically outperformed the competing models. One of the important issues in support vector machine is the parameter optimisation (variable selection). The variable selection of the support vector machine for financial distress prediction was empathised in the study. The support vector machine parameters are optimised by a using genetic algorithm.

Although the Moepya, Nelwamondo and Van der Walt (2014) study relates to financial statement fraud detection, the principle of the support vector machine application in the South African context is of importance to the current study.

Moepya *et al.* (2014:1) compared three support vector machine models, namely the radial-based function, quadratic and linear kernel choice, to the k-nearest neighbour (k-NN) and logistic regression. Based on their investigation and in this particular application of fraud detection in company financial statements, the support vector machine proved more effective compared with the statistical models applied in the study.

The results of the study showed the effectiveness of support vector machines as a tool to detect the manipulation of financial statements. Furthermore, the results indicated the robustness of the simple linear support vector machine compared with other models using the holdout sample. An important observation was the effect of variable selection on prediction accuracy, which was captured by an increase in model sensitivity and specificity.

In a recent study Fallahpour, Lakvan and Zadeh (2017) expanded on the original support vector machine and combined it with the sequential floating forward selection (SFFS) algorithm. Their study confirms the improved performance of the ensemble support vector machine as a financial distress prediction classifier.

6.3 CHARACTERISTICS OF SUPPORT VECTOR MACHINES

Support vector machines use a linear model to implement non-linear class boundaries by mapping input vectors non-linearly into a high-dimensional feature space (Chen, 2011a:4517). The support vector machine is described by Min and Lee (2005:604) as the algorithm that finds a special kind of linear model, namely the maximum margin hyperplane. The maximum margin hyperplane gives the maximum separation between two decision classes.

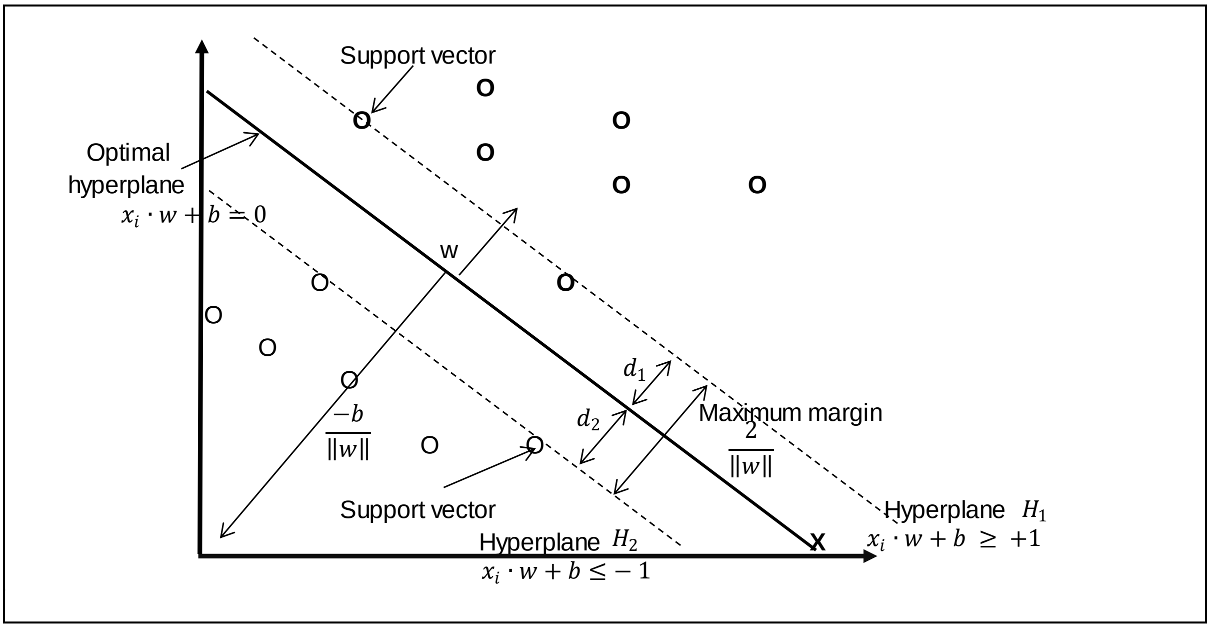
Lin, Liang and Chen (2011:15096) describe the basic procedure for applying the support vector machine to a classification problem, as follows:

- Firstly, the input vector is mapped into a feature space, which is possible with a higher dimension. The mapping is either linear or non-linear, depending on the kernel function selected.
- Secondly, within this feature space, an optimal separation or hyperplane between the two or more classes or vectors is established.

By applying structural risk minimisation, the objective of training support vector machines is to establish a globally optimised solution and avoiding overfitting. This allows the support vector machine to manage a large number of features.

Figure 6.1 is a graphical representation of support vector machine architecture. The circles represent two different classes: the circles left of H_2 represent financially distressed companies and the circles with thick boundaries, right of H_1, represent financially healthy[9] companies.

Figure 6.1: Architecture of a typical support vector machine



Source: Adapted from Yang (2003:48), Fletcher (2009:2), Őnder (2010) and Chen (2011c:3).

The objective of a support vector machine is to establish an optimal hyperplane, denoted by the line "X" in Figure 6.1. The optimal hyperplane separates the analysed sample companies into two groups of non-distressed (H_1) and financially distressed (H_2) companies. The largest distance between the nearest training data point of any class is called the margin, denoted by the distance between H_1 and H_2 in Figure 6.1. Geometrically the distance between these two hyperplanes is denoted by $\frac{2}{\|w\|}$. To maximise the distance between the two hyperplanes, $\|w\|$ has to be minimised.

The larger the margin the better the separation between the two decision classes, namely non-distressed and financial distress companies. Yang (2003:48) refers to the margin as a measure of the expected generalisation ability.

[9] Hereafter, the term *non-distressed* is used.

A set of features (outcome or result of a model) that describes a company as either financially distressed or non-distressed is called a vector and the vector closest to maximum hyperplanes H_1 or H_2 are called support vectors. Support vectors in Figure 6.1 are represented by the circles intersected by the hyperplanes H_1 and H_2.

The aim of the support vector machine in financial distress prediction is to maximise the margin between the support vectors - the distance between H_1 and H_2 in order to find the best separation between non-distressed and financially distressed companies (Őnder, 2010:9).

The following section describes the basic formulation of a support vector machine.

6.4 FORMULATION OF SUPPORT VECTOR MACHINES

Fletcher (2009) provided a simplification of the support vector machines initially conceived by Cortes and Vapnik (1995). The document provides the problem of classification for linearly separable data and introduces the concept of *margin* and the essence of support vector machines, namely margin maximisation. In addition, in the document, the methodology of the support vector machine extends to data, which are not fully linearly separable. The final section develops the concept of *support vector machines* further so that the model can be used for regression.

The following section explains and introduces the concept of *linearly separable data* and the essence of support vector machines (Fletcher 2009:1-5).

Assume L training points, where each input x_i has D attributes and is one of two classes $y_i = -1\ or + 1$. Training data are in the form:

$$\{x_i, y_i\} \tag{5}$$

where: $i = 1 \ldots L$

$y_i \in \{-1,+1\}$

$x \in R^D$

The support vector machine finds an optimal separating hyperplane that distinguishes an instance in one class from another. The hyperplane can be described as:

$$w \cdot x + b = 0 \qquad (6)$$

where: w is normal to the hyperplane, and
$\frac{b}{\|w\|}$ is the perpendicular distance from the hyperplane to the origin.

Referring to Figure 6.1, implementing a support vector machine requires the selection of variables w and b so that the training data can be described by:

$$x_i \cdot w + b \geq +1 \quad \text{for } y_i = +1 \qquad (7)$$
$$x_i \cdot w + b \leq -1 \quad \text{for } y_i = -1 \qquad (8)$$

Fletcher (2009:2) combines these equations into the following equation:

$$y_i(x_i \cdot w + b) - 1 \geq 0 \ \forall_i \qquad (9)$$

The two hyperplanes (H_1 and H_2), which intersect the points, i.e. the support vectors, can be described as:

$$x_i \cdot w + b = +1 \quad \text{for } H_1 \text{, and} \qquad (10)$$
$$x_i \cdot w + b = -1 \quad \text{for } H_2 \qquad (11)$$

In Figure 6.1, d_1 denotes the perpendicular distance between the optimal hyperplane (X) and H_1, and similarly for d_2 and H_2. The distance between the optimal hyperplane and H_1 is equal to the distance between the optimal hyperplane and H_2, which suggests that d_1 = d_2, and is known as the margin. The objective is to maximise the margin, which is an indication of how well the data are separated and whether there is an improvement in generalisation.

Geometrically, the distance between the optimal hyperplane (X) and the hyperplane (H_1 or H_2) is $\frac{1}{\|w\|}$, or between H_1 and H_2 is $\frac{2}{\|w\|}$. In order to maximise the distance between H_1 and H_2 , $\|w\|$ has to be minimised. Minimising $\|w\|$ is equivalent to minimising $\frac{1}{2}\|w\|^2$.

In order to cater for the constraints in this minimisation, Lagrange multipliers $\boldsymbol{\alpha}$ have to be allocated to it, where $\alpha_i \geq 0 \ \forall_i$:

$$L_p \equiv \frac{1}{2}\|w\|^2 - \alpha\ [y_i(x_i \cdot w + b) - 1\forall_i] \tag{12}$$

$$\equiv \frac{1}{2}\|w\|^2 - \sum_{i=1}^{L} \alpha_i[y_i(x_i \cdot w + b) - 1] \tag{13}$$

$$\equiv \frac{1}{2}\|w\|^2 - \sum_{i=1}^{L} \alpha_i\, y_i(x_i \cdot w + b) + \sum_{i=1}^{L} \alpha_i \tag{14}$$

In order to find the w and b, which maximise, and the $\boldsymbol{\alpha}$, which maximises (while keeping $\alpha_i \geq 0 \ \forall_i$) Fletcher (2009:3) proposes that it can be done by differentiating L_p with respect to w and b and setting the derivatives to zero:

$$\frac{\partial L_p}{\partial\, w} = 0 => w = \sum_{i=1}^{L} \alpha_i\, y_i x_i \tag{15}$$

$$\frac{\partial L_p}{\partial\, \mathrm{b}} = 0 => \sum_{i=1}^{L} \alpha_i\, y_i = 0 \tag{16}$$

By substituting the above two equations into:

$$L_p \equiv \frac{1}{2}\|w\|^2 - \sum_{i=1}^{L} \alpha_i\, y_i(x_i \cdot w + b) + \sum_{i=1}^{L} \alpha_i \tag{17}$$

This gives a new equation, which being dependent on $\boldsymbol{\alpha}$, needs to be maximised:

$$L_D \equiv \sum_{i=1}^{L} \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i\alpha_j\, y_i y_j \boldsymbol{x}_i \cdot \boldsymbol{x}_j \qquad \text{s.t. } \alpha_i \geq 0 \ \forall_i \ \sum_{i=1}^{L} \alpha_i\, y_i = 0 \tag{18}$$

$$\equiv \sum_{i=1}^{L} \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i H_{ij} \alpha_i \qquad (19)$$

where: $H_{ij} \equiv y_i y_j \boldsymbol{x}_i \cdot \boldsymbol{x}_j$

$$\equiv \sum_{i=1}^{L} \alpha_i - \frac{1}{2} \boldsymbol{\alpha}^T \boldsymbol{H\alpha}$$

$$\text{s.t. } \alpha_i \geq 0 \ \forall_i \sum_{i=1}^{L} \alpha_i y_i = 0 \qquad (20)$$

L_D is referred to as the dual form of the primary L_p. According to Fletcher (2009:4) the dual form requires only the dot product of each input vector x_i to be calculated.

Having moved from minimising L_p to maximising L_D, the following must be established:

$$\max_{\alpha} \left[\sum_{i=1}^{L} \alpha_i - \frac{1}{2} \boldsymbol{\alpha^T H \alpha} \right] \qquad \text{s.t. } \alpha_i \geq 0 \ \forall_i \qquad (21)$$

$$\text{and } \sum_{i=1}^{L} \alpha_i y_i = 0 \qquad (22)$$

This is a convex quadratic optimisation problem, and a quadratic optimisation solver is run, which will return $\boldsymbol{\alpha}$ and from $\frac{\partial L_p}{\partial w} = 0 \Rightarrow w = \sum_{i=1}^{L} \alpha_i y_i x_i$ will result in w.

b remains to be calculated.

Any data point satisfying the equation $\frac{\partial L_p}{\partial \mathrm{b}} = 0 \Rightarrow \sum_{i=1}^{L} \alpha_i y_i = 0$, which is a support vector x_s, will have the form:

$$y_s(x_s \cdot w + b) = 1 \qquad (23)$$

Substituting in $\frac{\partial L_p}{\partial w} = 0 \Rightarrow w = \sum_{i=1}^{L} \alpha_i y_i x_i$: (24)

$$y_s \left(\sum_{m \in S} \alpha_m y_m x_m \cdot x_s + b \right) = y_s \qquad (25)$$

$$b = y_s - \sum_{m \in S} \alpha_m y_m x_m \cdot x_s \qquad (26)$$

Instead of using an arbitrary support vector x_s, it is better to take an average over all of the support vectors in S:

$$b = \frac{1}{Ns} \sum_{m \in S} \left(y_s - \sum_{m \in S} \alpha_m y_m x_m \cdot x_s \right) \tag{27}$$

The result is the variables w and b, which define the separating hyperplane's optimal orientation and hence the support vector machine.

6.5 CONCLUSION

This chapter dealt with the development of the super vector machine model as a method to provide a more accurate and timeous financial distress prediction.

A number of studies, for example, those of Yang (2003), Min and Lee. (2004) and Shin *et al*. (2005), criticise statistical and artificial neural network models as reliable predictors of financial distress.

Based on constraints identified and experienced with statistical, and more recently, artificial neural network models, Cortes and Vapnik (1995) established the foundation of support vector machines. The support vector machine model can be categorised as one of the models within the broad domain of artificial intelligence and machine learning models.

Various subsequent studies, for example, by Henchiri *et al.* (2009), Gorgani *et al.* (2010) and Őnder (2010), tested the support vector machine model and found the accuracy ratio to be superior to statistical and artificial neural network models.

Although most studies consulted report superior results, Chen and Hsiao (2008: 1154) and Őnder (2010:41) highlight a number of factors that can affect the performance of the support vector machine. The classification performance of the support vector machine can potentially be improved by fine-tuning these factors.

Several enhancements were subsequently proposed to improve generalisation with the support vector machine. Ribeiro *et al.* (2010) proposed the support vector machine plus as a new paradigm, while still upholding the main principles of the support vector machine. This model incorporated privileged information. Chaudhuri and De (2011) developed the fuzzy support vector machine and Chen (2011a) introduced a particle swarm-inspired optimisation technique to the support vector machine.

In the South African context, Moepya *et al.* (2014) tested three variants of the support vector machine and found that, considering local variables, the simple linear support vector machine was sufficiently robust, forming the basis for the current study.

CHAPTER 7

RESEARCH DESIGN AND METHODOLOGY

7.1 INTRODUCTION

Two principal inferences can be drawn from the literature review. Firstly, artificial intelligence models generally outperform statistical models in predicting financial distress. Secondly, it is evident that quantitative research investigating whether the combination of financial and quantitative non-financial variables in an artificial intelligence model contributes to more accurate and timeous financial health and distress prediction is still in its infancy in South Africa.

7.2 RESEARCH DESIGN

7.2.1 Introduction

This section commences with an analysis of the study's theoretical foundation established in the previous chapters. The research design and methodology chosen for the study deal with the research problem, which emphasises the recognition of dynamic variables in financial distress prediction.

Research design and research methodology should be differentiated. Research design consists of the overall approach to be followed in testing the research hypotheses (Hofstee, 2006:113). Research methodology deals with the details of the research process to be followed, namely the research instruments, data collection and analysis.

Various studies over a number of years have attempted to establish whether a company is financially distressed or non-distressed, or will become financially distressed or non-distressed in the foreseeable future. These studies primarily used models based on financial variables. The primary source of these financial variables was the historical financial results of the company. Because these financial results are of a historical nature, the determination whether a company in its operating environment will become financially distressed or non-distressed in future becomes problematic and complex.

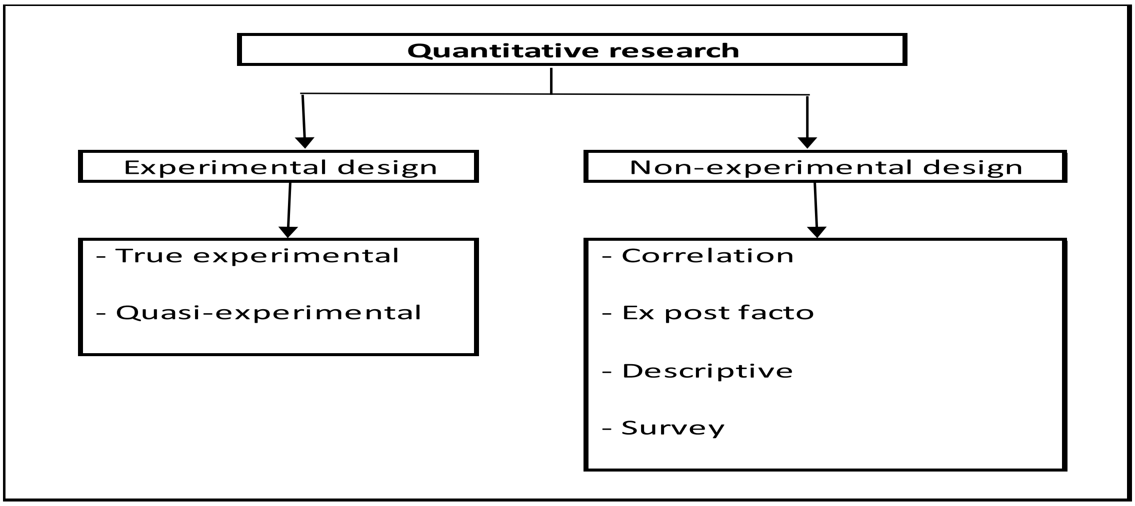
Using historical financial information is not criticised because it forms the foundation of the study. Non-financial variables, whether of a macro-, microeconomic and/or strategic nature, should also be incorporated into a model to determine where a company is positioned on the financial distress continuum.

In order to solve the research problem and meet the study objectives, the analysis and interpretation of financial and quantitative non-financial variables are relied upon. In achieving this objective, a quantitative research design approach is followed. Maree (2012:82) describes *quantitative research* as the application of empirical methods to explain phenomena by collecting numerical data, typically to be analysed by way of a statistical method. The key term in this description is "explain". Saunders, Lewis and Thornhill (2009:414) describe the term *explain* allowing the researcher to explore, present, describe and examine relationships and trends within the data sample.

7.2.2 Types of quantitative research designs

Maree (2012:82) differentiates between two types of quantitative research designs, illustrated in Figure 7.1.

Figure 7.1: Types of quantitative research designs



Source: Adapted from Maree (2012:82).

The main difference between experimental and non-experimental research design is that experimental research uses scientific methods to establish the cause-effect relationship among variables. The purpose of experimental research is to study causal links between variables; where an independent variable is manipulated to determine the effect on a dependent variable. Non-experimental research lacks the manipulation of an independent variable.

The distinction between experimental and non-experimental approaches is regarded as important. The reason for this is that, while the experimental research approach provides strong evidence that changes in an independent variable cause differences in a dependent variable, and non-experimental approaches cannot.

Maree (2012:84-85) states that a true experimental approach is characterised by the following:

- a comparison of two or more groups or sets of conditions;
- a design that allows maximum control of extraneous variables;
- the use of inferential statistics;
- the direct manipulation of the independent variable;
- the measurement of the dependent variable; and
- the random assignment of participants to treatment groups.

The choice between experimental and non-experimental approaches is subject to the nature of the research question. If it involves a causal relationship between variables and if the independent variable can be manipulated, the experimental approach would be preferred; otherwise the non-experimental approach would be suitable.

Maree (2012:86-88) identifies four types of non-experimental research designs, namely correlation, *ex post facto*, descriptive and survey designs. Firstly, the correlation design consists of the exploration of relationships between two or more phenomena and making predictions. The correlation design allows for the measurement of interrelationships among several variables simultaneously. Secondly, an *ex post facto* research design is used to explore causal relationships between variables that cannot be manipulated. Thirdly, a descriptive design describes an existing phenomenon by using numbers to characterise individuals or groups. Variables can also not be manipulated, but rather measure factors of interest. Lastly, in survey design, a sample of participants is selected on which a test or questionnaire is tested.

7.2.3 Quasi-experimental design

Based on the overview of research designs above, a quasi-experimental quantitative research approach is followed in this study, with the aim to prove the null hypothesis.

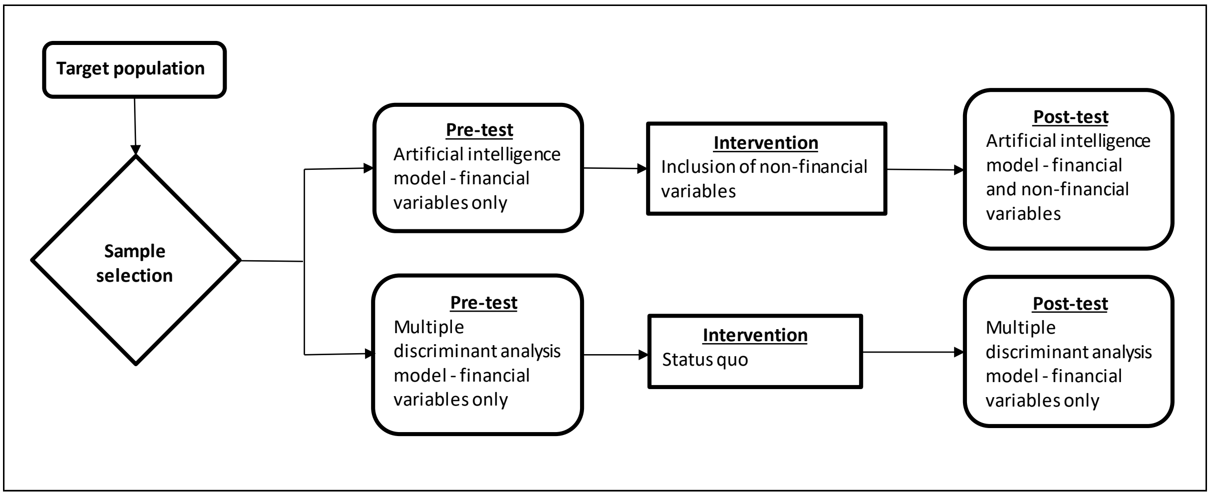
A quasi-experimental approach differs from an experimental approach in that no manipulation of an independent variable takes place. The most common form of a quasi-experimental approach consists of a pre-test post-test non-equivalent control group design (Maree, 2012:114).

A pre-test post-test approach allows for the application with or without a control group. By including a control group in the test more certainty in the interpretation of results will be obtained that the differences in the pre-test and post-test are not causally related to the intervention.

To establish the true effect of a particular intervention, a test as well as a control group is required. The test group receives the intervention and the status quo is maintained with the control group.

The quasi-experimental design adopted in this study is illustrated in Figure 7.2.

Figure 7.2: Quasi-experimental design



Source: Own compilation.

A pre-test of the test group and control group is conducted on the sample selection. An artificial intelligence model and a statistical model (multi-discriminant analysis model) are applied as test and control group respectively. Each group uses financial ratios as input variable, generating a specific result, whether the sample company is expected to be in financial distress or not.

In the intervention stage, quantitative non-financial variables are combined with the financial variables and a post-test is conducted. There will be no intervention with the control group. The pre-test and post-test results of the test group will be compared with that of the control group to determine whether:

- the pre-test results of the artificial intelligence model outperform the pre-test results of the statistical model;
- the post-test results of the artificial intelligence model outperforms the pre-test results of the same model; and

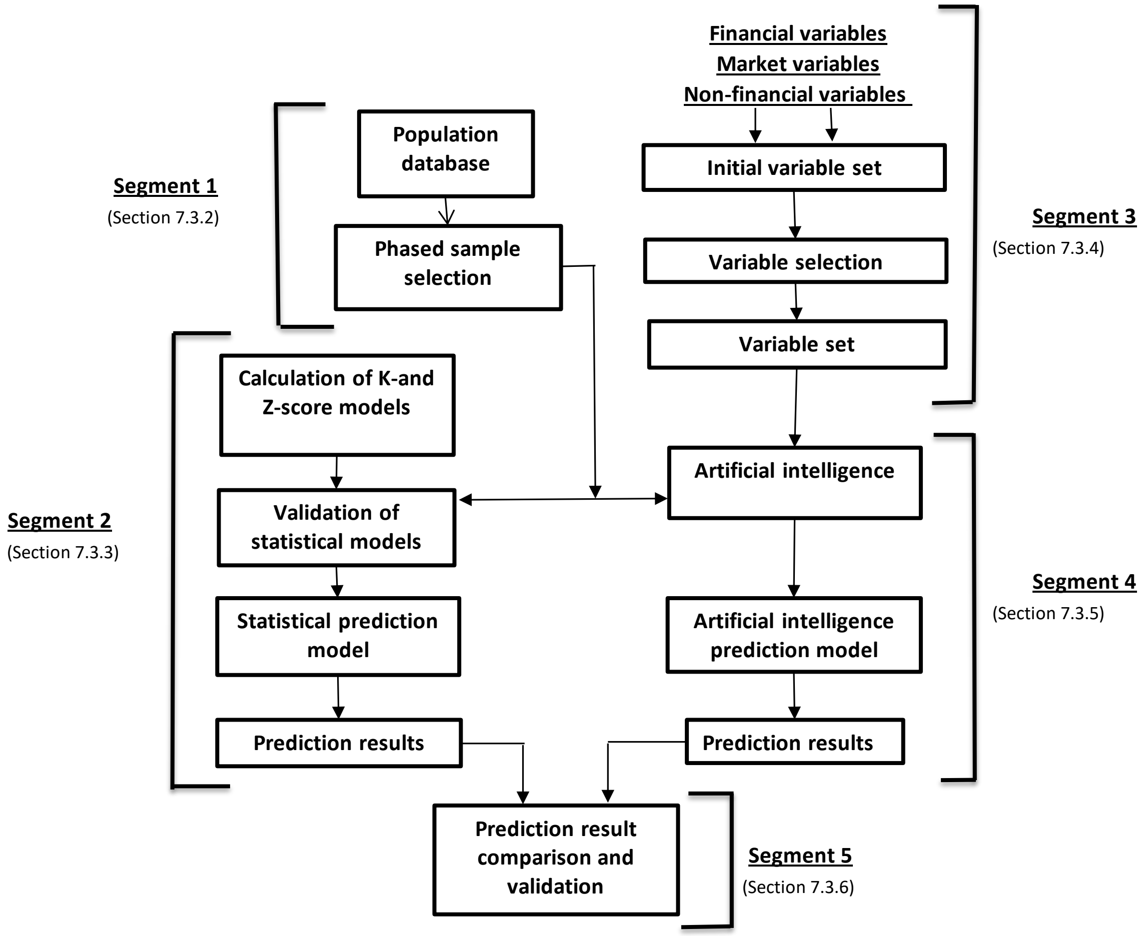
- the post-test results of the artificial intelligence model outperform the pre-test results of the statistical model.

7.3 METHODOLOGY

7.3.1 Background

The empirical process followed in the study is illustrated in Figure 7.3. The process consists of the identification and selection of database companies. This selected sample will be used as subjects in both the statistical and proposed artificial intelligence models. This is followed by the testing and validation of the statistical models based on financial variables. The subsequent step in the research process is the development of the artificial intelligence model and testing thereof on a combination of financial, market and quantitative non-financial variables. Finally, the test results of the statistical model and proposed artificial intelligence model are compared in order to determine the prediction accuracy and testing of the research hypotheses.

Figure 7.3: Empirical research process



Statistical and proposed artificial intelligence model
Segment 1: Sample identification and selection

Statistical model
Segment 2: Calculation, testing and validation of statistical models

Proposed artificial intelligence model
Segment 3: Identification and selection of financial, market and macroeconomic variables

Proposed artificial intelligence model
Segment 4: Development of the proposed artificial intelligence financial distress prediction model

Statistical and proposed artificial intelligence model
Segment 5: Validation of the proposed artificial intelligence financial distress prediction model

Source: Own compilation.

The empirical research process illustrated in Figure 7.3 is divided into five broad segments:

- **Segment 1** (Section 7.3.2) **-** the first step consists of the sample identification and selection through a phased selection process and the second step, the application of random sampling to reduce the target sample to a more manageable size.
- **Segment 2** (Section 7.3.3) - consists of the calculation, testing and validating of the Altman Z-score and De la Rey K-score statistical models in order to determine its prediction accuracy. The statistical model providing the most accurate prediction results based on South African data will be compared with the proposed artificial intelligence model.
- **Segment 3** (Section 7.3.4) **-** consists of the identification and selection of financial, market and quantitative non-financial variables to be applied to the development of the proposed artificial intelligence model in the study. No unified theory for an appropriate variable selection methodology could be identified from a review of empirical studies for a 13-year period. This review process resulted in the identification of an unfeasible sized variable set and a selection process had to be applied to identify the most often-applied variables. The final step in the variable identification and selection process will be a further reduction or narrowing down of the initial variable set through a principal component analysis process to select a final variable set. This final variable set is expected to be most feasible for application in the current study and will cover all characteristics for evaluating company performance, namely gearing, liquidity, profitability, efficiency and cash flow.
- **Segment 4** (Section 7.3.5) - consists of the development of the artificial intelligence model, where the model will be trained and tested on various combinations of financial, market and quantifiable non-financial variables selected in Segment 3.
- **Segment 5** (Section 7.3.6) - in the research process, illustrated in Figure 7.3, will be to compare and validate the test results of the statistical model with those of the artificial intelligence model to determine prediction accuracy and testing of the research hypotheses.

Each of the segments in the empirical research process is discussed in detail in the following section.

7.3.2 Sample identification and selection – Segment 1

The study sample is based on South African companies listed on the JSE. As at 12 January 2015, a total of 358 companies, categorised into 12 main sectors and 38 sub-sectors, were listed on the JSE.

Both primary data and secondary data were obtained from the INET BFA database. INET BFA is the pre-eminent provider of stock market, fundamental research data and news to the financial sector and the corporate market at large. Data provisioning is made available *via* web-based research and real-time delivery products as well as customised data sets for input into client-side systems, websites, print media and displays such as plasma screens.

Secondary data for the purposes of identifying the various sample companies according to its classification on the financial distress continuum – viable, in transition or in the distress category – have been identified by way of the K-score result, available from the INET BFA database.

The source of primary data will be the standardised historical financial statements of sample companies extracted from the INET BFA database. The ratios for the K-score model will be calculated separately based on the standardised historical financial statements of the sample companies. The result of each of these ratios will be used as input for the proposed artificial intelligence model and combined with various quantitative non-financial variables

A number of companies listed on the JSE such as mining and financial companies are deemed not suitable or eligible for consideration for this study due to, for example, differences in accounting convention and methodologies applied to the analysis and interpretation of financial results.

In order to identify the target population from this population, a filtering process is followed. The process is expected to result in a target population where each entity will have an equal opportunity of being included in the final sample. The filter approach will eliminate those companies not regarded as suitable for the study, and is described as follows:

- Filter 1: all companies listed within the Basic Materials (Industrial Metals sector and Mining sector), Oil and Gas (whole sector), Financials sectors (whole sector) are excluded.
- Filter 2: all companies with shares in a suspended status for one or the other reason are eliminated.
- Filter 3: only companies with its primary listing on the JSE will be considered for inclusion in the test sample.
- Filter 4: only companies listed for the last 10 years or listed prior to 2005 will be retained in the potential test sample. The purpose of this criterion is to consider the most recent accounting period, which reflects the current economic environment and changes that have taken place in accounting statements, which, in turn, have also changed certain financial requirements that would have had a serious impact on the company.

Table 7.1 provides a summary and reflects the number of entities listed on the JSE per sector and the target population identified per sector.

Table 7.1: Summary of the target population of companies listed on the JSE

Sector	Total number of companies per sector listed on the JSE	Number of companies per sector in the target population	Target population as % of total number of companies listed on the JSE
AltX	47	6	12.77
Basic materials	64	7	10.94
Consumer services	37	30	81.08
Consumer	23	15	65.22
Development capital	1	0	0.00
Financials	92	0	0.00
Health	8	4	50.00
Industrials	64	38	59.38
Oil and gas	5	0	0.00
Technology	11	9	81.82
Telecommunication	4	2	50.00
Venture capital	2	1	50.00
TOTAL	**358**	**112**	**31.28**

Source: Own compilation.

Based on the four-phase filtering process, a total of 112 companies were extracted, representing 31.28% of the 358 companies listed within the 12 main and 38 sub-sectors of the JSE.

Determining the financial health of 112 companies over a 10-year period would require substantial time and effort and may not be feasible for this study. A manageable sample is therefore selected in a manner to be representative of the target population. Based on a 95% confidence level and 5% confidence interval and a standard deviation of 0.5, the appropriate sample size is calculated to be 87 companies. A final sample size of 87 companies represents 77.68% of the target population.

To ensure that selection bias is minimized, each of the 112 companies in the target population (numbered from 1 to 112 in the first column) should have an equal chance of being selected, and each company should be selected independent from the other. To ensure compliance to this requirement, an electronically generated random number method was applied. The random number function of a Hewlett•Packard 12c® financial calculator was used to generate random numbers for this purpose.

Random numbers from 1 to 87 were generated and assigned to each of the 112 target population companies. The Excel® spreadsheet column with random numbers (second column) were then sorted in an ascending order from 1 to 112. The first 87 companies in this column were then selected as the sample representative of the target population.

The final sample of 87 companies, based on the selection process described above, is summarised in Table 7.2.

Table 7.2: Summary of sample companies selected from the target population

Sector	Number of companies per sector in the target population	Number of companies per sector in the sample	Sample number of companies as % of target population
AltX	6	6	100.00
Basic materials	7	5	71.43
Consumer services	30	23	76.67
Consumer	15	13	86.67
Health	4	3	75.00
Industrials	38	28	73.68
Technology	9	8	88.89
Telecommunication	2	1	50.00
Venture capital	1	0	0.00
TOTAL	**112**	**87**	**77.68**

Source: Own compilation.

AfroCentric Investment Corporation Ltd in the Health Equipment & Services sector of the JSE was included in the target population by error. Although the original company W.B. Holdings Ltd was listed in 1988, AfroCentric Investment Corporation Ltd reverse listed into the original company in 2006. Based on the phased elimination process described above, AfroCentric Investment Corporation Ltd has a financial history of seven years and not the required minimum of 10 years; it should not have been included in the target population. The error was discovered at the sample selection stage, and this company is therefore excluded from the final sample.

AfroCentric Investment Corporation Ltd was retained in the target population, but excluded from the final sample. With 86 companies remaining in the sample, the number of sample entities as a percentage of the target population reduced from 77.68% to 76.79% (see Table 7.3).

Table 7.3: Sample companies selected from the target population – excluding AfroCentric Investment Corporation Ltd.

Sector	Number of companies per sector in the target population	Number of companies per sector in the sample	Sample number of companies as % of target population
AltX	6	6	100.00
Basic materials	7	5	71.43
Consumer services	30	23	76.67
Consumer	15	13	86.67
Health	4	2	50.00
Industrials	38	28	73.68
Technology	9	8	88.89
Telecommunication	2	1	50.00
Venture capital	1	0	0.00
TOTAL	**112**	**86**	**76.79**

Source: Own compilation.

Should the company AfroCentric Investment Corporation Ltd be excluded from both the target population and the sample, it would result in the required sample size reducing to 86 companies. The 86 companies represent 76.79% of the target population. The effect of this error is deemed negligible.

In the sample selection process three areas were identified to refine the selected sample, namely: the identification of financially distressed and non-distressed companies, the identification and treatment of outliers and the identification and treatment of missing data values. Each of these areas are discussed below.

(i) Identification of financially distressed and non-distressed sample companies

A number of studies were reviewed in order to determine the methodology followed in the separation of financially distressed and non-distressed companies. It was established that only a limited number of studies disclosing the methodology followed.

Those studies that disclosed the methodology predominantly relied on information readily available and disclosed on either an applicable stock exchange, a database or defined by a company act. The remainder of studies, which disclosed a separation methodology, applied a user-defined proxy to differentiate between financially distressed and healthy companies.

Table 7.4 summarises those studies that used readily available information on financially distressed and healthy companies

Table 7.4: Data source and identification of financially distressed companies

Study	Data source	Definition of financial distress applied
Argyrou (2006:31-32)	Voitto data base and Finish Company's Act	Shareholders' equity less than 40% share capital
Zheng et al (2007:3)	Chinese Stock Exchange	Special Treatment "ST" notation - company reported sequential losses over a two year period
Gorgani et al (2010:223)	Teheran Stock Exchange (Iran), Accounting Research Database and Iran Trade Law	Special Treatment "ST" notation - accumulated loss more than twice shareholders' equity
Yanqing et al (2010:374)	Shanghai and Shenzen A-share markets	Special Treatment "ST" notation - company reported sequential losses over a two year period
Divsalar et al (2011:215)	Teheran Stock Exchange and Iran Trade Law	Special Treatment "ST" notation - accumulated loss more than twice shareholders' equity
Kim (2011:50)	Australian Stock Exchange, Securities Industry Research Centre of Asia-Pacific and Datastream	Companies that failed to pay its annual listing fee or appointment of a liquidator or administrator
Zhou & Lai (2012:15097)	Compustat North America, Wharton Data Service	Full-value delivery, stock transaction suspended, re-constrution, bankruptcy,withdrawal from the exchange
Lin et al (2013:87)	Taiwan Economic Journal (TEJ) database	Full-value delivery, stock transaction suspended, re-constrution, bankruptcy,withdrawal from the exchange
Tinoco et al (2013:397)	London Share Price Database (LSPD)	Suspension or cancellation of shares, liquidation, or receivership, cancellation or suspension of shares
Zhou (2013:19)	Compustat North America, Wharton Data Service	Status indicated as "bankrupt" or "non-bankrupt"
Bauer et al (2014:435)	Accounting data from Company Analysis, Exstat and Datastream. Data on failed companies from Fame (Bureau van Dijk) and London Business School Library	Liquidation, administration/receivership, valueless
Bredart (2014:3)	Bloomberg	File for bankruptcy
Xu et al (2014:62)	Shenzhen Stock Exchange & Shanghai Stock Exchange in China	Special Treatment "ST" notation - company reported sequential losses over a two year period or financial misstatements
Bao et al (2015:297)	Chinese Stock Exchange	Special Treatment "ST" notation - company reported sequential losses over a two year period or financial misstatements
Khademolqorani et al (2015:5)	Teheran Stock Exchange (Iran) and Iran Trade Law	Retained losses more than 50% of capital
Nagaraj et al (2015:3)	UCI Machine Learning Repository	Bankrupt and non-bankrupt
Sun et al (2015:12)	Chinese Stock Exchange	Special Treatment "ST" notation - company reported sequential losses over a two year period or financial misstatements
Zhou et al (2015:55)	Chinese Stock Exchange	Special Treatment "ST" notation - company reported sequential losses over a two year period or financial misstatements

Source: Own compilation.

From the review conducted on available studies, the availability of information on financially distressed companies is predominantly driven by regulatory requirements. For example, on the Chinese stock exchanges, a company is identified as financially distressed by way of an indicator, such as “ST” or “Special Treatment” or a specific financial variable. In other instances, companies filing for bankruptcy have an obligation to file final financial statements and are conveniently identified as "bankrupt" on an official publically available database. It is evident that regulatory requirements simplify the differentiation between financially healthy and distressed companies.

In instances where information on financially distressed companies was unavailable or not published, user-defined financial variables or methodologies were applied to identify and differentiate between financially distressed and healthy companies. A summary of these studies is depicted in Table 7.5.

Table 7.5: User-defined identification of financially distressed companies

Study	Definition of financial distress applied
Kidane (2004:16)	Delisted companies due to failure
Lam (2004:571)	Rate of return on shareholders' equity
Ooghe & Spaenjers (2005:11)	Request for judicial composition, temporary postponement of payments, bankruptcy
Cheng, Chen & Fu (2006:585)	Relied on public announcements regarding financial distress
Naidoo (2006:65)	Negative earnings
Muller (2008:43)	Relied on cash flow related information
Kim et al (2014:10)	Delisted companies due to non-payment of annual listing fees or appointment of liquidator/administrator

Source: Own compilation.

It is evident from the review of available literature that a user-defined process of identification of distressed companies is more complex. Where no specific information is published to identify financially distressed or bankrupt companies, a more complex process has to be followed. This would require a more in-depth investigation. For example, the fact that a company was delisted, does not necessarily imply it was the result of bankruptcy. The delisting could have been the result of a merger or acquisition. Should a researcher use a delisting as a proxy for the identification of distressed or bankrupt company, further investigation would be required to establish the actual reason for delisting and to identify the actual distressed or bankrupt company.

Some of the user-defined proxies for financially distressed companies, identified in Table 7.5, are probably questionable or debateable. However, where this information is simply not available, such as in the South African context a researcher has to improvise. However, to improvise, it is of importance that the methodology has to be simplistic and empirically justifiable.

From Table 7.5, Lam (2004) applied the return on equity (ROE) as a non-complex and empirically justifiable proxy for identification of financial performance prediction. The performance of neural networks was compared with the average return from the top one-third returns in the market (maximum benchmark) that approximate the return from perfect information as well as with the overall market average return (minimum benchmark).

In the current study, the process is adapted where the return on equity is used as a proxy for identification of financially distressed companies. Instead of focussing on the top one-third returns, the focus is on the bottom 50%, representing the poor performing companies. These poor performing companies serve as a proxy for financially distressed companies.

Return on equity (ROE) is a measure of a company's profitability by revealing the quantum of profit a company generates with funds invested by shareholders. In its simplistic format, this financial variable can be expressed as the net income divided by shareholders' equity. Reliance is placed on a refined version of the return on equity obtained from the INET BFA database. The inflation-adjusted return on equity is calculated by subtracting inflation-adjusted depreciation on depreciable fixed assets from net income and dividing it by the average shareholders' equity (Period T and Period T–1) after adding the additional value from inflation-adjusted depreciable fixed assets for each period.

The inflation-adjusted return on average shareholders' equity is therefore extracted from the INET BFA database for each of the selected sample companies over a 10-year period from 2005 to 2014. Results highlight the occurrence of outliers in some instances, and require further investigation.

(ii) Identification and treatment of outliers

Ghosh and Vogt (2012:3455-3456) describe an *outlier* as an observation far away (positive or negative) from the central point (median) or other observations. The rule of thumb is that an outlier is any observation whose removal from a sample would result in a change in the estimate of a parameter of interest by 10% or more.

The cause of outliers can be ascribed to either a measurement error or actual value. In this instance, the outlier can be ascribed to the latter, i.e. a drastic change in shareholders' equity (for example, additional ordinary shares issued, change in cost of control after an acquisition, change in non-distributable reserves and/or change in distributable reserves) and/or change in net income.

The first step in identifying both negative and positive outliers is to calculate the statistical midpoint of the range. The range (return on equity results for each company over the observation period) is divided into four quartiles. The first and third quartile are isolated and subtracted from each other to determine the interquartile range (IRQ). A 50% value is multiplied with the interquartile range (IQR*1.5) to determine a reasonable upper and lower fence:

- The lower fence is equal to the first quartile – IQR*1.5.
- The upper fence is equal to the third quartile + IQR*1.5.

Any value above or below the upper and lower fence is then identified as an outlier.

Ghosh and Vogt (2012) identified and evaluated the following three typical treatments of outliers:

- Keep the outlier and treat it like any other data point.
- Winsorise it (assign a lesser weight or modify the value so it is closer to the other range values.
- Eliminate the outlier.

However, irrespective of the treatment applied, the risk of Type I or Type II error cannot be excluded. The added difficulty is that there is no unanimously accepted theoretical framework for the treatment of outliers (Cousineau & Chartier, 2010:59).

Because there is no specific recommendation in the reviewed literature regarding the treatment strategy to follow, and based on the result of the quartile calculation and interquartile range, each outlier is assessed and treated by considering either of the first two treatments suggested by Ghosh and Vogt (2012). In order to minimise the risk of Type I or Type II errors, only extreme outliers are adjusted. The remainder of the outliers are retained and treated like any other data point.

Table 7.6 summarises the extreme outlier results and the reasons for being identified as an outlier. The identified extreme outliers are replaced with the average of inflation-adjusted return on equity value either side of the extreme value, and is presented in the last column of Table 7.6.

Table 7.6: Unadjusted and adjusted extreme outlier values for the inflation-adjusted return on equity

Company name	Short name	Code	Date of outlier value	Unadjusted outlier value: Inflation-adjusted ROE (%)	Reason	Adjusted outlier value: Inflation-adjusted ROE (%)
African Media Entertainment Ltd	AME	AME	2005	1 002.79	Negative non-distributable reserve	20.10
Famous Brands Ltd	FAMBRANDS	FBR	2011	6 629.27	Reduction in non-distributable reserve and cost of control	25.32
Howden Africa Holdings Ltd	HOWDEN	HWN	2007	-6 755.62	Reduction in non-distributable reserve and distributable reserves	131.04
			2008	-9453.38	Reduction in non-distributable reserves	179.86
Netcare Ltd	NETCARE	NTC	2013	2 081.18	Increase in non-distributable and distributable reserve and reduction in cost of control	5.97
Pick 'n Pay Stores Ltd	PICKNPAY	PIK	2006	-5 027.55	Reduction in distributable reserves and increase in cost of control	-43.51
			2009	-2 024.35	Reduction in non-distributable reserves and increase in distributable reserves and net income	-6.00

Source: Compiled from INET BFA database.

The treatment of the Howden Africa Holdings Ltd. extreme outlier, posed a problem because it occurred in two consecutive years (2007 and 2008). To overcome this problem, the outlier value of 2007 is adjusted by calculating the average of the values in 2006 and 2009, and the outlier value of 2008 is treated similarly by calculating the average of the values in 2007 and 2009.

Following the treatment of extreme outliers through the Winsorise method proposed by Ghosh and Vogt (2012), the mean of the range for a particular company is calculated. By applying an adaptation of the Lam (2004) method described above, the top 50% performers are separated from the bottom 50% poor performers. The reason for identifying the top and poor performers on this basis and not top and bottom third as in the Lam (2004) study is to achieve a balanced match between top and poor performers. In the validation subset the same process is followed with the separation of top and poor performers. The 50% poor performers serve as a proxy for financially distressed companies.

The testing and validation subsets are separated into equal and balanced financially distressed and non-distressed companies. However, the testing and validation subsets comprise a total number of 71 and 17 companies respectively. The 50% separation methodology applied results in the healthy company category holding one sample company more than the financially distressed category in both the testing and validation subset. To adjust the unbalanced subsets, a randomly selected company in the non-distressed company category (PPC Ltd) of the testing subset is transferred to the non-distressed company category of the validation subset. The poorest performer in the healthy category (Super Group Ltd) of the validation subset is transferred to the financially distressed category in the validation subset. The risk is that this adjustment could potentially result in either a Type I or Type II error.

(iii) Identification and treatment of missing data values

An additional anomaly was identified from the inflation-adjusted return on equity values obtained for the INET BFA database. A number of companies reflected missing data values during the observation period.

Table 7.7 displays the companies, reflecting missing values and reasons for the missing values.

Table 7.7: Missing values and reasons for missing values

Company name	Short name	Code	Date of missing value	Reason
Adcorp Holdings Ltd	ADCORP	ADR	2007	Change in financial year-end from December to March due to a significant change in group structure and alignment with individuals' tax year-end.
African Media Entertainment Ltd	AME	AME	2008	Change in financial year-end from October to March. No reason stated.
Awethu Breweries Ltd	AWETHU	AWT	2014	Late submission of financial statements.
KAP Industrial Holdings Ltd	KAP	KAP	2006	Change in financial year-end from December to June to synchronise its financial year-end with that of a significant shareholder.
MICROmega Holdings Ltd	MICROMEGA	MMG	2013	Change in financial year-end from December to March. No reason stated.
Primeserv Group Ltd	PRIMESERV	PMV	2010	Change in financial year-end from December to March to align financial reporting with underlying operating activities
Tongaat Hulett Ltd	TONGAAT	TON	2009	Change in financial year-end from December to March to align financial year-end with the sugar season in all operating countries.

Source: Compiled from INET BFA database.

Batista and Monard (n.d.:1) state that treatment of missing data should be carefully assessed otherwise bias might be introduced into the knowledge induced. In considering the treatment of missing data Batista and Monard (n.d.:12) caution against the imputation of mean or mode and internal algorithms, or even the most advanced imputation method.

In Table 7.7, all missing values relate to a change in financial year-end for one or the other legitimate reason, except those of Awethu Breweries Ltd. The inflation-adjusted return on equity for Awethu Breweries Ltd was not available due to the late submission of audited financial statements.

Subsequent to the selection of the final sample it became apparent that Awethu Breweries Ltd failed to submit its audited financial statements and was delisted from the JSE. Due to this company not complying with the minimum selection criterion of a 10-year financial history, it had to be eliminated from the final sample. The final sample for the purposes of the study therefore reduced from 86 to 85 companies. The final sample now reduces from the original 77.68% to 75.89% of the target population, which is regarded as acceptable.

With a change in financial year-end results, the new financial year subsequent to the year in which the financial year-end is changed includes the first 12 months plus the number of months in the new financial year. The year in which the financial year-end has changed should in fact not reflect as a missing value, because the value is incorporated into the subsequent financial year-end.

7.3.3 Calculation, testing and validation of statistical models – Segment 2

Prior studies have done various validation tests, such as an out-of-sample-period *ex ante* (forecast test) and the Lachenbruch jackknife procedure (Charitou *et al.,* 2004:487). The Lachenbruch jackknife procedure is useful in dealing with relative small samples and provides an almost unbiased estimate of a misclassification rate. The overall correct prediction over a one-, two- and three-year period prior to financial distress in the Charitou *et al.* (2004) study resulted in an 82%, 72% and 70% correct prediction respectively, and differs not much from the forecast classification rates of the logistic regression model tested in the study. These results were all statistically significant at the 1% level, indicating a strong association between the observed and predicted groups of companies.

In the Koh and Low (2004:473), study the model classification accuracy was 95.22%, 95.65% and 97.39% for the logistic regression, neural network model and decision tree models respectively. However, the same sample data was used for model construction and testing, which resulted in an upward bias. By providing an unbiased assessment, all three models were tested on a separate validation sample and resulted in a 95%, 94% and 91% accuracy rate for the decision tree, logistic regression and neural network model respectively. Based on these results, the decision tree model was selected as the best performing or accurate prediction model.

The Berg (2007:134-135) study results concurred with those of the Koh and Low (2004) study, namely to avoid embedding unwanted sample dependency, quantitative models should be validated on a different sample set from the sample set used for model building. Berg (2007) considered a power curve to visually compare the predictive performance of the various models in his study. An accuracy ratio (AR) was used as a single measure that summarised the predictive accuracy of each risk measure into a single statistic. This metric was obtained by comparing the power curve of a random financial distress prediction model under investigation with that of a perfect model. The closer the power curve was to the perfect power curve, the better the model performed. The results indicated that the linear discriminant analysis model (LDA), generalised linear model (GLM) and neural network model (NN) did not differ significantly. The generalised additive models (GAM) significantly outperformed all other models at all levels of risk.

In a subsequent study, Lensberg *et al.* (2006:692) compared a genetic programming model with two logistical regression models to determine the genetic programming model's financial distress predicting effectiveness. The logistic regression models, as a more standardised methodology, were 77% and 76% accurate in predicting financial distress on the training and testing samples respectively. This was compared with the 82% and 81% accuracy of the genetic programming model and the same training and testing samples.

Various validation tests were done within the South African context. Kidane (2004:121) tested the practical prediction ability of two multiple discriminant analysis models, namely the Altman Z-score and Springate financial distress prediction models. The binomial statistical technique was used to conduct the comparative test up to five years prior to distress. Descriptive statistical techniques such as means, medians, standard deviation and frequency distribution were used to determine the classification accuracy of the two models.

Muller (2008) applied four different financial predictive models to validate and determine the predictive ability of a selection of financial variables over a one- to five-year period prior to failure:

- multiple discriminant analysis;
- recursive partitioning;
- logistic regression analysis;
- neural networks.

Muller (2004:88) concluded that the different predictive models achieved different accuracies, summarised as follows:

- The multiple discriminant analysis and recursive partitioning models provided the better information on the number of distressed companies and hence had the best normalised cost of failure (NCF) value.
- The neural network and logistic regression models provided the best overall predictive accuracy.

Arens (2014:49) validated the Altman Z-score in medical schemes by applying the following tests:

- Mann-Whitney test to compare a selection of variables and Z-scores of failed and non-failed medical schemes;
- a correlation matrix of the independent variables in relation to the Z-score;
- classification of accuracy or error rates of the Z-scores (prediction scores) in South African medical schemes (the predictions were based on data one and two years prior to distress or failure).

Over the study period 2002 to 2011, Arens (2014:79) achieved an average accuracy and error rate of 82% and 17.9% respectively, which is consistent with the, 84%, 88% and 85% accuracy rates for other financial distress prediction models.

A validation test based on the first two steps in the procedure followed by Arens (2014) will be applied to examine the accuracy or error rates of the Altman Z-score and the De la Rey K-score models for periods T, T-1, T-2, T-3 and T-5. T (2014) represents the base year, T-1 (2013) represents the year prior to the base year, T-2 (2012) represents the two-year period prior to the base year, T-3 (2011) represents the three-year period prior to the base year and T-5 (2009) represents the five-year period prior to the base year[10].

Each test will be conducted at a 1%, 5% and 10% confidence level. In order to avoid embedding unwanted sample dependency an unbiased assessment of the two statistical models will be conducted separately on the training and validation subsets.

(i) The Mann-Whitney U test

This is a non-parametric test on ordinal data that is used to compare two sample means that originate from the same population, and is used to test whether the two samples are equal or not. The Mann-Whitney test is based on the following assumptions:

- The sample drawn from the population is random.
- Independence in the samples and mutual independence is assumed.
- Ordinal measurement scale is assumed.

The current study satisfies the Mann-Whitney test assumptions and is therefore implemented to establish whether there is a noticeable difference between the dependent variable (inflation-adjusted return on equity) and the two independent variables (Altman Z-score and the De la Rey K-score models).

The second step in the validation process is to conduct a correlation test between the two independent variables (Altman Z-score and the De la Rey K-score models) and the dependent variable (inflation-adjusted return on equity). To evaluate and compare

[10] These test periods will be applied throughout the study.

the two statistical models, the Spearman's correlation coefficient and accuracy rate will be applied.

(ii) Spearman's correlation coefficient (rho)

This is a statistical technique to measure strength of a monotonic relationship between two sample sets. In a sample, it is denoted by r_s and is constrained by the following equation:

$$-1 \leq r_s \leq 1$$

The interpretation of this equation is for example, the closer r_s to ± 1, the stronger the monotonic relationship. Alternatively, a perfect Spearman's correlation of +1 or -1 occurs.

The Spearman's correlation coefficient test will be applied to examine the ability of the two statistical models to predict financial distress over the periods T, T-1, T-2, T-3 and T-5. Each test will be conducted at a 1%, 5% and 10% confidence level.

The final step in the validation process is to test the accuracy or error rate.

(iii) Accuracy or error rate

A diversity of techniques is available to measure forecast accuracy.

Avenhuis (2013:24) proposes a classification matrix containing numbers that reveal the predictive ability. The overall accuracy rate is the percentage of correct classifications to total classifications. The overall accuracy rate can be separated into the accuracy rate of good predicted financially distressed companies and good predicted non-distressed companies.

The performance of a model can be measured by its accuracy, which can be expressed as the degree of correctly or successfully classifying a company's position on the distress continuum. Tofallis (2015:1) defines this degree or magnitude of relative error (MRE) as the absolute value of the ratio of the error to the observed value. When this value is multiplied by 100%, it gives the absolute percentage error (APE).

In an earlier South African study, Kornik (2004) described the Type I and Type II error rates as the probability of the error conditional on the actual status (point on the financial distress continuum).

Kornik (2004:101-102) evaluated the measurement of error rates. The calculation of error rates can be interpreted as the probability of error conditional on either the actual status of the company (the number of Type I errors divided by the actual number of distressed companies in the sample) or the prediction made (the number of Type 1 errors divided by the number of predicted distressed companies). It is noted that if the probability of failure for the sample differs from that of the total population, an inference of percentage of total number of distressed companies to the population may not be meaningful.

Kornik (2004:101) highlights the problem when using the number of available statistical classification techniques where the number of predicted financially distressed companies will depend on a selected cut-off point. For example, where a cut-off point is set at 0.1, a company for which the expected probability of financial distress exceeds 10% will be forecast as a company in impending financial distress. However, a company for which the expected probability of financial distress is less than 10% is expected to be healthy. It is further noted that by simply increasing the number of companies that the model predicts as impending financial distress by reducing the cut-off probability for the particular statistical model, creates a trade-off between Type I and Type II errors.

According to Kornik (2004:101-102), a perfectly accurate model would classify 100% of the companies correct. In considering a hypothetical situation in which two sample groups (financially distressed and non-distressed) are classified as significantly different in size, a situation may arise in which the proportion of the larger group (non-distressed) is classified almost completely correctly, while only a relatively small proportion of the smaller, but more crucial financially distressed group is accurately identified. In this instance, the percentage correctly classified may not be an informative measure.

Kornik (2004:102) proposes a "weighted efficiency" (WE) measure of model performance, which is adapted for the purposes of the current study, as follows:

$$WE = \frac{BWF}{VB} * \frac{BWF}{TWF} * CC \tag{28}$$

where:

CC	= percent classified correctly
BWF	= financial distressed companies correctly identified by the model
VB	= all financially distressed companies identified by the model
TWF	= total number of financially distressed companies in the sample

WE is sensitive to both the percentage of financially distressed companies classified correctly and the percentage of those companies identified as financially distressed by the model.

7.3.4 <u>Variable identification and selection – Segment 3</u>

The next segment in the research process, as depicted in Figure 7.3, relates to the identification and selection of financial, market and quantitative non-financial variables, solely for purposes of use in the application of the proposed artificial intelligence model. The source of data for the financial variables is based on published company information obtained from the INET BFA database. Macroeconomic data and other market variables are obtained from the South African Reserve Bank and Department of Finance, Statistical Services respectively.

The following section provides a theoretical background on variable identification and selection. This section is split into three sub-sections, namely: financial, market and quantitative non-financial variables.

(i) Financial variables

According to the available literature, there is no unified theory and no reason why a certain set of variables is preferable to any other. Three broad variable identification and selection groups were identified, as follows:

- Firstly, a number of studies based variable identification and selection on those variables most widely used and deemed contributive in previous studies. (Charalambous *et al.,* 2000:405; McKee, 2003:575; Lee, 2004:14; Ooghe *et al.,* 2005:5; Lensberg *et al.,* 2006:686; Altman, Zhang & Yen, 2007:10-11; Hossari, 2007:14; Sai *et al.,* 2007:3; Dakovic, Czado & Berg, 2010:1741; Chen, 2011b:11264; Shiri, Ahangary, Vaghfi & Kholousi, 2012:411; Yap, Munuswamy & Mohamed, 2012:334-335).

- Secondly, a number of studies identified variables based on its use in previous studies and then applied one or the other statistical selection technique. Lin and McClean (2000:47-48) selected variables based on a three-step process, firstly, financial theory and human judgement; secondly, analysis of variance; and thirdly, factor analysis. Atiya (2001:932) applied a pre-screening variable identification process based on individual indicator prediction accuracy and correlation matrix, and then a subsequent cross-validation procedure to select applicable variables.

  Charitou *et al.* (2004:474-478) selected the most significant variables by conducting a univariate logistic regression for each ratio. In addition, the forward selection and backward elimination methods, both in SPSS, were applied. Min and Lee (2005:606) identified several variables by the principal component analysis and t-test for graphical analysis and then applied stepwise logistic regression analysis in selecting variables. Wu, Fang and Goo (2006:3) stated that the result of the normality test revealed that most financial variables were not normally distributed as had been stated in previous research and therefore applied the Mann-Whitney-Wilcoxon test in selecting variables.

Masten and Masten (2009:10-12) expressed the opinion that in selecting variables most previous studies used an empirical approach of variable identification followed by a stepwise procedure in variable selection in the final logistic regression or discriminant model. This procedure is not statistically rigorous and different sequencing or initial ordering of variables need not result in a unique variable selection. Masten and Masten (2009) applied two approaches with the aim to improve their data set results – firstly, a three-stage approach consisting of a bivariate logistic regression, a correlation process and a logistic stepwise procedure. Secondly, a classification and regression tree (CART) approach was used in the final variable selection process. In another study, Muller (2008:66) identified the most significant variables from previous reputable research and based the variable selection on the Kruskal Wallis test.

Jian-guang *et al.* (2010:152) are of the view that different financial variables may reflect different information in financial distress prediction and even the discriminant ability of a particular variable may change with the concept characteristic of financial distress. The neighbourhood rough set method was applied in the variable selection process. Divsalar *et al.* (2011:215) identified variables based on its popularity in previous research. The final variables were selected by means of a sequential feature selection (SFS) analysis, which is a common procedure to reduce dimensionality in data.

Gunnersen *et al.* (2012:2) identified an initial set of variables based on a review of various publications. From this initial set of variables, a non-deterministic accuracy-based feature selection method was used to eliminate variables that did not contribute to the accuracy of the classification system. Once noncontributory inputs had been removed, a heuristic accuracy-based feature selection method was used to select the optimal variable combination.

Kim and Kang (2012:9311) identified a number of variables and categorised the variables as profitability, debt coverage, leverage, capital structure, liquidity, activity and size. The final variable set was selected by assessing the performance of each variable based on receiver operating characteristic (ROC) curve analysis. The performance criterion of each variable in ROC curve analysis is represented as the value of an area under the ROC curve (AUROC), which is the probability that a classifier will rank a randomly chosen positive variable higher than a randomly chosen negative variable. If the AUROC of a variable is 1 or closer to 1, it means it is a perfect or near perfect variable in financial distress prediction.

Mukhopadhyay, Tiwari, Narsaria and Karmaker (2012:74) confirmed the importance of the choice of variables to every financial distress prediction study. The authors identified variables based on the literature and selected a set of variables based on the Wilcoxon's Rank-Sum test.

- Finally, a number of studies used the Altman (1968) Z-score model as a standard for comparison for subsequent financial distress classification studies using discriminant analysis. (Kim & Yoo, 2006:4; Wu, 2007:24; Fu-yuan, 2008:543; Appiah & Abor, 2009:436; Jing-rong & Jun, 2009:277; Rui, 2010:559; Hauser & Booth, 2011:570; Olson *et al.,* 2012:467).

Quah and Srinivasan (2005:2) state that it is difficult to prescribe any guidelines for variable selection. Too few variables can introduce bias in the modelling process and can lead to generalisation error. On the other hand, too many variables may lead to overfitting and may have substantial multicollinearity.

Argyrou (2006:54) argues that the variable selection is both problematic and pivotal. It is pivotal because variables play an important role in enhancing the internal validity of the particular study together with the scope of the study's findings. Variable selection is problematic because the absence of a unified financial distress theory has led to the proliferation of the variables used in financial distress prediction. In addition, he argues that the heterogeneity of data used in financial distress prediction models has exasperated this problem. The existence of different financial distress procedures also contributes to the problem of selecting appropriate variables.

Argyrou (2006:54) expresses the view that the problem is beyond the scope of his study and it remains open for future research to identify those variables with the highest discriminatory power. In concurrence with Argyrou's view and evident from a literature review for the purposes of the current study, no unified approach could be identified to select financial variables as a basis for financial distress prediction.

(ii) Market variables

In this study, market variables are considered for inclusion in the proposed artificial intelligence model to determine whether they enhance the model's predictive accuracy.

Market variables are based on a company's share price information and performance. The share price performance reflects an investor's perception of the company's historical performance and future prospects – cash flow and earnings and/or financial position. Irrespective of an investor's investment strategy and risk appetite, whether it is income- or capital growth-driven, there appears to be an interdependency between a company's cash flow and earnings and financial position and share price performance.

An investor might assess a company's financial results as well as other random publicly available information in determining an investment strategy. Combining financial and market variables in an investment strategy may influence the particular company's share price movement and potentially contribute to market efficiency, compared with considering financial variables in isolation (Tinoco & Wilson, 2013:400).

Although there may not be an immediate and direct relationship between a company's share price movement and the probability of financial distress, the share price is an efficient processor of all publicly available information, according to Tinoco and Wilson (2013:400). Changes in share prices could result in the re-alignment of investment portfolios, and *vice versa*, which could be an early indicator of financial distress. Inclusion of market variables in a financial distress prediction model may equally enhance its efficiency and prediction accuracy.

A review of literature for the purposes of the study reveals that limited attention has been given to determine whether the inclusion of market variables enhances the prediction accuracy of financial distress prediction models. A limited number of studies reviewed, consider market variables in its financial distress modelling (see Section 5.2.2).

Those studies that do consider market variables focus predominately on earnings per share (EPS) – (Lin & McClean, 2000:46; Lee & Chen, 2007:4; Chen & Hsiao, 2008:1152; Huang *et al.,* 2008:1036; Lieu, Lin, & Yu, 2008:1066; Duan, Huang & Wang, 2010:1963; Ahmadi, Amjadian & Pardegi, 2012:35).

A number of financial distress prediction studies relied on the Altman (1968) Z-score ratio, the market value of equity to total debt, as a proxy of a market variable (Charitou *et al.,* 2004:477; Gepp, 2005:57; Altman *et al.,* 2007:12).

An additional number of studies include one or more market variables, other than the studies quoted above. Table 7.8 provides a summary of the market variables used in studies over and above the earnings per share and market value of equity to total liabilities discussed above.

Table 7.8: Additional market variables previously used in studies

Study	Market variable used in a particular study			
Atiya (2001:932)	Price to cash flow ratio	Rate of change of share price	Rate of change of cash flow per share	Stock price volatility
Koh & Low (2004:469)	Market value of equity to total assets			
Lam (2004:571)	Earnings per share	Dividends per share	Market capitilisation	Relative strength index
Leksrisakul & Evans (2005:10)	Excess returns			
Chen & Hsiao (2008:1152)	Earnings per share	Cash flow per share		
Kim & Partington (2008:2-3)	Market to book ratio			
Chen (2011a:4518)	Earnings per share	Dividendend payout ratio	Price to book ratio	
Sun & Li (2012:2260)	Earnings per share	Net assets per share		

Source: Own compilation.

From the literature review and Table 7.8, it is evident that no single or combination of market variables was consistently applied to financial distress prediction models. It appears that the prior mentioned earnings per share ratio, the Z-score variable and the market value of equity to total liabilities ratio were the only market variables researched in most studies.

(iii) Quantitative non-financial variables

In contrast to the high volume of research related to the relationship between financial variables and financial distress, as is evident from the section above, significantly less research has been conducted on the relationship between financial distress and the economic cycle. Even less research could be found, combining financial and quantitative non-financial variables, in the context of the current study, in determining financial distress.

Cybinski (2001:34) applied principal component analysis to reduce the number of macroeconomic variables from 43 to only five orthogonal factors. The five factors were identified, based on the series upon which each factor loaded, as follows: level of activity/demand or growth factor; cost of capital borrowing factor; labour market tightness factor; construction factor; and expenditure factor. These and their lagged values for up to three years were included for each data-year of financial variables on the financially distressed company. Cybinski (2001:40) concludes that some explanatory variables are consistently related to financial distress risk for many years before financial distress lends weight to the validity of past single-equation financial distress prediction model specifications with respect to those variables. Cybinski (2001) cautions that other variables need to be treated as endogenous to systems of simultaneous equation models because correlations with other important explanatory variables cause them to be miss-specified in single-equation formulations.

The Dunis and Triantafylllidis (2003:9) study includes the following macroeconomic variables based on existing empirical evidence: the real gross domestic product (RGDP), real money supply (RM4), rate of unemployment (RUN), and the output gap (GAP) as measures of the business cycle, the consumer price index (CPI) as a measure of inflation, the FTSE 100 stock index (FTINX), three-month treasury bill (TRB) as the short-term interest rate, the 10-year government bond yield (GB10Y) as the long-term interest rate, the terms of trade (TOFTR), and finally the number of insolvencies (NINS) as an endogenous variable. Dunis and Triantafylllidis (2003:24-25) established that a neural network regression model with macroeconomic variables provides an attractive alternative to models without macroeconomic variables included.

Lam (2004:571) uses the following macroeconomic variables as predictor attributes: federal budget/gross domestic product; government spending/gross domestic product; money supply 1; money supply 2; short-term interest rate; spread between short-term and long-term interest rate; consumer price index; trade balance/gross domestic product; current account balance/gross domestic product; effective exchange rate; and purchase price of crude oil. Lam (2004:578) shows that financial and macroeconomic variables cannot generate significantly higher returns than the average index.

Liu (2004:940) identifies macroeconomic variables as those variables having the most significant influence on a company's financial health. According to Liu's study, the econometric results show that company failure rates are responsive to changes in the nominal interest rate, price level, real credit and corporate birth rates.

Argyrou (2006:6) selected the following five macroeconomic variables calculated for five consecutive years preceding the year of financial distress: terms of trade (export prices/import prices); gross domestic product; 12-month interest rate; total household disposable income; and cost of living, price index. Argyrou (2006:88) established that, in the context of the particular study, the macroeconomic imbalance may have had no effect on the phenomenon of financial distress.

Hol (2007:80) expresses the view that many different variables are found to be significant in predicting financial distress under different circumstances. The most common variables are as follows: production (GDP), the monetary side (M, an interest rate of CPI) and variables such as unemployment and stock indices. Hol (2007:88) concludes that the gross domestic product gap, an industrial production index and the money supply (M1) are significant additional predictors of financial distress probability.

Within the South African context, Masekesa (2010) examined the interaction between corporate failures and macroeconomic aggregates, and especially the accounts of policy-induced changes in the macroeconomy in the period 1994 to 2009. Taking previous studies as a basis, Masekesa (2010:19-22) states that macroeconomic determinants of corporate failure are closely bound to the following: economic health as measured by gross domestic product; new incorporations; money market and credit conditions (money supply and interest rates); economic openness (measured by foreign exchange rates); uncertainty measured by inflation; financial crisis; industry-specific factors (this variable was set as a dummy); and corporate failure rate.

Masekesa's (2010) estimates corroborated the study's hypothesis that both short- and long-run relationships exist between corporate failure rate in South Africa and the selected macroeconomic determinants. In addition, he established that corporate failure rates were significantly and positively associated with the average lending rate, inflation rate, new corporation, exchange rate, 2007-2009 financial crisis and inversely related to gross domestic product and money supply both in the short and long run.

Zhou *et al*. (2010) explored the effect of macroeconomic information on improving financial distress prediction. The macroeconomic variables included gross domestic product; personal income index; consumer price index; and the money supply (M2) index, which reflects the amount of money supply in the economy. The test results showed that neural network models were slightly improved when macroeconomic variables were included.

Chen (2011a:4518) included the following macroeconomic variables in a ratio as a 'monitoring index', which combines and calculates a company's monetary aggregates, namely direct and indirect finance, stock price index, industrial production index, non-agricultural employment, customer-cleared exports, imports of machinery and electrical equipment, manufacturing sales, and wholesale, retail and food services sales. The study results showed that financial variables had a greater effect on financial prediction performance than non-financial ratios and macroeconomic indices did.

Although there appears to be a relationship between the studies reviewed regarding which macroeconomic variables could potentially be applied to financial distress prediction models, different conclusions were attained.

A limited number of studies combined financial and macroeconomic variables in a single financial distress prediction model. Dunis and Triantafylllidis (2003), Liu (2004), Hol (2007), Masekesa (2010) and Zhou *et al.* (2010) established that financial distress was positively associated with macroeconomic variables. However, Cybinski (2001), Lam (2004), Argyrou (2006) and Chen (2011c:11) had a contrary view in that no positive association between macroeconomic variables and financial distress could be established.

The next section discusses the selection of financial, market and quantitative non-financial variables to be tested for inclusion in the proposed artificial intelligence model.

7.3.5 <u>Development of the proposed artificial intelligence model – Segment 4</u>

Financial distress prediction model performance results can be highly sensitive to the data sample used for validation. To avoid embedding unwanted sample dependency, Berg (2007:134) proposes that the results of quantitative models should be validated on observations not included in the same sample used for model development, which is referred to as out-of-sample validation.

Based on this proposal, the final step in the sample selection process to ensure test sample estimation accuracy is to partition the data into two mutually exclusive subsets, termed training and validation subsets in this study.

A review of the available literature offered little guidance in selecting a training and validation sample. Some studies based the subset on an 80%-20% rule (Lin & Lee 2011:98; Min & Lee 2004:4; Sookhanaphibarn, Polsiri, Choensawat & Lin, 2007:94) and other studies based its subsets on a 70%-30% rule (Chen, 2011c:4; Marconi, Quaranta & Tartufoli, 2010:24; Merkevicius & Garsva, 2007:146).

The Min and Lee (2005:606) study based the weighting of the training and validation subsets on the applicable financial distress prediction model: the 80%-20% rule was applied to the support vector machine, multiple discriminant analysis and logistic regression models. The validation subset was used to test the results, not to develop the model. For application of the back-propagation neural network model, the sample was divided into three subsets: 60% for the training subset, 20% for the validation subset, and finally, 20% for the holdout subset.

The 80%-20% rule proposed by Min and Lee (2005:606) in their study will be applied to assessing the predictive accuracy of the proposed artificial intelligence financial distress prediction model in this study.

The earlier random selection process was applied for the purpose of selecting the training and validation subsets for this study. Firstly, the 86 selected sample companies were numbered from 1 to 86. In the subsequent step, a Hewlett•Packard

12c® financial calculator random number function was used to generate a new set of random numbers to each of the sample companies.

The first 69 companies represent the approximate 80% training subset and the remaining 17 companies represent the approximate 20% validation subset. Random numbers were allocated to each of the 86 sample companies.

The first 69 companies, or 80%, represent the training subset, and the remaining 17 companies represent the validation subset.

A summary of the sample entities divided into the training and validation subsets is shown in Table 7.9.

Table 7.9: Summary of the number of sample companies in the training and validation subset

Main sector	Total number sample companies	Training subset		Validation subset	
		Number of sample companies	Percentage of total sample companies	Number of sample companies	Percentage of total sample companies
AltX	6	4	66.66	2	33.33
Basic materials	5	3	60.00	2	40.00
Consumer services	23	20	86.95	3	13.04
Consumer	13	12	92.30	1	7.69
Health	2	2	100.00	0	0
Industrials	28	20	71.42	8	28.57
Technology	8	7	87.50	1	12.50
Telecommunications	1	1	100.00	0	0
Total	**86**	**69**	**80.23**	**17**	**19.77**

Source: Own compilation.

In Table 7.9, the 86 sample companies are divided into a training and validation subset and summarised into the eight main JSE sectors. Sample companies in the training and validation subset are reasonably represented in the various sectors, with the majority between 67% and 100%, adding up to approximately 80%, represented in the training subset and between zero and 40%, adding up to approximately 20% of the total sample.

The Health and Telecommunications sectors are not represented in the validation subset, due to only two and one sample companies represented in those sectors respectively. It is not expected that this anomaly would affect the overall results because the 80%-20% principle, based on a random subset selection is applicable.

7.3.6 Validation of the proposed artificial intelligence model – Segment 5

In order to validate the forecasting accuracy of the proposed artificial intelligence model against the validated statistical model reflecting the best forecasting ability, a similar testing procedure as in the validation of statistical financial distress prediction models will be followed (see Section 7.3.3).

It is expected that the study's empirical research (H_0 – null hypotheses) will conclude that financial variables in conjunction with quantitative non-financial variables will improve the ability of the developed artificial intelligence model to predict company financial health more accurately on a financial distress continuum than a statistical financial distress prediction model will do.

7.4 QUALITY AND RIGOUR OF THE RESEARCH DESIGN

The quality of the quantitative data obtained from the INET BFA database is subject to the original source of the data obtained from the sample company's audited and published financial information. The researcher relies on compliance by the company and its auditors to the prescribed rules of the International Financial Reporting Standards (IFRS)[11] and the generally accepted accounting practices (GAAP)[12]. In addition, as a listed entity, the company is subject to the rules and regulations of the

---

[11] The International Accounting Standards Board ("IASB") is responsible for setting International Accounting Standards.

[12] The Financial Accounting Standards Board ("FASB") is responsible for promulgating or amending the rules of Generally Accepted Accounting Practice ("GAAP") as occasion requires.

JSE and the Companies Act no. 71 of 2008. The quality of quantitative data can therefore be accepted as reliable and credible.

For quantitative non-financial variables, the researcher relies reliance on official sources such as the South African Reserve Bank and South African Revenue Services. The reliability of non-financial qualitative variables may be problematic. These variables are based on subjective interpretation and therefore not considered for the purpose of the study.

The XLSTAT® version 2016.3 is a statistical data analysis solution for Excel®, developed by Addinsoft, and is used for all statistical data analysis.

7.5 CONCLUSION

The research design and methodology in this chapter were based on the theoretical foundation established in the previous chapters.

It is evident that stakeholders require more sophisticated and dynamic financial distress prediction models due to increasing dynamics in a company's internal and external operating environment. This requirement is evident in the diversity and evolvement of financial distress prediction models from the basic statistical to the more sophisticated artificial intelligence models.

As the market dynamics changes, reliance on merely historical financial information is insufficient, and a wider perspective in the form of non-financial information is required to form part of the financial distress prediction process.

This chapter commenced with a motivation for the application of the quasi-experimental research approach. The research methodology to be followed in this study was divided in five segments and illustrated in Figure 7.3.

The first segment described the methodology to be followed in the identification and selection of the database of companies to be used in this study. This theoretical overview forms the basis of the research methodology to be followed in Segment 2, 4 and 5 of this study.

The methodology to be followed in the identification of financial, market and quantitative non-financial variables was described in Segment 3. These variables will be used in the development of the proposed artificial financial distress prediction model described in Segment 4.

Parallel to Segment 4, the methodology to calculate and validate the statistical financial distress prediction models were described in Segment 2. Finally, the methodology the compare and validate the results from Segment 2 and 4, was described in Segment 5.

The statistical distress prediction models are validated and compared with the inflation-adjusted return on equity in the next chapter.

CHAPTER 8

CALCULATION, VALIDATION AND COMPARISON OF THE K-SCORE AND Z-SCORE FINANCIAL DISTRESS PREDICTION MODELS

8.1 INTRODUCTION

The predictive ability of the K-score, Z-score and inflation-adjusted return on equity is calculated, validated and compared in this chapter in order to determine which of the models will be used as comparison against the proposed artificial intelligence model.

Various validation models were reviewed in Section 7.2 and each of these models was based on a unique selected sample and variable set, with each presenting a unique result set. In addition, unique circumstances applied, rendering a comparison of test results of the various studies problematic. Adding to this potential debate is the question of cross-border comparability of financial distress prediction models.

In an effort to resolve most of the questions and in particular the problem relating to cross-border comparability, the review and application of validation models therefore focus on South African studies, for example, the studies of Kidane (2004), Muller (2008) and Arens (2014).

Kidane (2004) used a binomial statistical technique to validate and compare the Altman Z-score and Springate financial distress prediction models. Muller (2008) used four different financial predictive models (multiple discriminant analysis, recursive partitioning, logistic regression and neural networks) to validate and determine the predictive ability of a selected financial variable set over a one- to five-year period prior to failure. Finally, Arens (2014) used a three-pronged approach in validating the Altman Z-score model and to determine its predictive ability, namely the Mann-Whitney test, a Spearman's rho correlation matrix and an accuracy ratio.

The Arens study (2014:3) achieved an accuracy rate of 82% and an error rate of 17.9%, which was consistent with the accuracy rate of most failure prediction models, namely 84%, 88% and 85% for statistical, artificial intelligence and theoretical models respectively. Against the background of the Arens (2014) study and the fact that the Altman Z-score model was the primary subject used in the validation methodology, it will broadly form the basis of the current study. The final step in Arens' methodology is adapted by applying a weighted efficiency test instead of the standard accuracy or error rates test.

A test commonly done to determine the accuracy of a financial distress prediction model is the percentage accuracy and error rates (Type I and Type II errors) test. Ideally, a perfect financial distress prediction model would classify financial distress 100% correctly. This assumption is unrealistic and not an informative measure of the usefulness of a model. In a random sample particularly, not all sample entities would proportionally be equally financially distressed or non-distressed. The sizes of these groups differ and change from year to year. Korobow and Stuhr (1985) formulated the weighted efficiency (WE) test as a practical application to resolve the differences and to provide increased sensitivity to both the accuracy and efficiency aspects of financial distress prediction models.

The results of the observations are described as follows:

- the Mann-Whitney test to compare the results of the Z-score, K-score and inflation-adjusted return on equity for T, T-1, T-2, T-3 and T-5;
- the Spearman's rho non-parametric test to determine the correlation between Z-score, K-score and inflation-adjusted return on equity for T, T-1, T-2, T-3 and T-5; and
- the classification accuracy and efficiency between the Z-score, K-score and inflation-adjusted return on equity for T, T-1, T-2, T-3 and T-5 based on the Korobow and Stuhr (1985) weighted efficiency test.

The first step in the presentation of the validation results is the presentation of the descriptive statistics. These are provided in the next section.

8.2 DESCRIPTIVE STATISTICS

The differences between the Z-score, K-score and inflation-adjusted return on equity models are identified through the calculation of descriptive statistics. To achieve this, several main descriptive statistics are calculated, i.e. mean, median, standard deviation, minimum and maximum.

Table 8.1 displays the descriptive statistics for the total sample of 85 companies listed on the JSE for the inflation-adjusted return on equity, Z-score and K-score over the 10-year test period.

The descriptive statistics represented in Appendix 8.1 include the total sample (Panels A and B), the distressed category (Panels C and D), the healthy category (Panels E and F), and finally, the depressed[13] category (Panels G and H).

In order to simplify a comparison of statistics between the three models, the average of each calculated statistic over the 10-year test period is presented in Table 8.1.

Table 8.1: Descriptive statistics for the inflation-adjusted return on equity, Z-score and K-score models - average for the period from 2005 to 2014

	Total sample			Distressed category			Healthy category			Depressed category		
Statistic	ROE	K	Z	ROE	K	Z	ROE	K	Z	ROE	K	Z
Number of observations	85	85	85	22	16	6	22	58	66	41	11	13
Minimum	-2.615	-3.777	-0.563	-2.615	-3.777	-0.563	0.436	0.204	3.024	0.129	-0.080	1.919
Maximum	4.190	5.780	25.964	0.125	-0.230	1.568	4.190	5.780	25.964	0.428	0.766	2.895
1st Quartile	0.123	0.003	3.089	-0.063	-1.430	0.453	0.527	0.611	3.923	0.185	0.199	2.258
Median	0.243	0.625	4.327	0.045	-0.733	0.821	0.674	1.106	5.040	0.245	0.274	2.498
3rd Quartile	0.438	1.349	6.053	0.090	-0.403	1.337	0.994	1.784	6.521	0.310	0.378	2.680
Mean	0.346	0.739	5.014	-0.143	-1.089	0.746	1.001	1.399	5.917	0.255	0.300	2.467
Variance (n-1)	0.550	2.142	20.753	0.569	1.121	0.727	0.880	1.350	21.750	0.007	0.083	0.095
Standard deviation (n-1)	0.718	1.453	3.902	0.609	1.013	0.806	0.894	1.147	3.831	0.084	0.237	0.305

Source: Own compilation.

Because a particular company can be positioned anywhere on a financial distress continuum (see Section 1.7.4), all three broad categories, namely distressed, healthy or depressed, including a combined category, are presented in Table 8.1.

[13] The term *depressed* refers to the intermediate, neutral or ignorance zone between 1.81 and 2.99 for the K-score model and -0.19 and 0.20 for the Z-score model.

A noticeable difference in values between the three models within each of the three broad categories can be noticed. Significant differences in the minimum, maximum median and mean values between the Z-score model and the K-score and inflation-adjusted return on equity models are evident and are confirmed in the variance and standard deviation of the models.

The large variance and standard deviation between the Z-score compared with the other two models can be ascribed to a particular variable in the Z-score model, which relates to a company's market capitalisation and value of total liabilities. This variable is a function of a company's share price, which is subject to stock market volatility and investor sentiment.

A number of outliers are detected, with particular reference to the Z-score model. These outliers are the result of drastic or abnormal movement in the volume-weighted average price (VWAP) of shares in City Lodge Hotels Ltd, Famous Brands Ltd, Italtile Ltd, Mr Price Group Ltd, Phumelela Gaming and Leisure Ltd, PPC Ltd and Spur Corporation Ltd. The volume-weighted average share price of each of the companies, as published on the INETBFA database, is verified with the share price at the company's financial year-end as published in its audited annual financial statements. The volume-weighted average share price is retained where the trend is in line with the financial year-end share price.

An unusual discrepancy was detected between the volume-weighted average share price and financial year-end share price of Italtile Ltd for 2005, 2006 and 2007, which related to an error on the INETBFA database. The published financial year-end share price for these three years was used as replacement and proxy for the volume-weighted average share price to remedy the discrepancy.

The descriptive statistics in Table 8.1 reveals a closer relationship between the inflation-adjusted return on the equity model and Z-score model within the distressed category based on the variance and standard deviation statistics. This relationship changes in the healthy category to a closer relationship between the inflation-adjusted return on equity and the K-score. The change can be ascribed to the bulk of observations falling into this category, and is affected by outliers due to volatile movement in the volume-weighted average share price. In the depressed category, a closer relationship between the volume-adjusted weighted average share price and the K-score is evident. The relationship between the various models in the total category is erratic because it is affected by the outliers from the healthy category.

An initial observation based on the descriptive statistics is unconvincing and necessitates further investigation. A three-pronged approach to the sample is adopted based on the Mann-Whitney, Spearman's rho and weighted efficiency tests, which are discussed in the following sections.

8.3 MANN-WHITNEY U TEST

The Mann-Whitney U test is a non-parametric test of the null hypothesis, which is used to test whether one sample has larger mean values than another sample from the same population against an alternative hypothesis, when there is an unknown distribution of explanatory variables selected from the distressed, healthy and depressed categories.

Because the Mann-Whitney U test is a non-paracontinuous-level test, it does not require a special distribution of dependent variables in the analysis. It is therefore an acceptable test to compare mean scores when the dependent variable is not normally distributed and at least of ordinal scale.

To determine the significance in this test, it is assumed that $n > 80$ or each of the two samples are at least > 30. The U-value calculated can be compared against the normal distribution to determine the confidence level.

The reason for using the Mann-Whitney U test is to rank the samples for various conditions, and then determine how the two rank totals differ. The Mann-Whitney test statistics "U" reflects the difference between the two rank totals. The smaller the U value, the less likely it is to have occurred by chance, and *vice versa*.

The Mann-Whitney U test results as presented in Appendix 8.2 were averaged to simplify the discussion and are shown in Table 8.2.

Table 8.2: Mann-Whitney U test results – average for the period from 2005 to 2014

	T		T-1		T-2		T-3		T-5	
Statistic	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z
Number of observations	84	84	83	84	83	84	83	84	84	85
Obervations with missing data	0	0	0	0	0	0	0	0	0	0
Observations without missing data	84	84	83	84	83	84	83	84	84	85
Minimum	-3.777	-0.563	-3.869	-0.498	-4.067	-0.511	-4.339	-0.620	-4.499	-0.959
Maximum	5.780	25.964	5.602	21.429	5.490	15.340	5.526	15.175	4.952	16.199
Mean	0.739	5.014	0.701	4.895	0.658	4.732	0.589	4.645	0.516	4.699
Standard deviation	1.453	3.902	1.433	3.384	1.424	2.734	1.428	2.713	1.418	2.840
U	2692	187	2769	177	2825	167	2949	176	3141	203
Expected value	3524	3562	3519	3556	3517	3560	3516	3558	3520	3561.6
Variance (U)	99201	100806	98991	100560	98929	100709	98864	100646	99022	100794
p-value (Two-tailed)	0.083	<0.0001	0.108	<0.0001	0.107	<0.0001	0.147	<0.0001	0.272	<0.0001

Source: Own compilation.

In interpreting the Mann-Whitney U test results, a distinction is made between the two independent sample groups, comparing the K-score and Z-score results with the inflation-adjusted return on equity. If D is to be assumed, the difference in position between the ROE-K and ROE-Z, and P1 – P2 to be the difference of position between ROE-K and ROE-Z, for the two-tailed test, the null hypothesis H_0 and alternative hypothesis H_1 are as follows:

- H_0: P1 – P2 = D (or the difference of location between the two sample groups is equal to 0); and
- H_1: P1 – P2 ≠ D (or the difference in location between the two sample groups is different from 0).

An examination of the findings in Table 8.2 reveals that the U value of the ROE-K is noticeably larger than the ROE-Z values. The larger U value of the ROE-K is an indication that its occurrence is less by chance than the lower U values of the ROE-Z. The U value of the ROE-K increases over the 10-year test period from T to T-5.

The U value of the ROE-Z in Table 8.2 appears to be erratic over the 10-year test period. The U value decreases between T and T-1, and then increases marginally onwards to T-5.

In applying the Mann-Whitney hypothesis test at 0.01, 0.05 and 0.10 significance levels (∝), H_0 cannot be rejected if the ρ-value is greater than 0.01, 0.05 or 0.10. If the ρ-value is greater than the significance levels (∝), the explanatory variables matter in the three categories (distressed, healthy and depressed). Should the ρ-value be less than the significance levels (∝), the explanatory variables are statistically different in the three categories.

Based on the summarised Mann-Whitney U-test results for ROE-K, Table 8.2 indicates a ρ-value consistently greater at the 0.01, 0.05 and 0.10 significance levels (∝) from T to T-5. The opposite is evident from the results for ROE-Z because the ρ-value is less than 0.0001 at the three significance levels (∝), namely the explanatory variables are statistically different in the distressed, healthy and depressed category.

There is a significant difference between the mean values of ROE-K and ROE-Z and these values increase marginally from T to T-5. The main reason for this difference is the high maximum values of ROE-Z, which are also reflected in the standard deviation between ROE-K and ROE-Z. In order to establish the reason for the consistently high mean, maximum values and standard deviation of ROE-Z, an examination of the Z-score results of the sample companies is required.

The result of each of the five variables in the Z-score model is examined and it is established that Variables 4 and 5 contribute most to the high Z-score results. Variable 4 determines the company's market capitalisation to total liabilities, multiplied by a factor value of 0.006. Variable 5 determines the number of times total assets are covered by revenue, multiplied by a factor of 0.999. Both variables have to be scrutinised in further detail to establish which element in the particular variable can be isolated in contributing most to its high value.

The results show that Variable 4 is the main contributor to the high Z-score values and in particular the market capitalisation component of the variable, which is a function of the company's volume-weighted average price. An increase in a company's share price, with other variable elements (ordinary shares in issue and total liabilities) being relatively stable during the study 10-year test period, will contribute to a high Variable

4 value. This will ultimately lead to a high Z-score for a particular company, where a Z-score higher than 2.99 is an indication of a healthy company.

In examining Variable 5, the second main contributor to the high Z-score values, it is important to establish the company's nature of business. For example, an information technology company may have a high turnover and low total asset value because most of its assets may be vested in unaccounted intellectual property, which may result in a high value for this particular variable. The opposite is also true, where a hotel company with investment in high value prime fixed property relative to its turnover may result in a low value for this variable.

Table 8.2 indicates Variable 4 having a significant effect on ROE-Z, throughout the test period from T to T-5 as evident from the high maximum value, mean and standard deviation. The number of companies in each of the depressed, healthy and distressed categories is compared to determine whether the addition of a market variable (Variable 4 of the Z-score model) to a pure financial model (K-score model) will lead to better financial distress prediction ability of the first-mentioned model.

Tables 8.3 and 8.4 summarise the number of sample companies in each category based on the depressed, healthy and distressed score parameters for the Z-score and K-score (see Sections 2.3 and 2.4 respectively).

Table 8.3: Companies in the depressed, healthy and distressed category as a percentage of total sample for the Z-score – annually for the period from 2005 to 2014

	2014	%	2013	%	2012	%	2011	%	2010	%	2009	%	2008	%	2007	%	2006	%	2005	%	Average
Depressed	15	18	12	14	18	21	16	19	15	18	13	15	14	17	11	13	10	12	6	7	15
Healthy	62	73	66	79	62	73	63	74	63	75	62	74	63	75	69	82	68	81	73	86	77
Distressed	8	9	6	7	5	6	6	7	6	7	9	11	7	8	4	5	6	7	6	7	7
Missing values	0		1		0		0		1		1		1		1		1		0		
Total excluding missing values	85	100	84	100	85	100	85	100	84	100	84	100	84	100	84	100	84	100	85	100	100

Source: Own compilation.

Table 8.4: Companies in the depressed, healthy and distressed category as a percentage of total sample for the K-score – annually for the period from 2005 to 2014

	2014	%	2013	%	2012	%	2011	%	2010	%	2009	%	2008	%	2007	%	2006	%	2005	%	Average
Depressed	13	15	11	13	10	12	14	16	12	14	16	19	13	15	6	7	11	13	9	11	14
Healthy	47	55	50	60	56	66	52	61	56	67	53	63	60	71	64	76	59	70	64	75	66
Distressed	22	26	23	27	19	22	18	21	15	18	14	17	10	12	13	15	14	17	11	13	19
Missing values	3		1		0		1		2		2		2		2		1				
Total excluding missing values	82	96	84	100	85	100	84	99	83	99	83	99	83	99	83	99	84	100	85	99	99

Source: Own compilation.

For the Z-score and K-score results, the depressed category represents 15% and 14% respectively of the 85 sample companies. A more noticeable difference is evident in the distressed category where the Z-score and K-score represent 7% and 19% respectively of the 85 sample companies.

However, based on the Mann-Whitney U test results in Table 8.2, it was expected that the effect of the fourth variable in the Z-score model on ROE-Z would have resulted in a higher percentage of the sample companies falling within the healthy category of the Z-score model. Thus, as a result, the Z-score model appears to be a better financial distress prediction model than the K-score model as the lead time increases. However, a comparison of the two models and percentage companies of the total sample in Tables 8.3 and 8.4 reveals that the contrary is true because the Z-score model represents only 11% more companies in the healthy category than the K-score model.

The Mann-Whitney U test results, therefore, indicate that the addition of a single market variable to a financial distress prediction model does not have a significant effect on determining the superiority of the K-score model over the Z-score model, and *vice versa*. This market variable, however, contributes to the erratic results over all time periods as it is subject to a volatile movement in share price as a result of changes in investors' sentiment and based on an interpretation of financial, market and/or macroeconomic variables.

The results of the Spearman's rho test are discussed in the following section.

8.4 SPEARMAN'S RHO TEST

The Spearman's rho test is a non-parametric measure of rank correlation and the purpose of this test is to assess how well the relationship between two variables can be described using a monotonic function (whether linear or not). If there are no repeated values, a perfect Spearman's correlation of +1 or -1 occurs when each of the variables is a perfect monotone function of the other.

The Spearman's rho test results in Appendix 8.3 were averaged to simplify the discussion and are presented in Table 8.5.

Table 8.5: Spearman's rho test results – average for the period from 2005 to 2014

	T		T-1		T-2		T-3		T-5	
Statistic	**ROE-K**	**ROE-Z**	**ROE-K**	**ROE-Z**	**ROE-K**	**ROE-Z**	**ROE-K**	**ROE-Z**	**ROE-K**	**ROE-Z**
Correlation matrix (Spearman)	0.479	0.429	0.395	0.413	0.284	0.371	0.234	0.336	0.205	0.300
p-values	0.000	0.001	0.006	0.003	0.056	0.009	0.068	0.019	0.177	0.034
Coefficients of determination (Spearman)	0.233	0.191	0.162	0.179	0.090	0.146	0.058	0.096	0.051	0.097

Source: Own compilation.

Table 8.5 represents results of the test conducted at 0.01, 0.05 and 0.10 significance levels ($\propto$). The values presented in Table 8.3 are significantly different from 0 at the three significance levels ($\propto$).

The correlation between the inflation-adjusted return on equity and the K-score (ROE-K) is positive and tends to be high in T, but deteriorates over time from T to T-5. The correlation between the inflation-adjusted return on equity and the Z-score (ROE-Z) is similarly positive and deteriorates as the lead time increases from T to T-5.

A comparison of the trend in the correlation values of the various years between ROE-K and ROE-Z is displayed in Table 8.6.

Table 8.6: Percentage change in correlation values between the years for ROE-K and ROE-Z

	T and T-1	T-1 and T-2	T-2 and T-3	T3 and T-5	T and T-5	Average T to T-5
ROE-K	-18	-28	-18	-13	-37	-19
ROE-Z	-4	-10	-9	-11	-30	-9

Source: Own compilation.

Table 8.6 indicates that the ROE-K correlation for T to T-5 deteriorates at a higher rate than the ROE-Z correlation. In addition, except for a stronger correlation value in T in favour of ROE-K, the ROE-Z correlation remains stronger than the ROE-K correlation from T to T-5. This suggests that ROE-K is a poorer year-on-year performer than ROE-Z over time.

The ROE-Z correlation deteriorating at a slower rate than the ROE-K correlation can be ascribed to the same effect that the market variable (Variable 4) has on the Z-score, as discussed in Section 8.4.

Based on the Spearman's rho test results, the inclusion of a single market variable in a financial distress prediction model appears to have a positive, but not a significant effect on its ability to predict financial distress over the longer term compared with a financial distress prediction model based purely on financial variables. This corresponds with the conclusion of the Mann-Whitney U test results.

The results of the final test in determining the predictive ability of the inflation-adjusted return on equity for the K-score and Z-score models are discussed in the following section.

8.5 WEIGHTED EFFICIENCY TEST

The weighted efficiency test is the final test in the validation of the K-score and Z-score financial distress prediction models.

The total percentage correctly classified is not an informative measure of a model's usefulness because it fails to highlight the variation in size between the three different categories, namely distressed, healthy and depressed. As indicated in Section 8.1, Korobow and Stuhr (1985) devised the weighted efficiency test to take cognisance of the variation in size between the three categories. The weighted efficiency test results in Appendix 8.4 were averaged to simplify the discussion and are presented in Table 8.7. The results of the percentage correct classification test are included in Table 8.7 for comparative purposes.

Table 8.7: Weighted efficiency and percentage correct classification test results for the distressed, healthy, depressed and combined categories – average for the period from 2005 to 2014

	K-score		Z-score	
Distressed	**WE**	**CC**	**WE**	**CC**
T	4.249	12.573	0.844	5.095
T-1	1.981	9.593	0.648	4.614
T-2	0.919	7.212	0.367	3.703
T-3	0.378	5.658	0.252	3.381
T-5	0.482	6.234	0.087	2.361
Healthy	**WE**	**CC**	**WE**	**CC**
T	5.379	20.958	5.902	22.632
T-1	5.463	20.770	6.242	22.661
T-2	4.342	19.492	6.251	22.962
T-3	3.983	18.834	5.774	22.335
T-5	3.574	17.939	5.586	21.983

	K-score		Z-score	
Depressed	**WE**	**CC**	**WE**	**CC**
T	0.852	6.957	0.620	6.639
T-1	0.829	7.856	0.654	7.369
T-2	0.606	6.754	0.686	7.997
T-3	0.976	8.053	1.014	9.130
T-5	0.297	5.738	0.904	8.737
Combined	**WE**	**CC**	**WE**	**CC**
T	5.101	40.488	1.464	34.366
T-1	8.274	38.219	7.545	34.644
T-2	5.718	33.458	7.329	34.662
T-3	5.337	32.545	7.040	34.846
T-5	4.354	29.912	6.577	33.081

Note: WE = weighted efficiency
CC = percentage correct classification

Source: Own compilation.

The results of the weighted efficiency and percentage correct tests in Table 8.4 are grouped into three categories, similar to the Mann-Whitney and Spearman's rho tests, namely financial distressed, healthy and depressed. A combination of all three categories is added to provide an overall perspective of the total correct classification. The financially distressed group for both the K-score and Z-score results deteriorates over the period from T to T-5. Movement in the weighted efficiency percentage values is erratic for the financially healthy group. In both the K-score and Z-score, the values between T and T-1 improve marginally, but then deteriorate from T-2 to T-5. A similar observation can be made for the depressed and combined groups.

More importantly, low weighted efficiency and percentage correct classification result values were obtained for both the K-score and Z-score models in all three categories. Although still low, the healthy category presents higher values for both the K-score and Z-score models. This low score result corresponds with the view held by Korobow and Stuhr (1985:270-272), namely that the weighted efficiency test distinguishes distinctly between failure prediction models and weakness prediction models. The K-score and Z-score are in effect failure prediction models and the inflation-adjusted return on equity is a proxy for a weakness prediction model.

There is an advantage in detecting financial weakness in advance of failure (Korobow & Stuhr, 1985:268). The inflation-adjusted return on equity model as a weakness prediction model could provide an indication that a company came close to failure at some point, but did not eventually fail, and *vice versa*.

8.6 CONCLUSION

The purpose of this chapter was to calculate, validate and compare the K-score and Z-score model and to determine each model's financial distress predictive ability over a number of years. Because companies operate in a dynamic environment, their financial well-being is determined by a number of internal and external variables, i.e. financial, market and quantitative non-financial variables, which can change over time. A particular company can therefore be positioned anywhere on a financial distress continuum; it can move back and forth on the continuum depending on the effect of a combination of variables on its financial results (see Section 1.7.4). The various tests were not limited in determining the ability to predict financial distress only, but were expanded to include a healthy and depressed category.

The Mann-Whitney U test was applied to determine whether one sample had larger mean values than another from the same population against an alternative hypothesis. The results for ROE-K indicated a ρ-value consistently greater at the various significance levels from T to T-5. The opposite was evident from the results for ROE-Z because the ρ-value was less than 0.0001 at the three significance levels, namely that the explanatory variables were statistically different in the distressed, healthy and depressed category.

Based on the Spearman's rho test results, the correlation between the inflation-adjusted return on equity and the K-score (ROE-K) was positive but deteriorated from T-1 to T-5. The correlation between the inflation-adjusted return on equity and the Z-score (ROE-Z) was similarly positive and deteriorated as the lead time increased from T-1 to T-5.

It was evident from the change in the correlation from T to T-5 that the ROE-K correlation deteriorated more than the ROE-Z correlation. Except for a stronger correlation value in T in favour of ROE-K, the ROE-Z correlation remained stronger than the ROE-K correlation from T-1 to T-5. Therefore, over time, the ROE-Z year-on-year performed better than the ROE-K.

The weighted efficiency test is an improvement on the popular accuracy or percentage error test. All three categories, namely distressed, healthy and depressed, presented erratic results from T to T-5. The disappointing low result values for both the weighted efficiency and percentage correct classification tests, due to the large number of companies incorrectly classified as either distressed, healthy or depressed, corroborated the conclusion by Korobow and Stuhr (1985).

The validation of the financial distress predictive ability of each of the K-score and Z-score models was probably hampered by the following factors:

- The country of origin of each model had to be taken into consideration. Factors could be attributed to possible differences in accounting conventions, assumptions, data sets, time periods and failure or financial distress definitions at the time of development.
- Furthermore, the results of each the three tests were affected by the inclusion of a market variable in the Z-score model. It was expected that this would improve the ability of the Z-score model to predict financial distress over a longer period. However, the results were erratic and therefore indicated that

none of the Mann-Whitney U test, Spearman's rho test or the weighted efficiency test provided a convincing case in accurate forecasting financial distress in favour of either the K-score or Z-score model against the inflation-adjusted return on equity model. Another argument was that the inclusion of a market variable in the Z-score model could most probably have contributed to improving its test results.

A company cannot only rely on a single market variable such as the company's share price because a volatile movement could have a significant effect on the results of a financial distress prediction model such as demonstrated in the Z-score through the Mann-Whitney U, Spearman's rho and the weighted efficiency tests.

Although neither of the models could convincingly be singled out as an ideal benchmark, both the K-score and Z-score models[14] will be used to validate and compare against the proposed financial distress prediction model based on the consideration of additional market and quantitative non-financial variables.

[14] The SVM-K-score and SVM-Z-score models will be compared with the proposed financial distress prediction model (see Section 10.1).

CHAPTER 9

VARIABLE SELECTION AND MODEL DEVELOPMENT

9.1 INTRODUCTION

In terms of the review of literature in Chapter 7, a general criticism is the approach followed in variable selection for financial distress prediction models. A number of approaches were followed in the selection of financial variables – some based on the application of rigorous statistical models such as principal component analysis (Chen, 2011c), and logistic regression, least squares, support vector machine and linear discriminant analysis (Van Gestel *et al*., 2006). A number of studies based variable selection on existing financial distress prediction models, predominantly the Altman Z-score model (Jing-ron *et al*., 2009; Peat *et al*., 2012; Rui, 2010). Another group based variable selection on financial variables commonly used in other financial distress prediction studies (Alfaro *et al*., 2008; Divsalar *et al*., 2011). Finally, some studies based their variable selection on formal institutional data sets such as the UCI Repository of Machine Learning Databases (Chen, Yang, Wang, Liu, Xu, Wang & Liu, 2011) or commercial banks (Min *et al*., 2006).

Despite justification for each methodology, no unified approach could be identified as a basis for financial variable selection. Irrespective of the variable selection methodology used, as highlighted above, it remains a fundamental step in the financial distress prediction process.

The following section is divided into three sections; first, the selection of financial variables, followed by the selection of market variables, and lastly, quantitative non-financial variables (see Figure 7.3, Segment 3).

For the purpose of the proposed artificial intelligence model, a hybrid approach was followed in the selection of financial and market variables. The first stage was to identify an initial variable set based on variables commonly used in financial distress prediction models. The final selection was based on the application of a mathematical model, described in detail below.

9.2 FINANCIAL VARIABLE SELECTION

9.2.1 Background

A three-stage process was followed. The first stage consisted of the identification of studies from a pool of financial distress prediction studies. The second stage consisted of the identification of financial variables applied to these financial distress prediction studies, and the third stage consisted of the selection of a specific set of financial variables for participation in the input vector of the principal component analysis.

9.2.2 Stage 1

For this stage, 233 financial distress prediction studies from 2000 to 2012 were reviewed in order to compile a candidate pool of financial variables. To be included, the particular study had to satisfy the following criteria (adapted from Hossari, 2007:14-15):

- The study had to provide a clearly identifiable financial variable-based financial distress prediction model.
- The study had to make an original contribution to the literature.
- Any study replicating an existing model such as the Altman Z-score was excluded.
- The data sample of the particular study had to exclude any company listed within the Basic Materials (Industrial Metals sector and Mining sector), Oil and Gas (whole sector), Financials sectors (whole sector).
- The study had to clearly indicate what specific financial variables were used in predicting financial distress and not merely mention a category of ratio tested.

The result of the first stage in the identification process is summarised in Table 9.1.

Table 9.1: Identification of studies for selection of financial variables

Year	Studies available	Studies selected	Studies selected as percentage of studies available (%)
2000	10	6	60.00
2001	5	1	20.00
2002	5	3	60.00
2003	11	3	27.27
2004	16	9	56.25
2005	21	12	57.14
2006	21	11	52.38
2007	25	16	64.00
2008	19	10	52.63
2009	17	5	29.41
2010	22	9	40.91
2011	27	13	48.15
2012	34	13	28.24
TOTAL	**233**	**111**	**47.64**

Source: Own compilation.

Based on the selection criteria set out above, 111 or 47.64% of the 233 available studies qualified for financial variable identification – Stage 2.

9.2.3 Stage 2

This stage in the financial variable identification and selection process consisted of the identification and listing of all variables used in financial distress prediction models in each of the 111 identified studies. After elimination of duplicated financial variables, 489 different financial variables were identified from the 111 studies reviewed.

Since testing each of the 489 or combinations thereof would be unfeasible, the following criteria to reduce the number of financial variables were applied (adapted from Alfaro *et al.*, 2008:114):

- the financial variable had to commonly be used in financial distress prediction. Financial variables applied less than 10 times over the 13-year review period were eliminated because it was not deemed to be useful in financial distress prediction;
- information needed to calculate the particular financial variable had to be readily available;
- financial variables had to be understandable and easy to calculate.

Table 9.2 summarises the number of times and percentage of total that a particular ratio was applied during the 13-year literature review period, from 2000 to 2012.

Table 9.2: Frequency of financial variables applied from 2000 to 2012

Application of financial variables	Number of times applied	Precentage of total (%)	First stage reduction (times)	First stage reduction (% of total)	Second stage reduction (times)	Second stage reduction (% of total)
1 to 5	425	86.91	0	0	0	0
6 to 10	34	6.95	34	53.13	0	0
11 to 15	10	2.04	10	15.63	10	33.33
16 to 20	3	0.61	3	4.69	3	10.00
21 to 30	7	1.43	7	10.94	7	23.33
31 to 40	4	0.82	4	6.25	4	13.33
41 to 50	3	0.61	3	4.69	3	10.00
51 to 60	1	0.20	1	1.56	1	3.33
61 to 70	1	0.20	1	1.56	1	3.33
71 to 80	1	0.20	1	1.56	1	3.33
Total	**489**	**100.00**	**64**	**100.00**	**30**	**100.00**

Source: Own compilation.

Table 9.2 provides information on the following: 425 occasions when a particular financial variable was applied between one to five times; 34 occasions when a particular financial variable was applied between six and 10 times; 10 occasions when a particular financial variable was applied between 11 and 15 times; one occasion when a particular financial variable was applied between 71 and 80 times, and so forth.

Based on the above reduction criteria, all financial variables applied more than 11 times over the 13-year literature review period were identified as potential predictors of financial distress.

Table 9.3 provides a list of these financial variables. In addition, the frequency of application of each financial variable is provided in the last column.

Table 9.3: Financial variables identified as potential predictors of financial distress

Variable set	Description	Year													
		2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	TOTAL
Profitability	**Net income / total assets** (ROA) (return on assets) (total assets = non-current assets + current assets)	2	1	0	2	2	7	8	9	6	3	5	12	11	**68**
	Net income / revenue (net profit margin)	2	0	0	0	5	3	1	5	3	1	6	6	8	**40**
	Net income / total equity (ROE) or (return on net worth)	1	0	1	2	2	4	6	6	5	3	2	4	8	**44**
	Ebit / total assets (ROA before interest and tax) (gross return on assets) (Ebit = profit on ordinary activities before interest & tax)	0	0	2	2	2	6	2	3	2	3	1	7	7	**37**
	Ebit / revenue (operating profit margin) or ebit margin	1	0	1	3	2	4	3	5	3	2	4	7	9	**44**
	Ebit / interest expense (interest coverage or times interest earned or income gearing)	1	0	0	1	2	3	1	4	3	1	3	3	3	**25**
	Retained earnings / total assets (earnings capitalisation)	0	0	2	0	1	3	3	2	0	1	1	5	8	**26**
Efficiency	**Revenue / gross fixed assets** (fixed asset turnover of operating standing)	1	0	0	2	0	2	2	0	1	0	2	3	0	**13**
	Revenue / total equity (equity turnover)	0	0	0	2	0	2	1	3	2	2	2	1	5	**20**
	Revenue / trade receivables (Accounts receivable turnover) (operating standing)	1	0	0	0	1	1	2	7	2	1	5	2	3	**25**
	Revenue / total assets (total asset turnover) or (capital turnover)	0	0	1	1	5	8	5	9	5	2	7	6	11	**60**
	Cost of sales / inventories (Inventory turnover) (operating standing)	0	0	0	0	0	1	1	4	2	0	3	1	1	**13**
	Stock ratio (Days) (Days inventory outstanding) (Inventory / turnover)	1	0	0	0	2	3	1	1	1	1	1	0	3	**14**
	Cash / total assets	0	0	0	0	1	1	0	2	0	1	2	4	1	**12**
Gearing	**Total equity / total assets** (solvency ratio or own funds ratio)	2	0	1	1	3	4	5	4	4	3	3	1	5	**36**
	Total liabilities / total assets (leverage) or (asset-liability ratio)	0	0	1	0	2	3	2	1	1	2	3	5	4	**24**
	Total debt / total assets (Debt ratio or total liability ratio)	1	0	0	0	2	3	1	1	1	1	1	0	3	**14**
	Total debt / total equity (interest-bearing debt or financial leverage)	2	0	0	2	1	1	1	1	3	1	1	3	4	**20**
Liquidity	**Working capital / total assets** (operating liquidity)	0	0	2	1	2	6	3	4	3	0	3	5	6	**35**
	Working capital / revenue ((ca-cl) /revenue) (working capital turnover)	0	0	0	0	2	4	0	3	2	1	3	3	4	**22**
	(Cash + trade receivables) / current liabilities (quick ratio) (credit standing)	2	0	1	2	1	0	3	6	3	1	1	3	5	**28**
	Cash / current liabilities (cash ratio) or cash position (cash flow ratio)	1	0	0	0	0	2	2	1	3	2	0	0	3	**14**
	Current assets / total liabilities	1	0	0	1	2	5	1	3	2	1	0	4	2	**22**
	Current assets / net assets	5	0	0	1	7	6	7	14	8	4	5	8	9	**74**
	Current assets / short-term debt	0	0	0	0	3	1	1	4	0	1	1	0	0	**11**
	Current liabilities / total liabilities	2	0	1	1	1	1	0	1	0	2	3	1	1	**14**
	Quick assets / revenue or (Cash + accounts receivable) / turnover	0	0	0	0	2	4	0	1	1	2	1	2	2	**15**
	Quick assets / short-term debt	0	0	0	0	3	3	1	1	0	1	1	2	1	**13**
Cash flow	**Cash flow / current liabilities**	1	0	0	0	1	1	0	2	1	0	3	4	3	**16**
	TOTAL	**27**	**1**	**13**	**24**	**57**	**92**	**63**	**107**	**67**	**43**	**73**	**102**	**130**	**799**

Source: Own compilation.

Twenty-nine financial variables within the profitability, efficiency, gearing liquidity and cash flow variable set were identified as initial financial variables.

In selecting financial variables to be included as input vectors for the principal component analysis, consideration was given to the risk of including too many or too few variables. As indicated above, no accepted or proven criteria for determining the ideal number of financial variables were available. Due to this limitation, the relative importance of a particular financial variable in the proposed financial distress prediction model will nevertheless only be determined through the application of an empirical method.

9.2.4 <u>Stage 3</u>

The final stage in the financial variable selection process consisted of the variable reduction, and final selection of financial variables.

Following on the process described in Stage 2, the financial variables identified were those applied more than 11 times over a 13-year literature review period. A total of 29 financial variables within the profitability, efficiency, gearing, liquidity and cash-flow categories were identified for potential inclusion in the current study

However, the inclusion of all 29 financial variables appeared to be impractical and unfeasible. Too many variables may lead to the risk of generalisation and overfitting. Consideration was therefore given to further reduce the number of financial variables to those able to contribute meaningful to prediction accuracy.

From the review of applicable literature in Chapter 7, no proven or unified method was available to identify and select the optimal number and appropriate financial variables for inclusion in a financial distress prediction model.

Based on this limitation, a mathematical technique was used to evaluate the relative contribution of each of the 29 variables to one another. Principal component analysis (PCA) and factor analysis (FA) were considered as potential techniques for variable reduction. Both these techniques could be used to identify highly correlated variables. In this context, highly correlated variables were those measuring the same construct, which implied redundancy. Because of this redundancy, some of the highly correlated variables could be eliminated leading to a reduced number of variables.

Although principal component analysis and factor analysis are both variable reduction techniques, the most important difference between these techniques is the assumption of an underlying causal structure. Factor analysis assumes an underlying relationship between variables compared with the principal component analysis, which makes no assumptions about the underlying causal model. Principal component analysis is simply a variable reduction technique to reduce a large number of variables in a relatively small number of variables that account for most of the variance in a set of identified variables.

The principal component analysis was conducted through XLSTAT® in order to reduce the 29 financial variables. The 29 financial variables were grouped into the following five broad categories:

- profitability – seven variables;
- efficiency – seven variables;
- gearing – four variables;
- liquidity – 10 variables;
- cash flow – one variable.

The principal component analysis was conducted on each of the 85 sample companies. The process was repeated for each of the four categories (profitability, efficiency, gearing and liquidity). The cash flow category was excluded as it contained only one financial variable.

The following steps were followed in conducting the principal component analysis on each of the 85 sample companies:

Step 1: The eigenvalue was determined from the eigenvector. The eigenvector with the highest eigenvalue was regarded as the "principal component" of a specific category of variables. Each eigenvalue related to a certain factor and each factor related to a certain dimension. Each factor was a linear combination of all the variables in a specific category. An eigenvalue represented the extent of a variance (diversion from the mean) for a specific factor.

Step 2: The number of meaningful factors that could be retained was determined.

- **Eigenvalue test** – in this test, the factor with an eigenvalue larger than one was retained and evaluated. This concept is explained as follows - each identified variable contributed one unit variance to the total variance in the specific category. Each factor with an eigenvalue larger than one represented a larger variance contribution than would have been contributed by a single variable. This factor then represented a meaningful portion of the variance and could therefore be considered for inclusion in the final model.

  The contrary argument was also valid – a factor with an eigenvalue smaller than one represented a smaller variance than would have been contributed by a single variable. As indicated previously, the primary objective of the principal component analysis was to reduce a large number of variables. This objective could not be achieved by retaining a factor representing a smaller variance than would have been contributed by a single variable. Therefore, the factor with an eigenvalue smaller than one was regarded as immaterial and was not considered for inclusion in the final model.

- **Scree test** – eigenvalues were ranked from the highest to lowest, representing components according to their importance or contribution. The first or first two factors were expected to represent the highest variances. By implication, these factors represented high projection values on the initial multidimensional graph.

However, this result was interpreted with caution as the ensuing factors could contain hidden information. The eigenvalues and cumulative variances were plotted on the Y- and X-axis of a scree graph. The biggest division between the factors with high eigenvalues and factors with low values was determined. The factors left of the division were recognised as the most valuable contributors. Factors on the right-hand side of the division were regarded as immaterial and were not considered for inclusion in the final model.

- **Factor loading** – the objective of factor loading in the context of the study was to identify uncorrelated (orthogonal) variables and factors. Variables with the highest given factor were considered for inclusion in the final model.

- **Contribution of each variable to the factor value (expressed as a percentage)** – the variable contributing the highest percentage to a factor. This was an alternative test to determine the relative importance of a variable in a specific category to determine the contribution. The variable with the highest percentage contribution was considered for inclusion in the final model.

- **Squared cosines of the variables** – the objective of this final step was to determine the meaningfulness by which a variable avoided projection errors. The variable would not be considered should the squared cosine be low or tend towards nil.

  A squared cosine table was considered in order to confirm a good fit between the variable and the graph axes – a larger squared cosine confirmed the proximity to the graph axes.

Following the principal component analysis process, the intention was to select five financial variables – one each per profitability, efficiency, gearing, liquidity and cash flow category.

(i) Profitability variables:

The results of all 85 sample companies were combined to calculate an average based on each of the steps described above, and are summarised in Table 9.4.

Table 9.4: Summary of principal component analysis in selection of profitability variables

	NI/TA	EBIT/TA	NI/TE	NI/REV	EBIT/REV	EBIT/INT EX	OTHER
Number	37	13	12	11	8	2	2
Percentage of total sample	44	15	14	13	9	2	2
Eigenvalue	4.950	5.017	5.644	5.035	5.177	4.205	4.627
Variability (F1)	70.921	71.661	73.061	71.936	73.953	60.067	66.001
Cumulative variability (F1 + F2)	89.466	90.270	90.115	90.764	90.567	82.396	92.478
Highest factor loading (F1)	0.978	0.966	0.983	0.981	0.970	0.955	0.981
Highest factor loading (F2)	0.830	0.825	0.852	0.846	0.739	0.945	0.912
Contribution of variable (F1)	19.789	18.998	19.686	19.677	18.652	24.045	20.806
Contribution of variable (F2)	59.750	57.967	67.711	63.371	51.001	65.909	45.212
Squared cosine (F1)	0.958	0.934	0.967	0.963	0.943	0.913	0.963
Squared cosine (F2)	0.702	0.696	0.752	0.735	0.592	0.894	0.832

Source: Own compilation.

Key to the abbreviations in Table 9.4:

- NI/TA = net income to total assets
- EBIT/TA = earnings before interest and tax to total assets
- NI/TE = net income to total equity
- NI/REV = net income to revenue
- EBIT/REV = earnings before interest and tax to revenue
- EBIT/INT EX = earnings before interest and tax to interest expense
- RI/TA = retained income to total assets

The net income to total equity displayed the highest eigenvalue, followed by earnings before interest and tax to revenue, net income to revenue and earnings before interest and tax to total assets. The net income to total assets ranked fifth out of the total of seven variables.

In terms of factor loading, net income to total equity displayed the highest value, followed by net income to revenue, retained income to total assets, net income to total assets earnings before interest and tax to revenue, and earnings before interest and tax to total assets. The net income to total assets variable ranked fourth out of the total of seven variables.

From a squared cosine perspective, the net income to total equity displayed the highest value, followed by the net income to revenue, and retained income to total assets, net income to total assets and earnings before interest and tax to revenue variables. In this instance, the net income to total assets variable ranked fourth out of a total of seven variables.

The contribution of the variable to the factor was the final test conducted. Earnings before interest and tax to interest expense and retained income to total assets ranked highest, with net income to total assets ranking third.

The deciding factor in the selection of a profitability variable was the number of times the variable was applied. Although the net income variable to total assets did not rank the highest in the summary table, based on average calculated values, it was the variable most often used based on the various tests in the principal component analysis – 37 out of 85 times or 44%, compared with 15% as the next highest percentage. As the ranking in Table 9.4 was based on an average calculation of the 85 sample companies, it could have been affected by both positive and negative outliers.

The net income to total asset variable was selected to represent the profitability category, with a contribution or weighting value of 19.789%.

(ii) Efficiency variables:

An identical process to the above was followed in selecting the efficiency variable. Table 9.5 represents the average values of all 85 sample companies.

Table 9.5: Summary of principal component analysis in selection of efficiency variables

	REV/TA	REV/GR FA	INV/REV	COS/INV	REV/TE	REV/TR REC	CASH/TA	COS/REV
Number	31	14	12	10	7	6	4	1
Precentage of total sample	36	16	14	12	8	7	5	1
Eigenvalue	3.865	3.480	3.586	3.426	3.577	3.620	3.548	4.395
Variability (F1)	56.983	52.247	51.227	48.947	52.274	52.653	50.676	62.784
Cumulative variability (F1 + F2)	80.765	79.679	75.949	75.145	81.175	78.329	79.398	82.097
Highest factor loading (F1)	0.954	0.923	0.920	0.934	0.928	0.869	0.915	0.923
Highest factor loading (F2)	0.842	0.848	0.862	0.829	0.802	2.124	0.849	0.950
Contribution of variable (F1)	24.307	25.498	24.305	26.137	24.317	25.006	24.185	19.399
Contribution of variable (F2)	46.688	41.408	44.102	38.534	34.200	41.958	35.954	66.800
Squared cosine (F1)	0.911	0.854	0.850	0.876	0.862	0.854	0.839	0.853
Squared cosine (F2)	0.717	0.726	0.756	0.697	0.649	0.678	0.740	0.903

Source: Own compilation.

Key to the abbreviations in Table 9.5:

- REV/TA = revenue to total assets
- REV/GR FA = revenue to gross fixed assets
- INV/REV = inventory to revenue
- COS/TE = cost of sales to total equity
- REV/TR REC = revenue to trade receivables
- CASH/TA = cash to total assets
- COS/REV = cost of sales to revenue

The selection of the efficiency variable was less complex than selecting the profitability ratio. The revenue to total assets variable ranked highest in the various principal component tests, and was therefore selected as the variable representing the efficiency category. The contribution or weighting value was calculated at 24.307%.

(iii) Gearing variables:

The following section represents the selection of the gearing variables. The average values of all 85 sample companies are displayed in Table 9.6.

Table 9.6: Summary of principal component analysis in selection of gearing variables

	TD/TE	TE/TA	TL/TA	TD/TA
Number	48	17	12	8
Percentage of total sample	56	20	14	9
Eigenvalue	3.354	3.299	2.980	3.374
Variability (F1)	83.851	85.450	74.494	84.387
Cumulative variability (F1 + F2)	99.739	98.389	98.050	99.822
Highest factor loading (F1)	0.924	0.969	0.922	0.948
Highest factor loading (F2)	0.480	0.462	0.793	0.529
Contribution of variable (F1)	26.830	29.832	31.811	26.893
Contribution of variable (F2)	41.608	59.924	69.665	53.193
Squared cosine (F1)	0.895	0.941	0.930	0.903
Squared cosine (F2)	0.255	0.307	0.659	0.301

Source: Own compilation.

Key to the abbreviations in Table 9.6:

- TD/TE = total debt to total equity
- TE/TA = total equity to total assets
- TL/TA = total liabilities to total assets
- TD/TA = total debt to total assets

The selection of the gearing variables presented a similar complication to the case of the profitability variable selection.

In the eigenvalue test, the total debt to total assets variable rank highest, followed by total debt to total equity, total equity to total assets and the total liabilities to total assets variables.

The factor loading test ranked the total equity to total assets variable highest, followed by the total debt to total assets, the total debt to total equity and then the total liabilities to total assets variables.

The squared cosine test ranked the total equity to total assets variable highest, followed by the total liabilities to total assets, total debt to total assets, and finally, the total debt to total equity variables.

As in the case of the selection of the profitability variable, the deciding factor in the selection of a gearing variable was the number of times this particular variable was applied. Although the total debt to total equity variable did not rank the highest in the summary table, based on average calculated values, it was the variable most often used based on the various tests in the principal component analysis – 48 out of 85 times or 56%, compared with 20% as the next highest percentage. Because the ranking in Table 9.6 was based on an average of the 85 sample companies, it could have been affected by positive or negative outliers.

The total debt to total equity variable was selected to represent the gearing category, with a contribution or weighting value of 26.830%.

(iv) Liquidity variables:

Next, a liquidity variable was selected. Table 9.7 represents the average values of all 85 sample companies.

Table 9.7: Summary of principal component analysis in selection of liquidity variables

	(CA-INV)/CL	WC/TA	CA/CL	WC/REV	CA/TL	CASH/CL	QA/ST-D
Number	22	18	16	15	10	3	1
Percentage of total sample	26	21	19	18	12	4	1
Eigenvalue	6.294	6.008	6.032	6.280	5.514	5.907	6.750
Variability (F1)	62.939	60.435	61.698	62.803	56.437	59.070	67.495
Cumulative variability (F1 + F2)	84.868	82.854	84.774	85.394	81.108	87.035	89.556
Highest factor loading (F1)	0.976	0.963	0.970	0.974	0.966	0.968	0.983
Highest factor loading (F2)	0.845	0.860	0.886	0.863	0.848	0.851	0.937
Contribution of variable (F1)	15.356	15.843	15.987	15.380	17.430	16.232	14.327
Contribution of variable (F2)	35.095	36.359	35.974	35.539	31.275	27.249	39.805
Squared cosine (F1)	0.953	0.928	0.941	0.948	0.933	0.937	0.967
Squared cosine (F2)	0.731	0.748	0.793	0.753	0.720	0.750	0.878

Source: Own compilation.

Key to the abbreviations in Table 9.7:

- (CA-INV)/CL = (current assets – inventory) to current liabilities
- WC/TA = working capital to total assets
- CA/TL = current assets to total liabilities

- WC/REV = working capital to revenue
- CA/TL = current assets to total liabilities
- Cash/CL = cash to current liabilities
- QA/ST-D = quick assets to short-term debt

Although the quick assets to short-term debt variable revealed the highest eigenvalue, it was used only once as a liquidity variable. Current assets less inventory to current assets ranked the next highest, followed by working capital to revenue, current assets to current liabilities, cash to current liabilities, and finally, current assets to total liabilities variable.

The quick assets to short-term debt variable ranked the highest in terms of the factor loading test, followed by the current assets less inventory to current assets variable and the rest of the variables in the liquidity category.

A similar pattern to the above was evident in the squared cosine test, which ranked the short-term debt variable highest, followed by the current assets less inventory to current assets variable and the rest of the variables in the liquidity category.

As in the case of the selection of the profitability and gearing variables, the deciding factor in the selection of a liquidity variable was the number of times the particular variable was applied. Although the current assets less inventory to current assets variable did not rank the highest in the summary table, based on average calculated values, it was the variable most often used based on the various tests in the principal component analysis – 22 out of 85 times or 26%, compared with 20% as the next highest percentage. The quick assets to short-term debt variable, which ranked the highest in terms of the eigenvalue, factor loading and squared cosine tests, was used only once. As previously, the ranking in Table 9.7 was based on an average of the 85 sample companies and could have been affected by positive or negative outliers.

The current assets less inventory to current liabilities variable was selected to represent the liquidity category, with a contribution or weighting value of 15.356%.

(v) Cash flow variables:

The final category focused on the selection of a cash flow variable. Because only one variable was identified in this category, a principal component analysis was not required.

The cash flow to current liabilities variable was identified as the most appropriate variable in the cash flow category

9.3 MARKET VARIABLE SELECTION

9.3.1 <u>Identification of market variables</u>

As indicated earlier, a limited number of studies considered the inclusion of market variables in a financial distress prediction model. Most of the studies reviewed included one or more market variables, but without considering whether its inclusion enhanced the prediction accuracy of the particular financial distress prediction model. It is evident that earnings per share and the Altman Z-score variable (market value of equity to total liabilities) are the most popular market variables applied to financial distress prediction models.

Table 9.8 is a compilation of market variables applied to the various studies reviewed. The table includes the two popular variables mentioned in the previous paragraph, market variables from studies reviewed and summarised in Table 7.8, and a number of market variables that relate to market perception and their effect on a company's share price movement. These market variables will be considered for testing and inclusion in the current study.

Table 9.8: Market variables identified as potential predictors of financial distress

Market variable	Abbreviation	Definition	Description of market variable
Earnings per share	EPS	Net income after tax to total number of ordinary shares issued	Measures earnings capability of an ordinary share
Price-earnings ratio	PE	Market price per share to earnings per share	Measures the amount an investor is willing to pay for each rand of earnings
Earnings yield	EY	Earnings per share to market price per share	Measures the earnings capability of an ordinary share realtive to its market value
Dividends per share	DPS	Total dividends to total number of ordinary shares issued	Measures the dividend return earned by a share
Dividend yield	DY	Dividend per share to market value per share	Measures the dividend earned by a share relative to its market value
Dividend payout ratio	DPO	Dividend per share to earnings per share	Indicates the percentage of earnings paid out as a dividend
Book value per share	BVPS	Total shareholders' equity to total number of ordinary shares issued	Measures equity on a per share basis
Price to book value	PBV	Market price per share to book value per share	Measures market price of an ordinary share relative to its book value
Market capitalisation	Market cap	Market price per share to total number of ordinary shares issued	Measures the market value of the company according to the stock market
Cash flow per share	CFPS	Cash flow from operating activities to total number of ordinary shares issued	Measures the cash flow generating ability of an ordinary share
Rate of change of share price	RoC(P)	Closing share price in period T - closing share price in period T-1 to closing share price in period T-1	Measures the percentage change between the most recent and the historical share price
Rate of change of cash flow per share	RoC(CF)	Cash flow per share in period T - cash flow per share in period T-1 to cash flow per share in period T-1	Measures the percentage change between the most recent and the historical share price
Market value of equity to total liabilities	MKVALLT	(Share price x total number of ordinary shares issued) to total liabilities	Measure how much a company's assets can decline in value before liabilities exceed assets and the company becomes insolvent
Relative strength index	RSI	100 - (100/(1+relative strength)), where relative strength is the average of m periods' up closes to average of m periods' down closes	A technical momentum indicator of market sentiment that compares the magnitude of a share prices' recent gains to losses in an attempt to determine overbought or oversold conditions
Earnings before interest and tax to total number of ordinary shares issued	EBITPS	Earnings before interest and tax to total number of ordinary shares issued	Measures the earnings capability per share, before interest and tax
Excess returns	ER	Share price in year T-1 less share price index return in year T-1	Measures the returns earned by a share that exceeds a benchmark index with a similar level of risk

Source: Compiled from Gulfbase investment tutorial database.

Table 9.8 contains 16 market variables with each having the potential to contribute to prediction accuracy of the proposed artificial intelligence model. However, it may be impractical to include all these market variables. Therefore, the number of market variables must be reduced to those contributing most significantly to prediction accuracy.

9.3.2 Selection of market variables

The principal component analysis was conducted through XLSTAT® in order to reduce the 16 market variables.

Market variables can be regarded as a single category, therefore, a single variable must be selected from the 16 market variables to be consistent with the financial variable selection process. However, because market variables form an essential element of the current study, the selection was expanded to include two variables instead of one variable. Table 9.9 represents the average values of all 85 sample companies.

Table 9.9: Summary of principal component analysis for selection of market variables

	EPS	MC	DPS	INF ADJ PPS	EBIT TO S O	CF PS	PE
Number	15	13	11	11	7	6	5
Percentage of total sample	18	15	13	13	8	7	6
Eigenvalue	8.079	8.368	8.574	8.079	7.533	8.198	7.422
Variability (F1)	50.537	52.303	53.589	50.491	48.615	51.234	46.387
Cumulative variability (F1 + F2)	73.707	76.096	76.385	74.262	71.623	75.680	73.452
Highest factor loading (F1)	0.961	0.950	0.967	0.951	0.957	0.963	0.943
Highest factor loading (F2)	0.876	0.869	0.877	0.879	0.855	0.877	0.903
Contribution of variable (F1)	11.564	10.855	11.090	10.888	12.422	11.538	12.486
Contribution of variable (F2)	21.051	20.907	21.757	20.947	20.999	20.024	19.938
Squared cosine (F1)	0.925	0.908	0.936	0.923	0.916	0.928	0.889
Squared cosine (F2)	0.767	0.763	0.775	0.780	0.739	0.775	0.816

Table 9.9: (continued)

	NAV PS	MV/ TL	PTB	DY	EY	BV/S	DPOUT
Number	5	3	3	2	2	1	1
Percentage of total sample	6	4	4	2	2	1	1
Eigenvalue	7.210	7.916	7.185	6.488	5.848	8.073	7.115
Variability (F1)	46.811	52.957	44.909	40.550	36.550	50.456	44.467
Cumulative variability (F1 + F2)	72.464	76.240	67.324	65.842	62.065	70.075	70.780
Highest factor loading (F1)	0.974	0.906	0.916	5.402	0.904	0.960	0.968
Highest factor loading (F2)	0.864	0.792	0.786	0.878	0.833	0.794	0.782
Contribution of variable (F1)	13.554	10.467	11.691	12.543	14.062	11.428	13.180
Contribution of variable (F2)	19.259	19.164	17.451	19.085	17.010	20.064	14.543
Squared cosine (F1)	0.952	0.822	0.839	0.815	0.817	0.923	0.938
Squared cosine (F2)	0.758	0.643	0.628	0.781	0.696	0.630	0.612

Source: Own compilation

Key to the abbreviations in Table 9.9:

- EPS = earnings per share
- MC = market capitalisation
- DPS = dividends per share

- INF ADJ PPS = inflation-adjusted price per share
- EBIT to SO = earnings before interest and tax to total number of shares issued
- CF PS = cash flow per share
- PE = price earnings
- NAV PS = net asset value per share
- MV / TL = market value of equity to total liabilities
- PTB = price to book value
- DY = dividend yield
- EY = earnings yield
- BV/S = book value per share
- DPOUT = dividend payout ratio

The eigenvalue, factor loading and squared cosine tests were conducted on the identified variables in Table 9.9.

The dividends per share ranked highest in the eigenvalue test, followed by the market capitalisation, cash flow per share, earnings per share and book value per share variables. The dividends per share ranked third and were used 13 times on the 85 sample companies.

The net asset value per share ranked the highest in terms of the factor loading test, followed by the dividend payout, dividend per share, cash flow per share and earnings per share variables. The net asset value per share was used five times on the 85 sample companies.

The net asset value per share ranked highest at a squared cosine test value of 0.952 and was used five times on the 85 sample companies. This was followed by the dividend payout, dividend per share, cash flow per share and earnings per share variable. The squared cosine test value for the earnings per share variable was 0.925.

The final selection of market variables was based on the highest average number of times they were identified in the principal component analysis of the 85 sample companies. This was similar to the final selection process used for the identification of financial variables. Although the earnings per share and market capitalisation variables did not rank the highest in the summary table, based on average calculated values, these were the variables most often used based on the various tests in the principal component analysis –15 and 13 times respectively. The quick assets to short-term debt variable, which ranked the highest in terms of the eigenvalue, factor loading and squared cosine tests, was used only once. As the ranking in Table 9.9 was based on an average of the 85 sample companies, it could have been affected by positive or negative outliers.

The earnings per share and market capitalisation variables were selected as the most appropriate variables to represent the market variables in the proposed financial distress model. Contribution or weighting values of 11.564% and 10.885% were calculated for these two market variables respectively.

9.4 QUANTITATIVE NON-FINANCIAL VARIABLE SELECTION

9.4.1 Background

In order to test the null hypothesis (H_0), combining financial and quantitative non-financial or macroeconomic variables, will improve the ability of an artificial intelligence model to predict company financial health more accurately than a statistical financial distress prediction model.

No consensus could be established from the literature review on which single or combination of quantitative non-financial variables would contribute and satisfy the null hypothesis.

A number of macroeconomic indicators were considered for inclusion as quantitative non-financial variables earlier in the study (see Section 7.3.4). However, specialist knowledge in the field of econometrics is required to identify the most appropriate single or combination of macroeconomic variables that may potentially have a meaningful effect on the proposed financial distress prediction model. The development of an econometric model is beyond the scope of the current study. An official composite business cycle indicator as published by the South African Reserve Bank on a quarterly basis was therefore relied on.

9.4.2 Composite business cycle indicators

The composite business cycle indicator is a composition of a number of individual economic indicators in a single time series. This composite economic indicator portrays the movement of and turning points in the business cycle (Venter & Pretorius, 2001:63).

Three composite indexes are published on a quarterly basis providing insight into the movement and changes in the business cycle:

- **Composite leading business indicator** consists of individual indicators, which tend to shift direction ahead of changes in the business cycle. A change in the direction of the composite leading business cycle indicator is usually an early indication that a turning point in the business cycle is imminent.
- **Composite coincident business cycle indicator** combines a number of economic time series, which usually move in harmony with the business cycle. A change in the direction of the composite coincident business cycle indicator, generally after the composite leading business cycle indicator has changed direction, indicates that a turning point may have been reached.
- **Composite lagging business cycle indicator** consists of a number of economic time series, which only change after the business cycle has begun to follow a particular pattern or trend. This indicator is normally used to confirm a trend.

9.4.3 Composition of the composite business cycle indicators

The following section lists the component time series included in the three composite business cycle indicators.

(i) Leading business cycle indicator:

- job advertisement space in the *Sunday Times* newspaper: percentage change over 12 months;
- number of residential building plans passed for flats, townhouses and houses larger than 80 m^2;
- interest rate spread: 10-year government bonds less 91-day Treasury bills;
- real M1 money supply (deflated with CPI): six-month smoothed growth rate;
- index of commodity prices (in United States dollar) for a basket of South African-produced export commodities;
- composite leading business cycle indicator of South Africa's major trading partner countries: percentage change over 12 months;
- gross operating surplus as a percentage of gross domestic product;
- Rand Merchant Bank (RMB)/Bureau for Economic Research (BER) Business Confidence Index;
- net balance of manufacturers observing an increase in average number of hours worked per factory worker (half weight);
- net balance of manufacturers observing an increase in the volume of domestic orders received (half weight);
- number of new passenger vehicles sold: percentage change over 12 months.

(ii) Coincident business cycle indicator

- gross value added at constant prices, excluding agricultural, forestry and fishing;
- total formal non-agricultural employment;
- value of retail and new vehicle sales at constant prices;
- industrial production index;

- utilisation of production capacity in manufacturing.

(iii) Lagging business cycle indicator

- cement sales (in tons);
- value of non-residential buildings completed at constant prices;
- ratio of gross fixed capital formation in machinery and equipment to final consumption expenditure on goods by households;
- ratio of inventories to sales in manufacturing and trade;
- nominal labour cost per unit of production in the manufacturing sector: percentage change over 12 months;
- predominant prime overdraft rate of banks;
- ratio of consumer instalment sale credit to disposable income of households.

9.4.4 <u>Business cycle indicators</u>

The business cycle indicators for the 10-year period from 2005 to 2014 for the leading, coincident and lagging indicators were obtained from the South African Reserve Bank database. The monthly data was annualised and standardised in order to be comparable with the annual company financial results and the applicable financial distress models.

The annualised and standardised results of the three business cycle indicators are displayed in Table 9.10.

Table 9.10: Annualised and standardised business cycle indicators

Year	Leading indicator	Coincident indicator	Lagging indicator
	Index: 2010=100	Index: 2010=100	Index: 2010=100
12/31/2005	0.929	0.909	1.028
12/31/2006	0.970	0.982	1.047
12/31/2007	0.955	1.053	1.105
12/31/2008	0.891	1.063	1.222
12/31/2009	0.855	0.957	1.123
12/31/2010	1.000	1.000	1.000
12/31/2011	1.003	1.056	1.019
12/31/2012	0.994	1.105	1.051
12/31/2013	0.994	1.133	1.082
12/31/2014	0.978	1.146	1.054

Source: Compiled from South African Reserve Bank Quarterly Report, December 2015.

Each of the business cycle indicators will be added to the proposed financial distress model. The model will first be tested with the leading indicator, followed by the coincident and the lagging indicator. The indictor providing the best result in terms of the support vector machine analysis will be selected as the quantitative non-financial variable for the proposed financial distress model. The result of the analysis is discussed in Chapter 10.

9.5 CONCLUSION

A rigorous process was followed in the identification and selection of financial and market variables. This process resulted in a large number of variables, which had to be reduced to a feasible number of variables.

A mathematical process, through principal component analysis, was followed in the final selection of financial and market variables. One financial variable each in the profitability, efficiency, gearing, liquidity and cash flow category was selected, and two market variables.

The selection of quantitative non-financial variables was more simplistic. Instead of constructing an econometric time series model, which is beyond the scope of the current study, an official model was used. The leading, coincidence and lagging indicators as published by the South African Reserve Bank, were annualised and standardised for the purposes of the study. The testing of each of these indicators to select the most appropriate quantitative non-financial variable is discussed in Chapter 10 and 11.

In Chapter 10, the proposed artificial intelligence financial distress prediction model[15] will be classified and the classification accuracy tested through the application XLSTAT®.

[15] Hereafter referred to as the F-score model.

CHAPTER 10

CLASSIFICATION AND CLASSIFICATION ACCURACY OF THE SVM-K-SCORE, SVM-Z-SCORE AND F-SCORE FINANCIAL DISTRESS PREDICTION MODELS

10.1 INTRODUCTION

In terms of the theoretical overview in Chapter 6, a *support vector machine* can be described as a binary classifier based on a supervised statistical learning model with associated learning algorithms, which are used for analysing and classifying data, recognising patterns and regression and density estimation.

The objective of a support vector machine is to construct a linear model to estimate a decision function that distinguishes between two classes in a data set (Moepya, Nelwamondo & Van der Walt, 2014:44). Two hyperplanes are selected that maximises the distance or the separation between the two classes: γ = -1 or +1. The larger the distance between the hyperplanes, referred to as the margin, the more reliable the classification.

In the following sections, the selection of options in XLSTAT® is described. This is followed by a classification of the K-score, Z and F-score models. The chapter is concluded by a discussion and comparison of the classification accuracy of each of the models.

In order to differentiate between the Mann-Whitney U and Spearman's rho and the statistical tests conducted when evaluating the original K-score and Z-score models as discussed in Chapter 8, and the support vector machine test described in this chapter and Chapter 11, the original K-score and Z-score models will be renamed as the SVM-K-score and SVM-Z score models. The SVM-K-score and SVM-Z-score models are in effect new models in that they are calculated using the support vector machine. The data input to calculate the SVM-K-score and SVM-Z-score models is the score calculated in the original K-score and Z-score models. The SVM-K-score and the SVM-Z-score models are therefore a proxy for the original K-score and Z-score models and are used as the benchmark in evaluating the F-score models.

10.2 SETTING UP OF THE SUPPORT VECTOR MACHINE CLASSIFIER

The support vector machine function within the machine learning module of XLSTAT® was used to conduct the analysis. XLSTAT® allows for a variety of options ranging from the selection of data to the display of results. The process followed in the current study in setting up and training of the support vector machine classifier is described in the following section:

- The first step consisted of the selection of the response variable. Two distinct binary values were selected representing the response variable: 0 was selected to indicate a positive class or non-distressed companies and 1 was selected to indicate a negative class or financially distressed companies.
- The second step consisted of the selection of the explanatory variable. The explanatory variable represented either the SVM-K-score, SVM-Z-score or F-score. The quantitative option was activated. The columns selected corresponded to the period fields, over the 10-year study period from 2005 to 2014. The rows selected corresponded to the 85 sample companies.
- In the third step, the Options tab allows for the preparation of the optimisation algorithm to a specific requirement. Three tuneable parameters were available:

    - "C" is the regularisation parameter. It translates to how much misclassification will be allowed during the optimisation process. A large C value implies a strong penalty on each misclassified observation. C was set at the default value one for all tests in this study.
    - The "epsilon" is the numerical precision parameter. The epsilon is machine dependent and was left at the default value of 1e-12.

- The “tolerance” parameter defines the tolerance or accuracy of the optimisation algorithm when comparing support vectors. The tolerance was left at the default value of 0.001.

Within the Options tab, the “pre-processing” field allows for three options to rescale the explanatory data. Firstly, “rescaling” – quantitative variables are rescaled between 0 and 1 using the observed minimum and maximum for each variable. Secondly, “standardisation”- the quantitative variable is standardised using the sample mean and variance for each variable. Thirdly, “none” – no transformation is applied. The standardisation field was selected for this study.

Lastly, within the Options tab, the “kernel” field allows for the selection of the kernel to be applied to the data set to extend the feature space. Four kernels are available. Firstly, “linear kernel” – this is the basic linear dot product. Secondly, “power kernel” – this kernel requires the selection of coefficient and gamma parameters. Thirdly, “radial basis function kernel” – this kernel requires the selection of gamma parameters. Lastly, “sigmoid kernel” - this kernel requires the selection of coefficient and gamma parameters.

The linear kernel was selected for this study based on the result achieved by the South African study conducted by Moepya, Nelwamondo and Van der Walt (2014:47-50). It was established that among various methods tested, the linear kernel support vector machine model showed the least amount of deterioration, and outperformed all other test models, including the radial basis function kernel support vector machine. The average accuracy score for the linear kernel support vector machine model was 89.66% compared with a 79.31% accuracy score for the radial basis function kernel support vector machine.

- In the final step, in the missing data tab, XLSTAT® allows for two options:

    - "Do not accept missing data" – This option is activated for XLSTAT® not to continue with calculations if missing values have been detected.
    - "Remove observations" – This option is selected to remove observations with missing data.

  The "remove observations" option was selected for this study. Where a missing value (financial, market or non-financial variable) due to unavailability of data for one or another reason is detected in an observation period, the sample company is removed from the total sample for the purposes of a particular calculation.

The classification and analysis of each of the SVM-K-score, SVM-Z-score and F-score models are discussed in the following section.

10.3 CLASSIFICATION OF THE SVM-K-SCORE FINANCIAL DISTRESS PREDICTION MODEL

10.3.1 <u>Background</u>

The support vector machine differentiates between two binary states of financial health – a company can either be financial distressed or non-distressed. Zero (0) and one (1) were used to identify non-distressed and financially distressed companies respectively. The non-distressed category includes both the healthy and depressed companies.

10.3.2 <u>Financially distressed companies</u>

The lower cut-off point or where the company's SVM-K-score was below -0.19 was determined to reflect a distressed state. All companies with a SVM-K-score below -0.19 in 2014, the final year of the test period, was assigned a "1". The following sample companies in Table 10.1 were identified as being distressed during 2014 (see Appendix 10.1 for details):

Table 10.1: SVM-K-score model – financially distressed companies in T (2014)

Adcorp	Delta	Moneyweb	RCL Food
Aveng	Datatec	Nictus	Rex Trueform
Altron	Gijima	Nampak	Seardel
AH-Vest	Grindrod	Naspers	Winhold
Astrapak	JD Group	P&P	
Beige	Jasco	Primeserv	

Source: Own compilation.

These 22 companies represented 25.9% of the total 85 sample companies.

10.3.3 Non-distressed companies[16]

The SVM-K-score model categorised companies to be depressed between -0.19 and +0.20 and any company with an SVM-K-score above +0.20 was regarded as relatively safe or healthy.

Because the support vector machine required a binary classification, the depressed and healthy states were combined into one single classification, denoted by "0".

The following sample companies in Table 10.2 were identified and categorised into this one single non-distressed state (see Appendix 10.1 for details):

[16] Hereafter, non-distressed companies include depressed and healthy companies.

Table 10.2: SVM-K-score model – non-distressed companies in T (2014)

Adaptit	Datacentrix	Micromega	Steinhoff
AECI	Digicore	Mr Price	Sovereign
Afrox	Distell	Massmart	Spanjaard
African Media	EOH	Mustek	Super Group
Astral Food	Famous Brands	Metair	Spar
Business Connextion	Hudaco	MTN	Stratcorp
Bell	Howden	M&R	Spur
Basil Read	Iliad	Netcare	Silverbridge
Bidvest	Illovo	Nu-World	Tiger Brands
Caxton	ISA	Onelogix	Tongaat
Crookes	Italtile	Omnia	Transpaco
City Lodge	KAP	Phumelela	Truworths
Clicks	Lewis	Pinnacle	Tsogo
Cargo	Masonite	PPC	Value
Cashbuild	Mediclinic	Remgro	Woolworths
Cullinan	Metrofile	Reunert	

Source: Own compilation.

These 63 companies represent 74.1% of the total 85 sample companies

10.4 CLASSIFICATION OF THE SVM-Z-SCORE FINANCIAL DISTRESS PREDICTION MODEL

10.4.1 Background

As described in the case of the SVM-K-score in Section 10.3.1, the support vector machine differentiates between two binary states of financial health – a company can either be financially distressed or non-distressed. Zero (0) and one (1) were used to identify non-distressed and financially distressed companies respectively.

10.4.2 Financially distressed companies

The lower cut-off or where the SVM-Z-score was below 1.81 was determined to reflect a company being in a distressed state. All companies with an SVM-Z-score below 1.81 in 2014, the final year of the test period, was assigned a "1".

The following sample companies in Table 10.3 were identified as being in a distressed state during 2014 (see Appendix 10.2 for details):

Table 10.3: SVM-Z-score model – financially distressed companies in T (2014)

AH-Vest	Grindrod
Beige	Nictus
Delta	Seardel
Gijima	Stratcorp

Source: Own compilation.

These eight companies represented 9.4% of the total 85 sample companies.

10.4.3 Non-distressed companies

Companies categorised in the depressed category were those that had SVM-Z-scores between 1.81 and 2.99. Companies with an SVM-Z-score higher than 2.99 were categorised as non-distressed companies.

The following sample companies in Table 10.4 were identified and categorised into this one single non-distressed state (see Appendix 10.2 for details):

Table 10.4: SVM-Z-score model – non-distressed companies in T (2014)

Adaptit	Cullinan	Metrofile	PPC
Adcorp	Datacentrix	Micromega	RCL Food
AECI	Digicore	Moneyweb	Remgro
Afrox	Distell	Mr Price	Reunert
African Media	Datatec	Massmart	Rex Trueform
Altron	EOH	Metair	Silverbridge
Astrapak	Famous Brands	MTN	Sovereign
Astral Food	Howden	Mustek	Spanjaard
Aveng	Hudaco	M&R	Spar
Basil Read	Iliad	Nampak	Spur
Business Connextion	Illovo	Naspers	Steinhoff
Bell	ISA	Netcare	Supergroup
Bidvest	Italtile	Nu-world	Tiger Foods
Cargo	JD Group	Onelogix	Tongaat
Caxton	Jasco	Omnia	Transpaco
Crookes	KAP	Phumelela	Truworths
City Lodge	Lewis	P&P	Tsogo
Clicks	Masonite	Primeserv	Value
Cashbuild	Mediclinic	Pinnacle	Winhold
			Woolworths

Source: Own compilation.

These 77 companies in the non-distressed state represented 90.6% of the total 85 sample companies.

The process followed in the current study in setting up and training of the support vector machine classifiers is described is the following section.

10.5 CLASSIFICATION OF THE F-SCORE FINANCIAL DISTRESS PREDICTION MODEL

10.5.1 Background

As described in the case of the SVM-K-score and SVM-Z-score in Sections 10.3.1 and 10.4.1, the support vector machine differentiates between two binary states of financial health – a company can either be financially distressed or non-distressed. Zero (0) and one (1) were used to identify non-distressed and financially distressed companies respectively.

In the test, a binary classification was done on each of the F-score models:

- financial variables (Model 1);
- financial and market variables (Model 2);
- financial and market variables plus the leading business cycle indicator as a non-financial variable (Model 3);
- financial and market variables plus the coincident business cycle indicator as a non-financial variable (Model 4);
- financial and market variables plus the lagging business cycle indicator as a non-financial variable (Model 5).

In determining the financial distressed state, the financial results for each of the sample companies over the 10-year period from 2005 to 2014 were ranked by applying a quartile ranking method. Particular attention was given to the first quartile or 25th percentile. All companies grouped within the 25th percentile of the current year (2014) were considered candidates for the distressed state. However, relying purely on the 25th percentile resulted in a too broad distressed category. The category was therefore narrowed and limited to all companies below the median value of the 25th percentile of the current year. Based on this view, a company reflecting a financial result below the median of the 25th percentile was regarded as a poor performer or distressed.

Reporting a result within the distressed category for a single financial year was regarded as adequate and sufficient to adversely affect a movement in a company's share price (Naidoo, 2006:650). For the purposes of the study, T was represented by the F-score for 2014.

A company was therefore categorised as distressed if:

T = F-score < median of the 25th percentile for year n

The results of the binary classification for each of the F-score models are described separately in the following section.

10.5.2 <u>Financial variables (Model 1)</u>

(i) Financially distressed companies

The 25th percentile of the F-score values were determined for 2014 and resulted in the lowest ranked value of -1.658 and the highest of 1.890. All companies with a value of less than 0.84 in the T were categorised to be within the distressed category.

The following sample companies in Table 10.5 with a median value of less than 0.84 were categorised as distressed (see Appendix 10.3 for details):

Table 10.5: F-score Model 1 – financially distressed companies in T (2014)

Aveng	Grindrod	Naspers	Seardel
AECI	Hudaco	Nu-world	Sovereign
Basil Read	Italtile	Omnia	Spanjaard
Caxton	Jasco	Phumelela	Spar
Crookes	Masonite	RCL Food	Stratcorp
Cullinan	Moneyweb	Remgro	Tiger
Distell	Mustek	Reunert	Tongaat
Gijima	Metair	Rex Trueform	

<u>Source</u>: Own compilation.

These 31 companies represented 36.5% of the total sample of 85 companies.

(ii) Non-distressed companies

Companies not categorised in the distressed category as described above were regarded as having a temporary depressed or positive growth rate in the F-score, and therefore categorised as being in the non-distressed category.

The following sample companies in Table 10.6 with a value of more than 0.84 were categorised as non-distressed (see Appendix 10.3 for details):

Table 10.6: F-score Model 1 – non-distressed companies in T (2014)

Adaptit	Cargo	Lewis	Pinnacle
Adcorp	Cashbuild	Mediclinic	PPC
Altron	Datacentrix	Metrofile	Steinhoff
Afrox	Digicore	Micromega	Super Group
AH-Vest	Delta	Mr Price	Spur
African Media	Datatec	Massmart	SilverBridge
Astrapak	EOH	MTN	Transpaco
Astral Foods	Famous Brands	M&R	Truworths
Business Connextion	Howden	Nictus	Tsogo
Beige	Iliad	Nampak	Value
Bell	Illovo	Netcare	Woolworths
Bidvest	ISA	OneLogix	Winhold
City Lodge	JD Group	P&P	
Clicks	KAP	Primeserv	

Source: Own compilation.

These 54 companies represented 63.5% of the total sample of 85 companies.

10.5.3 Financial and market variables (Model 2)

(i) Financially distressed companies

The 25th percentile of the F-score values for the combined financial and market variables were determined for 2014 and resulted in the lowest ranked value of -1.529 and the highest of 2.487. All companies with a median value of less than 1.073 in the T were categorised to be within the distressed state.

The following sample companies in Table 10.7 with a median value of less than 1.073 were categorised as distressed (see Appendix 10.4 for details):

Table 10.7: F-score Model 2 – financially distressed companies in T (2014)

Aveng	Cullinan	KAP	Sovereign
Afrox	Delta	Masonite	Spanjaard
AH-Vest	Gijima	MoneyWeb	Spar
Astrapak	Grindrod	Mustek	StratCorp
Business Connextion	Hudaco	Phumelela	Winhold
Bell	Italtile	RCL Foods	
Basil Read	JD Group	Rex Trueform	
Caxton	Jasco	Seardel	

Source: Own compilation.

These 29 companies represented 34.1% of the total sample of 85 companies.

(ii) Non-distressed companies

Companies not categorised in the distressed state as described above were regarded as having a temporary depressed and positive growth rate in the F-score, and therefore categorised as being in the non-distressed state.

The following sample companies in Table 10.8 with a value of more than 1.073 were categorised as non-distressed (see Appendix 10.4 for details):

Table 10.8: F-score Model 2 – non-distressed companies in T (2014)

Adaptit	Digicore	Massmart	PPC
Adcorp	Distell	Metair	Remgro
Altron	Datatec	MTN	Reunert
AECI	EOH	M&R	Steinhoff
African Media	Famous Brands	Nictus	Super Group
Astral Foods	Howden	Nampak	Spur
Beige	Iliad	Naspers	SilverBridge
Bidvest	Illovo	Netcare	Tiger Brands
Crookes	ISA	Nu-World	Tongaat
City Lodge	Lewis	OneLogix	Transpaco
Clicks	Mediclinic	Omnia	Truworths
Cargo	Metrofile	P&P	Tsogo
Cashbuild	Micromega	Primeserv	Value
Datacentrix	Mr Price	Pinnacle	Woolworths

Source: Own compilation.

These 56 companies represented 65.9% of the total sample of 85 companies.

10.5.4 Financial, market and non-financial variables (Leading business cycle indicator) (Model 3)

(i) Financially distressed companies

The leading business cycle indicators in standardised format were added to the financial and market indicators.

The 25th percentile of the F-score values for these combined were determined for 2014 and resulted in the lowest ranked value of -0.646 and the highest of 3.242. All companies with a median value of less than 2.011 in the T were categorised to be within the distressed state.

The following sample companies in Table 10.9 with a median value of less than 2.011 were categorised as distressed (see Appendix 10.5 for details):

Table 10.9: F-score Model 3 – financially distressed companies in T (2014)

Aveng	Cullinan	Jasco	Rex Trueform
Afrox	Delta	KAP	Seardel
AH-Vest	Gijima	Masonite	Sovereign
Business Connextion	Grindrod	MoneyWeb	Spanjaard
Bell	Hudaco	Mustek	Spar
Basil Read	Italtile	Phumelela	Stratcorp
Caxton	JD Group	RCL Foods	Winhold

Source: Own compilation.

These 28 companies represented 32.9% of the total sample of 85 companies.

(ii) Non-distressed companies

Companies not categorised in the distressed state as described above were regarded as having a temporary depressed and positive growth rate in the F-score, and therefore categorised as being in the non-distressed state.

The following sample companies in Table 10.10 with a value of more than 2.011 were categorised as non-distressed (see Appendix 10.5 for details):

Table 10.10: F-score Model 3 – non-distressed companies in T (2014)

Adaptit	Digicore	Metair	Reunert
Adcorp	Distell	MTN	Steinhoff
Altron	Datatec	M&R	Super Group
AECI	EOH	Nictus	Spur
African Media	Famous Brands	Nampak	SilverBridge
Astrapak	Howden	Naspers	Tiger Brands
Astral Foods	Iliad	Netcare	Tongaat
Beige	Illovo	Nu-World	Transpaco
Bidvest	ISA	OneLogix	Truworths
Crookes	Lewis	Omnia	Tsogo
City Lodge	Mediclinic	P&P	Value
Clicks	Metrofile	Primeserv	Woolworths
Cargo	Micromega	Pinnacle	
Cashbuild	Mr Price	PPC	
Datacentrix	Massmart	Remgro	

Source: Own compilation.

These 57 companies represented 67.1% of the total sample of 85 companies.

10.5.5 Financial, market and non-financial variables (Coincident business cycle indicator) (Model 4)

(i) Financially distressed companies

The coincident business cycle indicators in standardised format were added to the financial and market indicators.

The 25th percentile of the F-score values for these combined were determined for 2014 and resulted in the lowest ranked value of -0.566 and the highest of 3.503. All companies with a median value of less than 2.076 in the T were categorised to be within the distressed state.

The following sample companies in Table 10.11 with a median value of less than 2.076 were categorised as distressed (see Appendix 10.6 for details):

Table 10.11: F-score Model 4 – financially distressed companies in T (2014)

Aveng	Grindrod	MoneyWeb	Sovereign
Basil Read	Hudaco	Mustek	Spanjaard
Caxton	Italtile	Phumelela	Spar
Cullinan	JD Group	RCL Foods	Stratcorp
Delta	Jasco	Rex Trueform	Winhold
Gijima	Masonite	Seardel	

Source: Own compilation.

These 23 companies represented 27.1% of the total sample of 85 companies.

(ii) Non-distressed companies

Companies not categorised in the distressed state as described above were regarded as having a temporary depressed and positive growth rate in the F-score, and therefore categorised as being in the non-distressed state.

The following sample companies in Table 10.12 with a value of more than 2.076 were categorised as non-distressed (see Appendix 10.6 for details):

Table 10.12: F-score Model 4 – non-distressed companies in T (2014)

Adaptit	Cargo	Lewis	PPC
Adcorp	Cashbuild	Mediclinic	Remgro
Altron	Datacentrix	Metrofile	Reunert
AECI	Digicore	Micromega	Steinhoff
Afrox	Distell	Mr Price	Super Group
AH-Vest	Datatec	Massmart	Spur
African Media	EOH	Nictus	SilverBridge
Astrapak	Famous Brands	Nampak	Tiger Brands
Astral Foods	Howden	Naspers	Tongaat
Business Connextion	Iliad	Netcare	Transpaco
Beige	Illovo	Nu-World	Truworths
Bell	ISA	OneLogix	Tsogo
Bidvest	KAP	Omnia	Value
Crookes	Metair	P&P	Woolworths
City Lodge	MTN	Primeserv	
Clicks	M&R	Pinnacle	

Source: Own compilation.

These 62 companies represented 72.9% of the total sample of 85 companies.

10.5.6 Financial, market and non-financial variables (Lagging business cycle indicator) (Model 5)

(i) Financially distressed companies

The lagging business cycle indicators in standardised format were added to the financial and market indicators.

The 25th percentile of the F-score values for these combined were determined for 2014 and resulted in the lowest ranked value of -0.425 and the highest of 3.596. All companies with a median value of less than 2.151 in the T were categorised to be within the distressed state.

The following sample companies in Table 10.13 with a median value of less than 2.151 were categorised as distressed (see Appendix 10.7 for details):

Table 10.13: F-score Model 5 – financially distressed companies in T (2014)

Aveng	Cullinan	KAP	Rex Trueform
Afrox	Delta	Masonite	Seardel
AH-Vest	Gijima	Metrofile	Sovereign
Astrapak	Grindrod	MoneyWeb	Spanjaard
Business Connextion	Hudaco	Mustek	Spar
Bell	Italtile	Nu-World	Stratcorp
Basil Read	JD Group	Phumelela	Winhold
Caxton	Jasco	RCL Foods	

Source: Own compilation.

These 31 companies represented 36.47% of the total sample of 85 companies.

(ii) Non-distressed companies

Companies not categorised in the distressed state as described above were regarded as having a temporary depressed and positive growth rate in the F-score, and therefore categorised as being in the non-distressed state.

The following sample companies in Table 10.1 with a value of more than 2.151 were categorised as non-distressed (see Appendix 10.7 for details):

Table 10.14: F-score Model 4 – non-distressed companies in T (2014)

Adaptit	Digicore	Metair	Reunert
Adcorp	Distell	MTN	Steinhoff
Altron	Datatec	M&R	Super Group
AECI	EOH	Nictus	Spur
African Media	Famous Brands	Nampak	SilverBridge
Astral Foods	Howden	Naspers	Tiger Brands
Beige	Iliad	Netcare	Tongaat
Bidvest	Illovo	OneLogix	Transpaco
Crookes	ISA	Omnia	Truworths
City Lodge	Lewis	P&P	Tsogo
Clicks	Mediclinic	Primeserv	Value
Cargo	Micromega	Pinnacle	Woolworths
Cashbuild	Mr Price	PPC	
Datacentrix	Massmart	Remgro	

Source: Own compilation.

These 54 companies represented 63.5% of the total sample of 85 companies.

10.5.7 Summary

The number of distressed on non-distressed sample companies per financial distress prediction model is summarised in Table 10.15. A binary classification, as required by the support vector machine, was used to categorise distressed and non-distressed companies: one (1) for distressed companies and zero (0) for non-distressed companies.

Table 10.15: Number of companies in a distressed and non-distressed state for the SVM-K-score, SVM-Z-score and F-score models

Financial distress prediction model	Distressed (1)	Non-distressed (0)	TOTAL
SVM-K-score	22	63	85
SVM-Z-score	8	77	85
F-score (Model 1)	31	54	85
F-score (Model 2)	29	56	85
F-score (Model 3)	28	57	85
F-score (Model 4)	23	62	85
F-score (Model 5)	31	54	85

Source: Own compilation.

Sample companies identified as distressed by the SVM-K and SVM-Z-score models, and also identified by the F-score models (Model 1 to Model 5) are summarised in Table 10.16. The table reflects the number of distressed companies by the SVM-K and SVM-Z models and also this number as a percentage of the total number of distressed companies identified by the various F-score models.

The analysis in Table 10.16 produced a disappointing result similar to the result achieved in Chapter 8. The single difference in the F-score model comparison with the SVM-K-score and SVM-Z-score in Chapter 8 was that the F-score model was based on financial variables only – Model 1 in the analysis.

Table 10.16: Comparison of the number of companies identified as distressed by the F-score model and the SVM-K-score and SVM-Z-score models

F-score	F-score distressed companies (number)	Distressed companies identified by F- and SVM-K-score models (number)	Distressed companies identified by F- and SVM-K-score models (%)	Distressed companies identified by F- and SVM-Z-score models (number)	Distressed companies identified by F- and SVM-Z-score models (%)
Model 1	31	8	25.8	3	9.7
Model 2	29	13	44.8	6	20.7
Model 3	28	12	42.9	6	21.4
Model 4	23	11	47.8	5	21.7
Model 5	31	13	41.9	6	19.4

Source: Own compilation.

The F-score model was expanded in the above analysis to include:

- financial variables (Model 1);
- financial plus market variables (Model 2);
- financial and market variables plus a leading business cycle indicator (Model 3);
- financial and market variables plus a coincident business cycle indicator (Model 4);
- financial and market variables plus a lagging business cycle indicator (Model 5).

A noticeable improvement was achieved in the result, albeit somewhat erratic, when the test was conducted on Models 2, 3, 4 and 5. The best result was achieved with a comparison between the SVM-K-score distressed companies and the F-score Model 4. A similar acceptable result was achieved between the SVM-Z-score and the F-score Model 4. Both SVM-K-score and SVM-Z-score deteriorated again with the Model 5 test.

In the next section, the classification accuracy of the various financial distress prediction models will be examined by using the support vector machine. The results of the SVM-K-score model will be examined, followed by an examination of the SVM-Z-score results. The classification accuracy of each of the five F-score models will be examined in the final section and compared with the results of the SVM-K and SVM-Z-score models.

10.6 CLASSIFICATION ACCURACY OF THE SVM-K-SCORE AND SVM-Z-SCORE FINANCIAL DISTRESS PREDICTION MODEL

10.6.1 Background

XLSTAT® was applied in this test and divided into two subsections. The first test determined the classification accuracy for the full 10-year test period, followed by testing T-1, T-2, T-3 and T-5 periods to determine whether there were changes in the classification accuracy. The SVM-K-score and SVM-Z-score models were validated in the second test, first, over the full 10-year test period, followed by testing T-1, T-2, T-3 and T-5 periods.

The following parameters were applied to both tests:

- The regularisation parameter "C" was set at the default value one.
- The numerical precision parameter "epsilon" was set at the default value of 1e-12.
- The "tolerance" parameter defining the tolerance or accuracy of the optimisation algorithm when comparing support vectors was set at the default value of 0.001.

The "standardisation" and "linear kernel" options were selected in the "pre-processing" and "kernel" fields respectively.

10.6.2 Optimised classifiers for the SVM-K-score and SVM-Z-score models

Tables 10.17 and 10.18 display the optimised support vector machine classifiers. The binary classifiers classified each company as either zero (0) or one (1). One represented the companies in financial distress; and zero, the positive class, represented the non-distressed companies.

Table10.17: Summary of optimised support vector machine classifiers for the SVM-K-score model

	T	T-1	T-2	T-3	T-5
Positive class	0	0	0	0	0
Number of observations in the training set	76	78	79	79	80
Number of removed observations	9	7	6	6	5
Bias	-1.364	-1.228	-1.510	-1.506	-1.000
Number of support vectors	41	41	43	42	54

Source: Own compilation.

Table10.18: Summary of optimised support vector machine classifiers for the SVM-Z-score model

	T	T-1	T-2	T-3	T-5
Positive class	0	0	0	0	0
Number of observations in the training set	79	79	80	80	81
Number of removed observations	6	6	5	5	4
Bias	-1.001	-0.999	-1.106	-1.099	-1.163
Number of support vectors	23	24	20	19	18

Source: Own compilation.

The first column in both tables, denoted by T, tested the full 10-year observation period against the 2014 base year, denoted by T. In Table 10.17, nine companies or observations were removed due to missing data and only 76 companies were absorbed and used to train the classifiers out of which 41 support vectors were identified. In Table 10.18, six companies or observations were removed due to missing data and only 79 companies were absorbed and used to train the classifiers out of which 23 support vectors were identified.

The second column in Table 10.17 and Table 10.18, denoted by T-1, tested the first nine years (2005 to 2013) against T. In Table 10.17, seven companies or observations were removed due to missing data and 78 companies were absorbed and used to train the classifiers out of which 41 support vectors were identified. In Table 10.18, six companies or observations were removed due to missing data and 79 companies were absorbed and used to train the classifiers out of which 24 support vectors were identified.

The third column in both tables, denoted by T-2, tested the first eight years (2005 to 2012) against T. In Table 10.17, six companies or observations were removed due to missing data and 79 companies were absorbed and used to train the classifiers out of which 43 support vectors were identified. In Table 10.18, five companies or observations were removed due to missing data and 80 companies were absorbed and used to train the classifiers out of which 20 support vectors were identified.

The fourth column in Table 10.17 and Table 10.18, denoted by T-3, tested the first seven years (2005 to 2011) against T. In Table 10.17, six companies or observations were removed due to missing data and 79 companies were absorbed and used to train the classifiers out of which 42 support vectors were identified. In Table 10.18, five companies or observations were removed due to missing data and 80 companies were absorbed and used to train the classifiers out of which 19 support vectors were identified.

Finally, the fifth column of Table 10.17 and Table 10.18, denoted by T-5, tested the first seven years (2005 to 2009) against T. In Table 10.17, five companies or observations were removed due to missing data and 80 companies were absorbed and used to train the classifiers out of which 54 support vectors were identified. In Table 10.18, four companies or observations were removed due to missing data and 81 companies were absorbed and used to train the classifiers out of which 18 support vectors were identified.

For the SVM-K-score, the bias produced erratic results; it moved towards an unbiased position (zero) between T and T-1, but then deteriorated during the T-2 and T-3 test period, and then recorded -1.000 during the T-5. As in the instance of the SVM-K-score model, the bias produced erratic results for the SVM-Z-score model; it moved towards an unbiased position (zero) between T and T-1, but then deteriorated during the T-2 and T-3 test period, and then recorded -1.163 during the T-5.

10.6.3 Performance of the SVM-K-score and SVM-Z-score classifiers

The performance of the SVM-K-score and SVM-Z-score classifiers is reviewed in this section. The tables in this section display the confusion matrix for the SVM-K-score and SVM-Z-score training sample for the full test period (T), followed by the periods (T-1), (T-2), (T-3) and (T-5).

The effect of the variable selection on prediction accuracy will reflect in an increase or decrease in the "Sensitivity" and "Specificity" (see Section 6.2.2).

The first line of each table displays the "Sensitivity" – the true positive. This is the proportion of the sample companies correctly or positively identified as distressed companies.

The second line in each table displays the "Specificity" – the true negative. This is the proportion of sample companies correctly classified as non-distressed companies.

The third line displays the overall classification accuracy.

Table 10.19 displays an extract from the confusion matrix for the SVM-K-score training sample for the full test period (T), followed by the periods (T-1), (T-2), (T-3) and (T-5). The period (T) was used as the base period, whereafter the forecast accuracy of each period was compared with that of the base year (T).

Table 10.19: SVM-K-score training sample (distressed/non-distressed – 0/1) for T to T-5 – extract from the confusion matrix

	T	T-1	T-2	T-3	T-5
	Correct (%)	Correct (%)	Correct (%)	Correct (%)	Correct (%)
Sensitivity	16.67	11.11	21.05	21.05	0.00
Specificity	100.00	100.00	96.67	96.67	100.00
Overall accuracy	80.26	79.49	78.48	78.48	73.75
Type I error	83.33	88.89	78.95	78.95	100.00
Type II error	0.00	0.00	3.33	3.33	0.00

Source: Own compilation.

In Table 10.19, "Sensitivity" was 16.67% in T, deteriorated during T-1, and increased again to 21.05% in T-2 and T-3. "Specificity" was 100% in T, T-1 and T-5 and lower at 96.67% in T-2 and T-3. The overall accuracy of 80.26% was recorded in T and reduced gradually to 73.75% in T-5. The Type I error increased from 83.33% in T to 100% in T-5. The Type II error was zero in T, T-1 and T-5 and marginally higher at 3.33% in T-2 and T-3.

Table 10.20 displays an extract from the confusion matrix for the SVM-Z training sample for the full test period (T), followed by the periods (T-1), (T-2), (T-3) and (T-5). T was used as the base period, whereafter the forecast accuracy of each period was compared with T.

Table 10.20: SVM-Z-score training sample (distressed/non-distressed – 0/1) for T to T-5 – extract from the confusion matrix

	T	T-1	T-2	T-3	T-5
	Correct (%)	Correct (%)	Correct (%)	Correct (%)	Correct (%)
Sensitivity	0.00	0.00	12.50	12.50	12.50
Specificity	100.00	100.00	98.61	98.61	98.63
Overall accuracy	89.87	89.87	90.00	90.00	90.12
Type I error	100.00	100.00	87.50	87.50	87.50
Type II error	0.00	1.39	1.39	1.39	1.37

Source: Own compilation.

In Table 10.20 for T, “Sensitivity” was 0%, “Specificity” 100% and the overall accuracy recorded at 89.87%. Of the total eight companies, none were correctly identified as distressed. The Type I error was recorded at 100%, where all eight non-distressed companies were incorrectly classified as distressed.

10.7 CLASSIFICATION ACCURACY OF THE F-SCORE FINANCIAL DISTRESS PREDICTION MODEL

10.7.1 <u>Background</u>

As in determining the classification accuracy of the SVM-K-score and SVM-Z-score models, a similar process was followed with the F-score model.

For the F-score, the test on the SVM-K-score and SVM-Z-score model was based on the following separate categories:

- financial variables (Model 1);
- financial and market variables (Model 2);
- financial and market variables plus leading business cycle indicators (Model 3);
- financial and market variables plus coincident business cycle indicators (Model 4);
- financial and market variables plus lagging business cycle indicators (Model 5).

XLSTAT® was used in the test to determine the classification accuracy for the full 10-year test period, followed by testing T-1, T-2, T-3 and T-5 periods to determine whether there were changes in the classification accuracy.

The following parameters were applied to each of the above categories:

- The regularisation parameter “C” was set at the default value one.
- The numerical precision parameter “epsilon” was set at the default value of 1e-12.

- The “tolerance” parameter defining the tolerance or accuracy of the optimisation algorithm when comparing support vectors was set at the default value of 0.001.

The “standardisation” and “linear kernel” options were selected in the “pre-processing” and “kernel” fields respectively.

10.7.2 Optimised classifiers for the F-score model

A summary of the optimised support vector machine classifiers for the F-score is discussed in this section. Each category of the F-score model referred to in Section 10.7.1 is discussed separately, supported by a table.

The first column (T) to the fifth column (T-5) in each table tested each period against the 2014 base year, starting with the full 10-year test period followed by one year forward to five years forward. Each column reflects the number of companies or observations removed due to missing data where the remaining companies were absorbed and used to train the classifiers out of which the support vectors were identified in the last row of each table.

The binary classifiers classified each observation or company as either zero (0) or one (1). One represented the companies in financial distress; and zero, the positive class, represented the non-distressed companies.

(i) F-score based on financial variables (Model 1)

Table 10.21 provides a summary of the optimised support vector machine classifiers. The binary classifiers classified each observation or company as either zero (0) or one (1). One represented the companies in financial distress; and zero, the positive class, represented the non-distressed companies.

Table 10.21: Summary of optimised support vector machine classifiers for the F-score Model 1

	T	T-1	T-2	T-3	T-5
Positive class	0	0	0	0	0
Number of observations in the training set	79	79	80	80	81
Number of removed observations	6	6	5	5	4
Bias	-0.173	-0.633	-0.991	-1.333	-1.036
Number of support vectors	51	51	58	56	61

Source: Own compilation.

Seventy-nine companies or observations in T, increasing to 81 companies in T-5, were absorbed and used to train classifiers after between four and six companies were removed due to missing data. The number of support vectors identified varied between 51 in T and 61 in T-5. The bias deteriorated between T and T-3 and improved marginally in T-5.

(ii) F-score based on financial and market variables (Model 2)

Table 10.22 provides a summary of the optimised support vector machine classifiers. The binary classifiers classified each observation or company as either zero (0) or one (1). One represented the companies in financial distress; and zero, the positive class, represented the non-distressed companies.

Table 10.22: A summary of optimised support vector machine classifiers for the F-score Model 2

	T	T-1	T-2	T-3	T-5
Positive class	0	0	0	0	0
Number of observations in the training set	85	85	85	85	85
Number of removed observations	0	0	0	0	0
Bias	4.384	2.431	1.218	1.329	-0.998
Number of support vectors	30	53	59	59	65

Source: Own compilation.

In this test, market variables were added to the financial variables. From T to T-5, all 85 sample companies were absorbed and used to train classifiers. The number of support vectors identified varied drastically between 30 in T and 65 in T-5. The bias moved from a high 4.384 in T towards -0.998 in T-5.

(iii) F-score based on financial and market variables plus the leading business cycle indicator (Model 3)

Table 10.23 provides a summary of the optimised support vector machine classifiers. The binary classifiers classified each observation or company as either zero (0) or one (1). One represented the companies in financial distress; and zero, the positive class, represented the non-distressed companies.

Table 10.23: A summary of optimised support vector machine classifiers for the F-score Model 3

	T	T-1	T-2	T-3	T-5
Positive class	0	0	0	0	0
Number of observations in the training set	85	85	85	85	85
Number of removed observations	0	0	0	0	0
Bias	8.835	4.656	1.874	2.786	-0.986
Number of support vectors	29	53	57	58	61

Source: Own compilation.

In this test, a leading business cycle indicator representing a quantitative non-financial variable was added to the financial and market variables. From T to T-5, all 85 sample companies were absorbed and used to train classifiers. The number of support vectors identified varied drastically between 29 in T and 61 in T-5. The bias moved from a high 8.835 in T towards -0.986 in T-5.

(iv) F-score based on financial and market variables plus the coincident business cycle indicator (Model 4)

Table 10.24 provides a summary of the optimised support vector machine classifiers. The binary classifiers classified each observation or company as either zero (0) or one (1). One represented the companies in financial distress; and zero, the positive class, represented the non-distressed companies.

Table 10.24: A summary of optimised support vector machine classifiers for the F-score Model 4

	T	T-1	T-2	T-3	T-5
Positive class	0	0	0	0	0
Number of observations in the training set	85	85	85	85	85
Number of removed observations	0	0	0	0	0
Bias	7.066	1.417	-0.990	-0.992	-1.000
Number of support vectors	30	48	52	53	56

Source: Own compilation.

In this test, a coincident business cycle indicator representing a quantitative non-financial variable was added to the financial and market variables. From T to T-5, all 85 sample companies were absorbed and used to train classifiers. The number of support vectors identified varied drastically between 30 in T and 56 in T-5. As in the previous tests, the bias moved from a high 7.066 in T towards -1.000 in T-5.

(v) F-score based on financial and market variables plus the lagging business cycle indicator (Model 5)

Table 10.25 provides a summary of the optimised support vector machine classifiers. The binary classifiers classified each observation or company as either zero (0) or one (1). One represented the companies in financial distress; and zero, the positive class, represented the non-distressed companies.

Table 10.25: A summary of optimised support vector machine classifiers for the F-score Model 5

	T	T-1	T-2	T-3	T-5
Positive class	0	0	0	0	0
Number of observations in the training set	85	85	85	85	85
Number of removed observations	0	0	0	0	0
Bias	7.624	4.056	3.249	2.870	-0.228
Number of support vectors	31	56	62	62	65

Source: Own compilation.

In this test, a lagging business cycle indicator representing a quantitative non-financial variable was added to the financial and market variables. From T to T-5, all 85 sample companies were absorbed and used to train classifiers. The number of support vectors identified varied drastically between 31 in T and 65 in T-5. As in the previous tests, the bias moved from a high 7.624 in T towards -0.228 in T-5.

10.7.3 Performance of the F-score classifiers

The performance of the F-score classifiers is reviewed in this section. The tables display the confusion matrix for the F-score training sample for the full test period (T), followed by the periods (T-1), (T-2), (T-3) and (T-5).

The effect of the variable selection on prediction accuracy will reflect in an increase or decrease in the "Sensitivity" and "Specificity" (see Section 6.2.2).

The first line of each table displays the "Sensitivity" – the true positive. This is the proportion of the sample companies correctly or positively identified as non-distressed companies.

The second line in each table displays the "Specificity" – the true negative. This is the proportion of sample companies correctly classified as distressed companies.

The third line displays the overall classification accuracy.

(i) F-score based on financial variables (Model 1)

Table 10.26 displays an extract from the confusion matrix for the F-score Model 1 training sample for the full test period (T), followed by the periods (T-1), (T-2), (T-3) and (T-5). T was used as the base period, whereafter the forecast accuracy of each period was compared with T.

Table 10.26: F-score training sample (distressed/non-distressed – 0/1) for T to T-5 – extract from the confusion matrix

	T	T-1	T-2	T-3	T-5
	Correct (%)	Correct (%)	Correct (%)	Correct (%)	Correct (%)
Sensitivity	25.93	25.93	10.71	17.86	3.45
Specificity	100.00	98.08	100.00	96.15	100.00
Overall accuracy	74.68	73.42	68.75	68.75	65.43
Type I error	74.07	74.07	89.29	82.14	96.55
Type II error	0.00	1.92	0.00	3.85	0.00

Source: Own compilation.

In Table 10.26, "Sensitivity" recorded poor results with 25.93% for T and T-1 and deteriorated to 3.45% in T-5. "Specificity" was 100% for T, T-2 and T-5 and marginally deteriorated for T-1 and T-3. The overall accuracy deteriorated gradually from 74.68% in T to 65.43% in T-5. Type I errors increased gradually from 74.07% to 96.55% in T-5. A low Type II error was recorded for the full test period from T to T-5.

(ii) F-score based on financial and market variables (Model 2)

Table 10.27 displays an extract from the confusion matrix for the F-score Model 2 training sample for the full test period (T), followed by the periods (T-1), (T-2), (T-3) and (T-5). T was used as the base period, whereafter the forecast accuracy of each period was compared with T.

Table 10.27: F-score training sample (distressed/non-distressed – 0/1) for T to T-5 – extract from the confusion matrix

	T	T-1	T-2	T-3	T-5
	Correct (%)	Correct (%)	Correct (%)	Correct (%)	Correct (%)
Sensitivity	93.10	62.07	34.48	37.93	0.00
Specificity	100.00	89.29	89.29	89.29	89.29
Overall accuracy	97.65	80.00	70.59	71.76	65.88
Type I error	6.90	37.93	65.52	62.07	100.00
Type II error	0.00	10.71	10.71	10.71	0.00

Source: Own compilation.

In Table 10.27, “Sensitivity” deteriorated from 93.10% in T to 0.00% in T-5. “Specificity” was 100% for the T and T-5 and static at 89.29% for T-1, T-2 and T-3. The overall accuracy deteriorated from 97.65% in T to 65.88% in T-5. Type I errors increased gradually from 6.90% as the forecast period lengthened to 100.00% in T-5. A low Type II error was recorded for the full test period from T to T-5.

(iii) F-score based on financial and market variables plus a leading business cycle indicator (Model 3)

Table 10.28 displays an extract from the confusion matrix for the F-score Model 3 training sample for the full test period (T), followed by the periods (T-1), (T-2), (T-3) and (T-5). T was used as the base period, whereafter the forecast accuracy of each period was compared with T.

Table 10.28: F-score training sample (distressed/non-distressed – 0/1) for T to T-5 – extract from the confusion matrix

	T	T-1	T-2	T-3	T-5
	Correct (%)	Correct (%)	Correct (%)	Correct (%)	Correct (%)
Sensitivity	100.00	50.00	17.86	28.57	0.00
Specificity	100.00	91.23	94.74	91.23	100.00
Overall accuracy	100.00	77.65	69.41	70.59	67.06
Type I error	0.00	50.00	82.14	71.43	100.00
Type II error	0.00	8.77	5.26	8.77	0.00

Source: Own compilation.

In Table 10.28, “Sensitivity” deteriorated drastically as the forecast period lengthened from 100.00% in T to 0.00% in T-5. The change in “Specificity” was less volatile than in the case of the financial and market variables in Table 10.27, but marginally more volatile than for the financial variables in Table 10.26. The overall accuracy deteriorated from 100.00% in T to 67.06% in T-5. Type I errors increased gradually from 0.00% as the forecast period lengthened to 100.00% in T-5. A low Type II error was recorded for the full test period from T to T-5.

(iv) F-score based on financial and market variables plus a coincident business cycle indicator (Model 4)

Table 10.29 displays an extract from the confusion matrix for the F-score Model 4 training sample for the full test period (T), followed by the periods (T-1), (T-2), (T-3) and (T-5). T was used as the base period, whereafter the forecast accuracy of each period was compared with T.

Table 10.29: F-score training sample (distressed/non-distressed – 0/1) for T to T-5 – extract from the confusion matrix

	T	T-1	T-2	T-3	T-5
	Correct (%)	**Correct (%)**	**Correct (%)**	**Correct (%)**	**Correct (%)**
Sensitivity	100.00	13.04	0.00	0.00	0.00
Specificity	98.39	95.16	100.00	100.00	100.00
Overall accuracy	98.82	72.94	72.94	72.94	72.94
Type I error	0.00	86.96	100.00	100.00	100.00
Type II error	1.61	4.84	0.00	0.00	0.00

Source: Own compilation.

In Table 10.29, "Sensitivity" deteriorated drastically as the forecast period lengthened from 100.00% in T to 0.00% in T-5. The change in "Specificity" improved from 98.39% in T to 100.00% in T-5. The overall accuracy deteriorated from 98.82% in T and remained static on 72.94% from T-1 to T-5. Type I errors increased from 0.00% in T to 86.96% in T-1 and increased to 100% from T-2 to T-5. A low Type II error was recorded for the full test period from T to T-5.

(v) F-score based on financial and market variables plus a lagging business cycle indicator (Model 5)

Table 10.30 displays an extract from the confusion matrix for the F-score Model 5 training sample for the full test period (T), followed by the periods (T-1), (T-2), (T-3) and (T-5). T was used as the base period, whereafter the forecast accuracy of each period was compared with T.

Table 10.30: F-score training sample (distressed/non-distressed – 0/1) for T to T-5 – extract from the confusion matrix

	T	T-1	T-2	T-3	T-5
	Correct (%)	Correct (%)	Correct (%)	Correct (%)	Correct (%)
Sensitivity	100.00	58.06	48.39	48.39	0.00
Specificity	100.00	83.33	79.63	81.48	100.00
Overall accuracy	100.00	74.12	68.42	69.41	63.53
Type I error	0.00	41.94	51.61	51.61	100.00
Type II error	0.00	16.67	20.37	18.52	0.00

Source: Own compilation.

In Table 10.30, “Sensitivity” deteriorated drastically as the forecast period lengthened from 100.00% in T to 0.00% in T-5. A volatile trend in “Specificity” was recorded throughout the test period from T to T-5. The overall accuracy deteriorated from 100.00% in T to 63.53% in T-5. The Type I errors increased from 0.00% in T to 100.00% in T-5 as the forecast period lengthened. A higher Type II error was recorded for the full test period from T to T-5.

10.8 CONCLUSION

The confusion matrix for the training sample of each of the models for the full test period discussed above is summarised and reviewed in this section. The results of the models are ranked. Firstly, the overall results are ranked from high (best performer) to low (worst performer), and secondly, the Type I and II errors are ranked from low (best performer with least errors) to high (worst performer with most errors). Table 10.31 displays a summary of the overall accuracy of the SVM-K-score, SVM-Z-score and F-score (Model 1 to 5) models, ranked from the best to worst performer. Only the top three performers are reviewed.

Table 10.31: Ranked overall accuracy of the SVM-K-score, SVM-Z score and F-score (Model 1 to Model 5) models for T to T-5

	T		T-1		T-2		T-3		T-5		T to T-5	
	Correct (%)	Rank	Correct (%)	Rank	Correct (%)	Rank	Correct (%)	Rank	Correct (%)	Rank	Average	Rank
SVM-K-score	80.26	6	79.49	3	78.48	2	78.48	2	73.75	2	78.09	3
SVM-Z-score	89.87	5	89.87	1	90.00	1	90.00	1	90.12	1	89.97	1
F-score Model 1	74.68	7	73.42	6	68.75	6	68.75	7	65.43	6	70.21	7
F-score Model 2	97.65	4	80.00	2	70.59	4	71.76	4	65.88	5	77.18	4
F-score Model 3	100.00	1	77.65	4	69.41	5	70.59	5	67.06	4	76.94	5
F-score Model 4	98.82	3	72.94	7	72.94	3	72.94	3	72.94	3	78.12	2
F-score Model 5	100.00	2	74.12	5	68.42	7	69.41	6	63.53	7	75.10	6

Source: Own compilation.

In period T, the F-score - Model 3 ranked highest, followed by the F-score – Model 5 and F-score – Model 2. All the F-score models deteriorated over the test period.

Based on an average calculation (T to T-5), the SVM-Z-score ranked highest, followed by the F-score – Model 4 and SVM-K-score. The SVM-Z-score model consisted of a weighted compilation of financial and a single market variable and the second-ranked model, F-score, consisted of a weighted compilation of financial, market and non-financial variables. The third-ranked performer SVM-K-score consisted of a weighted compilation of financial variables only.

However, the overall accuracy as discussed above should not be reviewed in isolation.

Table 10.32 displays a summary of the Type I errors of the SVM-K-score, SVM-Z-score and F-score (Model 1 to 5) models, ranked from the best to worst performer. The model with the lowest Type I error is ranked first, and the highest Type I error ranked last. Only the top three performers are reviewed.

Table 10.32: Ranked Type I errors of the SVM-K-score, SVM-Z-score and F-score (Model 1 to Model 5) models for T to T-5

	T		T-1		T-2		T-3		T-5		T to T-5	
	Correct (%)	Rank	Correct (%)	Rank	Correct (%)	Rank	Correct (%)	Rank	Correct (%)	Rank	Average	Rank
SVM-K-score	83.33	6	88.89	6	78.95	3	78.95	4	100.00	7	86.02	6
SVM-Z-score	100.00	7	100.00	7	87.50	5	87.50	6	87.50	1	92.50	7
F-score Model 1	74.07	5	74.07	4	89.29	6	82.14	5	96.55	2	83.22	5
F-score Model 2	6.90	4	37.93	1	65.52	2	62.07	2	100.00	3	54.48	2
F-score Model 3	0.00	3	50.00	3	82.14	4	71.43	3	100.00	4	60.71	3
F-score Model 4	0.00	2	86.96	5	100.00	7	100.00	7	100.00	5	77.39	4
F-score Model 5	0.00	1	41.94	2	51.61	1	51.61	1	100.00	6	49.03	1

Source: Own compilation.

In period T, the F-score - Model 5 ranked highest, followed by the F-score – Model 4 and F-score – Model 3. The F-score – Model 5 ranking remained static over the full test period.

Based on an average calculation (T to T-5), the F-score – Model 5 ranked highest, followed by the F-score – Model 2 and F-score – Model 3. All the top performers, where non-distressed companies were incorrectly classified as distressed (i.e. Type I error), were based on one of the F-score models. Although the SVM-K-score and SVM-Z-score models recorded the highest overall accuracy, they recorded the highest Type I errors.

Table 10.33 displays a summary of the Type II errors (where a distressed company is incorrectly classified as a non-distressed company) of the SVM-K-score, SVM-Z-score and F-score (Model 1 to 5) models, ranked from the best to worst performer. The model with the lowest Type II error is ranked first, and the model with the highest Type II error ranked last. Only the top three performers are reviewed.

Table 10.33: Ranked Type II errors of the SVM-K-score, SVM-Z score and F-score (Model 1 to Model 5) models for T to T-5

	T		T-1		T-2		T-3		T-5		T to T-5	
	Correct (%)	Rank	Correct (%)	Rank	Correct (%)	Rank	Correct (%)	Rank	Correct (%)	Rank	Average	Rank
SVM-K-score	0.00	1	0.00	1	3.33	4	3.33	3	0.00	6	1.33	4
SVM-Z-score	0.00	2	1.39	2	1.39	3	1.39	2	1.37	7	1.11	1
F-score Model 1	0.00	3	1.92	3	0.00	1	3.85	4	0.00	1	1.15	2
F-score Model 2	0.00	4	10.71	6	10.71	6	10.71	6	0.00	2	6.43	6
F-score Model 3	0.00	5	8.77	5	5.26	4	8.77	5	0.00	3	4.56	5
F-score Model 4	1.61	7	4.85	4	0.00	2	0.00	1	0.00	4	1.29	3
F-score Model 5	0.00	6	16.67	7	20.37	7	18.52	7	0.00	5	11.11	7

Source: Own compilation.

All the models, except the F-score – Model 4, recorded no errors in period T. Most models continued to record low errors throughout the test period.

Based on an average calculation, the SVM-Z-score model and F-score – Model 1 ranked highest, followed by the F-score – Model 4 in third position. All the top-performing models comprised a weighted compilation of financial, market and non-financial variables.

CHAPTER 11

VALIDATION OF THE SVM-K-SCORE, SVM-Z-SCORE AND F-SCORE FINANCIAL DISTRESS PREDICTION MODELS

11.1 INTRODUCTION

The validation module in the XLSTAT® statistical software package was selected to establish how well the classifiers for each of the financial distress prediction models performed. The process followed in this study in setting up the validation process of the support vector machine classifiers is described is the following section.

In the Validation tab, the “validation” field is selected if a subset is to be used to validate the model. XLSTAT® has four options for defining the method for obtaining the observations to be used in the validation:

- **Random** – The observation is randomly selected by the model. The “Number of observations”, N, must be specified.
- **N last row** – The N last observations are selected for the validation. The linear kernel was selected for the study on the basis of this result
- **N first row** – The N first observations are selected for the validation. The linear kernel was selected for the study on the basis of this result
- **Group variable** – If this option is selected, a binary variable should be selected, with 0s and 1s. The 1s identify the observations for use in the validation.

The random option was selected for the purposes of the study. A total of 17 sample companies were selected as the number of observations to be randomly selected from the 85 sample companies for validation of the model (see Section 7.3.5 for an explanation of the method followed in identifying 17 observations).

The training and validation sample of each financial distress prediction model is discussed separately in the following section.

11.2 VALIDATION OF THE SVM-K-SCORE MODEL

Table 11.1 provides a summary of the optimised support vector machine classifiers. The binary classifiers classified each observation or company as either zero (0) or one (1). One represented the companies in financial distress; and zero, the positive class, represented the non-distressed companies.

Table 11.1: A summary of optimised support vector machine classifiers for the SVM-K-score

	T	T-1	T-2	T-3	T-5
Positive class	0	0	0	0	0
Number of observations in the training set	59	61	62	62	63
Number of removed observations	9	7	6	6	5
Bias	-1.195	-1.290	-1.225	-1.006	-1.000
Number of support vectors	34	37	37	30	40

Source: Own compilation.

Tables 11.2 and 11.3 provide a summary of "Sensitivity", "Specificity" and Type I and II errors of the validation results for the training and validation samples of the SVM-K-score model.

Table 11.2: Validation results of the SVM-K-score training sample for T to T-5

	T	T-1	T-2	T-3	T-5
	Correct (%)	Correct (%)	Correct (%)	Correct (%)	Correct (%)
Sensitivity	18.75	6.67	12.50	0.00	0.00
Specificity	100.00	97.83	100.00	100.00	100.00
Overall accuracy	77.97	75.41	77.42	79.03	76.19
Type I error	81.25	93.33	87.50	100.00	100.00
Type II error	0.00	2.17	0.00	0.00	0.00

Source: Own compilation.

The SVM-K-score training sample was based on 68 sample companies. In Table 11.3, "Sensitivity" followed an erratic trend during the first three periods and dropped to 0.00% in T-3 and T-5. The change in "Specificity" remained static at 100.00% for most of the test period, except for a once-off drop to 97.83% in T-1. The overall accuracy similarly followed an erratic trend from T to T-5. Type I errors varied on a high level in T to T-2 and increased to 100.00% in T-3 and T-5. No Type II errors were recorded in all periods, except for 2.17% in T-1.

Table 11.3: Validation results of the SVM-K-score validation sample for T to T-5

	T	T-1	T-2	T-3	T-5
	Correct (%)	Correct (%)	Correct (%)	Correct (%)	Correct (%)
Sensitivity	0.00	33.33	0.00	0.00	0.00
Specificity	100.00	85.71	92.86	100.00	100.00
Overall accuracy	88.24	76.47	76.47	64.71	64.71
Type I error	100.00	66.67	100.00	100.00	100.00
Type II error	0.00	14.29	7.14	0.00	0.00

Source: Own compilation.

The validation sample was based on 17 sample companies randomly selected by XLSTAT®. In Table 11.3, "Sensitivity" performed poorly at 0.00% for most of the test period. "Specificity" remained static at 100.00% for most of the test period, except for a deterioration to 85.71% in T-1 and to 92.86% in T-2. The overall accuracy deteriorated gradually from T to T-5. Type I errors remained at a high 100.00% for most of the test period, except for 66.67% in T-1. No Type II errors were recorded in most periods, except for 14.29% in T-1 and 7.14% in T-2.

11.3 VALIDATION RESULTS OF THE SVM-Z-SCORE MODEL

Table 11.4 provides a summary of the optimised support vector machine classifiers. The binary classifiers classified each observation or company as either zero (0) or one (1). One represented the companies in financial distress; and zero, the positive class, represented the non-distressed companies.

Table 11.4: A summary of optimised support vector machine classifiers for the SVM-Z-score

	T	T-1	T-2	T-3	T-5
Positive class	0	0	0	0	0
Number of observations in the training set	62	62	63	63	64
Number of removed observations	6	6	5	5	4
Bias	-1.996	-1.033	-1.002	-1.118	-1.219
Number of support vectors	18	19	18	17	16

Source: Own compilation

Tables 11.5 and 11.6 provide a summary of "Sensitivity", "Specificity" and Type I and II errors of the validation results for the training and validation samples of the SVM-Z-score model.

Table 11.5: Validation results of the SVM-Z-score training sample for T to T-5

	T	T-1	T-2	T-3	T-5
	Correct (%)	**Correct (%)**	**Correct (%)**	**Correct (%)**	**Correct (%)**
Sensitivity	25.00	0.00	0.00	14.29	14.29
Specificity	98.15	100.00	100.00	100.00	98.25
Overall accuracy	88.72	90.32	93.65	90.48	89.06
Type I error	75.00	100.00	50.00	85.71	85.71
Type II error	1.85	0.00	0.00	0.00	1.75

Source: Own compilation.

The SVM-Z training sample was based on 68 sample companies. In Table 11.5, "Sensitivity" of 25% was recorded in T and deteriorated to 0.00% in T-1 and T-2, to recover marginally to 14.29% in T-4 and T-5. "Specificity" was recorded at 98.15% in T and remained static at 100.00% for most of the test period, except for a once-off deterioration to 98.25% in T-5. The overall accuracy followed an erratic trend from T to T-5. Type I errors deteriorated from 75% in T to 100% in T-1, improved to 50% in T-2 and deteriorated again to 85.71% in T-3 and T-5. No Type II errors were recorded in all periods, except for 1.85% in T-1 and 1.75% in T-5.

Table 11.6: Validation results of the SVM-Z-score validation sample for T to T-5

	T	T-1	T-2	T-3	T-5
	Correct (%)	Correct (%)	Correct (%)	Correct (%)	Correct (%)
Sensitivity	0.00	0.00	0.00	0.00	0.00
Specificity	94.12	100.00	100.00	100.00	100.00
Overall accuracy	94.12	88.24	82.35	94.12	94.12
Type I error	0.00	100.00	100.00	100.00	100.00
Type II error	5.88	0.00	0.00	0.00	0.00

Source: Own compilation.

The validation sample was based on 17 sample companies randomly selected by XLSTAT®. In Table 11.6, "Sensitivity" performed poorly at 0.00% for the whole test period. "Specificity" remained static at 100.00% for most of the test period, except for 94.12% recorded in T. The overall accuracy performed well, a relatively low 82.35% in T-2 and a high 94.12% in T, T-3 and T-5. Type I errors remained at a high 100.00% for most of the test period, except for 0.00% in T. No Type II errors were recorded in most periods, except for 5.88% in T.

11.4 VALIDATION OF THE F-SCORE MODEL

This section is divided into five categories – one for each F-score model.

11.4.1 F-score based on financial variables (Model 1)

Table 11.7 provides a summary of the optimised support vector machine classifiers. The binary classifiers classified each observation or company as either zero (0) or one (1). One represented the companies in financial distress; and zero, the positive class, represented the non-distressed companies.

Table 11.7: A summary of optimised support vector machine classifiers for the F-score Model 1

	T	T-1	T-2	T-3	T-5
Positive class	0	0	0	0	0
Number of observations in the training set	62	62	63	63	64
Number of removed observations	6	6	5	5	4
Bias	0.853	0.071	-0.823	-1.088	-0.876
Number of support vectors	48	41	54	52	53

Source: Own compilation.

Tables 11.8 and 11.9 provide a summary of "Sensitivity", "Specificity" and Type I and II errors of the validation results for the training and validation samples of the F-score Model 1.

Table 11.8: Validation results of the F-score Model 1 training sample for T to T-5

	T	T-1	T-2	T-3	T-5
	Correct (%)	**Correct (%)**	**Correct (%)**	**Correct (%)**	**Correct (%)**
Sensitivity	30.77	27.77	21.43	15.38	15.38
Specificity	97.22	100.00	97.14	100.00	97.37
Overall accuracy	69.35	74.19	63.49	65.08	64.06
Type I error	69.23	72.73	78.57	84.62	84.62
Type II error	2.78	0.00	2.86	0.00	2.63

Source: Own compilation.

The F-score training sample was based on 68 sample companies. In Table 11.8, "Sensitivity" deteriorated gradually from 30.77% in T to 15.38% in T-5. "Specificity" was recorded at 97.22% in T, recovered to 100.00% in T-1, deteriorated in T-2 to 97.14%, recovered to 100.00% in T-3, and finally, deteriorated to 97.37% in T-5. The overall accuracy followed an erratic trend from T to T-5. Type I errors increased gradually from 69.23% in T as the forecast period lengthened, to 84.62% in T-5. A low Type II error was recorded for the full test period.

Table 11.9: Validation results of the F-score Model 1 validation sample for T to T-5

	T	T-1	T-2	T-3	T-5
	Correct (%)	Correct (%)	Correct (%)	Correct (%)	Correct (%)
Sensitivity	33.33	0.00	0.00	0.00	0.00
Specificity	81.82	90.00	92.86	100.00	63.64
Overall accuracy	64.71	52.94	76.47	64.71	41.18
Type I error	66.67	100.00	100.00	100.00	100.00
Type II error	18.18	10.00	7.14	0.00	36.36

Source: Own compilation.

The validation sample was based on 17 sample companies randomly selected by XLSTAT®. In Table 11.9, "Sensitivity" was recorded at 33.33% correctness in T, but performed poorly at 0.00% for the remaining test period. "Specificity" improved from 81.82% in T to 100.00% in T-3, but deteriorated to 63.64% in T-5. The overall accuracy performed relatively poor and followed an erratic trend from T to T-5. Type I errors remained at a high 100.00% for most of the test period, except for 66.67% in T. A relatively low Type II error was recorded for most of the test period.

11.4.2 F-score based on financial and market variables (Model 2)

Table 11.10 provides a summary of the optimised support vector machine classifiers. The binary classifiers classified each observation or company as either zero (0) or one (1). One represented the companies in financial distress; and zero, the positive class, represented the non-distressed companies.

Table 11.10: A summary of optimised support vector machine classifiers for the F-score Model 2

	T	T-1	T-2	T-3	T-5
Positive class	0	0	0	0	0
Number of observations in the training set	68	68	68	68	68
Number of removed observations	0	0	0	0	0
Bias	3.633	2.084	1.001	1.224	-0.998
Number of support vectors	26	41	49	45	51

Source: Own compilation.

Tables 11.11 and 11.12 provide a summary of “Sensitivity”, “Specificity” and Type I and II errors of the validation results for the training and validation samples of the F-score Model 2.

Table 11.11: Validation results of the F-score Model 2 training sample for T to T-5

	T	T-1	T-2	T-3	T-5
	Correct (%)	Correct (%)	Correct (%)	Correct (%)	Correct (%)
Sensitivity	86.36	60.87	41.67	30.43	0.00
Specificity	100.00	91.11	88.64	93.33	100.00
Overall accuracy	95.59	80.88	72.06	72.06	69.12
Type I error	13.64	39.19	58.33	69.57	100.00
Type II error	0.00	8.89	11.36	6.67	0.00

Source: Own compilation.

The F-score training sample was based on 68 sample companies. In Table 11.11, “Sensitivity” deteriorated gradually from 86.36% in T to 30.43% in T-3 and 0.00% in T-5. “Specificity” was recorded at 100.00% in T, deteriorated in T-1 to T-3 and recovered to 100.00% in T-5. The overall accuracy deteriorated gradually as the forecast period lengthened from 95.59% in T to 69.12% in T-5. Type I errors increased gradually as the forecast period lengthened from 13.64% in T to 100.00% in T-5. A low Type II error was recorded for the full test period.

Table 11.12: Validation results of the F-score Model 2 validation sample for T to T-5

	T	T-1	T-2	T-3	T-5
	Correct (%)	Correct (%)	Correct (%)	Correct (%)	Correct (%)
Sensitivity	100.00	42.86	20.00	0.00	0.00
Specificity	100.00	80.00	83.33	75.00	100.00
Overall accuracy	100.00	64.71	64.71	52.94	64.71
Type I error	0.00	57.14	80.00	100.00	100.00
Type II error	0.00	20.00	16.67	25.00	0.00

Source: Own compilation.

The validation sample was based on 17 sample companies randomly selected by XLSTAT®. In Table 11.12, "Sensitivity" performed poorly and deteriorated drastically as the forecast period lengthened. "Specificity" deteriorated from 100.00% in T to 75% in T-3 and recovered to 100.00% in T-5. The overall accuracy deteriorated from 100.00% in T and stabilised at 64.71% from T-1 onwards. Type I errors started at 0.00% in T and increased to 100.00% in T-3 and T-5. A relatively low Type II error was recorded for most of the test period.

11.4.3 F-score based on financial and market variables plus a leading business cycle indicator (Model 3)

Table 11.13 provides a summary of the optimised support vector machine classifiers. The binary classifiers classified each observation or company as either zero (0) or one (1). One represented the companies in financial distress; and zero, the positive class, represented the non-distressed companies

Table 11.13: A summary of optimised support vector machine classifiers for the F-score Model 3

	T	T-1	T-2	T-3	T-5
Positive class	0	0	0	0	0
Number of observations in the training set	68	68	68	68	68
Number of removed observations	0	0	0	0	0
Bias	6.647	1.305	2.930	3.416	-0.991
Number of support vectors	26	43	45	47	45

Source: Own compilation.

Tables 11.14 and 11.15 provide a summary of "Sensitivity", "Specificity" and Type I and II errors of the validation results for the training and validation samples of the F-score Model 3.

Table 11.14: Validation results of the F-score Model 3 training sample for T to T-5

	T	T-1	T-2	T-3	T-5
	Correct (%)	Correct (%)	Correct (%)	Correct (%)	Correct (%)
Sensitivity	100.00	23.81	44.00	54.17	0.00
Specificity	97.78	95.74	90.70	88.64	100.00
Overall accuracy	98.53	73.53	73.53	76.47	70.59
Type I error	0.00	76.19	56.00	45.83	100.00
Type II error	2.22	4.26	9.30	11.36	0.00

Source: Own compilation.

The F-score training sample was based on 68 sample companies. In Table 11.14, "Sensitivity" deteriorated drastically from 100.00% in T to 0.00% in T-5. "Specificity" was recorded at 97.78% in T, deteriorated in T-1 to T-3 and recovered to 100.00% in T-5. The overall accuracy deteriorated gradually from 98.53% in T and varied between 70.59% and 76.47% in the remaining forecast period. Type I errors increased gradually as the forecast period lengthened from 0.00% in T to 100.00% in T-5. A low Type II error was recorded for the full test period.

Table 11.15: Validation results of the F-score Model 3 validation sample for T to T-5

	T	T-1	T-2	T-3	T-5
	Correct (%)	Correct (%)	Correct (%)	Correct (%)	Correct (%)
Sensitivity	83.33	0.00	0.00	50.00	0.00
Specificity	90.91	90.00	84.62	76.92	100.00
Overall accuracy	88.24	52.94	64.71	70.59	58.82
Type I error	16.67	100.00	100.00	50.00	100.00
Type II error	9.09	10.00	15.38	23.08	0.00

Source: Own compilation.

The validation sample was based on 17 sample companies randomly selected by XLSTAT®. In Table 11.15, "Sensitivity" performed poorly and deteriorated drastically as the forecast period lengthened. "Specificity" of 90.91% and 90.00% was recorded in T and T-1 respectively, deteriorated to 76.92% in T-3 and recovered to 100.00% in T-5. The overall accuracy deteriorated from 88.24% in T and varied in the remaining forecast period. A 100.00% Type I error was recorded in T-1, T-2 and T-5, after a

16.67% error was recorded in T. A relatively low Type II error was recorded for most of the test period.

11.4.4 F-score based on financial and market variables plus a coincident business cycle indicator (Model 4)

Table 11.16 provides a summary of the optimised support vector machine classifiers. The binary classifiers classified each observation or company as either zero (0) or one (1). One represented the companies in financial distress; and zero, the positive class, represented the non-distressed companies.

Table 11.16: A summary of optimised support vector machine classifiers for the F-score Model 4

	T	T-1	T-2	T-3	T-5
Positive class	0	0	0	0	0
Number of observations in the training set	68	68	68	68	68
Number of removed observations	0	0	0	0	0
Bias	5.097	0.727	-0.502	-0.485	-1.001
Number of support vectors	26	37	44	40	50

Source: Own compilation.

Tables 11.17 and 11.18 provide a summary of “Sensitivity”, “Specificity” and Type I and II errors of the validation results for the training and validation samples of the F-score Model 4.

Table 11.17: Validation results of the F-score Model 4 training sample for T to T-5

	T	T-1	T-2	T-3	T-5
	Correct (%)	**Correct (%)**	**Correct (%)**	**Correct (%)**	**Correct (%)**
Sensitivity	94.44	11.76	10.00	11.11	0.00
Specificity	98.00	96.08	97.92	98.00	100.00
Overall accuracy	97.06	75.00	72.06	75.00	70.59
Type I error	5.56	88.24	90.00	88.89	100.00
Type II error	2.00	3.92	2.08	2.00	0.00

Source: Own compilation.

The F-score training sample was based on 68 sample companies. In Table 11.17, "Sensitivity" deteriorated drastically from 94.44% in T to 0.00% in T-5. "Specificity" was recorded at 98.00% in T, deteriorated marginally in T-1 to T-3 and recovered to 100.00% in T-5. The overall accuracy deteriorated gradually from 97.08% in T and varied between 70.59% and 75.00% in the remaining forecast period. Type I errors increased drastically from 5.56% in T as the forecast period lengthened to 100.00% in T-5. A low Type II error was recorded for the full test period.

Table 11.18: Validation results of the F-score Model 4 validation sample for T to T-5

	T	T-1	T-2	T-3	T-5
	Correct (%)	Correct (%)	Correct (%)	Correct (%)	Correct (%)
Sensitivity	80.00	16.67	0.00	0.00	0.00
Specificity	100.00	100.00	100.00	100.00	100.00
Overall accuracy	94.12	70.59	82.35	70.59	76.47
Type I error	20.00	83.33	100.00	100.00	100.00
Type II error	0.00	0.00	0.00	0.00	0.00

Source: Own compilation.

The validation sample was based on 17 sample companies randomly selected by XLSTAT®. In Table 11.18, "Sensitivity" deteriorated from 80.00% in T to 0.00% in T-5. "Specificity" was recorded at 100.00% for the full test period. The overall accuracy deteriorated gradually from 94.12% in T to 76.47% in T-5. Type I errors started at 20.00% in T and increased to 100.00% from T-2 to T-5. No Type II errors were recorded for the full test period.

11.4.5 F-score based on financial and market variables plus a lagging business cycle indicator (Model 5)

Table 11.19 provides a summary of the optimised support vector machine classifiers. The binary classifiers classified each observation or company as either zero (0) or one (1). One represented the companies in financial distress; and zero, the positive class, represented the non-distressed companies

Table 11.19: A summary of optimised support vector machine classifiers for the F-score Model 5

	T	T-1	T-2	T-3	T-5
Positive class	0	0	0	0	0
Number of observations in the training set	68	68	68	68	68
Number of removed observations	0	0	0	0	0
Bias	6.541	4.663	0.519	2.346	1.668
Number of support vectors	24	46	48	47	57

Source: Own compilation.

Tables 11.20 and 11.21 provide a summary of "Sensitivity", "Specificity" and Type I and II errors of the validation results for the training and validation samples of the F-score Model 5.

Table 11.20: Validation results of the F-score Model 5 training sample for T to T-5

	T	T-1	T-2	T-3	T-5
	Correct (%)	Correct (%)	Correct (%)	Correct (%)	Correct (%)
Sensitivity	96.30	71.43	4.35	41.67	25.93
Specificity	95.12	85.00	97.78	84.09	92.68
Overall accuracy	95.59	79.41	66.18	69.12	66.18
Type I error	3.70	28.57	95.65	58.33	74.07
Type II error	4.88	15.00	2.22	15.91	7.32

Source: Own compilation.

The F-score training sample was based on 68 sample companies. In Table 11.20, "Sensitivity" deteriorated drastically from 96.30% in T to 25.93% in T-5. "Specificity" was recorded at 95.12% in T, deteriorated marginally in T-1 to T-3 and recovered to 92.68% in T-5. The overall accuracy deteriorated gradually from 95.59% in T and varied between 79.41% and 66.18% in the remaining forecast period. Type I errors varied over the full test period from a low 3.70% in T to 95.68% in T-5. A low Type II error was recorded for the full test period.

Table 11.21: Validation results of the F-score Model 5 validation sample for T to T-5

	T	T-1	T-2	T-3	T-5
	Correct (%)	Correct (%)	Correct (%)	Correct (%)	Correct (%)
Sensitivity	100.00	0.00	14.29	28.57	50.00
Specificity	85.71	92.31	100.00	70.00	84.62
Overall accuracy	88.24	70.59	64.71	52.94	76.47
Type I error	0.00	100.00	85.71	71.43	50.00
Type II error	14.29	7.69	0.00	30.00	15.38

Source: Own compilation.

The validation sample was based on 17 sample companies randomly selected by XLSTAT®. In Table 11.21, "Sensitivity" accuracy of 100.00% was recorded in T and varied drastically in the remaining test period. "Specificity" varied during the full test period – between 100.00% in T-2 and 84.62% in T-5. The overall accuracy deteriorated marginally from 88.24% in T to 76.47% in T-5. Type I errors started at 0.00% in T, increased to 100.00% in T-1 and improved again gradually to 50.00% in T-5. Type II errors fluctuated during the full test period.

11.5 REVIEW OF THE SVM-K-SCORE, SVM-Z-SCORE AND F-SCORE MODEL TRAINING AND VALIDATION SAMPLE RESULTS

11.5.1 Review of training results

The confusion matrix for the training and validation samples of each of the models for the full test period discussed above is summarised and reviewed in this section. The results of the models are ranked. Firstly, the overall results are ranked from high (best performer) to low (worst performer), and secondly, the Type I and II errors are ranked from low (best performer with least errors) to high (worst performer with most errors).

Table 11.22 displays a summary of the overall accuracy of the SVM-K-score, SVM-Z-score and F-score (Model 1 to 5) models, ranked from the best to worst performer. As previously, only the top three performers were reviewed.

Table 11.22: Ranked overall accuracy of the training results for SVM-K-score, SVM-Z-score and F-score (Model 1 to Model 5) models for T to T-5

	T		T-1		T-2		T-3		T-5		T to T-5	
	Correct (%)	Rank	Correct (%)	Rank	Correct (%)	Rank	Correct (%)	Rank	Correct (%)	Rank	Average	Rank
SVM-K-score	77.97	6	75.41	4	77.42	2	79.03	2	76.19	2	77.20	5
SVM-Z-score	88.72	5	90.32	1	93.65	1	90.48	1	89.06	1	90.45	1
F-score Model 1	69.35	7	74.19	6	63.49	7	65.08	7	64.06	7	67.23	7
F-score Model 2	95.59	3	80.88	2	72.06	5	72.06	5	69.12	5	77.94	3
F-score Model 3	98.53	1	73.53	7	73.53	3	76.47	3	70.59	4	78.53	2
F-score Model 4	97.06	2	75.00	5	72.06	4	75.00	4	70.59	3	77.94	3
F-score Model 5	95.59	4	79.41	3	66.18	6	69.12	6	66.18	6	75.30	6

<u>Source</u>: Own compilation.

In T, the F-score - Model 3 ranked highest, followed by the F-score – Model 4 and F-score – Model 2. The F-score – Model 3 deteriorated to the worst position in T-2 and recovered to third and fourth position in T-3 and T-5.

Based on an average calculation (T to T-5), the SVM-Z-score ranked highest, followed by the F-score – Model 3. The F-score - Model 2 and F-score – Model 4 ranked equally in third position. The SVM-Z-score model had a weighted compilation of financial and market variables and the second- and third-ranked models, F-score, had a weighted compilation of financial, market and non-financial variables. The worst-ranked performers SVM-K-score and F-score – Model 1 had a weighted compilation of financial variables only. This excluded the F-score – Model 5, which ranked number six out of seven.

Table 11.23: Ranked Type I errors of the training results for SVM-K-score, SVM-Z-score and F-score (Model 1 to Model 5) models for T to T-5

	T		T-1		T-2		T-3		T-5		T to T-5	
	Correct (%)	Rank	Correct (%)	Rank	Correct (%)	Rank	Correct (%)	Rank	Correct (%)	Rank	Average	Rank
SVM-K-score	81.25	7	93.33	6	87.50	5	100.00	7	100.00	7	92.42	7
SVM-Z-score	75.00	6	100.00	7	50.00	1	85.71	5	85.71	3	79.28	6
F-score Model 1	69.23	5	72.73	3	78.57	4	84.62	4	84.62	2	77.95	5
F-score Model 2	13.64	4	39.19	2	58.33	3	69.57	3	100.00	4	56.15	3
F-score Model 3	0.00	1	76.19	4	56.00	2	45.83	1	100.00	5	55.60	2
F-score Model 4	5.56	3	88.24	5	90.00	6	88.89	6	100.00	6	74.54	4
F-score Model 5	3.70	2	28.57	1	95.65	7	58.33	2	74.07	1	52.06	1

<u>Source</u>: Own compilation.

In T, the F-score - Model 3 ranked highest, followed by the F-score – Model 5 and F-score – Model 4. The F-score – Model 3 ranking remained reasonably static over the full test period.

Based on an average calculation (T to T-5), the F-score – Model 5 ranked highest, followed by the F-score – Model 3 and F-score – Model 2. All the top performers, where non-distressed companies were incorrectly classified as distressed (i.e. Type I error), were based on one of the F-score models. Although the SVM-Z-score model ranked the highest in overall accuracy, it and the SVM-K-score had the highest Type I errors.

Table 11.24 displays a summary of the Type II errors (where a distressed company was incorrectly classified as a non-distressed company) of the SVM-K-score, SVM-Z-score and F-score (Model 1 to 5) models, ranked from the best to worst performer. The model with the lowest Type II error was ranked first, and the highest Type II error ranked last. Only the top three performers were reviewed.

Table 11.24: Ranked Type II errors of the training results for SVM-K-score, SVM-Z-score and F-score (Model 1 to Model 5) models for T to T-5

	T		T-1		T-2		T-3		T-5		T to T-5	
	Correct (%)	Rank	Correct (%)	Rank	Correct (%)	Rank	Correct (%)	Rank	Correct (%)	Rank	Average	Rank
SVM-K-score	0.00	2	2.17	3	0.00	1	0.00	1	0.00	1	0.43	1
SVM-Z-score	1.85	3	0.00	1	0.00	2	0.00	2	1.75	4	0.72	2
F-score Model 1	2.78	6	0.00	2	2.86	5	0.00	3	2.63	6	1.65	3
F-score Model 2	0.00	1	8.89	6	11.36	7	6.67	5	0.00	2	5.38	5
F-score Model 3	2.22	5	4.26	5	9.30	6	11.36	6	0.00	3	5.43	6
F-score Model 4	2.00	4	3.92	4	2.08	3	2.00	4	0.00	4	2.00	4
F-score Model 5	4.88	7	15.00	7	2.22	4	15.91	7	7.32	7	9.07	7

Source: Own compilation.

In T, the F-score - Model 2 ranked highest, followed by the SVM-K-score and SVM-Z-score models. Both the SVM-K-score and SVM-Z-score models performed consistently throughout the test period with the least Type II errors.

Based on an average calculation (T to T-5), the SVM-K-score and SVM-Z-score models ranked highest. The top-performing SVM-K-score model had a weighted compilation of financial variables, compared with all the other models, which had a combination of financial, market and non-financial variables.

11.5.2 Review of validation results

In the final test, following the training stage, the validation results of the SVM-K-score, SVM-Z-score and various F-score models were reviewed

As indicated above, the validation module in XLSTAT® was selected to establish how well classifiers of each of the SVM-K-score, SVM-Z-score and various F-score models perform. A random validation sample of 17 companies was selected from the training sample of 85 companies

Table 11.25 displays a summary of the overall accuracy of the validation results of the SVM-K-score, SVM-Z-score and F-score (Model 1 to 5) models, ranked from the best to worst performer. As previously, only the top three performers were reviewed.

Table 11.25: Ranked overall accuracy of the validation results for SVM-K-score, SVM-Z-score and F-score (Model 1 to Model 5) models for T to T-5

	T		T-1		T-2		T-3		T-5		T to T-5	
	Correct (%)	Rank	Correct (%)	Rank	Correct (%)	Rank	Correct (%)	Rank	Correct (%)	Rank	Average	Rank
SVM-K-score	88.24	4	76.47	2	76.47	3	64.71	4	64.71	4	74.12	3
SVM-Z-score	94.12	3	88.24	1	82.35	2	94.12	1	94.12	1	90.59	1
F-score Model 1	64.71	7	52.94	7	76.47	4	64.71	5	41.18	7	60.00	7
F-score Model 2	100.00	1	64.71	5	64.71	5	52.94	6	64.71	5	69.41	5
F-score Model 3	88.24	6	52.94	6	64.71	6	70.59	2	58.82	6	67.06	6
F-score Model 4	94.12	2	70.59	3	82.35	1	70.59	3	76.47	2	78.82	2
F-score Model 5	88.24	5	70.59	4	64.71	7	52.94	7	76.47	3	70.59	4

Source: Own compilation.

In T, the F-score - Model 2 ranked highest at 100.00%, followed by the F-score – Model 4 and SVM-Z-score. The F-score – Model 2 deteriorated over the full length of the test period. The second-ranked F-score – Model 4 remained consistent throughout the test period. The third-ranked SVM-Z-score improved to first overall position.

Based on an average calculation, the SVM-Z-score ranked highest, followed by the F-score – Model 4 and SVM-K-score model. The SVM-Z-score model had a weighted compilation of financial and market variables and the second-ranked models, F-score, had a weighted compilation of financial, market and non-financial variables. The third-ranked SVM-K-score model had a weighted compilation of financial variables only. Both the worst-ranked performers, F-score – Model 3 and Model 1, had a weighted compilation of financial variables, market and non-financial variables.

However, the overall accuracy should not be evaluated in isolation. A high overall accuracy would not necessarily result in a low Type I and/or Type II error. The Type I and Type II errors are ranked in Tables 11.26 and 11.27.

Table 11.26: Ranked Type I errors of the validation results for SVM-K-score, SVM-Z-score and F-score (Model 1 to Model 5) models for T to T-5

	T		T-1		T-2		T-3		T-5		T to T-5	
	Correct (%)	Rank	Correct (%)	Rank	Correct (%)	Rank	Correct (%)	Rank	Correct (%)	Rank	Average	Rank
SVM-K-score	100.00	7	66.67	2	100.00	3	100.00	3	100.00	2	93.33	7
SVM-Z-score	0.00	1	100.00	4	100.00	4	100.00	4	100.00	3	80.00	4
F-score Model 1	66.67	6	100.00	5	100.00	5	100.00	5	100.00	4	93.33	7
F-score Model 2	0.00	3	57.14	1	80.00	1	100.00	6	100.00	5	67.43	2
F-score Model 3	16.67	4	100.00	6	100.00	6	50.00	1	100.00	6	73.33	3
F-score Model 4	20.00	5	83.33	3	100.00	7	100.00	7	100.00	7	80.67	5
F-score Model 5	0.00	2	100.00	7	85.71	2	71.43	2	50.00	1	61.43	1

Source: Own compilation.

In T, the SVM-Z-score ranked highest, followed by the F-score – Model 5 and F-score – Model 2. The F-score – Model 5 ranking deteriorated to seventh ranked in T-1, but recovered to second and first ranked in T-3 and T-5 respectively.

Based on an average calculation (T to T-5), the F-score – Model 5 ranked highest, followed by the F-score – Model 2 and F-score – Model 3. All the top performers, where non-distressed companies were incorrectly classified as distressed (i.e. Type I error), were based on one of the F-score models. Although the SVM-Z-score model ranked the highest in overall accuracy, it ranked fourth in Type I errors.

Table 11.27 displays a summary of the Type II errors, where a distressed company was incorrectly classified as a non-distressed company, for the SVM-K-score, SVM-Z-score and F-score (Model 1 to 5) models, which ranked from the best to worst performer. The model with the lowest Type II error is ranked first, and the highest Type II error ranked last. As previously, only the top three performers were reviewed.

Table 11.27: Ranked Type II errors of the validation results for SVM-K-score, SVM-Z-score and F-score (Model 1 to Model 5) models for T to T-5

	T		T-1		T-2		T-3		T-5		T to T-5	
	Correct (%)	Rank	Correct (%)	Rank	Correct (%)	Rank	Correct (%)	Rank	Correct (%)	Rank	Average	Rank
SVM-K-score	0.00	1	14.29	6	7.14	4	0.00	1	0.00	1	4.29	3
SVM-Z-score	5.88	4	0.00	1	0.00	1	0.00	2	0.00	2	1.18	2
F-score Model 1	18.18	7	10.00	5	7.14	5	0.00	3	36.36	7	14.34	7
F-score Model 2	0.00	2	20.00	7	16.67	7	25.00	6	0.00	3	12.33	5
F-score Model 3	9.09	5	10.00	4	15.38	6	23.08	5	0.00	4	11.51	4
F-score Model 4	0.00	3	0.00	2	0.00	2	0.00	4	0.00	5	0.00	1
F-score Model 5	14.29	6	7.69	3	0.00	3	30.00	7	15.38	6	13.47	6

Source: Own compilation.

In T, the SVM-K-score, F-score - Model 2 and Model 4 ranked highest, followed by the Z and F-score – Model 4 models. The SVM-K-score and F-score – Models 2 and 4 recorded no Type II errors.

Based on an average calculation (T to T-5), the F-score – Model 4 and SVM-Z-score models ranked highest. The top-performing F-score – Model 4 had a weighted compilation of financial, market and non-financial variables, compared with the SVM-K-score model having a weighted compilation of financial variables only and the F-score – Model 3 and 2 having a compilation of financial, market and non-financial variables.

11.5.3 Summary of ranking results

The ranking results of the training and validation sample reviewed in Sections 11.5.1 and 11.5.2 respectively are summarised in order to simplify an interpretation of the results.

In Table 11.28 and Table 11.29, the overall, Type I and Type II average ranking scores for each model are added to display a total score. The overall accuracy was ranked in descending order – the best overall score ranked first and the worst score ranked seventh. The Type I and II errors were ranked in ascending order – the lowest error ranked first and the worst score ranked seventh.

Table 11.28: A summary of the overall accuracy, Type I and Type II errors and overall ranking results of the training sample for the SVM-K-score, SVM-Z-score and F-score (Model 1 to Model 5) models

	Overall accuracy	Type I error	Type II error	Total	Overall ranking
SVM-K-score	5	7	1	13	5
SVM-Z-score	1	6	2	9	1
F-score Model 1	7	5	3	15	7
F-score Model 2	4	3	5	12	4
F-score Model 3	2	2	6	10	2
F-score Model 4	3	4	4	11	3
F-score Model 5	6	1	7	14	6

Source: Own compilation.

In Table 11.28, based on the training sample, the SVM-Z-score model achieved the highest score, followed by the F-score – Model 2 second and the F-score – Model 4 third.

The SVM-Z-score achieved the highest ranking in terms of overall accuracy and classified the second least financially distressed companies as non-distressed (Type II error). Similarly, the F-score – Model 3 ranked second, recorded the second best in overall accuracy and Type I errors, but ranked poorly on Type II errors. The third-ranked model in terms of overall accuracy, F-score – Model 4, recorded average Type I and Type II errors.

The validation test is repeated for each model based on the same variables in the training sample applied to a sample of companies randomly selected by XLSTAT®. The validation results are displayed in Table 11.29.

Table 11.29: A summary of the overall accuracy, Type I and Type II errors and overall ranking results of the validation sample for the SVM-K-score, SVM-Z-score and F-score (Model 1 to Model 5) models

	Overall accuracy	Type I error	Type II error	Total	Overall ranking
SVM-K-score	3	7	3	13	5
SVM-Z-score	1	4	2	7	1
F-score Model 1	7	7	7	21	7
F-score Model 2	5	2	5	12	4
F-score Model 3	6	3	4	13	5
F-score Model 4	2	5	1	8	2
F-score Model 5	4	1	6	11	3

Source: Own compilation.

The SVM-Z-score maintained its first-ranked position, followed by the F-score – Model 2 and F-score – Model 5. The SVM-K-score and F-score – Model 3 performed poorly and ranked equally fifth. The F-score – Model 1, based on financial variables only, performed the worst in seventh position in both the training and validation results.

11.6 RESULTS AND PRESENTATION OF HYPOTHESIS

11.6.1 Background

The training and validation results were reviewed in Section 11.5 in order to determine whether financial distress prediction models based on a combination of financial, market and quantitative non-financial variables outperformed financial distress models based on financial variables only.

This section discusses the results of the various models, contrasting it with the formulated hypothesis. The research hypothesis and alternative hypothesis are either confirmed or rejected by the empirical findings.

11.6.2 Results and presentation of hypothesis

(i) Hypothesis 1 is defined as follows:

H_1: The Altman Z-score statistical financial distress prediction model has higher Type I and Type II errors than the South African-based De la Rey K-score statistical financial distress model.

H_0: The Altman Z-score statistical financial distress prediction model does not have higher Type I and Type II errors than the South African-based De la Rey K-score statistical financial distress model.

Model results:

Hypothesis 1 is based on the distinction between the non-South African-developed Z model and the South African-developed K-score model. The Z-score model was developed against different dynamics (economic variables and financial debt and equity instruments) available to South African companies on which the K-score model was based.

In the current study, both the Z-score and K-score models were tested on South African companies' financial results. The Z-score ranked higher than the K-score based on overall rankings (combined overall accuracy, Type I and Type II errors). However, the result could have been affected by the inclusion of a market variable in the Z-core model.

The research hypothesis is therefore rejected in favour of the alternative hypothesis.

(ii) Hypothesis 2 is defined as follows:

H_1: The artificial intelligence model based on financial variables only has higher Type I and Type II errors than the proposed artificial intelligence model based on a combination of financial, market and quantitative non-financial variables.

H_0: The artificial intelligence model based on financial variables only does not have higher Type I and Type II errors than the proposed artificial intelligence model based on a combination of financial, market and quantitative non-financial variables.

Model results:

Five versions of the F-score model were developed and tested against the SVM-K-score and SVM-Z-score models. The first version of the model consisted of financial variables only. In the second to fifth version, market and three versions of the business cycle indicators were added to determine whether the addition of these variables would improve the F-score prediction results.

Higher Type I and Type II errors in the F-score - Model 1 based on financial variables only would result in the acceptance of the research hypothesis. Should the F-score – Model 1 have lower Type I and Type II errors than those of the F-score - Models 2 to 5, the alternative hypothesis is accepted.

The results of the validation test displayed in Table 11.29 indicate that the F-score Models 2 to 5 with the added, market and quantitative non-financial variables enhanced the prediction accuracy of the F-score model. No meaningful improvement was evident by adding market variables (Model 2) and leading business indicators (Model 3) to financial variables (Model 1). The best improvement in the overall ranking was evident following the addition of coincident business cycle indicators (Model 4), followed by the addition of lagging business cycle indicators (Model 5).

The addition of the coincident business cycle indicator as a quantitative non-financial variable is in support of the research hypothesis.

The research hypothesis is therefore accepted and the alternative hypothesis is rejected.

(iii) Hypothesis 3 is defined as follows:

H_1: The statistical financial distress prediction models have higher Type I and Type II errors than the proposed artificial intelligence model based on a combination of financial, market and quantitative non-financial variables.

H_0: The statistical financial distress prediction models do not have higher Type I and Type II errors than the proposed artificial intelligence model based on a combination of financial, market and quantitative non-financial variables.

Model results:

The prediction accuracy of each of the five F-score models was tested against the internationally developed SVM-Z-score model and the South African-developed SVM-K-score model. The overall ranking of the models is displayed in Table 11.30.

Table 11.30: A summary of the overall ranking results of the validation sample for the SVM-K-score, SVM-Z-score and F-score (Model 1 to Model 5) models

	SVM-Z-score	F-score Model 4	F-score Model 5	F-score Model 2	F-score Model 3	SVM-K-score	F-score Model 1
Overall ranking	1	2	3	4	5	5	7

<u>Source</u>: Own compilation.

The SVM-Z-score model achieved the highest ranking followed by the F-score – Model 4 and the F-score – Model 5. Both the F-score models outperformed the SVM-K-score model.

However, these results should be interpreted in perspective and in the context of the support vector machine. Following a training process, the overall ranking results of the validation test should be an improvement on the overall ranking results of the training set. The model remaining at the same ranking or deteriorating in its ranking position should therefore not be considered in the results. Table 11.31 displays the overall ranking of the training results.

Table 11.31: A summary of the overall ranking results of the training sample for the SVM-K-score, SVM-Z-score and F-score (Model 1 to Model 5) models

	SVM-Z-score	F-score Model 4	F-score Model 5	F-score Model 2	F-score Model 3	SVM-K-score	F-score Model 1
Overall ranking	1	3	6	4	2	5	7

Source: Own compilation.

The SVM-Z-score, SVM-K-score, F-score – Model 1 and F-score – Model 2 did not improve in the validation test as the overall ranking results remained unchanged. The F-score – Model 3 deteriorated in the validation test from second to fifth overall ranking. The only models that improved in the validation test were the F-score – Model 3 and F-score – Model 5.

A further test indicates that the F-score – Model 5 displayed the most significant improvement in overall ranking, from sixth ranked in the training results to third ranked in the validation results. This is followed by the F-score – Model 4, which improved the least in ranking from third to second ranked.

The results of this analysis indicate that the financial distress prediction model, with added market variables and lagging business cycle indicator as a quantitative non-financial variable (F-score - Model 5), outperformed the SVM-K-score and SVM-Z-score statistical models and all the remaining F-score models. This is based on the significant improvement of the F-score – Model 5 from the training test to the validation test and it has the second-least combined ranking (highest overall accuracy and least Type I and Type II errors).

The research hypothesis is therefore accepted and the alternative hypothesis rejected.

11.7 CONCLUSION

The empirical results of the study indicated that a financial distress prediction model based on financial and market variables and a lagging business cycle indicator, as a proxy for a quantitative non-financial variable, ranked higher in terms of overall accuracy and had less Type I and Type II errors than a financial distress prediction model based purely on financial variables.

In addition, in the final selection of the best ranking model, consideration was given to whether there was a significant improvement in the validation result following a training process. Those models with validation results remaining the same or deteriorating were discarded and only those models with improved results were retained.

The first hypothesis was tested to establish whether an internationally developed financial distress prediction model performed worse than a South African-developed model. The research hypothesis was rejected in favour of the alternative hypothesis.

The second hypothesis was tested to establish whether an artificial intelligence model based on financial variables performed worse than an artificial intelligence model based on a combination of financial, market and quantitative non-financial variables. The research hypothesis was accepted and the alternative hypothesis rejected.

The third and certainly most important hypothesis was to determine whether a financial distress prediction model based on a unique combination of financial, market and quantitative non-financial variables would have a higher overall accuracy and the least Type I and II errors than a statistical financial distress prediction model based on financial variables only. The empirical results confirmed the acceptance of the research hypothesis.

The next and final chapter concludes the study with recommendations for future research.

CHAPTER 12

SUMMARY, CONCLUSION AND RECOMMENDATIONS

12.1 INTRODUCTION

The study is concluded with a review of the research methodology followed in dealing with the problem statement and the research objectives. In this concluding chapter, the focus is on a synopsis of the findings based on the research objectives for the study, the contribution to the existing body of knowledge and recommendations for further research.

12.2 RESEARCH SUMMARY

12.2.1 Background

In an increasing dynamic and globalised operating environment, it is an essential requirement for company stakeholders to adapt their decision-making process accordingly. In this operating environment, it is simply not sufficient to rely on a simple statistical financial distress prediction model. A more sophisticated mechanism or instrument is required in order to allow stakeholders to make informed decisions.

Due to the changing environment, the reliability, popularity and further development of a statistically based financial distress prediction model were constrained. Constraints such as reliance on historical financial information in this highly dynamic operating environment and the advent of computer technology and artificial intelligence contributed to a new era in financial distress prediction.

The literature review in Chapters 2, 3 and 4 provided a perspective of the evolvement of both the statistical and more sophisticated artificial intelligence financial distress prediction model.

12.2.2 Research problem statement

The combined changing environment and the evolvement of this decision-making mechanism or instrument required an increased demand for better and more sophisticated information. Against this background, the research problem statement was formulated as follows:

Informed decision-making in a dynamic operating environment is considered important in maintaining company financial health and proactively avoiding financial distress. However, most financial distress prediction models still rely on static historical financial information and do not take cognisance of both financial and quantitative non-financial variables.

Arising from the research problem, a number of research objectives were formulated. The process followed to achieve each objective is summarised in the following subsections.

12.2.3 Primary research objective

The primary research objective was stated as follows:

The primary objective of this study is to develop an artificial intelligence-based financial distress prediction model that incorporates a unique combination of financial and quantitative non-financial variables from a South African perspective. The intention with the proposed financial distress prediction model is to provide a more accurate and timeous company financial health and distress prediction on a financial distress continuum compared with a statistical financial distress prediction model.

Based on the literature review in Chapters 2, 3, 4 and 5 on the evolvement of financial distress prediction, it was established that financial distress prediction based on artificial intelligence models was still in its infancy in South Africa. Much reliance was placed on simplistic and reliable statistical financial distress prediction models.

Based on the research problem, a continuously changing operating environment requires a more sophisticated decision-making process to allow for informed decisions. Relying purely on simplistic and historical financial information as input for a statistical financial distress prediction model was questioned.

The following step, detailed in Chapter 8, was to identify a financial distress prediction model to be used as benchmark for testing of the proposed F-score financial distress prediction model. The original K-score and Z-score and the inflation-adjusted return on equity were validated and compared to determine which of these models would potentially be suitable as a benchmark. This test, however, did not achieve the anticipated results and it was therefore decided to compare the SVM-K-score and SVM-Z-score models with the F-score models (see Section 10.1).

In order to deal with the primary and overall objective of the study, Chapter 9 described the process to identify and select the input variables for the proposed artificial intelligence financial distress prediction model. First, a phased process was followed to identify a sample of JSE-listed companies to be used as subjects for the study. Second, financial and quantitative non-financial variables were selected through a phased process. The selection of financial variables was split into two subsections, namely financial and market variables. Quantitative non-financial variables consisted of macroeconomic variables.

In Chapter 10, the support vector machine, as a machine learning model within the broader artificial intelligence spectrum, was used to classify and determine the classification of the SVM-Z-score, SVM-K-score and F-score financial distress prediction models. XLSTAT® was used to execute the testing process.

Chapter 11 detailed the final step of the study; XLSTAT® was used to validate and compare the results of the SVM-Z-score, SVM-K-score and F-score financial distress prediction models. XLSTAT® allowed for the separation of the sample in a training and validation sample on a random basis. The final section of Chapter 11 reviewed and compared the validation results of each model in order to determine the acceptance or rejection of the research hypotheses.

The F-score financial distress prediction models were developed based on the identification and selection of financial, market and quantitative non-financial variables. The results of the F-score model were validated and compared with the results of the SVM-K-Score and SVM-Z-score models. The test results demonstrated that the F-score financial distress prediction model achieved superior results compared with the SVM-K-Score and SVM-Z-score models.

The primary objective was therefore achieved.

The primary objective was separated into four secondary objectives. Each of the secondary objectives is discussed below with reference to the main hypotheses of the study.

12.2.4 Secondary research objectives

Each secondary objective is discussed separately in the following subsections.

(i) First secondary objective:

To identify and select a sample representative of South African-listed companies. Predetermined criteria will be applied to identify the sample from a population group of industrial companies listed on the JSE through the INET BFA database, covering a 10-year test period from 2005 to 2014.

From the database, 385 companies were identified as eligible for inclusion as research subjects in the study. A filter approach was followed to eliminate those companies regarded as unsuitable for the study due to differences in accounting conventions and methodologies applied to the analysis and interpretation of financial results. These companies were, for example, related to the mining, financial and property sectors.

Of the original database, 112 companies were identified as the target population, 85 were selected as final sample companies based on a 5% confidence interval. These 85 companies were randomly selected form the target population, representing 76% of the target population.

The financial results of these 85 sample companies over the 10-year test period from 2005 to 2014 were used in all tests in the study.

This objective was therefore achieved.

(ii) Second secondary objective:

To identify and select financial and quantitative non-financial variables based on predetermined criteria and review of applicable literature.

The selection of variables was divided into three main categories, namely financial and market variables within the financial variable category and macroeconomic variables within the quantitative non-financial variable category.

Approximately 233 academic articles based on financial distress prediction studies were reviewed to identify financial and market variables. Of the 233 articles, 111 provided sufficient information to identify variables in this study. Within the 111 articles, 489 financial variables were identified. Following a filter process whereby, for instance, duplicated variables were eliminated, 11 financial variables remained for potential inclusion in the study. These 11 variables were grouped in five categories, namely profitability, efficiency, gearing, liquidity and cash flow.

The inclusion of the 11 variables in a model was regarded as unfeasible and the principal component analysis was applied as a variable reduction process. The principal component analysis process consisted of five individual tests, namely the eigenvalue test, scree test, factor loading test, contribution to factor test and a squared cosine test.

Based on this composite test, the following financial variables and weightings were identified for inclusion in the F-score model:

- Profitability – net income to total assets (denoted by X_1) with a 0.19789 weighting.
- Efficiency – revenue to total assets (denoted by X_2) with a 0.24307 weighting.
- Gearing – total debt to total equity (denoted by X_3) with a 0.26830 weighting.
- Liquidity – (current assets less inventory) to current liabilities (denoted by X_4) with a 0.15356 weighting.
- Cash flow – cash flow to current liabilities (denoted by X_5) with a 0.1111 weighting.

A similar process to the selection of financial variables was followed in the selection of market variables. Out of the 111 articles, 16 market variables were identified for potential inclusion in the F-score model. The principal component analysis process was used and the following market variables were selected:

- Earnings per share (denoted by X_6) with a weighting of 0.11564.
- Market capitalisation (denoted by X_7) with a weighting of 0.10885.

A different approach was followed in the selection of quantitative non-financial variables or macroeconomic variables. It was initially envisaged to select a number of macroeconomic variables for inclusion in the F-score model based on a review of applicable literature. However, this was regarded as problematic. The problem related to which macroeconomic variable should be included in the model, what combination variables should be included and what the weightings should be. The development of such a model would have required specialist economic knowledge and would not have been within the scope of this study. It was therefore decided to rely on the business cycle indicators compiled and published by the South African Reserve Bank.

The business cycle indicators consist of a leading, coincident and lagging indicator, and each indicator is a composite of macroeconomic variables. The data was obtained for a 10-year period from 2005 to 2014 and annualised and standardised for the purposes of the study.

Each of the business cycle indicators was evaluated in the final model development to determine which of the three indicators provided the desired result. The F-score model having a composite of financial and market variables was tested with each of the business cycle indicators. The final F-score model had a composite of financial variables, a composite of market variables and the best-performing business cycle indicator, and can be denoted as follows:

$$F = 0.19789X_1 + 0.34307X_2 + 0.26830X_3 + 0.15356X_4 + 0.1111X_5 + 0.11564X_6 + 0.10885X_7 + \in_8 \quad (29)$$

where: F = overall index

X_1 = net income / total assets

X_2 = revenue / total assets

X_3 = total debt / total equity

X_4 = (current assets – inventory) / current liabilities

X_5 = cash flow / current liabilities

X_6 = earnings per share

X_7 = market capitalisation

$\in_8$ = business cycle indicator (leading, coincident or lagging)

This objective was therefore achieved.

(iii) Third secondary objective:

To evaluate and validate the De la Rey K-score and Altman Z-score models representative of statistical financial distress prediction models and based on South African data to determine their predictive accuracy.

The objective was to validate the original K-score and Z-score models in order to determine each model's financial distress predictive ability over a number of lead periods, and to consider its potential use as a benchmark model for testing against the proposed F-score model.

The Mann-Whitney U and Spearman's rho tests were applied to the original K-score and Z-score results and tested against an inflation-adjusted return on equity for each sample company. A weighted efficiency test was finally conducted on the results to determine the accuracy and percentage error. The Mann-Whitney U test was done to determine whether one sample had larger mean values than another from the same population against an alternative hypothesis. Spearman's rho is a non-parametric measure of rank correlation and the purpose of this test was to assess the strength of the relationship between two variables described using a monotonic function (whether linear or not). In the final test, a weighted efficiency test developed by Korobow and Stuhr (1985), was done to take cognisance of the difference between samples and to provide increased sensitivity to both the accuracy and efficiency aspects of financial distress prediction models.

None of the Mann-Whitney U test, Spearman's rho or weighted efficiency tests to select a benchmark model to validate against the proposed F-score model achieved the desired results due to a number of constraints:

- The country of origin of each model. The Altman Z-score model originated in the United States of America and was based on a local sample and variables. The De la Rey K-score model, on the other hand, originated in South Africa and was based on a South African sample and variables. These models were strictly not comparable, as they have been based on, for example, different accounting conventions and with different funding instruments available to companies at the time.
- The Z-score model included a market variable, which the K-score did not have. Not considering the difference in country of origin, the Z-score could have had an advantage over the K-score model, because the Z-score model result could have been more responsive to market sentiment. This also is primarily the

premise on which the study was based – the fact that a non-financial variable would have a positive effect on the financial distress prediction result.

- The inflation-adjusted return on equity might not have been an appropriate proxy for differentiation between distressed and non-distressed companies. This conclusion corroborates with the conclusion of the Korobow and Stuhr (1985) study in that the inflation-adjusted return on equity is a financial performance measurement compared with the Z-score and K-score models, which were financial distress prediction models. Different concepts were measured.

This objective was not achieved and it was therefore decided to use both the K-score and Z-score models, referred to as the SVM-K-score and SVM-Z-score models, in comparison with the F-score model.

(iv) Fourth secondary objective:

To test the null hypothesis by establishing whether the addition of quantitative non-financial variables to financial variables in an artificial intelligence-based financial distress prediction model benchmarked against the SVM-K-score and SVM-Z-score models.

As foundation to solve the research problem and findings, reference is made to each of the hypotheses.

Hypotheses 1 - The Altman Z-score statistical financial distress prediction model had higher Type I and Type II errors than the South African-based De la Rey K-score statistical financial distress model.

Both the Z-score and K-score models were tested on South African company financial results. The Z-score ranked higher than the K-score based on overall rankings (combined overall accuracy, Type I and Type II errors). However, the result could have been affected by the inclusion of a market variable in the Z-score model.

The research hypothesis was rejected in favour of the alternative hypothesis.

Hypotheses 2 - The artificial intelligence model based on financial variables only has higher Type I and Type II errors than the proposed artificial intelligence model based on a combination of financial, market and quantitative non-financial variables.

The results of the validation test indicated that the F-score Models 2 to 5 with the added, market and quantitative non-financial variables enhanced the prediction accuracy of the F-score model. No meaningful improvement was evident by adding market variables (Model 2) and leading business indicators (Model 3) to financial variables (Model 1). The best improvement in the overall ranking was evident following the addition of coincident business cycle indicators (Model 4), followed by the addition of lagging business cycle indicators (Model 5).

The addition of the coincident business cycle indicator as a quantitative non-financial variable was in support of the research hypothesis and the hypothesis was therefore accepted.

Hypotheses 3 - The statistical financial distress prediction models had higher Type I and Type II errors than the proposed artificial intelligence model based on a combination of financial, market and quantitative non-financial variables.

The results of this analysis indicated that the financial distress prediction model, with added market variables and lagging business cycle indicator as a quantitative non-financial variable (F-score - Model 5), outperformed the SVM-K-score and SVM-Z-score statistical models and all the remaining F-score models. This was based on the significant improvement of the F-score – Model 5 from the training test to the validation test and it recorded the second least combined ranking (highest overall accuracy and least Type I and Type II errors).

It was concluded that the research hypothesis be accepted.

12.4 CONTRIBUTION TO EXISTING RESEARCH

The contribution of the current study is based on a South African-based model developed within the broad ambit of artificial intelligence or more specific, machine learning. The study empirically established that a combination of financial and quantitative non-financial variables in a financial distress prediction model enhanced the ability of a particular company stakeholder to identify financial distress earlier and more accurately, and where applicable, to take the appropriate remedial action to avoid default, and ultimately, bankruptcy. The earlier financial distress was detected, the better the likelihood of avoiding bankruptcy ultimately.

Based on an overview of the historical evolvement of financial distress prediction models, from the basic statistical models to the more sophisticated artificial intelligence models, the study came to the conclusion that there was no unified approach in financial distress prediction. In addition, it was evident that financial distress prediction was still in its infancy in South Africa, both in academic and practical applications.

The study added and expanded on the existing knowledge base in the academic community by comparing the financial distress predictive power of the De la Rey K-score model with that of the Altman Z-score model and also both these models with the results of an artificial intelligence financial prediction model (referred to as the F-score model). Furthermore, the study introduced the benefits derived from combining financial, market and various macroeconomic variables in the models.

The study proved that there is unlimited scope for the successful application of an artificial intelligence financial distress prediction model and its derivatives in South Africa. This study, therefore, introduced a modular financial distress prediction model which any stakeholder could adapt to their unique requirements and circumstances, by adding different combinations of financial, market and quantitative non-financial variables into one financial distress prediction model.

12.5 RECOMMENDATIONS FOR FURTHER RESEARCH

The field of financial distress prediction has significant scope for further research in South Africa. The following are recommendations for further research:

- The vast number of financial distress prediction models, each based on a unique set of variables, is proof that there is little or no uniformity in company financial distress prediction. First, a clear understanding of the terms *financial distress* and *non-distress* must be reached. Distress prediction is not a dichotomous event, but an event that can be pinpointed anywhere on a financial distress continuum.
- The availability of information on financially distressed companies must be improved by following the Chinese stock exchange's model. In terms of this model, companies not complying with certain financial parameters will be identified with a "Special Treatment" (ST) code on the JSE share information boards. In addition, South Africa must comply with international standards requiring companies in bankruptcy to file a final set of financial statements.
- Reliance is currently placed on internationally pre-developed software. This restricts the selection of certain parameters, which could affect the outcome and accuracy of test results. Relying on internationally developed software creates considerable scope for research in the field of software development, with application to company financial distress prediction.
- The application of machine learning models, within the wider artificial intelligence domain, to the prediction of company financial distress in South Africa is still in its infancy. This study attempted to provide an overview of the evolvement of financial distress prediction models and hopefully initiated further research in the field of artificial intelligence, machine learning and deep learning with application to financial distress prediction.
- Finally, as indicated above, the next potential field for research into financial distress prediction is *deep learning*. Deep learning is part of the broader domain of artificial intelligence and machine learning and consists of applications to learning tasks of artificial intelligence networks that contain more than one hidden layer.

12.6 CONCLUSION

This chapter provided a synopsis of the entire study. A broad overview was presented of the development of financial prediction models, which identified shortcomings unique to South Africa. A financial distress prediction model was developed based on machine learning principles, enhanced with market and quantitative non-financial variables and compared with existing financial prediction models. The empirical results proved that there certainly was value in the enhanced financial distress prediction model. However, research on the financial distress prediction model is still in its infancy, confirming the scope for further research into the broadly defined artificial intelligence financial distress prediction field.

References

Abid, F. & Zouari, A. 2002. Predicting corporate financial distress: a new neural network approach. *Finance India*, 16(2):601-612.

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

Ahmadi, F., Amjadian, S. & Pardegi, K. 2012. New approach to bankruptcy prediction using genetic algorithm. *International Journal of Computer Applications*, 44(4):34-38.

Ahn, B.S., Cho, S.S. & Kim, C.Y. 2000. The integrated methodology of rough set theory and artificial neural network for business failure prediction. *Expert Systems with Applications*, 18:65-74.

Ahn, H. & Kim, K. 2009. Bankruptcy prediction modelling with hybrid case-based reasoning and genetic algorithms approach. *Applied Soft Computing*, 9:599-607.

Alfaro, E., García, N., Gámez, M. & Elizondo, D. 2008. Bankruptcy forecasting: an empirical comparison of AdaBoost and neural networks. *Decision Support Systems*, 45:110-122.

Altman, E.I. 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance*, 24:589-609.

Altman, E.I., Iwanicz-Drozdowska, M., Laitinen, E.K. & Suvas, A. 2017, Financial distress prediction in an international context: a review and empirical analysis of Altman’s Z-score model. *Journal of International Financial Management and Accounting,* 28(2):131-171.

Altman, E.I., Sabato, G. & Wilson, N. 2008. *The value of qualitative information in SME risk management*. Working paper, Credit Management Research Centre. Leeds: Leeds University Business School.

Altman, E.I., Zhang, L. & Yen, J. 2007. *Corporate financial distress diagnosis in China.* Salomon Centre. Working Paper. New York: New York University.

Appiah, K.O. & Abor, J. 2009. Predicting corporate failure: some empirical evidence from the UK. *Benchmarking: An International Journal*, 16(3):432-444.

Arens, F.J. 2014. *The Altman corporate failure prediction model: Applied amongst South African medical schemes.* Unpublished master's dissertation. Department of Finance and Tax. Cape Town: University of Cape Town.

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

Argyrou, A. 2006. *Predicting financial distress using neural networks: another episode to the serial?* Unpublished master's dissertation. Department Accounting Hanken, Sweden: The Swedish School of Economics and Business Administration.

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

Avenhuis, J.O. 2013. *Testing the generalizability of the bankruptcy prediction models of Altman, Ohlson and Zmijewski for Dutch listed and large non-listed firms.* Unpublished master's dissertation. School of Management and Governance. Enschede, the Netherlands: University of Twente,

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

Baek, J. & Cho, S. 2003. Bankruptcy prediction for credit risk using an auto-associative neural network in Korean firms. *Proceedings of the 2003 IEEE International Conference on Computational Intelligence and Financial Engineering,* Seoul, 20-23 March:25-29.

Balcaen, S. & Ooghe, H. 2004. *Alternative methodologies in studies on business failure: do they produce better results than the classical statistical methods?* Working Paper, Faculteit Economie en Bedrijfskunde. Gent: Universiteit Gent: 249:1-33.

Balcaen, S. & Ooghe, H. 2006. 35 years of studies on business failure: an overview of the classic statistical methodologies and their related problems. *The British Accounting Review, 38:63-93.*

Baixauli, J.M. & Módica-Milo, A. 2010. The bias of unhealthy SMEs in bankruptcy prediction models. *Journal of Small Business and Enterprise Development*, 17(1):60-77.

Batista, G. & Monard, M.C. n.d. *An analysis of four missing data treatment methods for supervised learning*. Working paper. Institute of Mathematics and Computer Science. São Paulo: University of São Paulo.

Bauer, J. & Agarwal, V. 2014. Are hazard models superior to traditional bankruptcy prediction approaches? A comprehensive test. *Journal of Banking and Finance*, 40:432-442.

Beaver, W.H. 1966. Financial ratios as predictors of failure. *Supplement to Journal of Accounting Research,* 71-111.

Becerra, V.M., Galvão, R.K.H. & Abou-Seada, M. 2002. On the utility of input selection and pruning for financial distress prediction models. *Proceedings of the 2002 International Joint Conference on Neural Networks,* Honolulu, Hl. 12-17 May 2002:1328-1333.

Becchetti, L. & Sierra, J. 2003. Bankruptcy risk and productive efficiency in manufacturing firms. *Journal of Banking and Finance*, 27:2099-2120.

Behbood, V., lu, J. & Zhang, G. 2010. Adaptive inference-based learning and rule generation algorithms in fuzzy neural network for failure prediction. *Proceedings of the 2010 International Conference on Intelligent Systems and Knowledge Engineering,* Hangzhou, 15-16 November: 33-38.

Bellovary, J., Giacomino, D. & Akers, M. 2007. A review of bankruptcy prediction studies: 1930 to present. *Journal of Financial Education*, 33:1-42.

Berg, D. 2007. Bankruptcy prediction by generalized additive models. *Applied Stochastic Models in Business and Industry*, 23(1):129-143.

Bioch, J.C. & Popova, V. 2003. *Bankruptcy prediction with rough sets. Report Series*, Reference number ERS-2001-11-LIS. Research in Management. Erasmus Research Institute of Management, Rotterdam: Erasmus University.

Blum, M. 1974. Failing company discriminant analysis. *Journal of Accounting Research*, 12(1):1-15.

Brabazon, A. & O'Neill, M. 2004. Diagnosing corporate stability using grammatical evolution. *International Journal of Applied Mathematical and Computer Science*, 14(3):363-374.

Brédart, X. 2014. Bankruptcy prediction model: the case of the United States. *International Journal of Economics and Finance*, 6(3):1-7.

Bunyaminu, A. & Issah, M. 2012. Predicting corporate failure of UK's listed companies: comparing multiple discriminant analysis and logistic regression. *International Research Journal of Finance and Economics*, 94:6-22.

Casey, C.J. & Bartczak, N.J. 1984. Cash flow – It's not the bottom line. *Harvard Business Review,* 62(4):61-66.

Cleofas-Sánches, L., García, V., Marqués, A.I. & Sánchez, J.S. 2016. Financial distress prediction using hybrid associative memory with translation. *Applied Soft Computing*, 44:144-152.

Charalambakis, E.C. & Garrett, I. 2016. On the prediction of financial distress in developed and emerging markets: does the choice of accounting and market information matter? A comparison of UK and Indian firms. *Review of Quantitative Finance and Accounting,* 47:1-28.

Charalambous, C., Charitou A. & Kaourou, F. 2000. Comparative analysis of artificial neural network models: application in bankruptcy prediction. *Annals of Operations Research*, 99:403-425.

Charitou, A., Neophytou, E. & Charalambous, C. 2004. Predicting corporate failure: empirical evidence for the UK. *European Accounting Review*, 13(3):465-497.

Chaudhuri, A. & De, K. 2011. Fuzzy support vector machine for bankruptcy prediction. *Applied Soft Computing*, 11:2472-2486.

Chen, H., Yang, B., Wang, G., Liu, J., Xu, X., Wang, S. & Liu, D. 2011. A novel bankruptcy prediction model based on an adaptive fuzzy k-nearest neighbour method. *Knowledge-Based Systems*, 24:1348-1359.

Chen, M. 2011a. Bankruptcy prediction in firms with statistical and intelligent techniques and a comparison of evolutionary computation approaches. *Computers and Mathematics with Applications*, 62:4514-4524.

Chen, M. 2011b. Predicting corporate financial distress based on integration of decision tree classification and logistic regression. *Expert Systems with Applications*, 38:11261-11272.

Chen, M. 2011c. Using a hybrid evolution approach to forecast financial failures for Taiwan-listed companies. *Quantitative Finance*, 1-12.

Chen, L. & Hsiao, H. 2008. Feature selection to diagnose a business crisis by using a real GA-based support vector machine: an empirical study. *Expert Systems with Applications*, 35:1145-1155.

Chen, N., Ribeiro, B., Vieira, A. & Chen, A. 2012. Clustering and visualization of bankruptcy trajectory using self-organizing map. *Expert Systems with Applications,* 40:385-393.

Chen, N. & Vieira, A. 2009. Bankruptcy prediction based on independent component analysis. *Proceedings of the International Conference on Agents and Artificial Intelligence,* Porto, Portugal*,* 19-21 January:150-155.

Cheng, C., Chen, C. & Fu, C. 2006. Financial distress prediction by a radial basis function network with logit analysis learning. *Computers and Mathematics with Applications*, 51:579-588.

Coats, P.K. & Fant, L.F. 1993. Recognizing financial distress patterns using a neural network tool. *Financial Management,* 22(3):142-155.

Cortés, E.A., Martínez, M.G. & Rubio, N.G. 2007. A boosting approach for corporate failure prediction. *Journal for Applied Intelligence*, 27(1):29-37.

Cortes, V. & Vapnik, V. 1995. Support-vector networks. *Machine Learning*, 20:273-297.

Cousineau, D. & Chartier, S. 2010. Outliers detection and treatment: a review. *International Journal of Psychological Research*, 3(1):58-67.

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

Dakovic, R., Czado, C. & Berg, D. 2010. Bankruptcy prediction in Norway: a comparison study. *Applied Economic Letters*, 17:1739-1746.

Deakin, M. 1972. A discriminant analysis of predictors of business failures. *Journal of Accounting Research*, 10(1):167-179.

De la Rey, J.H. 1981. *Finansiële verhoudingsgetalle en die voorspelling van finansiële mislukking by nywerheidsondernemings in die Republiek van Suid Afrika.* Verslag E1. Buro vir Finansiële Analise, Pretoria: Universiteit van Pretoria.

Divsalar, M., Roodsaz, H., Vahdatinia, F., Norouzzadeh, G. 7 Behrooz, H. 2012. A robust data-mining approach to bankruptcy prediction. *Journal of Forecasting*, 31:504-523.

Duan, G., Huang, Z. & Wang, J. 2010. Extreme learning machines for financial distress prediction for listed company. *2010 International Conference on Logistic Systems and Intelligent Management,* Harbin, 9-10 January:1961-1965.

Dunis, C.L. & Triantafyllidis, J.A. 2003. Alternative forecasting techniques for predicting company insolvencies: the UK Example (1980-2001). *Neural Network World*, 13(4):326-360.

Fallahpour, S., Lakvan, E.N. & Zadeh, M.H. 2017. Using an ensemble classifier based on sequential floating forward selection for financial distress prediction problem. *Journal of Retailing and Consumer Services*, 34:159-167.

Fanping, K. & Shiwei, Z. 2010. Corporate financial distress prediction based on rough sets and PCA-RBFN model. *Proceedings of the Third IEEE International Conference on Computer Science and Information Technology*, Chengdu, 9-11 July:310-314.

Fletcher, T. 2009. *Support vector machines explained*. Department Computer Science. University College, London, March:1-18.

Fu-yuan, H. 2008. A genetic fuzzy neural network for bankruptcy prediction. *Proceedings of the 2008 International Conference on Risk Management & Engineering Management*, Zhanjiang City, 542-546.

Gepp, A.C.B. 2005. *An evaluation of decision tree and survival analysis techniques for business failure prediction*. Unpublished master's dissertation, Queensland: Bond University.

Gepp, A., Kumar, K. & Bhattacharya, S. 2010. Business failure prediction using decision trees. *Journal of Forecasting*, 29:536-555.

Ghosh, D. & Vogt, A. 2012. Outliers: an evaluation of methodologies. Section on Survey Research Methods. Joint Statistical Meetings, San Diego, July 28-August 2:3455-3460.

Gorgani, M.E., Moradi, M. & Yazdi, H.S. 2010. An empirical modelling of companies using support vector data description. *International Journal of Trade, Economics and Finance*, 1(2):221-224.

Grunert, J., Norden, L. & Weber, M. 2005. The role of non-financial factors in internal credit ratings. *Journal of Banking & Finance,* 29:509-531.

Gulfbase. *Investment tutorial, stock market ratios.* [Online] Available from: http://gulfbase.com/ [Accessed: 2015-12-15].

Gunnersen, S., Smith-Miles, K. & Le, V. 2012. Towards objective data selection in bankruptcy prediction. Proceedings of the WCCI 2012 IEEE World Congress on Computational Intelligence, Brisbane, Australia, June 10-15:1-8.

Guthikonda, S.M. 2005. *Kohonen self-organizing maps*. Working paper. Ohio: Wittenberg University, December:1-19.

Haber, J.R. 2005. Assessing how bankruptcy prediction models are evaluated. *Journal of Business & Economics Research*, 3(1):87-92.

Hall, S.G. 2002. *The macroeconomic and microeconomic factors influencing financial distress: a comparison of Altman's 1983 and 1993 bankruptcy prediction models.* Unpublished doctoral thesis. H. Wayne Huizenga School of Business and Entrepreneurship. Florida: Nova South-Eastern University.

Hanson, R.O. 2002. *A study of Altman's revised four-variable Z-score bankruptcy prediction model as it applies to the service industry.* Unpublished doctoral thesis. H. Wayne Huizenga School of Business and Entrepreneurship, Florida: Nova South-Eastern University.

Härdle, W.F., Prastyo, D.D. & Hafner, C. 2012. Support vector machines with evolutionary feature selection for default prediction. *SFB 649 Discussion Paper 2012-030*, 1-23.

Hauser, R.P. & Booth, D. 2011. Predicting bankruptcy with robust logistic regression. *Journal of Data Science*, 9:565-584.

He, J. 2002. *An empirical investigation of financial and market performance in the prediction of business failure for small (public) firm: an over-the-counter (OTC) experience.* Unpublished doctoral thesis. Cleveland: Cleveland State University.

Henchiri, Y., Benammou, S. & Hamza, T. 2009. Using support vector machine approach for forecasting the failure of the Tunisian companies. *Proceedings of the Seventh Scientific Meeting of Classification and Data Analysis Group (CLADAG) of the Italian Statistical Society.* Catania, Italy, 9-11 September:537-540.

Hofstee, E. 2006. *Constructing a good dissertation: A practical guide to finishing a master's, MBA or PhD on schedule.* Sandton, South Africa: EPE.

Hol, S. 2007. The influence of the business cycle on bankruptcy probability. *International Transactions in Operational Research,* 14:75-90.

Hossari, G. 2007. A heuristic framework for ratio-based modelling of corporate collapse. *Proceedings of the Twelfth Finsia-Melbourne Centre for Financial Studies Banking and Finance Conference*, Melbourne, 24-25 September 2007: 1-29.

Hossari, G. 2009. Signalling corporate collapse using a dual-classification scheme: Australian evidence. *International Review of Business Research Papers*, 5(3):333-345.

Hossari, G. 2010. Enhancing procedural efficiency in multiple-industry ratio-based modelling of corporate collapse. *International Review of Business Research Papers*, 6(1):99-107.

Hu, Y. 2008. Incorporating a non-additive decision making method into multi-layer neural networks and its application to financial distress analysis. *Knowledge-Based Systems*, 21:383-390.

Huang, S., Tsai, C., Yen, D.C. & Cheng, Y. 2008. A hybrid financial analysis model for business failure prediction. *Expert Systems with Applications*, 35:1034-1040.

Huang, S., Tang, Y., Lee, C. & Chang, M. 2012. Kernel local Fisher discriminant analysis based manifold-regularized SVM model for financial distress prediction. *Expert Systems with Applications*, 39:3855-3861.

Hung, C., Chen, J. & Wermter, S. 2007. *Hybrid probability based ensembles for bankruptcy prediction*. Working paper. Taiwan: Chung Yuan Christian University, n.d:1-6.

Jian-guang, H., Xiao-feng, H. & Jie, S. 2010. Dynamic financial distress prediction modelling based on slip time window and multiple classifiers. *Proceedings of the Seventeenth International Conference on Management Science & Engineering,* Melbourne, November 24-26:148-155.

Jing-rong, D & Jun, C. 2009. An evolutionary based wavelet network for business failure prediction. *WRI Global Congress on Intelligent Systems*, People's Republic of China: Chongqing Normal University.

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

Khademolqorani, S., Hamadani, A.Z. & Rafeie, F.M. 2015. A hybrid analysis approach to improve financial distress forecasting: empirical evidence from Iran. *Mathematical Problems in Engineering*: 1-9.

Kidane, H.W. 2004. *Predicting financial distress in IT and service companies in South Africa.* Unpublished master's dissertation. Faculty of Economic and Management Sciences. Bloemfontein: University of the Free State.

Kim, M.H. 2011.*The dynamic prediction of company failure: the influence of time, the economy and non-linearity*. Unpublished DPhil thesis. Faculty of Business and Economics. Sydney: University of Sydney.

Kim, M. & Kang, D. 2012. Classifiers selection in ensembles using genetic algorithms for bankruptcy prediction. *Expert Systems with Applications*, 39:9308-9314.

Kim, M.H. & Partington, G. 2008. The dynamic prediction of company failure. *Submitted for the Twenty-first Australasian Finance and Banking Conference,* Sydney, 16-18 December:41-77.

Kim, M.H. & Partington, G. 2014. Dynamic forecasts of financial distress of Australian firms. *Australian Journal of Management*,1-26.

Kim, M.H. & Yoo, P.D.A. 2006. Semiparametric model approach to financial bankruptcy prediction. *Proceedings of the 2006 IEEE International Conference on Engineering and Intelligent Systems.* Islamabad, 22-23 April:1-6.

Ko, P. & Lin, P. 2006. An evolution-based approach with modularized evaluations to forecast financial distress. *Knowledge-Based Systems*,19:84-91.

Koh, H.C. & Low, C.K. 2004. Going concern prediction using data mining techniques. *Managerial Auditing Journal*, 19(3):462-476.

Kornik, S. 2004. *Machine learning for corporate failure prediction: an empirical study of South African studies.* Unpublished master's dissertation. Department of Accounting and Department of Electrical Engineering. Cape Town: University of Cape Town.

Korobow, L. & Stuhr, D. 1985. Performance measurement of early warning models: comments on West and other weakness/failure prediction models. *Journal of Banking & Finance*, 9:267-273.

Kumar, K. & Tan, C. 2004. *Artificial intelligence in financial distress prediction.* Working Paper. Queensland: Bond University.

Lam, M. 2004. Neural network techniques for financial performance prediction: Integrating fundamental and technical analysis. *Decision Support Systems*, 37:567-581.

Lee, T-S. 2004. Incorporating financial ratios and intellectual capital in bankruptcy predictions. *International Journal of Business Systems*, 1-32.

Lee, W. 2008. *An empirical comparison of bankruptcy models – evidence from Taiwan*. Working Paper. Tamsui: Aletheia University.

Lee, M. & Chen, J. 2007. *Developing financial distress classification approach based on fuzzy-rough nearest-neighbor approach*. Working Paper. Department of Information Management. Taiwan: Fooyin University.

Lee, T. & Chen, I. 2005. A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines. *Expert Systems with Applications*, 28:743-752.

Leksrisakul, P. & Evans, M. 2005. A model of corporate bankruptcy in Thailand using multiple discriminant analysis. *Journal of Economic and Social Policy*, 10(1):1-36.

Lensberg, T., Eilifsen, A. & McKee, T.E. 2006. Bankruptcy theory development and classification via genetic programming. *European Journal of Operational Research*, 169, 677-697.

Li, H. & Sun, J. 2011. Predicting business failure using forward ranking-order case-based reasoning. *Expert Systems with Applications, 38*:3075-3084.

Li, H., Sun, J., Li, J. & Yan, X. 2013. Forecasting business failure using two-stage ensemble of multivariate discriminant analysis and logistic regression. *Expert Systems*, 30(5):385-397.

Libby, R. 1975. Accounting ratios and the prediction of failures: some behavioural evidence. *Journal of Accounting Research*, 13(Spring):150-161.

Lieu, P., Lin, C. & Yu, H. 2008. Financial early-warning models on cross-holding groups. *Industrial Management & Data Systems*, 108(8):1060-1080.

Lin, F., Liang, D. & Chu, W. 2010. The role of non-financial features related to corporate governance in business crisis prediction. *Journal of Marine Science and Technology*, 18(4):504-513.

Lin, F., Liang, D. & Chen, E. 2011. Financial ratio selection for business crisis prediction. *Expert Systems with Applications*, 38:15094-15102.

Lin, F.Y. & McClean, S. 2000. The prediction of financial distress using structured financial data from the internet. *International Journal of Computer Science and Security*, 1(1):43-57.

Lin, F., Yeh, C. & Lee, M. 2011. The use of hybrid manifold learning and support vector machines in the prediction of business failure. *Knowledge-based Systems*, 24: 95-101.

Lin, F., Yeh, C. & Lee, M 2013. A hybrid business failure prediction model using locally linear embedding and support vector machines. *Romanian Journal of Economic Forecasting*, 1:82-97.

Liu, J. 2004. Macroeconomic determinants of corporate failures: evidence from the UK. *Applied Economics,* 36:939-945.

Lussier, R.N. 1995. A nonfinancial business success model versus failure prediction models for young firms. *Journal of Small Business,* January:8-20.

Maditinos, D. & Chatzoglou, P. 2004. The use of neural networks in forecasting. *Review of Economic Sciences*, 6:161-176.

Marconi, M., Quaranta, A.G. & Tartufoli, S. 2010. Evaluating the risk of insolvency. *Open Journal of Economic Research*, 15-32.

Maree, J.G. ed. 2012. *Complete your thesis or dissertation successfully: practical guidelines.* Cape Town: Juta.

Martin, A., Aswathy, V. & Venkatesan, V.P. 2012. Framing qualitative bankruptcy prediction rules using ant colony algorithm. *International Journal of Computer Applications*, 41(21):27-31.

Martin, A., Gayathri, V., Saranya, G., Gayathri, P. & Venkatesan, P. 2011. A hybrid model for bankruptcy prediction using genetic algorithm, fuzzy c-means and mars. *International Journal on Soft Computing (UCS),* 2(1):12-24.

Masekesa, W.M. 2010. *Macroeconomic determinants of corporate failures in South Africa (1994 – 2009)*. Unpublished master's dissertation, Faculty of Commerce and Law, Wits Business School. Johannesburg: University of the Witwatersrand.

Masten, A.B. & Masten, I. 2009. *Comparison of parametric, semi-parametric and non-parametric methods in bankruptcy prediction*. Working Paper. Faculty of Economics. Ljubljana, Slovenia: University of Ljubljana.

Mayer-Schönberger, V. & Cukier, K. 2013. *Big data: a revolution that will transform how we live, work, and think*. London: John Murray.

McKee, T.E. 2003. Rough sets bankruptcy prediction models versus auditor signalling rates. *Journal of Forecasting*, 22:569-586.

McKee, T.E. & Greenstein, M. 2000. Predicting bankruptcy using recursive partitioning and a realistically proportioned data set. *Journal of Forecasting*, 19:219-230.

Merkevicius, E. & Garsva, G. 2007. Prediction of changes of bankruptcy classes with neuro-discriminate model based on self-organizing maps. *Information Technology and Control*, 36(1A):145-151.

Min, J.H. & Lee, Y-C. 2004. *Business failure prediction with support vector machines and neural networks: a comparative study*. Working Paper. College of Business Administration, Seoul: Sogang University.

Min, J.H. & Lee, Y. 2005. Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. *Expert Systems with Applications*, 28:603-614.

Min, S., Lee, J. & Han, I. 2006. Hybrid genetic algorithms and support vector machines for bankruptcy prediction. *Expert Systems with Applications*, 31:662-660.

Moepya, S.O., Nelwamondo, F.V. & Van der Walt, C. 2014. A support vector machine approach to detect financial statement fraud in South Africa: a first look. *Intelligent Information and Database Systems*, 8398:42-51.

Mukhopadhyay, A., Tiwari, S., Narsaria, A. & Karmaker, B.R. 2012. A new approach to predicting bankruptcy: combining DEA and multi-layer perceptron. *International Journal of Computer Science*, 9(4)(2):71-78.

Muller, G.H. 2008. *Development of a model to predict financial distress of companies listed on the JSE.* Research report presented in partial fulfilment of the requirements for the degree of Master of Business Administration. Stellenbosch: University of Stellenbosch.

Muller, G.H., Steyn-Bruwer, B.W, & 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):21-32.

Nagaraj, K. & Sridhar, A. 2015. A predictive system for detection of bankruptcy using machine learning techniques. *International Journal of Data Mining & Knowledge Management Process*, 5(1):29-40.

Naidoo, S.R. *A predictive model of the states of financial health in South African Businesses*. Unpublished doctoral thesis. Pretoria: University of South Africa.

Neophytou, E., Charitou, A. & Charalambous, C. 2000. Predicting corporate failure: empirical evidence for the UK. *European Accounting Review*, 13(3):1-30.

Neves, J.C. & Vieira, A. 2004. *Estimating bankruptcy using neural networks trained with hidden layer learning vector quantization.* Working Paper series 2-2004. Departamento de Gestão, Lisbon: School of Economics & Management.

Neves, J.C. & Vieira, A. 2006. Improving bankruptcy prediction with hidden layer learning vector quantization. *European Accounting Review,* 15(2):253-271.

Odam, M.D. & Sharda, R. 1990. A neural network model for bankruptcy prediction. *Proceedings of the IJCNN International Joint Conference on Neural Networks,* San Diego, 17-20 June:163-168.

Olson, D.L., Delen, D. & Meng, Y. 2012. Comparative analysis of data mining methods for bankruptcy prediction. *Decision Support Systems*, 52:464-473.

Őnder, C. 2010. *Bankruptcy prediction with support vector machines: An application to German companies.* Unpublished master's dissertation, Centre for Applied Statistics and Economics. Berlin: Humboldt-Universität zu Berlin.

Ooghe, H., Spaenjers. C. & Vandermoere, P. 2005. *Business failure prediction: simple-intuitive models versus statistical models.* Working Paper. Faculteit Economie en Bedrijfskunde. Gent: Universiteit Gent.

Peat, M. & Jones, S. 2012. Using neural nets to combine information sets in corporate bankruptcy prediction. *Intelligent Systems in Accounting, Finance and Management*, 19:90-101.

Pompe, P.P.M. & Bilderbeek, J. 2005. The prediction of bankruptcy of small- and medium-sized industrial firms. *Journal of Business Venturing*, 20:847-868.

Prieto, I. & Revilla, E. 2006. Learning capability and business performance: a non-financial and financial assessment. *The Learning Organization*, 13(2):166-185.

Quah, T. & Srinivasan, B. 2005. *Early detection of corporate failure using neural network*. Working Paper. School of Electrical and Electronic Engineering and School of Accountancy and Business. Singapore: Nanyang Technological University.

Rees, B. 1995. *Financial analysis*. 2nd ed. Hertfordshire: Prentice Hall Europe.

Ribeiro, B., Silva, C., Chen, N., Vieira, A. & Das Neves, J.C. 2012. Enhances default risk models with SVM+. *Expert Systems with Applications,* 39:10140-10152.

Ribeiro, B., Silva, C., Vieira, A., Gaspar-Cunha, A. & Das Neves, J. 2010. Financial distress model prediction using SVM+. *Proceedings of the 2010 International Joint Conference on Neural Networks,* Barcelona, 18-23 July:1-7.

Rini, D.P., Shamsuddin, S.M. & Yuhaniz, S.S. 2011. Particle swarm optimization: technique, system and challenges. *International Journal of Computer Applications (0975-8887)*, 14(1):19-27.

Rui, L. 2010. A particle swarm optimized fuzzy neural network for bankruptcy prediction. *Proceedings of the 2010 International Conference on Future Information Technology and Management Engineering*, Chanzhou, 9-10 October:557-560.

Sai, Y., Zhong, C. & Qu, L. 2007. A hybrid GA-BP model for bankruptcy prediction. *Eighth International Symposium on Autonomous Decentralized Systems (ISAD'07)*, Sedona, 21-23 March:473-477.

Santos, M.F., Cortez, P., Pereira, J. & Quintela, H. 2006. Corporate bankruptcy prediction using data mining techniques. In Zanasi, A., Brebbia, C,A, & Ebecken, N.F.F. eds. *Data Mining VII: Data, text and web mining and their business applications.* Southampton: WITpress, 349-357.

Saunders, P.L., Lewis, P. & Thornhill, A. 2009. *Research methods for business students*. 5th ed. Essex: Pearson Education Ltd.

Scarlat, E. Delcea, C. 2010. Hybrid grey-fuzzy-neural networks model for enterprise bankruptcy, *Proceedings of the 2010 International Joint Conference on Neural Networks*, Barcelona, 18-23 July:1-8.

Shi, H. 2009. Integration of unascertained method and neural networks in financial early warning. *Proceedings of the Second 2009 International Conference on Business Intelligence and Financial Engineering,* Beijing, 24-26 July:57-60.

Shi, L., Xi, L., Ma, X. & Hu, X. 2009. Bagging of artificial neural networks for bankruptcy prediction. *Proceedings of the 2009 International Conference on Information and Financial Engineering*, Singapore, 17-20 April:154-156.

Shin, K., Lee, T.S. & Kim, H. 2005. An application of support vector machines in bankruptcy prediction model. *Expert Systems with Application*, 28:127-135.

Shin, K-S. & Lee, Y-J. 2002. A genetic algorithm application in bankruptcy prediction modelling. *Expert Systems with Applications*. 23(1):321-328

Shiri, M.M., Ahangary, M., Vaghfi, S.h. & Kholousi, A. 2012. Corporate bankruptcy prediction using data mining techniques: evidence from Iran. *African Journal of Scientific Research*, 8(1):404-426.

Shumway, T. 2001. Forecasting bankruptcy more accurately: A sample hazard model. *Journal of Business*, 74:101-124.

Simić, D., Kovačević, I. & Simić, S. 2011. Insolvency prediction for assessing corporate financial health. *Logic Journal of IGPL Advanced Access*, 1-14.

Sookhanaphibarn, K., Polsiri, P., Choensawat, W. & Lin, F.C. 2007. Application of neural networks in business bankruptcy analysis in Thailand. *International Journal of Computational Intelligence Research*, 3(1):91-96.

South Africa. Companies Act, 71 of 2008, & Regulations. Cape Town: Juta Law.

South African Reserve Bank. 2015. *Quarterly Bulletin - June 2015 (No. 276).* Pretoria: South African Reserve Bank.

South African Reserve Bank. 2015. *Quarterly Bulletin - December 2015 (No. 278).* Pretoria: South African Reserve Bank.

Sprengers, M.A. 2005. *Bankruptcy prediction using classification and regression trees.* Unpublished bachelor dissertation. Informatics & Economics, Faculty of Economics. Rotterdam: Erasmus University.

Steyn, B.L., Warren, B.O. & Jonker, W.D. 2000. *Analysis, Valuations and Restructuring*. Pretoria: Renall.

Sun, L. 2007. A re-evaluation of auditors' opinions versus statistical models in bankruptcy prediction. *Review of Quantitative Finance and Accounting*, 28:55-78.

Sun, J. & Li, H. 2009. Financial distress early warning based on group decision making. *Computers & Operations Research,* 36:885-906.

Sun, J. & Li, H. 2012. Financial distress prediction using support vector machines: ensemble vs. individual. *Applied Soft Computing,* 12:2254-2265.

Sun, J., Li, H., Chang. P. & He, K. 2015. The dynamic financial distress prediction method of EBW-VSTW-SVM. *Enterprise Information Systems*,: 1-28

Thevnin, C. 2003. *A comparative examination of bankruptcy prediction; Altman MDA study versus Luther ANN study: a test of predictive strength between two techniques.* Unpublished doctoral thesis. Florida: Nova Southeastern University.

Tinoco, M.H. & Wilson, N. 2013. Financial distress and bankruptcy prediction among listed companies using accounting, market and macroeconomic variables. *International Review of Financial Analysis*, 30:394-419.

Tofallis, C. 2015. A better measure of relative prediction accuracy for model selection and model estimation. *Journal of the Operational Research Society*, 66:1352-1362.

Trippi, R.R. & Turban, E. (eds.) 1993. *Neural networks in finance and investing: using artificial intelligence to improve real-world performance.* Chicago: Probus.

Tsakonas, A., Dounias, G., Doumpos, M. & Zopounidis, C. 2006. Bankruptcy prediction with neural logic networks by means of grammar-guided genetic programming. *Expert Systems with Applications*, 30:449-461.

Van Gestel, T., Baesens, B., Suykens, J., Espinoza, M., Baestaens, D., Vanthienen, J. & De Moor, B. 2003. Bankruptcy prediction with least squares support vector machine classifiers. *Proceedings of the 2003 IEEE International Conference on Computational Intelligence for Financial Engineering*, Hong Kong, 20-23 March 2003:1-8.

Van Gestel, T., Baesens, B., Suykens, J.A.K., Van den Poel, D., Baestaens, D. & Willekens, M. 2006. Bayesian kernel based classification for financial distress detection. *European Journal of Operational Research*, 172:979-1003.

Van der Colff, F. 2012. *Company financial failure and distress: a perspective.* Unpublished master's dissertation. Faculty of Economic and Management Sciences. Pretoria: University of Pretoria.

Venter, J.C. & Pretorius, W.S. 2001. A note on the business cycle in South Africa during the period 1997 to 1999. *South African Reserve Bank Quarterly Bulletin,* 63-69

Wang, S., Wu, L., Zhang, Y. & Zhou, Z. 2009. Ant colony algorithm used for bankruptcy prediction. *Proceedings of the Second International Symposium on Information Science and Engineering*, Shanghai, China, 26-28 December:137-139.

Wang, Z. & Li, H. 2007. Financial distress prediction of Chinese listed companies: A rough set methodology. *Chinese Management Studies*, 1(2):93-110.

Wen-tsao, P. & Wei-yuan, L. 2008. Use probabilistic neural network to construct early warning model for business financial distress. *Proceedings of the Fifteenth International Conference on Management Science & Engineering.* Long Beach, USA. September 10-12:134-139.

Woodward, J.R. n.d. *GA or GP? That is not the question*. Working paper. School of Computer Science. Birmingham: The University of Birmingham.

Wu, J. 2007. *Do corporate governance factors matter for financial distress prediction of firms?* Evidence from Taiwan. Unpublished master's dissertation. Nottingham: University of Nottingham.

Wu, C., Fang, W. & Goo, Y. 2006. Variable selection method affects SVM approach in bankruptcy prediction. [Online] Available from www.researchgate.com [Downloaded: 2012-12-10].

Wu, W., Lee, V.C.S. & Tan, T.Y. 2006. Data preprocessing and data parsimony in corporate failure forecast models: evidence from Australian materials industry. *Accounting and Finance*, 46: 327-345.

Xiangguang, S. & Xiaozhong, S. 2010. Analysis of Chinese enterprise financial crisis prewarning based on BP neural network. *2010 International conference on Management and Service Science*, Wuhan, China, 24-26 August:1-3.

Xu, W., Xiao, Z., Dang, X., Yang, D. & Yang, X. 2014. Financial ratio selection for business failure prediction using soft set theory. *Knowledge-Based Systems*, 63:59-67.

Yang, Z.R. 2003. Support vector machines for company failure prediction. *Proceedings of the 2003 International Conference on Computational Intelligence for Financial Engineering*, Hong Kong, 20-23 March:47-54.

Yanqing, Z., Shiwei, Z., Junfeng, Y. & Lei, W. 2010. Predicting corporate financial distress by PCA-based support vector machines. *Proceedings of the 2010 International Conference on Networking and Information Technology*, Manila, Philippines, 11-12 June:373-376.

Yap, B.C.F., Munuswamy, S. & Bin Mohamed, Z. 2012. Evaluating company failure in Malaysia using financial ratios and logistic regression. *Asian Journal of Finance & Accounting*, 4(1):330-344.

Yazdipour, R. & Constand, R.L. 2010. Predicting firm failure: a behavioural perspective. *The Journal of Entrepreneurial Finance*, 14(3):90-104.

Yi, J. & Prybutok, V.R. 2001. Neural networks forecasting model as an alternative to OLS regression model for handling messy data. *Commerce and Economic* Review, 17(1):137-158.

Zanganeh, T., Rabiee, M. & Zarei, M. 2011. Applying adaptive neuro-fuzzy model for bankruptcy prediction. *International Journal of Computer Applications*, 20(3):15-21.

Zhang, G.P. 2007. Avoiding pitfalls in neural network research. *IEEE Transactions on Systems, Man, and Cybernetics*, 37(1):3-16.

Zhang, J. 2006. *Three essays on business failure: causality and prediction.* Unpublished doctoral thesis. Texas: Texas A&M University.

Zheng, Q. & Yanhui, J. 2007. Financial distress prediction based on decision tree models*. Proceedings of the IEEE Conference on Service Operations and Logistics and Informatics,* Philadelphia, PA*,* 27-29 August:1-6.

Zhou, L. 2013. Performance of corporate bankruptcy prediction models on imbalanced dataset: the effect of sampling methods. *Knowledge-Based Systems*, 41:16-25.

Zhou, Y. & Elhag, T.M.S. 2007a. Apply logit analysis in bankruptcy prediction. *Proceedings of the Seventh WSEAS International Conference on Simulation, Modelling and Optimization,* Beijing, China, 15-17 September:301-308.

Zhou, Y. & Elhag, T. 2007b. Financial assessment using neural networks. *In. Boyd, D (ed) Proceedings of the Twenty-third Annual ARCOM Conference.* Association of Researchers in Construction Management. Belfast, UK, 3-5 September:45-54.

Zhou, L., Lu, D. & Fujita, H. 2015. The performance of corporate financial distress prediction models with features selection by domain knowledge and data mining approaches. *Knowledge-Based Systems*, 85:52-61.

Zhou, J. & Tian, J. 2007. Predicting corporate financial distress based on rough sets and wavelet support vector machine. *Proceedings of the 2007 International Conference on Wavelet Analysis and Pattern Recognition,* Beijing, China, 2-4 November:602-607.

Zhou, J., Bai, T., Zhang, A. & Tian, J. 2008. The integrated methodology of rough set theory and fuzzy SVM for customer classification. *Proceedings of the 2008 IEEE Conference on Cybernetics and Intelligent Systems,* Chengdu, China, *21-24 September:353-357.*

Zhou, L. & Lai, K.K. 2012. Corporate financial crisis prediction using SVM models with direct search for features selection and parameters optimization. *Proceedings of the 5th International Joint Conference on Computational Sciences and Optimization*, Harbin, Heilongjiang, China, 23-26 June:760-764.

Zhou, L., Lai, K.K. & Yen, J. 2010. Bankruptcy prediction incorporating macroeconomic variables using neural network. *Proceedings of the 2010 International Conference on Technologies and Applications of Artificial Intelligence*, Hsinchu City, Taiwan, 18-20 November:80-85.

Zong-jun, W., Hiong-xia, L. & Xiao-lan, D. 2006. *Empirical study of business failure prediction based on rough set.* Working paper: School of Management. Huazhong, People’s Republic of China: University of Science and Technology.

APPENDIX 7.1

Selected sample

Sector	Company number	Company name	Short name	Code	Year of listing	Years listed	Latest financials available
AltX		**AltX**					
	1	AH-Vest Ltd	AH-VEST	AHL	1998	16	Jun-14
	2	Beige Holdings Ltd	BEIGE	BEG	1997	17	Jun-14
	3	ISA Holdings Ltd	ISA	ISA	1998	16	Feb-14
	4	MoneyWeb Holdings Ltd	MONEYWB	MNY	1999	15	Jun-14
	5	SilverBridge Holdings Ltd	SILVERB	SVB	1999	15	Jun-14
	6	StratCorp Ltd	STRATCORP	STA	2001	13	Feb-14
Basic Materials		**Basic Materials-Chemicals-Chemicals**					
	7	AECI Ltd	AECI	AFE	1966	48	Dec-13
	8	African Oxygen Ltd	AFROX	AFX	1963	51	Dec-13
	9	Delta EMD Ltd	DELTA	DTA	1983	31	Dec-13
	10	Omnia Holdings Ltd	OMNIA	OMN	1980	34	Mar-14
	11	Spanjaard Ltd	SPANJAARD	SPA	1987	27	Feb-14
Consumer Services		**Consumer Services-Media-Media**					
	12	African Media Entertainment Ltd	AME	AME	1997	17	Mar-14
	13	Caxton and CTP Publishers and Printers Ltd	CAXTON	CAT	1948	66	Jun-14
	14	Naspers Ltd	NASPERS-N	NPN	1994	20	Mar-14
		Consumer Services-Retail-Food & Drug					
	15	Clicks Group Ltd	CLICKS	CLS	1996	18	Aug-14
	16	Pick n Pay Stores Ltd	PICKNPAY	PIK	1968	46	Feb-14
	17	The Spar Group Ltd	SPAR	SPP	2004	10	Sep-14
		Consumer Services-Retail-General Retailers					
	18	Mr Price Group Ltd	MRPRICE	MPC	1952	62	Mar-14
	19	Rex Trueform Clothing Company Ltd	REX TRUE	RTO	1945	69	Jun-14
	20	Truworths International Ltd	TRUWTHS	TRU	1998	16	Jun-14
	21	Massmart Holdings Ltd	MASSMART	MSM	2000	14	Dec-13
	22	Nictus Ltd	NICTUS	NCS	1969	45	Mar-14
	23	Woolworths Holdings Ltd	WOOLIES	WHL	1997	17	Jun-14
	24	Cashbuild Ltd	CASHBIL	CSB	1986	28	Jun-14
	25	Iliad Africa Ltd	ILIAD	ILA	1998	16	Dec-13
	26	Italtile Ltd	ITLTILE	ITE	1988	26	Jun-14
	27	JD Group Ltd	JDGROUP	JDG	1986	28	Jun-14
	28	Lewis Group Ltd	LEWIS	LEW	2004	10	Mar-14
		Consumer Services-Travel & Leisure-Travel & Leisure					
	29	Phumelela Gamings and Leisure Ltd	PHUMELELA	PHM	2002	12	Jul-14
	30	Tsogo Sun Holdings Ltd	TSOGO SUN	TSH	1994	20	Mar-14
	31	City Lodge Hotels Ltd	CITYLDG	CLH	1992	22	Jun-14
	32	Famous Brands Ltd	FAMBRANDS	FBR	1994	20	Feb-14
	33	Spur Corporation Ltd	SPURCORP	SUR	1999	15	Jun-14
	34	Cullinan Holdings Ltd	CULLINAN	CUL	1947	67	Sep-14
Consumer		**Consumer-Automobiles & Parts-Automobiles & Parts**					
	35	Metair Investments Ltd	METAIR	MTA	1949	65	Dec-13
		Consumer-Food & Beverages-Beverages					
	36	Awethu Breweries Ltd	AWETHU	AWT	1997	17	Jun-14
	37	Distell Group Ltd	DISTELL	DST	1988	26	Jun-14

Annexure C continued

	Company number	Company name	Short name	Code	Year of listing	Years listed	Latest financials available
		Consumer-Food & Beverages-Food Producers					
	38	Astral Foods Ltd	ASTRAL	ARL	2001	13	Sep-14
	39	Crookes Brothers Ltd	CROOKES	CKS	1948	66	Mar-14
	40	RCL Foods Ltd	RCL	RCL	1989	25	Jun-14
	41	Sovereign Food Investments Ltd	SOVFOOD	SOV	1995	19	Feb-14
	42	Illovo Sugar Ltd	ILLOVO	ILV	1992	22	Mar-14
	43	Tiger Brands Ltd	TIGBRANDS	TBS	1944	70	Sep-14
	44	Tongaat Hulett Ltd	TONGAAT	TON	1952	62	Mar-14
		Consumer-Personal & Household-Household					
	45	Steinhoff International Holdings Ltd	STEINHOFF	SHF	1998	16	Jun-14
		Consumer-Personal & Household-Leisure					
	46	Nu-World Holdings Ltd	NUWORLD	NWL	1987	27	Aug-14
		Consumer-Personal & Household-Personal					
	47	Seardel Investment Corporation Ltd	SEARDEL	SER	1968	46	Mar-14
Health		**Health-Health-Health Equipment & Services**					
	48	Mediclinic International Ltd	MEDCLIN	MDC	1986	28	Mar-14
	49	Netcare Ltd	NETCARE	NTC	1996	18	Sep-14
Industrials		**Industrials-Construction & Materials-Construction & Materials**					
	50	Masonite (Africa) Ltd	MASONITE	MAS	1952	62	Dec-14
	51	PPC Ltd	PPC	PPC	1910	104	Sep-14
	52	Aveng Ltd	AVENG	AEG	1999	15	Jun-14
	53	Basil Read Holdings Ltd	BASREAD	BSR	1987	27	Dec-13
	54	Murray & Roberts Holdings Ltd	M&R-HLD	MUR	1948	66	Jun-14
		Industrials-Industrial Goods & Services-Electronic & Electrical Equipment					
	55	Allied Electronics Corporation Ltd	ALTRON	AEL	1979	35	Feb-14
	56	Jasco Electronics Holdings Ltd	JASCO	JSC	1987	27	Jun-14
	57	Reunert Ltd	REUNERT	RLO	1948	66	Sep-14
	58	Digicore Holdings Ltd	DIGICORE	DGC	1998	16	Jun-14
		Industrials-Industrial Goods & Services-General Industrials					
	59	Astrapak Ltd	ASTRAPAK	APK	1997	17	Feb-14
	60	Nampak Ltd	NAMPAK	NPK	1969	45	Sep-14
	61	Transpaco Ltd	TRNPACO	TPC	1987	27	Jun-14
	62	The Bidvest Group Ltd	BIDVEST	BVT	1990	24	Jun-14
	63	KAP Industrial Holdings Ltd	KAP	KAP	1994	20	Jun-14
	64	Remgro Ltd	REMGRO	REM	2000	14	Jun-14
		Industrials-Industrial Goods & Services-Industrial Engineering					
	65	Howden Africa Holdings Ltd	HOWDEN	HWN	1996	18	Dec-13
	66	Hudaco Industries Ltd	HUDACO	HDC	1985	29	Nov-13
	67	Bell Equipment Ltd	BELL	BEL	1995	19	Dec-13
		Industrials-Industrial Goods & Services-Industrial Transportation					
	68	Grindrod Ltd	GRINDROD	GND	1986	28	Dec-13
	69	OneLogix Group Ltd	ONELOGIX	OLG	2000	14	May-14
	70	Super Group Ltd	SUPRGRP	SPG	1996	18	Jun-14
	71	Value Group Ltd	VALUE	VLE	1998	16	Feb-14
	72	Cargo Carriers Ltd	CARGO	CRG	1987	27	Feb-14
		Industrials-Industrial Goods & Services-Support Services					
	73	Metrofile Holdings Ltd	METROFILE	MFL	1995	19	Jun-14
	74	MICROmega Holdings Ltd	MICROMEGA	MMG	1998	16	Mar-14
	75	Winhold Ltd	WINHOLD	WNH	1946	68	Sep-14
	76	Adcorp Holdings Ltd	ADCORP	ADR	1987	27	Feb-14
	77	Primeserv Group Ltd	PRIMESERV	PMV	1998	16	Mar-14
Technology		**Technology-Technology-Software & Computer Services**					
	78	Adapt IT Holdings Ltd	ADAPTIT	ADI	1998	16	Jun-14
	79	Business Connextion Group Ltd	BCX	BCX	2004	10	Aug-14
	80	Datatec Ltd	DATATEC	DTC	1994	20	Feb-14
	81	Datacentrix Holdings Ltd	DCENTRIX	DCT	1998	16	Feb-14
	82	EOH Holdings Ltd	EOH	EOH	1998	16	Jun-14
	83	Gijima Group Ltd	GIJIMA	GIJ	1999	15	Jun-14
		Technology-Technology-Technology Hardware & Equipment					
	84	Mustek Ltd	MUSTEK	MST	1997	17	Jun-14
	85	Pinnacle Holdings Ltd	PINNACLE	PNC	1987	27	Jun-14
Telecommunicatios		**Telecommunications-Telecommunications-Mobile Telecommunications**					
	86	MTN Group Ltd	MTN GROUP	MTN	1995	19	Dec-13

APPENDIX 8.1

Descriptive statistics

(a) Descriptive statistics (Year 10 to Year 6) – Total sample

ALL	J10			J9			J8			J7			J6		
Statistic	ROE	K	Z	ROE	K	Z	ROE	K	Z	ROE	K	Z	ROE	K	Z
Nbr. of observations	85	85	85	85	85	85	85	85	85	85	85	85	85	85	85
Minimum	-1.202	-4.993	-1.361	-0.611	-5.452	-1.929	-2.556	-5.450	-1.747	-0.331	-4.414	0.193	-4.466	-2.188	0.052
Maximum	3.222	5.414	17.438	6.189	5.224	16.641	4.719	5.352	18.242	3.560	4.598	14.632	2.952	4.169	14.043
1st Quartile	0.060	-0.261	2.934	0.102	-0.220	3.020	0.087	0.005	2.947	0.113	-0.141	2.704	0.124	-0.081	3.006
Median	0.185	0.382	4.360	0.200	0.485	4.290	0.201	0.415	4.223	0.215	0.440	3.904	0.228	0.526	4.197
3rd Quartile	0.353	1.005	5.749	0.448	1.093	6.089	0.409	1.159	5.938	0.401	1.135	6.347	0.370	1.087	5.645
Mean	0.294	0.424	4.772	0.428	0.460	4.842	0.301	0.511	4.767	0.318	0.547	4.638	0.289	0.638	4.478
Variance (n-1)	0.336	2.584	9.242	0.737	2.585	9.773	0.533	2.196	8.891	0.221	1.658	7.005	0.487	1.222	5.794
Standard deviation (n-1)	0.580	1.607	3.040	0.859	1.608	3.126	0.730	1.482	2.982	0.470	1.288	2.647	0.698	1.105	2.407

(b) (continued) (Year 5 to Year 1)

	J5			J4			J3			J2			J1		
Statistic	ROE	K	Z	ROE	K	Z	ROE	K	Z	ROE	K	Z	ROE	K	Z
Nbr. of observations	85	85	85	85	85	85	85	85	85	85	85	85	85	85	85
Minimum	-7.695	-3.686	0.193	-3.072	-4.187	0.259	-2.786	-2.165	0.256	-1.020	-2.289	-0.397	-2.410	-2.946	-1.147
Maximum	5.742	9.936	11.670	3.128	3.989	13.556	4.664	5.238	16.497	5.637	6.494	70.138	2.085	7.381	66.787
1st Quartile	0.106	0.029	2.875	0.153	0.150	2.986	0.175	0.211	3.582	0.176	0.096	3.272	0.134	0.246	3.563
Median	0.233	0.522	4.100	0.297	0.778	3.894	0.330	0.969	4.965	0.300	0.858	4.718	0.246	0.878	4.623
3rd Quartile	0.432	1.330	5.601	0.464	1.558	5.957	0.553	1.730	6.312	0.509	1.740	6.495	0.442	1.653	6.402
Mean	0.279	0.670	4.403	0.358	0.874	4.618	0.417	1.143	5.338	0.445	1.045	6.201	0.336	1.074	6.086
Variance (n-1)	1.357	2.719	4.964	0.449	1.580	6.576	0.484	1.950	8.289	0.633	2.265	73.659	0.268	2.659	73.334
Standard deviation (n-1)	1.165	1.649	2.228	0.670	1.257	2.564	0.696	1.396	2.879	0.795	1.505	8.582	0.517	1.631	8.564

(c) Descriptive statistics (Year 10 to Year 6) – Distressed category

DISTRESSED	J10			J9			J8			J7			J6		
Statistic	ROE	K	Z	ROE	K	Z	ROE	K	Z	ROE	K	Z	ROE	K	Z
Nbr. of observations	22	22	8	21	23	7	22	19	5	22	18	6	21	15	6
Minimum	-1.202	-4.993	-1.361	-0.611	-5.452	-1.929	-2.556	-5.450	-1.747	-0.331	-4.414	0.193	-4.466	-2.188	0.052
Maximum	0.060	-0.246	1.785	0.098	-0.195	1.526	0.087	-0.270	1.535	0.113	-0.213	1.699	0.120	-0.207	1.742
1st Quartile	-0.236	-1.693	-0.028	-0.004	-1.443	0.395	-0.151	-1.688	0.368	0.013	-1.297	0.853	-0.007	-1.034	0.637
Median	-0.003	-0.759	0.397	0.021	-0.564	0.658	0.018	-0.809	0.608	0.058	-0.663	1.503	0.053	-0.589	1.108
3rd Quartile	0.042	-0.456	1.717	0.060	-0.366	1.232	0.068	-0.415	1.206	0.095	-0.429	1.610	0.074	-0.338	1.624
Mean	-0.188	-1.334	0.521	-0.012	-1.256	0.501	-0.195	-1.295	0.394	0.004	-1.019	1.197	-0.208	-0.737	1.048
Variance (n-1)	0.139	1.819	1.380	0.023	2.170	1.386	0.358	1.672	1.648	0.019	1.011	0.392	0.972	0.299	0.476
Standard deviation (n-1)	0.373	1.349	1.175	0.153	1.473	1.177	0.598	1.293	1.284	0.137	1.005	0.626	0.986	0.547	0.690

(d) (continued) (Year 5 to Year 1)

	J5			J4			J3			J2			J1		
Statistic	ROE	K	Z	ROE	K	Z	ROE	K	Z	ROE	K	Z	ROE	K	Z
Nbr. of observations	21	15	9	21	10	7	21	13	4	22	14	6	26	11	6
Minimum	-7.695	-3.686	0.193	-3.072	-4.187	0.259	-2.786	-2.165	0.256	-1.020	-2.289	-0.397	-2.410	-2.946	-1.147
Maximum	0.098	-0.206	1.714	0.151	-0.260	1.791	0.174	-0.194	1.606	0.180	-0.208	1.203	0.169	-0.298	1.075
1st Quartile	-0.149	-1.853	1.106	-0.015	-1.273	0.550	-0.003	-1.044	0.397	-0.102	-1.468	0.405	0.027	-1.510	-0.150
Median	-0.008	-1.013	1.271	0.085	-1.055	1.012	0.087	-0.360	0.906	0.081	-0.687	0.479	0.057	-0.832	0.266
3rd Quartile	0.077	-0.481	1.609	0.106	-0.454	1.485	0.145	-0.256	1.427	0.125	-0.420	0.749	0.114	-0.413	0.715
Mean	-0.489	-1.310	1.195	-0.195	-1.189	1.019	-0.077	-0.635	0.918	-0.021	-0.930	0.497	-0.051	-1.184	0.173
Variance (n-1)	2.911	1.217	0.263	0.548	1.297	0.353	0.399	0.360	0.446	0.073	0.420	0.286	0.245	0.946	0.643
Standard deviation (n-1)	1.706	1.103	0.513	0.740	1.139	0.594	0.632	0.600	0.668	0.270	0.648	0.534	0.495	0.973	0.802

(e) Descriptive statistics (Year 10 to Year 6) – Healthy category

HEALTHY	J10			J9			J8			J7			J6		
Statistic	ROE	K	Z	ROE	K	Z	ROE	K	Z	ROE	K	Z	ROE	K	Z
Nbr. of observations	22	50	62	22	50	66	22	56	62	22	53	63	22	58	64
Minimum	0.353	0.216	3.055	0.448	0.204	2.931	0.409	0.082	3.055	0.401	0.215	2.996	0.364	0.199	3.046
Maximum	3.222	5.414	17.438	6.189	5.224	16.641	4.719	5.352	18.242	3.560	4.598	14.632	2.952	4.169	14.043
1st Quartile	0.451	0.575	4.027	0.480	0.598	3.789	0.490	0.411	4.064	0.465	0.533	3.792	0.486	0.509	3.775
Median	0.744	0.984	5.292	0.724	0.939	5.052	0.596	0.945	4.983	0.597	1.032	5.006	0.622	0.947	4.832
3rd Quartile	1.228	1.935	6.372	1.130	1.495	6.449	0.825	1.445	6.653	0.703	1.650	6.701	0.939	1.380	6.067
Mean	0.953	1.364	5.849	1.244	1.308	5.759	0.965	1.183	5.779	0.810	1.253	5.556	0.863	1.154	5.310
Variance (n-1)	0.479	1.282	7.829	1.897	1.214	8.493	1.023	1.169	7.981	0.478	0.881	6.051	0.393	0.853	4.756
Standard deviation (n-1)	0.692	1.132	2.798	1.377	1.102	2.914	1.011	1.081	2.825	0.692	0.939	2.460	0.627	0.924	2.181

(f) (continued) (Year 5 to Year 1)

	J5			J4			J3			J2			J1		
Statistic	ROE	K	Z	ROE	K	Z	ROE	K	Z	ROE	K	Z	ROE	K	Z
Nbr. of observations	22	55	63	22	62	64	22	66	70	22	60	69	22	65	73
Minimum	0.430	0.238	3.027	0.463	0.209	3.009	0.549	0.205	3.109	0.507	0.234	3.014	0.442	0.241	3.002
Maximum	5.742	9.936	11.670	3.128	3.989	13.556	4.664	5.238	16.497	5.637	6.494	70.138	2.085	7.381	66.787
1st Quartile	0.541	0.557	3.837	0.560	0.720	3.762	0.641	0.659	4.171	0.578	0.781	4.121	0.575	0.765	3.891
Median	0.658	0.971	4.875	0.672	1.241	4.959	0.717	1.296	5.293	0.728	1.489	5.292	0.685	1.215	4.818
3rd Quartile	1.067	1.778	6.263	1.100	1.965	6.391	0.934	2.183	6.679	1.005	2.188	7.120	1.009	1.818	6.516
Mean	1.075	1.410	5.278	0.984	1.396	5.498	1.055	1.605	6.046	1.160	1.703	7.225	0.898	1.610	6.871
Variance (n-1)	1.319	2.178	3.589	0.516	0.848	5.438	0.832	1.490	7.130	1.614	1.567	85.364	0.246	2.013	80.865
Standard deviation (n-1)	1.148	1.476	1.894	0.718	0.921	2.332	0.912	1.221	2.670	1.270	1.252	9.239	0.496	1.419	8.992

(g) Descriptive statistics (Year 10 to Year 6) – Neutral category

NEUTRAL	J10			J9			J8			J7			J6		
Statistic	ROE	K	Z	ROE	K	Z	ROE	K	Z	ROE	K	Z	ROE	K	Z
Nbr. of observations	41	13	15	41	11	12	41	10	18	41	14	16	41	12	15
Minimum	0.062	-0.185	2.027	0.103	-0.181	1.826	0.088	-0.149	2.080	0.117	-0.306	1.938	0.126	-0.340	1.829
Maximum	0.353	0.360	2.969	0.447	1.920	2.963	0.386	1.355	2.988	0.394	0.156	2.704	0.362	0.168	2.889
1st Quartile	0.149	-0.113	2.375	0.158	-0.123	2.101	0.153	0.006	2.294	0.158	-0.141	2.135	0.181	-0.113	2.077
Median	0.185	-0.045	2.624	0.194	0.001	2.450	0.201	0.023	2.392	0.215	-0.100	2.334	0.225	-0.081	2.410
3rd Quartile	0.250	0.091	2.800	0.246	0.163	2.619	0.263	0.147	2.655	0.262	0.029	2.458	0.290	0.054	2.606
Mean	0.199	0.003	2.588	0.215	0.171	2.405	0.211	0.179	2.493	0.222	-0.066	2.314	0.236	-0.048	2.354
Variance (n-1)	0.006	0.024	0.076	0.007	0.354	0.134	0.007	0.192	0.073	0.005	0.015	0.049	0.005	0.019	0.123
Standard deviation (n-1)	0.080	0.156	0.275	0.084	0.595	0.366	0.083	0.438	0.270	0.073	0.124	0.222	0.069	0.139	0.351

(h) (continued) (Year 5 to Year 1)

	J5			J4			J3			J2			J1		
Statistic	ROE	K	Z	ROE	K	Z	ROE	K	Z	ROE	K	Z	ROE	K	Z
Nbr. of observations	41	16	13	41	13	14	41	6	11	41	10	10	37	9	6
Minimum	0.109	-1.251	1.822	0.154	-0.181	1.871	0.175	-0.063	1.925	0.182	1.998	1.998	0.169	-0.141	1.870
Maximum	0.429	0.191	2.892	0.452	0.195	2.915	0.506	0.179	2.902	0.507	2.977	2.977	0.440	0.164	2.756
1st Quartile	0.163	-0.088	2.101	0.220	-0.012	2.162	0.247	0.029	2.402	0.223	2.568	2.568	0.201	-0.029	2.363
Median	0.233	0.043	2.570	0.296	0.085	2.486	0.324	0.080	2.492	0.303	2.700	2.700	0.268	0.039	2.516
3rd Quartile	0.296	0.115	2.705	0.399	0.124	2.767	0.387	0.132	2.687	0.376	2.822	2.822	0.331	0.101	2.677
Mean	0.244	-0.051	2.453	0.306	0.051	2.457	0.328	0.073	2.504	0.313	2.656	2.656	0.273	0.027	2.450
Variance (n-1)	0.009	0.114	0.121	0.009	0.012	0.127	0.009	0.008	0.070	0.009	0.075	0.075	0.006	0.011	0.107
Standard deviation (n-1)	0.094	0.337	0.348	0.092	0.112	0.356	0.093	0.089	0.264	0.096	0.274	0.274	0.079	0.106	0.327

APPENDIX 8.2

MANN-WHITNEY TEST RESULTS

Same year (Y)	J10		J9		J8		J7		J6		J5		J4		J3		J2		J1	
	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z
Observations	82	85	84	84	85	85	84	85	83	84	83	84	83	84	83	84	84	84	84	85
Obs. with missing data	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Obs. without missing data	82	85	84	84	85	85	84	85	83	84	83	84	83	84	83	84	84	84	84	85
Minimum	-4.993	-1.361	-5.452	-1.929	-5.450	-1.747	-4.414	0.193	-2.188	0.052	-3.686	0.193	-4.187	0.259	-2.165	0.256	-2.289	-0.397	-2.946	-1.147
Maximum	5.414	17.438	5.224	16.641	5.352	18.242	4.598	14.632	4.169	14.043	9.936	11.670	3.989	13.556	5.238	16.497	6.494	70.138	7.381	66.787
Mean	0.424	4.772	0.460	4.842	0.511	4.767	0.547	4.638	0.638	4.478	0.670	4.403	0.874	4.618	1.143	5.338	1.045	6.201	1.074	6.086
Std. deviation	1.607	3.040	1.608	3.126	1.482	2.982	1.288	2.647	1.105	2.407	1.649	2.228	1.257	2.564	1.396	2.879	1.505	8.582	1.631	8.564
U	3147	276	3132	266	2878	189	3065	90	2734	121	2753	150	2363	136	2248	121	2454	252	2143	265
Expected value	3485	3613	3528	3528	3613	3613	3570	3613	3486	3528	3486	3528	3486	3528	3486	3528	3528	3528	3570	3613
Variance (U)	97580	102956	99372	99372	102956	102956	101150	102956	97608	99372	97608	99372	97608	99372	97608	99372	99372	99372	101150	102956
p-value (Two-tailed)	0.280	< 0.0001	0.210	< 0.0001	0.022	< 0.0001	0.113	< 0.0001	0.016	< 0.0001	0.019	< 0.0001	0.000	< 0.0001	< 0.0001	< 0.0001	0.001	< 0.0001	< 0.0001	< 0.0001
H0: Reject/Accept	A		A		A		A		A		A		R		R		R		R	
Ha: Reject/Accept	R		R		R		R		R		R		A		A		A		A	

One-year lead (Y1)	J9/J10		J8/J9		J7/J8		J6/J7		J5/J6		J4/J5		J3/J4		J2/J3		J1/J2	
	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z
Observations	82	85	84	84	85	85	84	85	83	84	83	84	83	84	83	84	84	84
Obs. with missing data	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Obs. without missing data	82	85	84	84	85	85	84	85	83	84	83	84	83	84	83	84	84	84
Minimum	-4.993	-1.361	-5.452	-1.929	-5.450	-1.747	-4.414	0.193	-2.188	0.052	-3.686	0.193	-4.187	0.259	-2.165	0.256	-2.289	-0.397
Maximum	5.414	17.438	5.224	16.641	5.352	18.242	4.598	14.632	4.169	14.043	9.936	11.670	3.989	13.556	5.238	16.497	6.494	70.138
Mean	0.424	4.772	0.460	4.842	0.511	4.767	0.547	4.638	0.638	4.478	0.670	4.403	0.874	4.618	1.143	5.338	1.045	6.201
Std. deviation	1.607	3.040	1.608	3.126	1.482	2.982	1.288	2.647	1.105	2.407	1.649	2.228	1.257	2.564	1.396	2.879	1.505	8.582
U	3253	366	3015	211	2919	152	3060	88	2779	165	2883	131	2448	156	2179	162	2385	162
Expected value	3444	3570	3570	3570	3613	3613	3528	3570	3486	3528	3486	3528	3486	3528	3486	3528	3570	3570
Variance (U)	95858	101150	101150	101150	102956	102956	99372	101150	97608	99372	97608	99372	97608	99372	97608	99372	101150	101150
p-value (Two-tailed)	0.538	< 0.0001	0.081	< 0.0001	0.031	< 0.0001	0.138	< 0.0001	0.024	< 0.0001	0.054	< 0.0001	0.001	< 0.0001	< 0.0001	< 0.0001	0.000	< 0.0001
H0: Reject/Accept	A		A		A		A		A		A		R		R		R	
Ha: Reject/Accept	R		R		R		R		R		R		A		A		A	

Two-year lead (Y2)	J8/J10		J7/J9		J6/J8		J5/J7		J4/J6		J3/J5		J2/J4		J1/J3	
	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z
Observations	82	85	84	84	85	85	84	85	83	84	83	84	83	84	83	84
Obs. with missing data	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Obs. without missing data	82	85	84	84	85	85	84	85	83	84	83	84	83	84	83	84
Minimum	-4.993	-1.361	-5.452	-1.929	-5.450	-1.747	-4.414	0.193	-2.188	0.052	-3.686	0.193	-4.187	0.259	-2.165	0.256
Maximum	5.414	17.438	5.224	16.641	5.352	18.242	4.598	14.632	4.169	14.043	9.936	11.670	3.989	13.556	5.238	16.497
Mean	0.424	4.772	0.460	4.842	0.511	4.767	0.547	4.638	0.638	4.478	0.670	4.403	0.874	4.618	1.143	5.338
Std. deviation	1.607	3.040	1.608	3.126	1.482	2.982	1.288	2.647	1.105	2.407	1.649	2.228	1.257	2.564	1.396	2.879
U	3137	318	3059	181	2914	140	3084	135	2929	140	2976	155	2399	195	2103	69
Expected value	3485	3613	3570	3570	3570	3570	3528	3570	3486	3528	3486	3528	3486	3528	3528	3570
Variance (U)	97580	102956	101150	101150	101150	101150	99372	101150	97608	99372	97608	99372	97608	99372	99358	101150
p-value (Two-tailed)	0.266	< 0.0001	0.108	< 0.0001	0.039	< 0.0001	0.159	< 0.0001	0.075	< 0.0001	0.103	< 0.0001	0.001	< 0.0001	< 0.0001	< 0.0001
H0: Reject/Accept	A		A		A		A		A		A		R		R	
Ha: Reject/Accept	R		R		R		R		R		R		A		A	

Three-year lead (Y3)	J7/J10		J6/J9		J5/J8		J4/J7		J3/J6		J2/J5		J1/J4	
	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z
Observations	82	85	84	84	85	85	84	85	83	84	83	84	83	84
Obs. with missing data	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Obs. without missing data	82	85	84	84	85	85	84	85	83	84	83	84	83	84
Minimum	-4.993	-1.361	-5.452	-1.929	-5.450	-1.747	-4.414	0.193	-2.188	0.052	-3.686	0.193	-4.187	0.259
Maximum	5.414	17.438	5.224	16.641	5.352	18.242	4.598	14.632	4.169	14.043	9.936	11.670	3.989	13.556
Mean	0.424	4.772	0.460	4.842	0.511	4.767	0.547	4.638	0.638	4.478	0.670	4.403	0.874	4.618
Std. deviation	1.607	3.040	1.608	3.126	1.482	2.982	1.288	2.647	1.105	2.407	1.649	2.228	1.257	2.564
U	3179	291	3051	180	2966	193	3207	122	3015	166	2923	183	2306	97
Expected value	3485	3613	3528	3528	3570	3570	3528	3570	3486	3528	3486	3528	3528	3570
Variance (U)	97580	102956	99372	99372	101150	101150	99372	101150	97608	99372	97608	99372	99358	101150
p-value (Two-tailed)	0.327	< 0.0001	0.131	< 0.0001	0.058	< 0.0001	0.309	< 0.0001	0.132	< 0.0001	0.072	< 0.0001	0.000	< 0.0001
H0: Reject/Accept	A		A		A		A		A		A		R	
Ha: Reject/Accept	R		R		R		R		R		R		A	

Five-year lead (Y5)	J5/J10		J4/J9		J3/J8		J2/J7		J1/J6	
	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z
Observations	82	85	84	84	85	85	84	85	83	84
Obs. with missing data	0	0	0	0	0	0	0	0	0	0
Obs. without missing data	82	85	84	84	85	85	84	85	83	84
Minimum	-4.993	-1.361	-5.452	-1.929	-5.450	-1.747	-4.414	0.193	-2.188	0.052
Maximum	5.414	17.438	5.224	16.641	5.352	18.242	4.598	14.632	4.169	14.043
Mean	0.424	4.772	0.460	4.842	0.511	4.767	0.547	4.638	0.638	4.478
Std. deviation	1.607	3.040	1.608	3.126	1.482	2.982	1.288	2.647	1.105	2.407
U	3187	326	3180	209	3233	189	3237	176	2869	113
Expected value	3444	3570	3528	3528	3570	3570	3528	3570	3528	3570
Variance (U)	95858	101150	99372	99372	101150	101150	99372	101150	99358	101150
p-value (Two-tailed)	0.406	< 0.0001	0.270	< 0.0001	0.290	< 0.0001	0.357	< 0.0001	0.037	< 0.0001
H0: Reject/Accept	A		A		A		A		A	
Ha: Reject/Accept	R		R		R		R		R	

APPENDIX 8.3

SPEARMAN RHO TEST RESULTS

Same year (Y)	J10		J9		J8		J7		J6		J5		J4		J3		J2		J1	
	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z
Correlation matrix (Spearman)	0.521	0.560	0.472	0.520	0.575	0.554	0.513	0.435	0.423	0.411	0.504	0.394	0.521	0.434	0.370	0.312	0.505	0.336	0.386	0.329
p-values	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.000	<0.0001	0.000	<0.0001	<0.0001	0.001	0.004	<0.0001	0.002	0.000	0.002
Coefficients of determination (Spearman)	0.271	0.313	0.223	0.271	0.330	0.306	0.263	0.189	0.179	0.169	0.254	0.155	0.272	0.188	0.137	0.098	0.255	0.113	0.149	0.108

One-year lead (Y1)	J9/J10		J8/J9		J7/J8		J6/J7		J5/J6		J4/J5		J3/J4		J2/J3		J1/J2	
	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z
Correlation matrix (Spearman)	0.500	0.471	0.449	0.539	0.460	0.557	0.217	0.387	0.396	0.407	0.440	0.420	0.409	0.364	0.337	0.326	0.345	0.248
p-values	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	0.048	0.000	0.000	0.000	<0.0001	<0.0001	0.000	0.001	0.002	0.003	0.001	0.023
Coefficients of determination (Spearman)	0.250	0.222	0.202	0.290	0.212	0.310	0.047	0.150	0.157	0.166	0.194	0.176	0.167	0.133	0.113	0.106	0.119	0.062

Two-year lead (Y2)	J8/J10		J7/J9		J6/J8		J5/J7		J4/J6		J3/J5		J2/J4		J1/J3	
	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z
Correlation matrix (Spearman)	0.448	0.47	0.317	0.473	0.182	0.48	0.354	0.375	0.335	0.376	0.237	0.287	0.287	0.292	0.114	0.213
p-values	<0.0001	<0.0001	0.003	<0.0001	0.098	<0.0001	0.001	0	0.002	0.001	0.032	0.009	0.009	0.008	0.306	0.052
Coefficients of determination (Spearman)	0.201	0.221	0.101	0.223	0.033	0.23	0.125	0.141	0.112	0.141	0.056	0.082	0.082	0.085	0.013	0.045

Three-year lead (Y3)	J7/J10		J6/J9		J5/J8		J4/J7		J3/J6		J2/J5		J1/J4	
	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z
Correlation matrix (Spearman)	0.323	0.414	0.243	0.449	0.256	0.409	0.221	0.361	0.234	0.267	0.252	0.274	0.109	0.179
p-values	0.003	<0.0001	0.027	<0.0001	0.019	0	0.045	0.001	0.034	0.015	0.023	0.012	0.326	0.104
Coefficients of determination (Spearman)	0.104	0.172	0.059	0.021	0.065	0.167	0.049	0.131	0.055	0.071	0.064	**0.075**	0.012	**0.032**

Five-year lead (Y5)	J5/J10		J4/J9		J3/J8		J2/J7		J1/J6	
	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z	ROE-K	ROE-Z
Correlation matrix (Spearman)	0.318	0.385	0.218	0.365	0.107	0.321	0.304	0.274	0.076	0.156
p-values	0.004	0	0.048	0.001	0.334	0.003	0.005	0.012	0.492	0.155
Coefficients of determination (Spearman)	0.101	0.148	0.047	0.133	0.011	0.103	0.092	0.075	0.006	0.024

APPENDIX 8.4

WEIGHTED EFFICIENCY TEST RESULTS

K-score

Financial distress	J10	J9	J8	J7	J6	J5	J4	J3	J2	J1	Average
Same year (Y)	6.914	5.415	6.184	3.006	1.958	5.454	2.937	2.260	2.818	5.540	**4.249**
One-year lead (Y1)		4.561	4.065	0.965	1.080	0.826	2.098	0.717	1.445	2.073	**1.981**
Two-year lead (Y2)			3.354	0.807	0.189	0.680	0.826	0.262	0.685	0.552	**0.919**
Three-year lead (Y3)				0.544	0.532	0.080	0.394	0.478	0.250	0.367	**0.378**
Five-year lead (Y5)						0.905	0.532	0.080	0.649	0.245	**0.482**
Financial health											
Same year (Y)	3.236	2.970	6.550	5.113	7.824	5.077	7.302	6.846	4.721	4.154	**5.379**
One-year lead (Y1)		3.236	4.433	5.569	4.262	6.708	7.087	6.260	6.264	5.349	**5.463**
Two-year lead (Y2)			3.236	3.653	3.911	4.262	6.708	3.487	4.484	4.990	**4.342**
Three-year lead (Y3)				3.236	3.653	3.911	4.262	4.006	3.487	5.323	**3.983**
Five-year lead (Y5)						3.236	4.433	3.223	3.679	3.301	**3.574**
Neutral											
Same year (Y)	3.045	0.905	1.469	0.259	0.157	1.837	0.488	0.039	0.322	0.000	**0.852**
One-year lead (Y1)		0.785	0.905	0.984	0.711	0.529	1.339	1.648	0.005	0.557	**0.829**
Two-year lead (Y2)			0.494	0.570	0.359	0.056	0.306	1.339	1.689	0.038	**0.606**
Three-year lead (Y3)				0.494	0.330	0.984	0.259	0.306	0.646	3.813	**0.976**
Five-year lead (Y5)						0.286	0.169	0.620	0.259	0.153	**0.297**
Combined											
Same year (Y)	9.959	6.320	7.653	3.266	2.115	7.291	3.426	2.299	3.140	5.540	**5.101**
One-year lead (Y1)		8.582	9.404	7.519	6.054	8.063	10.524	8.626	7.714	7.979	**8.274**
Two-year lead (Y2)			7.084	4.954	4.670	5.301	7.590	4.056	6.508	5.580	**5.718**
Three-year lead (Y3)				4.275	4.515	4.975	4.915	4.790	4.383	9.503	**5.337**
Five-year lead (Y5)						4.428	5.134	3.922	4.588	3.698	**4.354**

Z-score

Financial distress	J10	J9	J8	J7	J6	J5	J4	J3	J2	J1	Average
Same year (Y)	2.293	1.012	0.289	0.241	0.076	0.787	1.749	0.383	1.127	0.483	**0.844**
One-year lead (Y1)		0.875	1.670	0.289	0.075	0.076	1.361	0.518	0.365	0.605	**0.648**
Two-year lead (Y2)			1.444	0.062	0.090	0.075	0.255	0.403	0.495	0.113	**0.367**
Three-year lead (Y3)				0.836	0.065	0.090	0.252	0.076	0.385	0.065	**0.252**
Five-year lead (Y5)						0.189	0.008	0.090	0.071	0.076	**0.087**
Financial health											
Same year (Y)	6.900	5.710	6.900	4.950	7.955	5.986	6.871	5.379	4.096	4.272	**5.902**
One-year lead (Y1)		5.916	7.710	6.900	5.822	6.871	6.982	5.009	6.330	4.641	**6.242**
Two-year lead (Y2)			7.988	6.660	6.900	4.950	7.955	5.090	5.891	4.574	**6.251**
Three-year lead (Y3)				6.900	6.660	5.916	4.950	5.009	5.090	5.891	**5.774**
Five-year lead (Y5)						5.916	5.710	6.900	5.186	4.220	**5.586**
Neutral											
Same year (Y)	0.656	0.523	0.199	0.615	0.418	0.482	0.448	0.021	2.834	0.005	**0.620**
One-year lead (Y1)		0.656	0.830	1.162	1.307	0.242	1.144	0.448	0.021	0.077	**0.654**
Two-year lead (Y2)			0.656	0.523	1.162	0.387	0.991	1.144	0.459	0.165	**0.686**
Three-year lead (Y3)				2.546	0.830	0.816	0.918	0.664	0.286	1.037	**1.014**
Five-year lead (Y5)						0.979	0.155	2.122	0.615	0.648	**0.904**
Combined											
Same year (Y)	2.949	1.535	0.488	0.856	0.494	1.270	2.197	0.404	3.962	0.488	**1.464**
One-year lead (Y1)		7.447	10.209	8.351	7.204	7.189	9.486	5.976	6.716	5.322	**7.545**
Two-year lead (Y2)			10.088	7.378	7.512	6.187	8.597	6.484	7.530	4.852	**7.329**
Three-year lead (Y3)				10.282	7.555	6.822	6.121	5.749	5.761	6.993	**7.040**
Five-year lead (Y5)						7.085	5.873	9.112	5.873	4.944	**6.577**

APPENDIX 10.1

SVM-K-SCORE SAMPLE COMPANIES

Identification of distressed and non-distressed sample companies

Company name	Short name	Dist/non-dist	Code	2014	2013	2012	2011	2010	2009	2008	2007	2006	2005
Adapt IT Holdings Ltd	ADAPTIT	0	ADI	2.094	1.556	1.062	0.435	0.168	4.097	3.529	2.896	3.218	4.049
Adcorp Holdings Ltd	ADCORP	1	ADR	-0.468	-0.125	0.240	1.006	0.934	1.406			1.411	1.165
Aveng Ltd	AVENG	1	AEG	-0.654	-0.624	-0.665	0.064	0.360	0.324	0.471	5.238	-0.165	-0.832
Allied Electronics Corporation Ltd	ALTRON	1	AEL	-0.246	-0.995	0.312	0.671	0.392	0.757	1.240	1.350	0.903	0.934
AECI Ltd	AECI	0	AFE	0.338	0.204	0.010	0.215	0.223	-0.075	-0.345	0.216	1.575	0.405
African Oxygen Ltd	AFROX	0	AFX	-0.097	0.353	0.363	0.156	-0.104	0.125	0.576	0.678	2.117	3.514
AH-Vest Ltd	AH-VEST	1	AHL	-3.808	-0.366	0.455	-1.306	-0.421	-2.362	-4.187	-1.044	-1.091	0.626
African Media Entertainment Ltd	AME	0	AME	1.863	1.920	2.430	2.303	2.424			3.781	5.421	7.381
Astrapak Ltd	ASTRAPAK	1	APK	-0.922	0.598	-0.344	0.265	0.376	0.172	0.124	0.627	0.951	0.870
Astral Foods Ltd	ASTRAL	0	ARL	0.403	0.191	0.599	1.061	0.766	0.896	0.786	2.302	2.749	1.972
Business Connextion Group Ltd	BCX	0	BCX	0.884	0.872	0.399	-0.077	0.526	-0.161	0.034	0.444	0.160	0.988
Beige Holdings Ltd	BEIGE	1	BEG	-2.914	-3.432	-0.809	-0.143	-0.279	0.050	0.939	0.679	-1.545	-1.552
Bell Equipment Ltd	BELL	0	BEL	0.284	0.692	1.113	1.120	0.273	-1.327	1.242	1.573	1.489	-0.366
Basil Read Holdings Ltd	BASREAD	0	BSR		-0.232	-1.915	-0.516	-0.068	0.017	0.213	0.318	0.158	-0.141
The Bidvest Group Ltd	BIDVEST	0	BVT	0.083	0.315	0.415	0.350	0.363	0.191	0.195	0.319	0.394	0.854
Caxton and CTP Publishers and Printers Ltd	CAXTON	0	CAT	0.728	0.888	0.809	0.963	0.657	2.021	1.397	1.449	1.485	1.640
Crookes Brothers Ltd	CROOKES	0	CKS	2.567	1.495	1.358	1.825	-0.094	1.694	1.146	0.912	0.930	-0.029
City Lodge Hotels Ltd	CITYLDG	0	CLH	1.802	1.639	1.355	0.926	1.687	1.778	3.173	3.028	2.640	2.342
Clicks Group Ltd	CLICKS	0	CLS	1.006	0.979	1.242	1.120	1.064	0.709	0.731	0.618	0.638	0.493
Cargo Carriers Ltd	CARGO	0	CRG	-0.057	-0.181	-0.149	-0.124	-0.106	-0.341	-0.260	0.036	0.418	0.401
Cashbuild Ltd	CASHBIL	0	CSB	0.970	1.377	1.421	0.362	0.658	0.901	0.887	1.148	0.791	0.835
Cullinan Holdings Ltd	CULLINAN	0	CUL	0.435	0.170	0.024	-0.173	0.006	-0.416	-0.961	-0.408	-0.351	0.039
Datacentrix Holdings Ltd	DCENTRIX	0	DCT	0.639	0.493	1.026	1.229	1.266	1.873	1.581	1.442	0.699	1.300
Digicore Holdings Ltd	DIGICORE	0	DGC	0.655	0.647	0.175	1.359	1.110	2.175	3.989	3.503	3.639	1.929
Distell Group Ltd	DISTELL	0	DST	1.152	0.928	1.159	1.543	1.389	1.721	1.965	1.789	1.127	0.938
Delta EMD Ltd	DELTA	1	DTA	-3.820	-1.422	0.430	0.728	2.545	2.362	1.440	-2.165	-0.778	6.557
Datatec Ltd	DATATEC	1	DTC	-0.700	-0.536	-0.270	-0.516	-0.590	-0.257	-0.293	-0.063	-0.053	0.278
EOH Holdings Ltd	EOH	0	EOH	0.834	0.939	0.882	1.020	0.374	0.131	0.936	0.904	1.093	0.748
Famous Brands Ltd	FAMBRANDS	0	FBR	5.414	5.224	5.352	4.598	3.319	2.157	2.440	1.817	2.259	2.482
Gijima Group Ltd	GIJIMA	1	GIJ	-4.993	-5.452	-2.372	-4.414	0.849	0.422	0.693	-0.360	-0.264	-0.684
Grindrod Ltd	GRINDROD	1	GND	-0.452	-0.533	-0.484	-0.431	-0.147	-0.011	1.535	0.637	0.911	1.531
Hudaco Industries Ltd	HUDACO	0	HDC	-0.044	1.084	0.082	0.024	-0.207	-0.128	0.065	-0.203	1.602	1.701
Howden Africa Holdings Ltd	HOWDEN	0	HWN	2.007	2.653	0.969	0.797	0.330	0.971	0.497	1.101	-0.544	0.267
Iliad Africa Ltd	ILIAD	0	ILA	0.309	-0.464	-0.097	-0.702	-0.340	0.045	1.306	1.644	1.401	1.458
Illovo Sugar Ltd	ILLOVO	0	ILV	0.687	0.946	0.387	0.481	0.894	0.546	0.604	0.987	0.637	-0.460
ISA Holdings Ltd	ISA	0	ISA	2.708	2.053	2.727	2.708	2.576	2.044	1.661	2.703	3.488	4.978
Italtile Ltd	ITLTILE	0	ITE	2.119	2.044	1.788	1.635	1.439	1.481	2.071	2.373	2.365	2.267
JD Group Ltd	JDGROUP	1	JDG	-1.834	-0.564	0.241	-0.141	0.783	0.411	0.091	1.240	1.546	1.558

(Appendix 10.1 continued)

Company name	Short name	Dist/non-dist	Code	2014	2013	2012	2011	2010	2009	2008	2007	2006	2005
Jasco Electronics Holdings Ltd	JASCO	1	JSC	-0.524	-2.059	-0.325	-0.213	-0.375	0.048	1.274	1.784	1.695	1.781
KAP Industrial Holdings Ltd	KAP	0	KAP	0.091	0.001	0.063	0.550	0.199	-0.705	-0.012			1.404
Lewis Group Ltd	LEWIS	0	LEW	1.108	1.296	1.517	1.180	1.377	1.582	2.113	2.055	1.671	
Masonite (Africa) Ltd	MASONITE	0	MAS	0.544	0.919	0.713	0.030	0.052	0.833	2.190	1.122	0.749	0.425
Mediclinic International Ltd	MEDCLIN	0	MDC	-0.135	-1.145	-0.752	-0.683	-0.838	-1.007	-0.778	1.042	0.687	1.757
Metrofile Holdings Ltd	METROFILE	0	MFL	2.651	1.853	1.643	1.219	0.768	0.111	0.842	-0.256	-1.676	-2.946
MICROmega Holdings Ltd	MICROMEGA	0	MMG			0.416	-0.235	-0.252	0.457	1.952	2.430	2.331	1.363
MoneyWeb Holdings Ltd	MONEYWB	1	MNY	-0.817	0.401	-1.193	-1.609	0.837	0.040	1.339	1.241	0.521	-2.924
Mr Price Group Ltd	MRPRICE	0	MPC	3.321	3.584	3.169	2.928	1.332	1.893	1.966	1.638	1.620	1.185
Massmart Holdings Ltd	MASSMART	0	MSM		-0.315	-0.420	-0.126	0.459	0.698	0.924	0.519	0.234	-0.117
Mustek Ltd	MUSTEK	0	MST	-0.127	-0.122	0.005	0.271	-0.133	-0.206	-0.181	-0.300	-0.208	0.247
Metair Investments Ltd	METAIR	0	MTA	0.999	0.243	1.730	2.278	1.631	0.510	-0.017	1.376	1.821	1.569
MTN Group Ltd	MTN GROUP	0	MTN	1.442	1.589	1.393	1.580	1.224	1.076	1.679	1.237	1.785	0.885
Murray & Roberts Holdings Ltd	M&R-HLD	0	MUR	-0.113	-0.195	-1.461	-1.913	0.128	0.660	0.778	0.027	0.073	0.401
Nictus Ltd	NICTUS	1	NCS	-1.269	-1.575	-1.113	-1.203	-1.301	-1.251	-1.148	-1.063	-1.238	-1.184
Nampak Ltd	NAMPAK	1	NPK	-0.266	0.054	0.399	0.345	0.352	-0.675	0.113	0.511	0.684	0.616
Naspers Ltd	NASPERS-N	1	NPN	-1.007	-0.026	0.021	1.107	0.651	1.254	0.271	0.666	1.854	1.693
Netcare Ltd	NETCARE	0	NTC	0.261	1.285	-2.033	-1.271	-1.230	-1.121	-1.449	-1.351	-1.606	1.108
Nu-World Holdings Ltd	NUWORLD	0	NWL	1.148	1.193	0.783	0.612	1.223	0.522	0.735	1.485	1.410	1.242
OneLogix Group Ltd	ONELOGIX	0	OLG	0.573	0.477	0.761	0.783	0.961	0.471	1.204	0.893	1.725	3.581
Omnia Holdings Ltd	OMNIA	0	OMN	1.000	1.120	1.016	0.697	-0.487	0.845	0.430	0.585	0.239	0.770
Phumelela Gamings and Leisure Ltd	PHUMELELA	0	PHM	0.984	0.892	0.764	0.942	0.995	1.197	2.029	3.257	2.656	1.722
Pick n Pay Stores Ltd	PICKNPAY	1	PIK	-0.317	-0.367	0.220	-0.141	0.330	0.299	0.341	0.134	-0.024	0.029
Primeserv Group Ltd	PRIMESERV	1	PMV	-0.624	-1.463	-0.410			0.498	0.881	0.969	-0.420	0.164
Pinnacle Holdings Ltd	PINNACLE	0	PNC	0.299	0.796	1.068	1.043	1.018	0.801	0.729	0.536	0.142	0.834
PPC Ltd	PPC	0	PPC	0.134	0.868	0.921	1.021	1.524	1.624	2.980	2.551	2.729	2.538
RCL Foods Ltd	RCL	1	RCL	-1.067	-0.995	0.422	0.334	0.964	0.578	1.811	1.676	1.597	1.138
Remgro Ltd	REMGRO	0	REM	0.737	0.948	2.117	2.015	1.053	9.936	1.484	1.352	2.012	1.948
Reunert Ltd	REUNERT	0	RLO	3.009	1.418	1.958	2.043	0.864	1.250	1.460	0.416	1.142	1.376
Rex Trueform Clothing Company Ltd	REX TRUE	1	RTO	-0.324	-0.350	1.127	1.786	1.589	1.819	2.174	1.240	0.814	-1.468
Seardel Investement Corporation Ltd	SEARDEL	1	SER	-1.923	-0.216	0.240	-0.306	-1.240	-2.029	-1.311	-0.194	0.072	1.187
Steinhoff International Holdings Ltd	STEINHOFF	0	SHF	-0.045	-0.152	-0.320	-0.041	0.226	0.238	0.085	0.206	0.009	0.412
Sovereign Food Investments Ltd	SOVFOOD	0	SOV	0.360	0.443	0.232	-0.326	-0.589	-1.019	0.490	1.636	1.898	1.635
Spanjaard Ltd	SPANJAARD	0	SPA	0.216	0.541	0.345	0.333	0.062	0.501	0.628	0.124	-0.420	-0.311
Super Group Ltd	SUPRGRP	0	SPG	0.511	0.397	0.297	-0.248	-0.750	-3.182	-1.157	-0.416	-0.595	0.101
The Spar Group Ltd	SPAR	0	SPP	-0.185	0.402	0.293	0.348	0.451	0.443	0.310	0.246	0.382	0.714
StratCorp Ltd	STRATCORP	0	STA	0.553	-5.095	-5.450	-0.643	-2.188	-3.686	1.117	5.131	6.494	-0.298
Spur Corporation Ltd	SPURCORP	0	SUR	3.839	4.371	4.480	3.220	3.782	3.131	2.947	4.000	4.845	4.902
SilverBridge Holdings Ltd	SILVERB	0	SVB	2.768	2.254	-3.202	-1.685	4.169	1.429	2.594	2.190	-2.289	0.053
Tiger Brands Ltd	TIGBRANDS	0	TBS	0.729	0.879	1.701	1.933	1.911	2.315	2.122	2.180	2.329	1.055
Tongaat Hulett Ltd	TONGAAT	0	TON	0.176	0.155	0.291	0.079			0.251	4.485	0.728	0.483
Transpaco Ltd	TRNPACO	0	TPC	1.132	1.033	1.256	1.441	1.232	0.971	0.209	0.179	0.293	0.777
Truworths International Ltd	TRUWTHS	0	TRU	3.528	4.075	3.974	3.821	3.758	4.034	3.815	3.847	3.370	2.652
Tsogo Sun Holdings Ltd	TSOGO SUN	0	TSH	0.947	0.905	1.235	0.481	0.585	0.714	0.637	0.281	1.487	2.307
Value Group Ltd	VALUE	0	VLE	0.577	0.542	0.564	0.445	0.583	0.557	0.133	-0.228	0.772	0.615
Woolworths Holdings Ltd	WOOLIES	0	WHL	1.334	2.957	2.316	1.696	0.985	0.861	0.166	0.205	0.103	0.241
Winhold Ltd	WINHOLD	1	WNH	-0.397	-0.483	-1.068	-0.428	-0.302	-0.126	-0.133	-0.267	-0.140	0.141

APPENDIX 10.2

SVM-Z-SCORE SAMPLE COMPANIES

Identification of distressed and non-distressed sample companies

Company name	Short name	Distr /Non-distr	Code	2014	2013	2012	2011	2010	2009	2008	2007	2006	2005
Adapt IT Holdings Ltd	ADAPTIT	0	ADI	9.218	5.918	4.323	4.034	3.221	7.693	7.926	10.738	8.814	8.205
Adcorp Holdings Ltd	ADCORP	0	ADR	6.226	5.233	6.301	7.363	8.511	7.745	8.960		8.421	8.168
Aveng Ltd	AVENG	1	AEG	2.857	3.128	3.055	3.387	3.408	3.379	3.755	5.174	3.287	2.729
Allied Electronics Corporation Ltd	ALTRON	1	AEL	2.743	2.963	3.623	3.875	4.357	4.042	4.966	4.482	4.608	4.343
AECI Ltd	AECI	0	AFE	3.430	3.013	2.988	3.184	2.889	2.788	3.847	4.057	3.638	3.563
African Oxygen Ltd	AFROX	0	AFX	3.285	3.853	4.003	3.518	3.725	3.634	3.642	5.363	5.683	7.319
AH-Vest Ltd	AH-VEST	1	AHL	1.738	2.576	3.227	2.233	1.829	1.697	0.729	2.435	2.644	3.960
African Media Entertainment Ltd	AME	0	AME	6.640	6.073	6.722	6.674	6.266	6.316		5.920	8.174	9.298
Astrapak Ltd	ASTRAPAK	1	APK	2.475	2.931	2.659	3.302	3.653	3.441	3.117	3.862	4.099	4.095
Astral Foods Ltd	ASTRAL	0	ARL	5.236	4.594	5.211	5.906	5.945	5.831	5.391	6.077	6.475	6.527
Business Connextion Group Ltd	BCX	0	BCX	4.453	4.258	4.223	3.203	4.150	4.526	3.546	4.465	3.877	3.306
Beige Holdings Ltd	BEIGE	1	BEG	0.323	0.950	2.288	2.136	1.658	2.089	2.520	2.684	0.838	-0.261
Bell Equipment Ltd	BELL	0	BEL	3.084	3.183	4.143	3.159	3.062	1.958	3.238	4.969	4.486	3.347
Basil Read Holdings Ltd	BASREAD	1	BSR	2.300	1.526	2.406	2.549	2.410	2.570	3.686	3.429	2.846	1.075
The Bidvest Group Ltd	BIDVEST	0	BVT	4.837	4.982	4.986	5.006	5.148	5.205	4.644	5.496	5.073	5.706
Caxton and CTP Publishers and Printers Ltd	CAXTON	0	CAT	5.276	6.521	6.426	7.081	5.608	6.536	5.243	7.319	5.926	6.035
Crookes Brothers Ltd	CROOKES	0	CKS	5.749	5.099	3.489	4.697	4.391	4.833	6.901	7.316	5.588	3.529
City Lodge Hotels Ltd	CITYLDG	0	CLH	6.851	8.874	5.943	4.706	5.092	6.983	13.228	16.497	11.310	9.293
Clicks Group Ltd	CLICKS	0	CLS	6.369	6.711	7.251	6.733	6.799	5.549	6.044	5.257	5.134	4.707
Cargo Carriers Ltd	CARGO	1	CRG	2.865	2.070	2.286	2.429	2.499	2.502	2.451	2.706	2.977	3.035
Cashbuild Ltd	CASHBIL	0	CSB	5.175	6.341	6.849	4.715	4.937	4.952	4.111	5.798	5.188	4.747
Cullinan Holdings Ltd	CULLINAN	0	CUL	4.525	3.865	3.620	2.996	2.822	2.575	1.791	2.479	1.998	2.520
Datacentrix Holdings Ltd	DCENTRIX	0	DCT	5.739	5.098	5.938	7.103	6.548	6.105	6.445	6.718	5.397	6.991
Digicore Holdings Ltd	DIGICORE	0	DGC	5.476	4.165	4.075	6.347	5.909	7.425	9.327	8.938	7.783	5.724
Distell Group Ltd	DISTELL	0	DST	5.478	4.698	5.957	6.413	5.908	5.821	6.211	5.767	4.365	3.657
Delta EMD Ltd	DELTA	1	DTA	1.709	3.585	4.233	4.445	6.111	4.917	3.850	2.691	6.553	4.754
Datatec Ltd	DATATEC	0	DTC	5.950	6.619	6.848	6.957	6.196	5.754	5.929	7.411	6.882	5.368
EOH Holdings Ltd	EOH	0	EOH	5.360	5.397	5.025	5.008	4.244	3.732	5.119	5.274	5.104	4.623
Famous Brands Ltd	FAMBRANDS	0	FBR	17.438	15.155	14.370	13.098	10.462	8.038	7.826	6.190	6.557	6.515
Gijima Group Ltd	GIJIMA	1	GIJ	-0.694	0.194	2.492	1.494	3.103	2.892	3.117	2.604	2.543	1.870
Grindrod Ltd	GRINDROD	1	GND	0.470	0.658	0.608	0.640	0.694	0.588	0.372	0.444	0.382	0.348
Hudaco Industries Ltd	HUDACO	0	HDC	5.198	5.820	2.625	2.363	2.199	2.101	2.054	1.925	5.590	5.629
Howden Africa Holdings Ltd	HOWDEN	0	HWN	6.069	5.890	3.886	2.704	3.399	3.582	3.636	4.961	3.014	3.930
Iliad Africa Ltd	ILIAD	0	ILA	5.308	5.052	4.332	3.717	4.248	5.317	5.042	6.244	5.693	5.671
Illovo Sugar Ltd	ILLOVO	0	ILV	3.529	3.789	3.531	3.816	4.832	2.826	3.262	3.533	3.226	2.313
ISA Holdings Ltd	ISA	0	ISA	8.074	7.320	8.115	6.565	4.905	3.994	3.873	5.721	5.507	6.406
Italtile Ltd	ITLTILE	0	ITE	11.252	10.756	7.064	6.384	5.528	4.569	6.572	343.409	247.525	163.573
JD Group Ltd	JDGROUP	1	JDG	2.310	1.895	2.712	2.059	4.126	4.354	3.773	5.229	4.740	5.443

(Appendix 10.2 continued)

Company name	Short name	Distr /Non-distr	Code	2014	2013	2012	2011	2010	2009	2008	2007	2006	2005
Jasco Electronics Holdings Ltd	JASCO	1	JSC	2.324	1.515	2.363	2.011	2.081	2.705	4.534	4.796	5.001	5.872
KAP Industrial Holdings Ltd	KAP	1	KAP	2.624	2.416	2.080	3.053	2.788	2.469	2.735	3.912		3.842
Lewis Group Ltd	LEWIS	0	LEW	3.623	3.609	4.845	4.090	4.599	4.358	4.769	6.791	7.035	6.209
Masonite (Africa) Ltd	MASONITE	0	MAS	3.646	4.039	2.909	3.846	2.648	5.411	3.996	4.754	2.889	3.907
Mediclinic International Ltd	MEDCLIN	1	MDC	2.027	1.826	1.535	1.511	1.523	1.219	1.012	4.247	5.084	6.385
Metrofile Holdings Ltd	METROFILE	0	MFL	6.996	6.405	4.696	3.052	2.262	1.271	1.413	1.368	-0.397	-1.147
MICROmega Holdings Ltd	MICROMEGA	0	MMG	9.995		4.031	3.347	2.565	3.818	4.158	5.371	5.841	3.395
MoneyWeb Holdings Ltd	MONEYWB	0	MNY	8.175	10.847	5.344	6.701	5.756	5.423	9.723	8.269	7.899	6.279
Mr Price Group Ltd	MRPRICE	0	MPC	11.149	12.835	12.055	10.046	7.798	6.570	6.338	7.551	7.374	5.652
Massmart Holdings Ltd	MASSMART	0	MSM	4.360	6.854	5.480	6.048	6.187	5.843	6.043	5.056	5.091	4.987
Mustek Ltd	MUSTEK	1	MST	2.969	3.478	3.312	3.769	3.306	3.027	2.778	3.246	3.016	3.366
Metair Investments Ltd	METAIR	0	MTA	4.309	3.068	5.368	5.790	5.067	3.894	3.915	5.457	70.138	66.787
MTN Group Ltd	MTN GROUP	0	MTN	4.019	4.205	4.478	4.088	4.149	3.714	3.596	4.800	4.357	6.516
Murray & Roberts Holdings Ltd	M&R-HLD	1	MUR	2.934	2.717	2.229	2.547	3.371	3.401	3.770	3.986	2.664	3.002
Nictus Ltd	NICTUS	1	NCS	0.193	0.595	0.368	0.193	0.052	0.193	0.259	0.256	0.474	0.837
Nampak Ltd	NAMPAK	0	NPK	3.200	3.063	3.681	3.644	3.597	3.052	2.854	3.757	3.498	4.109
Naspers Ltd	NASPERS-N	0	NPN	6.297	4.373	4.624	5.084	5.275	4.151	3.009	4.836	4.695	3.746
Netcare Ltd	NETCARE	0	NTC	4.212	4.322	1.206	1.643	1.742	1.822	1.556	1.606	1.203	4.710
Nu-World Holdings Ltd	NUWORLD	0	NWL	5.203	5.782	5.418	5.445	5.502	4.982	5.827	7.887	5.891	6.402
OneLogix Group Ltd	ONELOGIX	0	OLG	3.252	3.424	3.546	3.423	3.046	3.117	3.730	3.598	4.463	4.681
Omnia Holdings Ltd	OMNIA	0	OMN	4.784	4.397	3.868	3.844	3.371	3.767	3.665	4.144	4.024	4.000
Phumelela Gamings and Leisure Ltd	PHUMELELA	0	PHM	4.794	4.973	4.865	5.451	5.140	6.320	10.487	11.961	11.611	11.380
Pick n Pay Stores Ltd	PICKNPAY	0	PIK	6.373	6.567	7.294	7.033	7.005	6.920	7.178	7.442	7.819	7.554
Primeserv Group Ltd	PRIMESERV	0	PMV	6.884	6.449	6.445	8.465		6.518	6.647	6.515	8.491	8.762
Pinnacle Holdings Ltd	PINNACLE	0	PNC	4.050	4.727	5.120	5.123	4.608	4.628	4.572	4.668	3.084	3.753
PPC Ltd	PPC	0	PPC	3.055	3.763	4.063	3.904	4.407	4.731	5.706	6.656	43.198	49.589
RCL Foods Ltd	RCL	1	RCL	2.527	2.363	4.065	4.909	4.754	4.519	5.188	5.293	4.438	4.451
Remgro Ltd	REMGRO	0	REM	7.420	6.136	11.223	10.193	9.471	10.492	13.556	13.643	7.399	8.737
Reunert Ltd	REUNERT	0	RLO	5.699	6.429	7.168	6.702	4.718	4.511	4.342	6.212	4.481	4.818
Rex Trueform Clothing Company Ltd	REX TRUE	0	RTO	3.167	3.372	4.232	4.258	3.848	4.405	4.389	3.991	3.595	3.470
Seardel Investement Corporation Ltd	SEARDEL	1	SER	1.785	2.111	2.313	2.282	1.914	1.714	1.992	2.492	2.751	3.252
Steinhoff International Holdings Ltd	STEINHOFF	1	SHF	2.710	2.484	2.362	2.330	3.182	3.152	2.915	3.358	3.183	2.756
Sovereign Food Investments Ltd	SOVFOOD	0	SOV	3.139	3.070	2.947	2.134	2.072	1.609	2.230	4.127	3.796	3.840
Spanjaard Ltd	SPANJAARD	0	SPA	3.642	3.714	2.640	3.208	3.304	3.248	2.912	3.409	3.854	4.094
Super Group Ltd	SUPRGRP	0	SPG	3.281	3.022	3.242	1.938	1.892	1.106	1.871	2.253	2.737	2.512
The Spar Group Ltd	SPAR	0	SPP	5.349	7.785	7.139	7.361	7.263	7.289	6.804	6.587	7.447	8.092
StratCorp Ltd	STRATCORP	1	STA	-1.361	-1.929	-1.747	1.699	0.618	1.355	4.959	5.388	5.658	0.185
Spur Corporation Ltd	SPURCORP	0	SUR	14.293	13.577	12.708	12.495	12.801	11.122	8.384	12.963	13.297	9.396
SilverBridge Holdings Ltd	SILVERB	0	SVB	6.750	7.217	5.254	5.282	6.802	5.346	6.146	4.171	0.484	3.878
Tiger Brands Ltd	TIGBRANDS	0	TBS	5.549	5.266	7.602	6.870	7.513	7.491	5.927	6.679	6.451	5.602
Tongaat Hulett Ltd	TONGAAT	1	TON	2.426	2.586	2.378	2.338	3.200		2.139	4.985	4.129	3.919
Transpaco Ltd	TRNPACO	0	TPC	5.012	5.071	4.980	4.999	4.480	4.050	3.275	3.109	3.459	3.788
Truworths International Ltd	TRUWTHS	0	TRU	12.951	16.641	18.242	14.632	14.043	11.670	8.965	12.592	10.855	9.132
Tsogo Sun Holdings Ltd	TSOGO SUN	0	TSH	3.954	4.389	4.870	2.611	3.825	3.764	3.105	4.017	5.397	7.968
Value Group Ltd	VALUE	0	VLE	3.452	3.614	3.366	3.191	3.345	3.199	2.546	2.902	3.537	3.704
Woolworths Holdings Ltd	WOOLIES	0	WHL	5.357	9.928	8.911	7.305	6.024	5.064	3.459	4.170	3.822	4.080
Winhold Ltd	WINHOLD	1	WNH	2.726	2.853	2.201	2.356	2.441	2.588	2.397	2.369	2.509	3.891

APPENDIX 10.3

F-SCORE SAMPLE COMPANIES

Model 1 – Financial variables

Identification of distressed and non-distressed sample companies

Company name	Short name	Code	Dist/non-dist	2014	2013	2012	2011	2010	2009	2008	2007	2006	2005
Adapt IT Holdings Ltd	ADAPTIT	ADI	0	1.332	1.071	0.951	0.855	1.186	1.853	1.314	1.217	1.232	1.064
Adcorp Holdings Ltd	ADCORP	ADR	0	3.158	2.256	1.933	1.910	1.922	2.589	2.094		1.671	1.645
Aveng Ltd	AVENG	AEG	1	0.683	0.693	0.642	0.571	0.708	0.748	1.247	1.077	1.143	0.958
Allied Electronics Corporation Ltd	ALTRON	AEL	0	1.159	0.999	0.854	1.025	1.099	1.018	1.106	0.791	1.005	1.052
AECI Ltd	AECI	AFE	1	0.794	0.838	0.807	0.995	1.174	0.790	0.669	0.902	1.016	1.217
African Oxygen Ltd	AFROX	AFX	0	0.959	0.835	0.764	0.778	0.721	1.128	0.564	0.918	0.513	-0.504
AH-Vest Ltd	AH-VEST	AHL	0	1.015	1.076	0.819	1.317	0.924	0.258	1.422	0.809	0.552	0.677
African Media Entertainment Ltd	AME	AME	0	0.869	0.867	1.217	0.980	0.779	1.306		1.110	1.755	2.090
Astrapak Ltd	ASTRAPAK	APK	0	1.082	0.760	0.717	0.844	1.020	3.338	0.852	0.840	1.100	1.189
Astral Foods Ltd	ASTRAL	ARL	0	1.001	0.770	0.764	1.069	1.103	1.050	1.089	0.962	1.386	1.511
Business Connextion Group Ltd	BCX	BCX	0	0.939	0.932	0.993	0.909	0.913	0.816	0.989	0.617	0.913	0.737
Beige Holdings Ltd	BEIGE	BEG	0	3.498	1.354	0.851	0.935	0.934	1.075	0.681	1.508	0.826	0.846
Bell Equipment Ltd	BELL	BEL	0	1.029	0.442	0.988	0.518	1.105	0.843	0.453	0.454	0.874	0.781
Basil Read Holdings Ltd	BASREAD	BSR	1	0.578	0.568	0.813	0.647	0.690	0.729	1.028	1.098	1.074	1.487
The Bidvest Group Ltd	BIDVEST	BVT	0	1.228	1.145	1.210	1.292	1.311	1.346	1.345	1.341	1.357	1.354
Caxton and CTP Publishers and Printers Ltd	CAXTON	CAT	1	0.739	0.862	1.159	0.898	1.079	1.083	0.788	0.691	0.790	1.050
Crookes Brothers Ltd	CROOKES	CKS	1	0.499	1.706	0.803	0.472	0.375	0.507	0.574	0.515	0.977	4.435
City Lodge Hotels Ltd	CITYLDG	CLH	0	1.316	1.129	0.911	0.925	1.508	1.010	1.318	1.457	1.236	1.080
Clicks Group Ltd	CLICKS	CLS	0	1.264	1.378	1.340	1.629	1.284	1.397	1.233	1.492	1.033	0.871
Cargo Carriers Ltd	CARGO	CRG	0	0.948	0.721	0.854	0.761	1.047	0.879	0.448	0.834	1.037	0.704
Cashbuild Ltd	CASHBIL	CSB	0	1.354	0.817	0.883	1.061	1.160	0.926	1.056	0.959	0.901	0.982
Cullinan Holdings Ltd	CULLINAN	CUL	1	0.619	0.499	0.653	0.789	0.775	0.712	0.842	0.901	1.103	0.654
Datacentrix Holdings Ltd	DCENTRIX	DCT	0	1.066	0.996	1.025	1.351	1.260	1.152	1.272	1.130	0.945	1.519
Digicore Holdings Ltd	DIGICORE	DGC	0	1.722	1.335	0.885	1.226	1.338	0.797	1.309	0.913	1.241	0.811
Distell Group Ltd	DISTELL	DST	1	0.613	0.502	0.685	0.831	0.753	0.560	0.468	0.616	0.611	0.560
Delta EMD Ltd	DELTA	DTA	0	1.102	0.324	1.104	0.708	-1.393	0.476	1.305	0.528	-8.954	0.197
Datatec Ltd	DATATEC	DTC	0	1.057	1.123	1.013	0.950	1.112	1.396	1.108	0.929	1.114	1.006
EOH Holdings Ltd	EOH	EOH	0	7.912	1.842	2.435	2.585	1.979	1.570	1.044	1.093	0.997	5.052
Famous Brands Ltd	FAMBRANDS	FBR	0	1.432	1.796	1.963	6.461	0.136	0.038	24.344	1.806	2.480	-3.595
Gijima Group Ltd	GIJIMA	GIJ	1	0.385	-0.301	-3.421	3.291	1.031	1.696	2.008	7.247	1.804	2.098
Grindrod Ltd	GRINDROD	GND	1	0.658	0.804	0.878	1.063	1.005	1.017	5.200	0.997	0.853	1.017
Hudaco Industries Ltd	HUDACO	HDC	1	0.802	0.691	1.146	1.286	1.396	1.716	0.952	1.364	0.819	0.802
Howden Africa Holdings Ltd	HOWDEN	HWN	0	1.317	1.026	0.493	1.017	1.113	1.406	-1.712	8.059	-2.323	0.770
Iliad Africa Ltd	ILIAD	ILA	0	1.049	1.083	0.925	0.896	1.023	1.000	1.221	1.084	1.041	1.015
Illovo Sugar Ltd	ILLOVO	ILV	0	0.937	0.643	0.662	0.668	0.552	0.721	0.686	0.696	0.749	0.659
ISA Holdings Ltd	ISA	ISA	0	1.900	0.094	0.899	0.753	0.990	1.258	1.035	0.711	1.067	1.238
Italtile Ltd	ITLTILE	ITE	1	0.198	1.438	1.348	1.536	-0.294	1.498	0.802	0.857	0.958	0.997
JD Group Ltd	JDGROUP	JDG	0	1.238	0.963	0.846	0.658	0.811	0.806	0.893	0.737	0.931	0.960

(Appendix 10.3 continued)

Company name	Short name	Code	Dist/non-dist	2014	2013	2012	2011	2010	2009	2008	2007	2006	2005
Jasco Electronics Holdings Ltd	JASCO	JSC	1	0.763	0.563	0.596	0.409	0.697	0.841	0.922	1.052	1.056	1.409
KAP Industrial Holdings Ltd	KAP	KAP	0	0.867	1.051	0.984	0.795	0.827	0.934	0.816	0.980		0.868
Lewis Group Ltd	LEWIS	LEW	0	0.867	0.807	1.013	0.785	0.649	0.841	0.672	0.637	0.961	0.846
Masonite (Africa) Ltd	MASONITE	MAS	1	0.541	0.915	0.835	1.051	1.020	0.931	1.005	0.909	1.184	0.911
Mediclinic International Ltd	MEDCLIN	MDC	0	1.069	1.117	1.400	1.376	1.859	1.860	1.242	1.017	-0.447	1.367
Metrofile Holdings Ltd	METROFILE	MFL	0	1.040	1.330	1.447	1.810	2.117	4.941	-2.428	-0.574	0.487	0.310
MICROmega Holdings Ltd	MICROMEGA	MMG	0	1.409		1.560	0.854	0.866	1.147	1.113	1.254	1.492	1.031
MoneyWeb Holdings Ltd	MONEYWB	MNY	1	0.688	0.761	1.123	0.523	0.861	0.780	1.002	0.834	0.754	-0.142
Mr Price Group Ltd	MRPRICE	MPC	0	1.415	1.256	1.178	1.321	1.312	1.173	1.066	0.959	1.043	1.184
Massmart Holdings Ltd	MASSMART	MSM	0	1.682	2.011	1.691	1.502	1.592	1.634	2.583	3.599	7.757	1.360
Mustek Ltd	MUSTEK	MST	1	0.523	0.867	0.760	0.824	0.954	0.824	0.883	0.705	0.779	0.765
Metair Investments Ltd	METAIR	MTA	1	0.820	0.722	1.013	0.946	1.030	1.190	0.892	0.732	0.917	1.046
MTN Group Ltd	MTN GROUP	MTN	0	0.843	0.915	0.950	0.958	1.280	1.249	1.248	1.921	9.797	2.202
Murray & Roberts Holdings Ltd	M&R-HLD	MUR	0	0.971	0.827	0.522	0.782	0.735	0.882	0.900	0.910	0.627	0.777
Nictus Ltd	NICTUS	NCS	0	1.574	0.447	0.365	0.463	0.667	0.707	0.488	0.520	0.631	0.510
Nampak Ltd	NAMPAK	NPK	0	1.171	0.907	0.849	0.794	1.094	1.023	0.938	0.874	0.769	0.889
Naspers Ltd	NASPERS-N	NPN	1	0.636	1.040	0.819	0.916	1.029	0.899	1.199	0.911	1.164	1.235
Netcare Ltd	NETCARE	NTC	0	1.015	1.078	-1.059	-1.155	-1.361	-1.686	-2.403	-3.763	-1.576	0.928
Nu-World Holdings Ltd	NUWORLD	NWL	1	0.682	1.551	0.322	1.114	0.473	0.987	0.887	1.460	1.153	1.264
OneLogix Group Ltd	ONELOGIX	OLG	0	1.082	1.097	1.365	1.282	1.170	1.461	1.312	1.471	1.571	1.601
Omnia Holdings Ltd	OMNIA	OMN	1	0.814	0.801	0.750	0.588	1.205	0.788	0.784	0.852	0.923	0.840
Phumelela Gamings and Leisure Ltd	PHUMELELA	PHM	1	0.709	0.828	0.854	0.872	0.959	0.563	1.825	2.366	2.200	2.358
Pick n Pay Stores Ltd	PICKNPAY	PIK	0	2.240	2.292	2.177	3.420	3.212	-1.553	0.686	1.178	1.175	28.149
Primeserv Group Ltd	PRIMESERV	PMV	0	1.849	1.698	1.525	2.102		1.820	1.518	1.451	1.954	2.224
Pinnacle Holdings Ltd	PINNACLE	PNC	0	1.022	0.939	0.979	1.142	0.880	1.083	1.079	0.865	1.092	1.146
PPC Ltd	PPC	PPC	0	1.292	1.269	1.238	1.142	1.309	1.291	1.006	0.838	0.824	0.463
RCL Foods Ltd	RCL	RCL	1	0.814	1.361	0.715	0.877	0.820	0.839	0.840	0.999	1.153	0.968
Remgro Ltd	REMGRO	REM	1	0.491	0.797	1.067	0.909	0.939	0.854	0.940	0.429	0.455	0.648
Reunert Ltd	REUNERT	RLO	1	0.767	0.781	0.856	0.994	1.003	1.079	0.877	0.640	0.941	1.072
Rex Trueform Clothing Company Ltd	REX TRUE	RTO	1	0.718	0.204	1.857	1.636	2.137	1.986	1.748	1.323	1.987	1.323
Seardel Investment Corporation Ltd	SEARDEL	SER	1	-0.231	0.656	0.543	0.415	0.704	0.313	0.565	0.421	0.742	0.835
Steinhoff International Holdings Ltd	STEINHOFF	SHF	0	1.448	1.770	1.918	1.599	1.366	1.444	1.913	1.216	1.414	0.859
Sovereign Food Investments Ltd	SOVFOOD	SOV	1	0.795	1.083	0.889	1.109	0.733	0.648	1.127	1.638	2.077	1.703
Spanjaard Ltd	SPANJAARD	SPA	1	0.456	0.764	0.641	0.807	0.715	0.637	1.177	0.753	1.029	0.798
Super Group Ltd	SUPRGRP	SPG	0	1.067	0.989	1.326	1.162	1.267	-3.099	3.356	2.074	2.228	1.332
The Spar Group Ltd	SPAR	SPP	1	-5.797	1.369	1.454	1.516	1.505	1.549	1.366	1.552	1.616	1.796
StratCorp Ltd	STRATCORP	STA	1	-1.276	0.109	5.103	0.781	0.591	-0.439	0.641	1.005	1.147	0.451
Spur Corporation Ltd	SPURCORP	SUR	0	1.044	1.142	1.243	1.340	0.857	1.403	1.024	1.072	0.886	1.194
SilverBridge Holdings Ltd	SILVERB	SVB	0	1.624	1.363	0.644	1.437	1.254	1.161	0.807	0.656	-0.905	1.266
Tiger Brands Ltd	TIGBRANDS	TBS	1	0.809	0.796	0.828	0.984	1.267	0.907	0.995	0.956	1.046	1.069
Tongaat Hulett Ltd	TONGAAT	TON	1	0.643	0.640	0.507	0.480	0.803		0.567	0.529	0.466	0.674
Transpaco Ltd	TRNPACO	TPC	0	0.984	0.972	0.933	1.254	1.102	1.515	0.994	0.686	1.026	0.928
Truworths International Ltd	TRUWTHS	TRU	0	1.310	1.324	1.109	1.272	1.418	1.294	1.387	1.092	0.931	1.044
Tsogo Sun Holdings Ltd	TSOGO SUN	TSH	0	1.212	1.256	1.110	1.013	1.447	1.674	1.285	1.002	1.314	1.131
Value Group Ltd	VALUE	VLE	0	1.113	1.169	0.900	0.880	0.960	1.030	0.975	0.720	0.996	1.252
Woolworths Holdings Ltd	WOOLIES	WHL	0	1.895	1.585	1.455	1.441	1.555	1.206	1.284	1.103	1.163	1.558
Winhold Ltd	WINHOLD	WNH	0	0.896	0.710	0.815	0.831	1.010	0.891	0.844	0.938	0.971	0.882

APPENDIX 10.4

F-SCORE SAMPLE COMPANIES

Model 2 – Financial and market variables

Identification of distressed and non-distressed sample companies

Company name	Short name	Code	Dist/ Non-dist	2014	2013	2012	2011	2010	2009	2008	2007	2006	2005
Adapt IT Holdings Ltd	ADAPTIT	ADI	0	1.267	1.232	1.054	0.934	1.197	1.841	1.311	1.251	1.312	1.107
Adcorp Holdings Ltd	ADCORP	ADR	0	3.385	2.574	2.203	2.142	2.178	2.881	2.094	-0.109	1.978	2.291
Aveng Ltd	AVENG	AEG	1	0.796	0.811	0.795	0.925	1.266	1.312	1.956	1.618	1.408	0.976
Allied Electronics Corporation Ltd	ALTRON	AEL	0	1.388	1.148	1.060	1.216	1.366	1.285	1.746	1.075	1.281	1.134
AECI Ltd	AECI	AFE	0	1.839	1.748	1.440	1.869	1.869	1.146	1.161	1.348	2.046	1.760
African Oxygen Ltd	AFROX	AFX	1	0.980	0.939	0.901	0.865	0.777	1.207	0.684	1.114	0.767	-0.331
AH-Vest Ltd	AH-VEST	AHL	1	1.026	1.078	0.859	1.354	0.853	0.373	1.349	0.812	0.534	0.685
African Media Entertainment Ltd	AME	AME	0	1.500	1.431	1.735	1.375	1.132	1.306	-0.109	1.483	2.083	2.351
Astrapak Ltd	ASTRAPAK	APK	1	1.040	0.780	0.709	0.922	1.199	3.401	0.887	0.987	1.236	1.386
Astral Foods Ltd	ASTRAL	ARL	0	2.068	1.271	1.661	2.406	2.221	2.083	2.040	2.604	2.888	2.640
Business Connextion Group Ltd	BCX	BCX	1	0.984	0.983	1.041	0.941	1.027	0.828	1.012	0.650	1.058	0.831
Beige Holdings Ltd	BEIGE	BEG	0	3.493	1.309	0.960	0.915	0.891	1.004	0.799	1.847	0.960	0.825
Bell Equipment Ltd	BELL	BEL	1	1.025	0.659	1.302	0.908	1.133	0.492	0.802	1.011	1.334	0.829
Basil Read Holdings Ltd	BASREAD	BSR	1	0.110	0.646	0.638	0.826	0.915	1.177	1.274	1.500	5.604	1.654
The Bidvest Group Ltd	BIDVEST	BVT	0	3.258	2.983	2.941	2.644	2.592	2.414	2.550	2.514	2.326	2.182
Caxton and CTP Publishers and Printers Ltd	CAXTON	CAT	1	0.832	1.023	1.277	1.033	1.187	1.201	0.901	0.859	0.965	1.181
Crookes Brothers Ltd	CROOKES	CKS	0	1.312	2.643	1.474	0.682	0.488	0.921	0.938	0.819	1.342	4.454
City Lodge Hotels Ltd	CITYLDG	CLH	0	2.063	1.883	1.398	1.260	2.029	1.428	1.920	2.009	1.664	1.469
Clicks Group Ltd	CLICKS	CLS	0	1.674	1.728	1.700	1.925	1.608	1.622	1.382	1.644	1.140	0.957
Cargo Carriers Ltd	CARGO	CRG	0	1.317	0.856	0.911	0.858	1.197	0.902	0.525	0.985	1.249	0.821
Cashbuild Ltd	CASHBIL	CSB	0	2.686	1.994	2.385	1.868	1.998	1.878	1.848	1.587	1.371	1.387
Cullinan Holdings Ltd	CULLINAN	CUL	1	0.648	0.630	0.700	0.845	0.802	0.741	0.794	0.968	1.086	0.662
Datacentrix Holdings Ltd	DCENTRIX	DCT	0	1.138	1.019	1.063	1.421	1.367	1.201	1.339	1.220	0.965	1.593
Digicore Holdings Ltd	DIGICORE	DGC	0	1.740	1.306	0.887	1.269	1.349	0.787	1.382	1.139	1.375	0.886
Distell Group Ltd	DISTELL	DST	0	1.474	1.164	1.265	1.392	1.318	1.124	1.002	1.121	0.977	0.915
Delta EMD Ltd	DELTA	DTA	1	0.719	0.363	1.124	0.764	-1.235	0.859	1.487	0.135	-9.287	0.421
Datatec Ltd	DATATEC	DTC	0	1.442	1.568	1.437	1.162	1.392	1.729	1.485	1.334	1.479	0.971
EOH Holdings Ltd	EOH	EOH	0	8.508	2.309	2.817	2.937	2.264	1.699	1.147	1.242	1.144	5.149
Famous Brands Ltd	FAMBRANDS	FBR	0	1.925	2.277	2.301	6.809	0.431	0.221	24.521	1.971	2.651	-3.482
Gijima Group Ltd	GIJIMA	GIJ	1	0.255	-0.384	-3.462	3.235	1.123	1.652	2.015	7.384	1.817	2.290
Grindrod Ltd	GRINDROD	GND	1	0.865	0.950	1.019	1.185	1.238	1.188	5.859	1.318	1.177	1.231
Hudaco Industries Ltd	HUDACO	HDC	1	0.802	1.758	2.417	2.468	2.359	2.658	2.029	2.131	1.473	1.308
Howden Africa Holdings Ltd	HOWDEN	HWN	0	1.786	1.625	0.892	1.280	1.247	1.610	-1.627	8.355	-2.323	3.689
Iliad Africa Ltd	ILIAD	ILA	0	1.166	1.140	0.986	0.864	1.059	1.116	1.357	1.299	1.217	1.186
Illovo Sugar Ltd	ILLOVO	ILV	0	1.145	0.882	0.811	0.784	0.803	0.958	0.943	0.896	0.971	0.713
ISA Holdings Ltd	ISA	ISA	0	1.908	0.079	0.931	0.807	1.094	1.230	0.988	0.858	1.064	1.445
Italtile Ltd	ITLTILE	ITE	1	0.310	1.506	1.419	1.604	-0.189	1.511	3.137	2.635	2.523	2.550
JD Group Ltd	JDGROUP	JDG	1	0.582	1.393	1.425	1.150	1.160	0.912	1.179	1.432	1.869	1.818

(Appendix 10.4 continued)

Company name	Short name	Code	Dist/ Non-dist	2014	2013	2012	2011	2010	2009	2008	2007	2006	2005
Jasco Electronics Holdings Ltd	JASCO	JSC	1	0.798	0.533	0.663	0.421	0.690	0.862	0.979	1.136	1.090	1.509
KAP Industrial Holdings Ltd	KAP	KAP	1	0.946	1.072	1.671	0.833	0.935	0.887	0.803	0.980	-0.109	0.949
Lewis Group Ltd	LEWIS	LEW	0	1.918	1.954	2.044	1.710	1.451	1.570	1.423	1.398	1.556	1.316
Masonite (Africa) Ltd	MASONITE	MAS	1	0.875	1.482	1.347	1.196	1.065	1.681	2.259	1.668	1.549	1.279
Mediclinic International Ltd	MEDCLIN	MDC	0	1.563	1.088	1.657	1.615	2.090	2.000	1.431	1.240	-0.310	1.582
Metrofile Holdings Ltd	METROFILE	MFL	0	1.087	1.401	1.548	1.874	2.212	4.946	-2.462	0.001	0.884	0.601
MICROmega Holdings Ltd	MICROMEGA	MMG	0	1.409	-0.109	1.585	1.119	0.849	1.153	1.129	1.278	1.555	1.163
MoneyWeb Holdings Ltd	MONEYWB	MNY	1	0.685	0.764	1.120	0.520	0.863	0.780	1.004	0.836	0.755	-0.145
Mr Price Group Ltd	MRPRICE	MPC	0	2.333	2.012	1.815	1.858	1.713	1.495	1.284	1.222	1.321	1.349
Massmart Holdings Ltd	MASSMART	MSM	0	2.242	2.763	2.400	2.059	2.275	2.304	3.439	4.234	8.289	1.749
Mustek Ltd	MUSTEK	MST	1	0.661	0.946	0.872	0.948	1.198	0.820	0.897	0.761	0.839	0.912
Metair Investments Ltd	METAIR	MTA	0	1.158	1.038	1.457	1.302	1.391	1.259	0.915	0.887	4.781	4.092
MTN Group Ltd	MTN GROUP	MTN	0	1.201	1.188	1.328	1.270	1.512	1.343	1.308	2.133	13.687	5.291
Murray & Roberts Holdings Ltd	M&R-HLD	MUR	0	1.233	1.034	0.263	0.168	1.108	1.625	1.586	1.480	0.881	0.946
Nictus Ltd	NICTUS	NCS	0	1.531	0.347	0.738	0.526	0.694	0.767	0.517	0.548	0.610	0.548
Nampak Ltd	NAMPAK	NPK	0	1.491	1.160	1.123	1.013	1.275	1.133	1.113	1.111	0.944	1.039
Naspers Ltd	NASPERS-N	NPN	0	2.513	3.083	2.333	2.242	2.163	1.868	2.428	1.986	2.109	2.166
Netcare Ltd	NETCARE	NTC	0	1.241	1.270	-0.904	-1.020	-1.214	-1.568	-2.370	-3.683	-1.414	0.995
Nu-World Holdings Ltd	NUWORLD	NWL	0	1.096	1.754	0.526	1.216	0.917	1.145	1.066	1.758	1.582	1.708
OneLogix Group Ltd	ONELOGIX	OLG	0	1.144	1.195	1.461	1.385	1.239	1.419	1.303	1.560	1.735	2.000
Omnia Holdings Ltd	OMNIA	OMN	0	2.505	2.409	1.891	1.556	1.344	2.034	1.637	1.551	1.325	1.519
Phumelela Gamings and Leisure Ltd	PHUMELELA	PHM	1	0.921	1.011	0.933	0.992	1.035	0.646	1.958	2.615	2.310	2.489
Pick n Pay Stores Ltd	PICKNPAY	PIK	0	2.395	2.427	2.343	3.621	3.487	-1.303	0.906	1.387	1.385	28.341
Primeserv Group Ltd	PRIMESERV	PMV	0	1.853	1.705	1.551	2.102	-0.109	1.777	1.510	1.577	2.050	2.224
Pinnacle Holdings Ltd	PINNACLE	PNC	0	1.161	1.239	1.261	1.365	1.071	1.124	1.142	1.026	1.391	1.289
PPC Ltd	PPC	PPC	0	1.506	1.493	1.448	1.310	1.548	1.487	1.296	1.167	3.474	2.522
RCL Foods Ltd	RCL	RCL	1	0.824	1.489	0.809	1.003	0.970	0.960	1.065	1.250	1.369	1.085
Remgro Ltd	REMGRO	REM	0	2.009	1.827	2.239	1.925	1.800	1.925	2.911	2.141	1.705	1.829
Reunert Ltd	REUNERT	RLO	0	1.338	1.461	1.642	1.675	1.600	1.818	1.618	0.955	1.620	1.601
Rex Trueform Clothing Company Ltd	REX TRUE	RTO	1	0.642	0.141	2.042	1.907	2.316	2.161	1.962	1.414	2.057	1.380
Seardel Investment Corporation Ltd	SEARDEL	SER	1	-0.159	0.684	0.583	0.552	0.678	0.431	0.386	0.458	0.906	0.898
Steinhoff International Holdings Ltd	STEINHOFF	SHF	0	2.140	2.227	2.303	1.945	1.719	1.711	2.194	1.489	1.663	1.066
Sovereign Food Investments Ltd	SOVFOOD	SOV	1	0.867	1.151	1.011	1.155	0.918	0.586	1.284	1.932	2.967	1.836
Spanjaard Ltd	SPANJAARD	SPA	1	0.474	0.832	0.713	0.890	0.768	0.713	1.320	0.745	1.065	0.832
Super Group Ltd	SUPRGRP	SPG	0	1.387	1.291	1.642	1.196	1.852	-3.387	3.299	2.266	2.354	1.485
The Spar Group Ltd	SPAR	SPP	1	-4.887	2.170	2.202	2.171	2.166	2.038	1.837	1.961	1.918	2.087
StratCorp Ltd	STRATCORP	STA	1	-1.255	0.057	5.081	0.771	0.503	-0.515	0.771	1.352	1.197	0.392
Spur Corporation Ltd	SPURCORP	SUR	0	1.242	1.397	1.410	1.469	1.002	1.537	1.066	1.227	0.998	1.273
SilverBridge Holdings Ltd	SILVERB	SVB	0	1.707	1.326	0.594	1.436	1.288	1.154	0.843	1.315	-0.727	1.266
Tiger Brands Ltd	TIGBRANDS	TBS	0	2.917	2.684	2.816	2.818	2.900	2.539	2.726	2.469	2.451	2.267
Tongaat Hulett Ltd	TONGAAT	TON	0	1.765	1.792	1.484	1.357	0.803	-0.109	1.174	0.575	1.286	1.191
Transpaco Ltd	TRNPACO	TPC	0	1.228	1.211	1.204	1.546	1.418	1.717	1.067	0.884	1.055	1.066
Truworths International Ltd	TRUWTHS	TRU	0	1.968	1.973	1.748	1.828	1.908	1.753	1.685	1.453	1.169	1.247
Tsogo Sun Holdings Ltd	TSOGO SUN	TSH	0	1.417	1.418	1.361	1.367	1.561	1.834	1.386	1.238	1.499	1.251
Value Group Ltd	VALUE	VLE	0	1.183	1.242	1.005	0.960	1.042	1.149	0.965	0.714	1.094	1.314
Woolworths Holdings Ltd	WOOLIES	WHL	0	2.340	2.011	1.839	1.711	1.847	1.345	1.363	1.312	1.322	1.686
Winhold Ltd	WINHOLD	WNH	1	0.884	0.698	0.779	0.863	1.043	0.913	0.851	0.983	0.984	0.940

APPENDIX 10.5

F-SCORE SAMPLE COMPANIES

Model 3 – Financial and market variables plus leading business cycle indicator

Identification of distressed and non-distressed sample companies

Company name	Short name	Code	Dist / Non-dist	2014	2013	2012	2011	2010	2009	2008	2007	2006	2005
Adapt IT Holdings Ltd	ADAPTIT	ADI	0	2.244	2.226	2.048	1.937	2.198	2.696	2.203	2.206	2.282	2.037
Adcorp Holdings Ltd	ADCORP	ADR	0	4.363	3.568	3.197	3.146	3.178	3.736	2.985	0.847	2.948	3.220
Aveng Ltd	AVENG	AEG	1	1.774	1.806	1.789	1.928	2.266	2.167	2.848	2.573	2.378	1.905
Allied Electronics Corporation Ltd	ALTRON	AEL	0	2.365	2.143	2.053	2.219	2.366	2.140	2.638	2.030	2.251	2.063
AECI Ltd	AECI	AFE	0	2.817	2.743	2.433	2.872	2.869	2.001	2.052	2.303	3.016	2.689
African Oxygen Ltd	AFROX	AFX	1	1.958	1.934	1.894	1.869	1.777	2.062	1.576	2.069	1.737	0.598
AH-Vest Ltd	AH-VEST	AHL	1	2.004	2.073	1.853	2.357	1.853	1.228	2.241	1.767	1.504	1.614
African Media Entertainment Ltd	AME	AME	0	2.478	2.426	2.728	2.379	2.132	2.161	0.783	2.438	3.053	3.280
Astrapak Ltd	ASTRAPAK	APK	0	2.018	1.774	1.702	1.925	2.200	4.256	1.778	1.942	2.206	2.315
Astral Foods Ltd	ASTRAL	ARL	0	3.046	2.265	2.654	3.409	3.221	2.938	2.931	3.560	3.858	3.570
Business Connextion Group Ltd	BCX	BCX	1	1.962	1.978	2.035	1.945	2.027	1.683	1.903	1.605	2.028	1.760
Beige Holdings Ltd	BEIGE	BEG	0	4.471	2.303	1.954	1.918	1.892	1.859	1.690	2.802	1.929	1.754
Bell Equipment Ltd	BELL	BEL	1	2.002	1.653	2.295	1.912	2.133	1.347	1.693	1.966	2.304	1.758
Basil Read Holdings Ltd	BASREAD	BSR	1	1.087	1.640	1.632	1.829	1.915	2.032	2.165	2.456	6.574	2.583
The Bidvest Group Ltd	BIDVEST	BVT	0	4.235	3.978	3.934	3.647	3.592	3.269	3.442	3.469	3.296	3.111
Caxton and CTP Publishers and Printers Ltd	CAXTON	CAT	1	1.809	2.017	2.270	2.036	2.187	2.056	1.792	1.814	1.935	2.110
Crookes Brothers Ltd	CROOKES	CKS	0	2.289	3.637	2.468	1.685	1.489	1.776	1.830	1.775	2.312	5.383
City Lodge Hotels Ltd	CITYLDG	CLH	0	3.041	2.877	2.392	2.263	3.029	2.283	2.812	2.965	2.634	2.398
Clicks Group Ltd	CLICKS	CLS	0	2.652	2.722	2.694	2.929	2.608	2.477	2.274	2.599	2.110	1.886
Cargo Carriers Ltd	CARGO	CRG	0	2.295	1.850	1.904	1.861	2.197	1.757	1.416	1.940	2.219	1.751
Cashbuild Ltd	CASHBIL	CSB	0	3.664	2.989	3.379	2.871	2.998	2.733	2.739	2.543	2.341	2.316
Cullinan Holdings Ltd	CULLINAN	CUL	1	1.625	1.625	1.693	1.848	1.802	1.596	1.686	1.923	2.056	1.591
Datacentrix Holdings Ltd	DCENTRIX	DCT	0	2.115	2.014	2.056	2.424	2.368	2.056	2.231	2.176	1.935	2.522
Digicore Holdings Ltd	DIGICORE	DGC	0	2.718	2.301	1.881	2.272	2.349	1.642	2.274	2.094	2.344	1.815
Distell Group Ltd	DISTELL	DST	0	2.452	2.158	2.258	2.396	2.318	1.979	1.894	2.076	1.947	1.844
Delta EMD Ltd	DELTA	DTA	1	1.697	1.357	2.118	1.768	-0.235	1.714	2.379	1.091	-8.317	1.350
Datatec Ltd	DATATEC	DTC	0	2.419	2.563	2.431	2.165	2.392	2.584	2.377	2.290	2.449	1.900
EOH Holdings Ltd	EOH	EOH	0	9.486	3.303	3.810	3.941	3.264	2.554	2.038	2.198	2.114	6.078
Famous Brands Ltd	FAMBRANDS	FBR	0	2.903	3.271	3.294	7.812	1.431	1.076	25.412	2.927	3.621	-2.553
Gijima Group Ltd	GIJIMA	GIJ	1	1.233	0.610	-2.469	4.239	2.123	2.507	2.907	8.339	2.787	3.219
Grindrod Ltd	GRINDROD	GND	1	1.842	1.944	2.012	2.188	2.238	2.043	6.751	2.273	2.147	2.160
Hudaco Industries Ltd	HUDACO	HDC	1	1.780	2.753	3.411	3.471	3.359	3.513	2.921	3.086	2.442	2.237
Howden Africa Holdings Ltd	HOWDEN	HWN	0	2.763	2.619	1.885	2.283	2.247	2.465	-0.735	9.310	-1.353	4.618
Iliad Africa Ltd	ILIAD	ILA	0	2.143	2.134	1.980	1.867	2.059	1.971	2.248	2.255	2.187	2.116
Illovo Sugar Ltd	ILLOVO	ILV	0	2.123	1.876	1.804	1.788	1.803	1.813	1.835	1.851	1.941	1.642
ISA Holdings Ltd	ISA	ISA	0	2.886	1.074	1.924	1.810	2.094	2.085	1.880	1.813	2.034	2.375
Italtile Ltd	ITLTILE	ITE	1	1.288	2.500	2.412	2.607	0.811	2.366	4.028	3.591	3.493	3.479
JD Group Ltd	JDGROUP	JDG	1	1.560	2.388	2.419	2.153	2.160	1.767	2.071	2.388	2.839	2.747

(Appendix 10.5 continued)

Company name	Short name	Code	Dist / Non-dist	2014	2013	2012	2011	2010	2009	2008	2007	2006	2005
Jasco Electronics Holdings Ltd	JASCO	JSC	1	1.775	1.527	1.656	1.425	1.690	1.717	1.871	2.091	2.060	2.438
KAP Industrial Holdings Ltd	KAP	KAP	1	1.924	2.066	2.664	1.836	1.935	1.742	1.694	1.936	0.861	1.879
Lewis Group Ltd	LEWIS	LEW	0	2.895	2.949	3.038	2.714	2.451	2.425	2.314	2.354	2.526	2.245
Masonite (Africa) Ltd	MASONITE	MAS	1	1.853	2.477	2.341	2.199	2.065	2.536	3.150	2.624	2.519	2.208
Mediclinic International Ltd	MEDCLIN	MDC	0	2.540	2.082	2.650	2.618	3.090	2.855	2.322	2.196	0.660	2.511
Metrofile Holdings Ltd	METROFILE	MFL	0	2.065	2.396	2.541	2.878	3.212	5.801	-1.570	0.957	1.854	1.530
MICROmega Holdings Ltd	MICROMEGA	MMG	0	2.387	0.886	2.579	2.122	1.849	2.008	2.020	2.233	2.525	2.092
MoneyWeb Holdings Ltd	MONEYWB	MNY	1	1.663	1.759	2.113	1.523	1.864	1.635	1.896	1.791	1.725	0.785
Mr Price Group Ltd	MRPRICE	MPC	0	3.311	3.007	2.809	2.861	2.713	2.350	2.175	2.178	2.291	2.279
Massmart Holdings Ltd	MASSMART	MSM	0	3.219	3.758	3.393	3.062	3.275	3.159	4.330	5.189	9.259	2.678
Mustek Ltd	MUSTEK	MST	1	1.639	1.941	1.866	1.952	2.199	1.675	1.789	1.716	1.809	1.841
Metair Investments Ltd	METAIR	MTA	0	2.136	2.033	2.450	2.305	2.391	2.114	1.807	1.843	5.751	5.021
MTN Group Ltd	MTN GROUP	MTN	0	2.179	2.183	2.322	2.273	2.512	2.198	2.200	3.088	14.657	6.220
Murray & Roberts Holdings Ltd	M&R-HLD	MUR	0	2.211	2.029	1.257	1.172	2.108	2.480	2.478	2.435	1.851	1.875
Nictus Ltd	NICTUS	NCS	0	2.509	1.341	1.731	1.529	1.694	1.622	1.408	1.503	1.580	1.477
Nampak Ltd	NAMPAK	NPK	0	2.468	2.154	2.117	2.016	2.275	1.988	2.005	2.067	1.913	1.968
Naspers Ltd	NASPERS-N	NPN	0	3.491	4.077	3.327	3.245	3.164	2.723	3.320	2.942	3.079	3.095
Netcare Ltd	NETCARE	NTC	0	2.219	2.265	0.089	-0.017	-0.214	-0.713	-1.479	-2.727	-0.444	1.924
Nu-World Holdings Ltd	NUWORLD	NWL	0	2.073	2.748	1.519	2.219	1.918	2.000	1.958	2.714	2.552	2.637
OneLogix Group Ltd	ONELOGIX	OLG	0	2.122	2.190	2.454	2.388	2.239	2.274	2.195	2.516	2.705	2.930
Omnia Holdings Ltd	OMNIA	OMN	0	3.482	3.403	2.884	2.560	2.345	2.889	2.528	2.506	2.295	2.448
Phumelela Gamings and Leisure Ltd	PHUMELELA	PHM	1	1.898	2.005	1.926	1.995	2.035	1.501	2.850	3.570	3.280	3.418
Pick n Pay Stores Ltd	PICKNPAY	PIK	0	3.372	3.422	3.336	4.625	4.487	-0.448	1.797	2.342	2.355	29.270
Primeserv Group Ltd	PRIMESERV	PMV	0	2.830	2.699	2.544	3.105	0.892	2.632	2.401	2.533	3.020	3.153
Pinnacle Holdings Ltd	PINNACLE	PNC	0	2.139	2.234	2.254	2.368	2.071	1.979	2.033	1.981	2.361	2.218
PPC Ltd	PPC	PPC	0	2.483	2.488	2.441	2.313	2.548	2.342	2.187	2.123	4.444	3.451
RCL Foods Ltd	RCL	RCL	1	1.802	2.484	1.802	2.007	1.970	1.815	1.956	2.205	2.339	2.014
Remgro Ltd	REMGRO	REM	0	2.986	2.821	3.233	2.929	2.800	2.780	3.803	3.097	2.675	2.759
Reunert Ltd	REUNERT	RLO	0	2.316	2.455	2.636	2.679	2.600	2.673	2.509	1.910	2.590	2.530
Rex Trueform Clothing Company Ltd	REX TRUE	RTO	1	1.620	1.135	3.036	2.911	3.316	3.016	2.853	2.369	3.027	2.309
Seardel Investment Corporation Ltd	SEARDEL	SER	1	0.819	1.678	1.576	1.555	1.679	1.286	1.278	1.413	1.876	1.827
Steinhoff International Holdings Ltd	STEINHOFF	SHF	0	3.118	3.221	3.296	2.948	2.719	2.566	3.085	2.444	2.633	1.995
Sovereign Food Investments Ltd	SOVFOOD	SOV	1	1.845	2.145	2.005	2.158	1.918	1.441	2.175	2.887	3.937	2.765
Spanjaard Ltd	SPANJAARD	SPA	1	1.452	1.827	1.706	1.893	1.768	1.568	2.211	1.701	2.035	1.761
Super Group Ltd	SUPRGRP	SPG	0	2.365	2.285	2.636	2.199	2.852	-2.532	4.190	3.221	3.323	2.414
The Spar Group Ltd	SPAR	SPP	1	-3.910	3.164	3.196	3.175	3.166	2.893	2.728	2.916	2.888	3.017
StratCorp Ltd	STRATCORP	STA	1	-0.278	1.051	6.074	1.774	1.503	0.340	1.663	2.307	2.167	1.322
Spur Corporation Ltd	SPURCORP	SUR	0	2.220	2.391	2.403	2.473	2.002	2.392	1.957	2.182	1.968	2.203
SilverBridge Holdings Ltd	SILVERB	SVB	0	2.684	2.321	1.587	2.440	2.288	2.009	1.734	2.271	0.243	2.195
Tiger Brands Ltd	TIGBRANDS	TBS	0	3.895	3.678	3.809	3.822	3.900	3.394	3.618	3.425	3.421	3.196
Tongaat Hulett Ltd	TONGAAT	TON	0	2.743	2.786	2.478	2.361	1.804	0.746	2.065	1.531	2.256	2.121
Transpaco Ltd	TRNPACO	TPC	0	2.205	2.205	2.197	2.549	2.418	2.572	1.958	1.839	2.025	1.995
Truworths International Ltd	TRUWTHS	TRU	0	2.946	2.968	2.741	2.831	2.908	2.608	2.577	2.408	2.139	2.176
Tsogo Sun Holdings Ltd	TSOGO SUN	TSH	0	2.394	2.413	2.355	2.371	2.561	2.689	2.278	2.194	2.469	2.181
Value Group Ltd	VALUE	VLE	0	2.161	2.237	1.998	1.963	2.043	2.004	1.856	1.669	2.064	2.243
Woolworths Holdings Ltd	WOOLIES	WHL	0	3.318	3.006	2.832	2.714	2.847	2.200	2.255	2.267	2.292	2.616
Winhold Ltd	WINHOLD	WNH	1	1.861	1.693	1.773	1.866	2.043	1.768	1.742	1.938	1.954	1.870

APPENDIX 10.6

F-SCORE SAMPLE COMPANIES
Model 4 – Financial and market variables plus coincident business cycle indicator

Identification of distressed and non-distressed sample companies

Company name	Short name	Code	Distr / non distr	2014	2013	2012	2011	2010	2009	2008	2007	2006	2005
Adapt IT Holdings Ltd	ADAPTIT	ADI	0	2.412	2.365	2.159	1.990	2.197	2.798	2.268	2.304	2.294	2.017
Adcorp Holdings Ltd	ADCORP	ADR	0	4.531	3.706	3.308	3.199	3.178	3.838	3.051	0.944	2.960	3.200
Aveng Ltd	AVENG	AEG	1	1.942	1.944	1.900	1.981	2.266	2.270	2.913	2.671	2.390	1.885
Allied Electronics Corporation Ltd	ALTRON	AEL	0	2.533	2.281	2.165	2.272	2.366	2.242	2.703	2.128	2.263	2.044
AECI Ltd	AECI	AFE	0	2.985	2.881	2.545	2.925	2.869	2.104	2.118	2.401	3.028	2.669
African Oxygen Ltd	AFROX	AFX	0	2.126	2.072	2.006	1.922	1.777	2.165	1.642	2.166	1.749	0.578
AH-Vest Ltd	AH-VEST	AHL	0	2.172	2.211	1.964	2.410	1.853	1.330	2.307	1.864	1.516	1.594
African Media Entertainment Ltd	AME	AME	0	2.646	2.564	2.840	2.432	2.132	2.263	0.849	2.536	3.065	3.260
Astrapak Ltd	ASTRAPAK	APK	0	2.186	1.912	1.814	1.978	2.199	4.358	1.844	2.039	2.218	2.295
Astral Foods Ltd	ASTRAL	ARL	0	3.214	2.404	2.766	3.462	3.221	3.041	2.997	3.657	3.870	3.550
Business Connextion Group Ltd	BCX	BCX	0	2.130	2.116	2.146	1.998	2.027	1.786	1.969	1.702	2.040	1.741
Beige Holdings Ltd	BEIGE	BEG	0	4.639	2.442	2.065	1.971	1.891	1.961	1.756	2.900	1.942	1.734
Bell Equipment Ltd	BELL	BEL	0	2.170	1.791	2.407	1.965	2.133	1.449	1.759	2.063	2.316	1.738
Basil Read Holdings Ltd	BASREAD	BSR	1	1.255	1.779	1.744	1.882	1.915	2.135	2.231	2.553	6.586	2.563
The Bidvest Group Ltd	BIDVEST	BVT	0	4.403	4.116	4.046	3.700	3.592	3.372	3.508	3.566	3.308	3.091
Caxton and CTP Publishers and Printers Ltd	CAXTON	CAT	1	1.977	2.155	2.382	2.089	2.187	2.158	1.858	1.911	1.947	2.090
Crookes Brothers Ltd	CROOKES	CKS	0	2.457	3.775	2.579	1.738	1.488	1.878	1.896	1.872	2.324	5.364
City Lodge Hotels Ltd	CITYLDG	CLH	0	3.209	3.015	2.503	2.316	3.029	2.385	2.878	3.062	2.646	2.379
Clicks Group Ltd	CLICKS	CLS	0	2.820	2.860	2.805	2.982	2.608	2.579	2.340	2.697	2.122	1.867
Cargo Carriers Ltd	CARGO	CRG	0	2.463	1.988	2.016	1.914	2.197	1.860	1.482	2.038	2.231	1.731
Cashbuild Ltd	CASHBIL	CSB	0	3.832	3.127	3.490	2.924	2.998	2.835	2.805	2.640	2.353	2.296
Cullinan Holdings Ltd	CULLINAN	CUL	1	1.793	1.763	1.805	1.901	1.802	1.699	1.752	2.020	2.068	1.571
Datacentrix Holdings Ltd	DCENTRIX	DCT	0	2.283	2.152	2.168	2.477	2.367	2.158	2.296	2.273	1.947	2.503
Digicore Holdings Ltd	DIGICORE	DGC	0	2.886	2.439	1.992	2.325	2.349	1.745	2.340	2.192	2.357	1.795
Distell Group Ltd	DISTELL	DST	0	2.620	2.296	2.370	2.449	2.318	2.082	1.960	2.173	1.959	1.825
Delta EMD Ltd	DELTA	DTA	1	1.865	1.495	2.229	1.821	-0.235	1.816	2.445	1.188	-8.305	1.330
Datatec Ltd	DATATEC	DTC	0	2.587	2.701	2.542	2.218	2.392	2.686	2.443	2.387	2.461	1.881
EOH Holdings Ltd	EOH	EOH	0	9.654	3.442	3.922	3.994	3.264	2.656	2.104	2.295	2.127	6.059
Famous Brands Ltd	FAMBRANDS	FBR	0	3.071	3.410	3.406	7.865	1.431	1.178	25.478	3.024	3.633	-2.573
Gijima Group Ltd	GIJIMA	GIJ	1	1.401	0.749	-2.357	4.292	2.123	2.609	2.973	8.437	2.799	3.199
Grindrod Ltd	GRINDROD	GND	1	2.010	2.082	2.124	2.241	2.238	2.145	6.817	2.370	2.159	2.140
Hudaco Industries Ltd	HUDACO	HDC	1	1.948	2.891	3.522	3.524	3.359	3.615	2.987	3.184	2.455	2.217
Howden Africa Holdings Ltd	HOWDEN	HWN	0	2.931	2.757	1.997	2.336	2.247	2.567	-0.670	9.408	-1.341	4.598
Iliad Africa Ltd	ILIAD	ILA	0	2.311	2.272	2.091	1.920	2.059	2.073	2.314	2.352	2.199	2.096
Illovo Sugar Ltd	ILLOVO	ILV	0	2.291	2.014	1.916	1.841	1.803	1.915	1.901	1.949	1.953	1.622
ISA Holdings Ltd	ISA	ISA	0	3.054	1.212	2.036	1.863	2.094	2.187	1.946	1.911	2.046	2.355
Italtile Ltd	ITLTILE	ITE	1	1.456	2.639	2.524	2.660	0.811	2.468	4.094	3.688	3.505	3.460
JD Group Ltd	JDGROUP	JDG	1	1.728	2.526	2.530	2.206	2.160	1.869	2.137	2.485	2.851	2.728

(Appendix 10.6 continued)

Company name	Short name	Code	Distr / non distr	2014	2013	2012	2011	2010	2009	2008	2007	2006	2005
Jasco Electronics Holdings Ltd	JASCO	JSC	1	1.943	1.665	1.768	1.478	1.690	1.819	1.936	2.188	2.072	2.418
KAP Industrial Holdings Ltd	KAP	KAP	0	2.092	2.204	2.776	1.889	1.935	1.844	1.760	2.033	0.873	1.859
Lewis Group Ltd	LEWIS	LEW	0	3.063	3.087	3.149	2.767	2.451	2.527	2.380	2.451	2.538	2.225
Masonite (Africa) Ltd	MASONITE	MAS	1	2.021	2.615	2.452	2.252	2.065	2.638	3.216	2.721	2.531	2.189
Mediclinic International Ltd	MEDCLIN	MDC	0	2.708	2.220	2.762	2.671	3.090	2.958	2.388	2.293	0.672	2.492
Metrofile Holdings Ltd	METROFILE	MFL	0	2.233	2.534	2.653	2.931	3.212	5.904	-1.504	1.054	1.866	1.510
MICROmega Holdings Ltd	MICROMEGA	MMG	0	2.555	1.024	2.690	2.175	1.849	2.110	2.086	2.330	2.537	2.072
MoneyWeb Holdings Ltd	MONEYWB	MNY	1	1.831	1.897	2.225	1.576	1.863	1.737	1.962	1.889	1.737	0.765
Mr Price Group Ltd	MRPRICE	MPC	0	3.479	3.145	2.920	2.914	2.713	2.453	2.241	2.275	2.303	2.259
Massmart Holdings Ltd	MASSMART	MSM	0	3.387	3.896	3.505	3.115	3.275	3.261	4.396	5.287	9.271	2.659
Mustek Ltd	MUSTEK	MST	1	1.807	2.079	1.977	2.005	2.198	1.777	1.855	1.814	1.821	1.821
Metair Investments Ltd	METAIR	MTA	0	2.304	2.171	2.562	2.358	2.391	2.216	1.872	1.940	5.763	5.002
MTN Group Ltd	MTN GROUP	MTN	0	2.347	2.321	2.433	2.326	2.512	2.300	2.266	3.185	14.669	6.200
Murray & Roberts Holdings Ltd	M&R-HLD	MUR	0	2.379	2.167	1.368	1.225	2.108	2.583	2.544	2.533	1.863	1.855
Nictus Ltd	NICTUS	NCS	0	2.677	1.479	1.843	1.582	1.694	1.725	1.474	1.600	1.592	1.457
Nampak Ltd	NAMPAK	NPK	0	2.636	2.293	2.228	2.069	2.275	2.091	2.071	2.164	1.926	1.948
Naspers Ltd	NASPERS-N	NPN	0	3.659	4.215	3.438	3.298	3.163	2.825	3.386	3.039	3.091	3.075
Netcare Ltd	NETCARE	NTC	0	2.387	2.403	0.201	0.036	-0.214	-0.611	-1.413	-2.630	-0.432	1.904
Nu-World Holdings Ltd	NUWORLD	NWL	0	2.241	2.887	1.631	2.272	1.917	2.103	2.024	2.811	2.564	2.618
OneLogix Group Ltd	ONELOGIX	OLG	0	2.290	2.328	2.566	2.441	2.239	2.377	2.261	2.613	2.717	2.910
Omnia Holdings Ltd	OMNIA	OMN	0	3.650	3.541	2.996	2.613	2.344	2.991	2.594	2.604	2.307	2.428
Phumelela Gamings and Leisure Ltd	PHUMELELA	PHM	1	2.066	2.144	2.038	2.048	2.035	1.603	2.916	3.668	3.292	3.399
Pick n Pay Stores Ltd	PICKNPAY	PIK	0	3.540	3.560	3.448	4.678	4.487	-0.346	1.863	2.439	2.367	29.250
Primeserv Group Ltd	PRIMESERV	PMV	0	2.998	2.837	2.656	3.158	0.891	2.735	2.467	2.630	3.032	3.133
Pinnacle Holdings Ltd	PINNACLE	PNC	0	2.307	2.372	2.366	2.421	2.071	2.081	2.099	2.079	2.373	2.198
PPC Ltd	PPC	PPC	0	2.651	2.626	2.553	2.366	2.548	2.444	2.253	2.220	4.456	3.432
RCL Foods Ltd	RCL	RCL	1	1.970	2.622	1.914	2.060	1.970	1.917	2.022	2.303	2.351	1.994
Remgro Ltd	REMGRO	REM	0	3.154	2.960	3.344	2.982	2.800	2.883	3.869	3.194	2.687	2.739
Reunert Ltd	REUNERT	RLO	0	2.484	2.593	2.747	2.732	2.600	2.776	2.575	2.007	2.602	2.511
Rex Trueform Clothing Company Ltd	REX TRUE	RTO	1	1.788	1.274	3.147	2.964	3.316	3.118	2.919	2.466	3.039	2.290
Seardel Investment Corporation Ltd	SEARDEL	SER	1	0.987	1.816	1.688	1.608	1.678	1.388	1.344	1.510	1.888	1.807
Steinhoff International Holdings Ltd	STEINHOFF	SHF	0	3.286	3.359	3.408	3.001	2.719	2.668	3.151	2.541	2.645	1.975
Sovereign Food Investments Ltd	SOVFOOD	SOV	1	2.013	2.283	2.117	2.211	1.918	1.543	2.241	2.985	3.950	2.745
Spanjaard Ltd	SPANJAARD	SPA	1	1.620	1.965	1.818	1.946	1.768	1.671	2.277	1.798	2.047	1.741
Super Group Ltd	SUPRGRP	SPG	0	2.533	2.423	2.747	2.252	2.852	-2.430	4.256	3.319	3.336	2.394
The Spar Group Ltd	SPAR	SPP	1	-3.742	3.303	3.307	3.228	3.166	2.996	2.794	3.014	2.900	2.997
StratCorp Ltd	STRATCORP	STA	1	-0.110	1.190	6.186	1.827	1.503	0.442	1.729	2.405	2.179	1.302
Spur Corporation Ltd	SPURCORP	SUR	0	2.388	2.530	2.515	2.526	2.002	2.495	2.023	2.280	1.980	2.183
SilverBridge Holdings Ltd	SILVERB	SVB	0	2.852	2.459	1.699	2.493	2.288	2.112	1.800	2.368	0.255	2.175
Tiger Brands Ltd	TIGBRANDS	TBS	0	4.063	3.816	3.921	3.875	3.900	3.496	3.684	3.522	3.433	3.177
Tongaat Hulett Ltd	TONGAAT	TON	0	2.911	2.924	2.589	2.414	1.803	0.849	2.131	1.628	2.268	2.101
Transpaco Ltd	TRNPACO	TPC	0	2.373	2.343	2.309	2.602	2.418	2.675	2.024	1.937	2.037	1.976
Truworths International Ltd	TRUWTHS	TRU	0	3.114	3.106	2.853	2.884	2.908	2.710	2.642	2.505	2.151	2.157
Tsogo Sun Holdings Ltd	TSOGO SUN	TSH	0	2.562	2.551	2.466	2.424	2.561	2.792	2.344	2.291	2.481	2.161
Value Group Ltd	VALUE	VLE	0	2.329	2.375	2.110	2.016	2.042	2.106	1.922	1.766	2.076	2.223
Woolworths Holdings Ltd	WOOLIES	WHL	0	3.486	3.144	2.944	2.767	2.847	2.302	2.320	2.365	2.304	2.596
Winhold Ltd	WINHOLD	WNH	1	2.029	1.831	1.885	1.919	2.043	1.870	1.808	2.035	1.966	1.850

APPENDIX 10.7

F-SCORE SAMPLE COMPANIES
Model 5 – Financial and market variables plus lagging business cycle indicator

Identification of distressed and non-distressed sample companies

Company name	Short name	Code	Distr / Non distr	2014	2013	2012	2011	2010	2009	2008	2007	2006	2005
Adapt IT Holdings Ltd	ADAPTIT	ADI	0	2.321	2.314	2.106	1.953	2.197	2.964	2.533	2.356	2.359	2.135
Adcorp Holdings Ltd	ADCORP	ADR	0	4.440	3.655	3.254	3.162	3.178	4.004	3.315	0.996	3.025	3.318
Aveng Ltd	AVENG	AEG	1	1.851	1.893	1.846	1.944	2.266	2.436	3.178	2.723	2.456	2.003
Allied Electronics Corporation Ltd	ALTRON	AEL	0	2.442	2.230	2.111	2.235	2.366	2.408	2.968	2.180	2.329	2.162
AECI Ltd	AECI	AFE	0	2.894	2.830	2.491	2.888	2.869	2.270	2.382	2.453	3.093	2.787
African Oxygen Ltd	AFROX	AFX	1	2.034	2.021	1.952	1.885	1.777	2.331	1.906	2.219	1.814	0.696
AH-Vest Ltd	AH-VEST	AHL	1	2.081	2.160	1.910	2.373	1.853	1.496	2.571	1.917	1.581	1.713
African Media Entertainment Ltd	AME	AME	0	2.555	2.513	2.786	2.395	2.132	2.429	1.113	2.588	3.130	3.379
Astrapak Ltd	ASTRAPAK	APK	1	2.094	1.862	1.760	1.941	2.199	4.524	2.108	2.092	2.283	2.413
Astral Foods Ltd	ASTRAL	ARL	0	3.122	2.353	2.712	3.425	3.221	3.206	3.261	3.709	3.935	3.668
Business Connextion Group Ltd	BCX	BCX	1	2.038	2.065	2.092	1.961	2.027	1.951	2.233	1.755	2.105	1.859
Beige Holdings Ltd	BEIGE	BEG	0	4.548	2.391	2.012	1.934	1.891	2.127	2.020	2.952	2.007	1.852
Bell Equipment Ltd	BELL	BEL	1	2.079	1.741	2.353	1.928	2.133	1.615	2.023	2.116	2.381	1.856
Basil Read Holdings Ltd	BASREAD	BSR	1	1.164	1.728	1.690	1.845	1.915	2.301	2.495	2.605	6.652	2.682
The Bidvest Group Ltd	BIDVEST	BVT	0	4.312	4.065	3.992	3.663	3.592	3.538	3.772	3.619	3.373	3.209
Caxton and CTP Publishers and Printers Ltd	CAXTON	CAT	1	1.886	2.104	2.328	2.052	2.187	2.324	2.122	1.963	2.013	2.209
Crookes Brothers Ltd	CROOKES	CKS	0	2.366	3.724	2.526	1.701	1.488	2.044	2.160	1.924	2.389	5.482
City Lodge Hotels Ltd	CITYLDG	CLH	0	3.118	2.965	2.449	2.279	3.029	2.551	3.142	3.114	2.711	2.497
Clicks Group Ltd	CLICKS	CLS	0	2.729	2.810	2.752	2.945	2.608	2.745	2.604	2.749	2.187	1.985
Cargo Carriers Ltd	CARGO	CRG	0	2.372	1.938	1.962	1.877	2.197	2.025	1.746	2.090	2.296	1.849
Cashbuild Ltd	CASHBIL	CSB	0	3.741	3.076	3.437	2.887	2.998	3.001	3.069	2.692	2.418	2.415
Cullinan Holdings Ltd	CULLINAN	CUL	1	1.702	1.712	1.751	1.864	1.802	1.864	2.016	2.073	2.133	1.690
Datacentrix Holdings Ltd	DCENTRIX	DCT	0	2.192	2.101	2.114	2.440	2.367	2.324	2.561	2.325	2.012	2.621
Digicore Holdings Ltd	DIGICORE	DGC	0	2.795	2.388	1.938	2.288	2.349	1.911	2.604	2.244	2.422	1.913
Distell Group Ltd	DISTELL	DST	0	2.529	2.246	2.316	2.412	2.318	2.247	2.224	2.225	2.024	1.943
Delta EMD Ltd	DELTA	DTA	1	1.773	1.445	2.176	1.784	-0.235	1.982	2.709	1.240	-8.240	1.449
Datatec Ltd	DATATEC	DTC	0	2.496	2.650	2.488	2.181	2.392	2.852	2.707	2.439	2.526	1.999
EOH Holdings Ltd	EOH	EOH	0	9.563	3.391	3.868	3.957	3.264	2.822	2.368	2.347	2.192	6.177
Famous Brands Ltd	FAMBRANDS	FBR	0	2.980	3.359	3.352	7.828	1.431	1.344	25.742	3.076	3.699	-2.455
Gijima Group Ltd	GIJIMA	GIJ	1	1.310	0.698	-2.411	4.255	2.123	2.775	3.237	8.489	2.864	3.317
Grindrod Ltd	GRINDROD	GND	1	1.919	2.032	2.070	2.204	2.238	2.311	7.081	2.423	2.224	2.259
Hudaco Industries Ltd	HUDACO	HDC	1	1.857	2.840	3.469	3.487	3.359	3.781	3.251	3.236	2.520	2.336
Howden Africa Holdings Ltd	HOWDEN	HWN	0	2.840	2.707	1.943	2.299	2.247	2.733	-0.405	9.460	-1.276	4.717
Iliad Africa Ltd	ILIAD	ILA	0	2.220	2.222	2.038	1.883	2.059	2.239	2.578	2.404	2.264	2.214
Illovo Sugar Ltd	ILLOVO	ILV	0	2.200	1.964	1.862	1.804	1.803	2.081	2.165	2.001	2.018	1.740
ISA Holdings Ltd	ISA	ISA	0	2.963	1.161	1.982	1.826	2.094	2.353	2.210	1.963	2.111	2.473
Italtile Ltd	ITLTILE	ITE	1	1.365	2.588	2.470	2.623	0.811	2.634	4.358	3.740	3.570	3.578
JD Group Ltd	JDGROUP	JDG	1	1.637	2.475	2.476	2.169	2.160	2.035	2.401	2.537	2.916	2.846

(Appendix 10.7 continued)

Company name	Short name	Code	Distr / Non distr	2014	2013	2012	2011	2010	2009	2008	2007	2006	2005
Jasco Electronics Holdings Ltd	JASCO	JSC	1	1.852	1.614	1.714	1.441	1.690	1.985	2.201	2.241	2.137	2.536
KAP Industrial Holdings Ltd	KAP	KAP	1	2.000	2.153	2.722	1.852	1.935	2.010	2.024	2.085	0.939	1.977
Lewis Group Ltd	LEWIS	LEW	0	2.972	3.036	3.095	2.730	2.451	2.693	2.644	2.503	2.603	2.344
Masonite (Africa) Ltd	MASONITE	MAS	1	1.929	2.564	2.398	2.215	2.065	2.804	3.480	2.773	2.597	2.307
Mediclinic International Ltd	MEDCLIN	MDC	0	2.617	2.170	2.708	2.634	3.090	3.123	2.652	2.345	0.737	2.610
Metrofile Holdings Ltd	METROFILE	MFL	1	2.142	2.483	2.599	2.894	3.212	6.069	-1.240	1.106	1.931	1.629
MICROmega Holdings Ltd	MICROMEGA	MMG	0	2.464	0.973	2.637	2.138	1.849	2.276	2.350	2.382	2.602	2.190
MoneyWeb Holdings Ltd	MONEYWB	MNY	1	1.740	1.846	2.171	1.539	1.863	1.903	2.226	1.941	1.802	0.883
Mr Price Group Ltd	MRPRICE	MPC	0	3.387	3.094	2.867	2.877	2.713	2.619	2.505	2.327	2.368	2.377
Massmart Holdings Ltd	MASSMART	MSM	0	3.296	3.845	3.451	3.078	3.275	3.427	4.660	5.339	9.336	2.777
Mustek Ltd	MUSTEK	MST	1	1.716	2.028	1.923	1.968	2.198	1.943	2.119	1.866	1.886	1.940
Metair Investments Ltd	METAIR	MTA	0	2.213	2.120	2.508	2.321	2.391	2.382	2.137	1.992	5.828	5.120
MTN Group Ltd	MTN GROUP	MTN	0	2.256	2.270	2.380	2.289	2.512	2.466	2.530	3.238	14.734	6.319
Murray & Roberts Holdings Ltd	M&R-HLD	MUR	0	2.287	2.116	1.314	1.188	2.108	2.749	2.808	2.585	1.928	1.973
Nictus Ltd	NICTUS	NCS	0	2.586	1.428	1.789	1.545	1.694	1.890	1.738	1.653	1.657	1.576
Nampak Ltd	NAMPAK	NPK	0	2.545	2.242	2.175	2.032	2.275	2.256	2.335	2.216	1.991	2.067
Naspers Ltd	NASPERS-N	NPN	0	3.567	4.164	3.384	3.261	3.163	2.991	3.650	3.091	3.157	3.193
Netcare Ltd	NETCARE	NTC	0	2.296	2.352	0.147	-0.001	-0.214	-0.445	-1.149	-2.578	-0.367	2.023
Nu-World Holdings Ltd	NUWORLD	NWL	1	2.150	2.836	1.577	2.235	1.917	2.268	2.288	2.863	2.629	2.736
OneLogix Group Ltd	ONELOGIX	OLG	0	2.198	2.277	2.512	2.404	2.239	2.542	2.525	2.665	2.782	3.028
Omnia Holdings Ltd	OMNIA	OMN	0	3.559	3.490	2.942	2.576	2.344	3.157	2.858	2.656	2.372	2.547
Phumelela Gamings and Leisure Ltd	PHUMELELA	PHM	1	1.975	2.093	1.984	2.011	2.035	1.769	3.180	3.720	3.357	3.517
Pick n Pay Stores Ltd	PICKNPAY	PIK	0	3.449	3.509	3.394	4.641	4.487	-0.180	2.127	2.492	2.432	29.369
Primeserv Group Ltd	PRIMESERV	PMV	0	2.907	2.786	2.602	3.121	0.891	2.901	2.731	2.682	3.097	3.252
Pinnacle Holdings Ltd	PINNACLE	PNC	0	2.216	2.321	2.312	2.384	2.071	2.247	2.363	2.131	2.438	2.317
PPC Ltd	PPC	PPC	0	2.560	2.575	2.499	2.329	2.548	2.610	2.517	2.272	4.522	3.550
RCL Foods Ltd	RCL	RCL	1	1.878	2.571	1.860	2.023	1.970	2.083	2.286	2.355	2.416	2.113
Remgro Ltd	REMGRO	REM	0	3.063	2.909	3.291	2.945	2.800	3.048	4.133	3.246	2.752	2.857
Reunert Ltd	REUNERT	RLO	0	2.393	2.543	2.693	2.695	2.600	2.941	2.839	2.060	2.667	2.629
Rex Trueform Clothing Company Ltd	REX TRUE	RTO	1	1.697	1.223	3.094	2.927	3.316	3.284	3.183	2.519	3.104	2.408
Seardel Investment Corporation Ltd	SEARDEL	SER	1	0.896	1.766	1.634	1.571	1.678	1.554	1.608	1.562	1.953	1.926
Steinhoff International Holdings Ltd	STEINHOFF	SHF	0	3.195	3.309	3.354	2.964	2.719	2.834	3.415	2.594	2.710	2.093
Sovereign Food Investments Ltd	SOVFOOD	SOV	1	1.921	2.233	2.063	2.174	1.918	1.709	2.505	3.037	4.015	2.864
Spanjaard Ltd	SPANJAARD	SPA	1	1.529	1.914	1.764	1.909	1.768	1.837	2.541	1.850	2.113	1.860
Super Group Ltd	SUPRGRP	SPG	0	2.441	2.373	2.693	2.215	2.852	-2.264	4.520	3.371	3.401	2.513
The Spar Group Ltd	SPAR	SPP	1	-3.833	3.252	3.253	3.191	3.166	3.162	3.058	3.066	2.965	3.115
StratCorp Ltd	STRATCORP	STA	1	-0.201	1.139	6.132	1.790	1.503	0.608	1.993	2.457	2.244	1.420
Spur Corporation Ltd	SPURCORP	SUR	0	2.296	2.479	2.461	2.489	2.002	2.660	2.287	2.332	2.045	2.301
SilverBridge Holdings Ltd	SILVERB	SVB	0	2.761	2.408	1.645	2.456	2.288	2.277	2.064	2.420	0.320	2.294
Tiger Brands Ltd	TIGBRANDS	TBS	0	3.972	3.766	3.867	3.838	3.900	3.662	3.948	3.574	3.498	3.295
Tongaat Hulett Ltd	TONGAAT	TON	0	2.820	2.874	2.535	2.377	1.803	1.015	2.395	1.680	2.333	2.219
Transpaco Ltd	TRNPACO	TPC	0	2.282	2.292	2.255	2.565	2.418	2.840	2.288	1.989	2.102	2.094
Truworths International Ltd	TRUWTHS	TRU	0	3.023	3.055	2.799	2.847	2.908	2.876	2.907	2.558	2.216	2.275
Tsogo Sun Holdings Ltd	TSOGO SUN	TSH	0	2.471	2.500	2.412	2.387	2.561	2.958	2.608	2.343	2.546	2.279
Value Group Ltd	VALUE	VLE	0	2.238	2.324	2.056	1.979	2.042	2.272	2.186	1.819	2.141	2.341
Woolworths Holdings Ltd	WOOLIES	WHL	0	3.395	3.093	2.890	2.730	2.847	2.468	2.585	2.417	2.369	2.714
Winhold Ltd	WINHOLD	WNH	1	1.938	1.780	1.831	1.882	2.043	2.036	2.073	2.088	2.031	1.968

APPENDIX 12.1

Appendices not included in the thesis

The following appendices were not included due to space constraints and are available on request.

1. Companies listed on the JSE - target population grouped per sector
2. Numbered target population (assigned sequential and random numbers)
3. Selected sample (random numbers assigned)
4. Selected sample (separated in testing and validation subsets)
5. Inflation-adjusted return on average shareholders' equity
6. Testing and validation subsets separated into financially healthy and distressed groups
7. Financial variables applied to financial distress prediction models from 2000 to 2012
8. PROFITABILITY VARIABLES. Results from the principal component analysis
9. EFFICIENCY VARIABLES. Results from the principal component analysis
10. GEARING VARIABLES. Results from the principal component analysis
11. LIQUIDITY VARIABLES. Results from the principal component analysis
12. MARKET VARIABLES. Results from the principal component analysis
13. QUANTITATIVE NON-FINANCIAL VARIABLES

    Leading, coincident and lagging business cycle indicators (Extract from the South African Reserve Bank, Quarterly Bulletin June 2015)
14. SVM-K-SCORE SUPPORT VECTOR COMPANIES (Period T to period T-5)
15. SVM-Z-SCORE SUPPORT VECTOR COMPANIES (Period T to period T-5)
16. SVM-K-SCORE CONFUSION MATRIX FOR TRAINING SAMPLE (Period T to period T-5)
17. SVM-Z-SCORE CONFUSION MATRIX FOR TRAINING SAMPLE (Period T to period T-5)
18. F-SCORE MODEL 1 SUPPORT VECTOR COMPANIES (Period T to period T-5)
19. F-SCORE MODEL 2 SUPPORT VECTOR COMPANIES (Period T to period T-5)
20. F-SCORE MODEL 3 SUPPORT VECTOR COMPANIES (Period T to period T-5)