Determinants of Corporate Failure: The Case of the Johannesburg Stock Exchange

Mabe, Queen Magadi and Lin, Wei

University of Johannesburg, Ohio University

8 August 2018

Determinants of Corporate Failure:
The Case of the Johannesburg Stock Exchange

W.Lin

Ohio University

Department of Mathematics

linw@ohio.edu

Q.M. Mabe

University of Johannesburg

School of Economics

magadim@uj.ac.za

Abstract: The aim of this paper is to estimate the probability of default for JSE listed companies. Our distinctive contribution is to use the multi-sector approach in estimating corporate failure instead of estimating failure in one sector, as failing companies are faced with the same challenge regardless of the sectors they operate in. The study creates a platform to identify the effect of Book-value to Market-value ratio on the probability to default, as this variable is often used as a proxy for corporate default in asset pricing models. Moreover, the use of Classification and Regression Trees uncovers other variables as reliable predictors to estimate corporate failure as the model is designed to choose the covariates with respect to classification ability. Our study also serves to add to the literature on how Logistic model performance compares to Machine Learning methods such as Classification and Regression Trees and Support Vector Machines. The study is the first to apply Support Vector Machines to predict failure on South African listed companies.

Keywords: Corporate default, Logistic Regression, Support Vector Machines, Classification and Regression Trees.

1. Introduction

Well functioning corporates are essential for economic growth, social welfare and an efficient banking sector. Banks make money from offering credit and the sizeable percentage of the demand of credit is attributable to the corporate sector. Indeed this is so, since the total credit that is extended to the private sector has been on an increasing trajectory between 2010-2015. According to the Reserve Bank

Quarterly Bulletin (2015), total credit extended to the private sector in 2013 was R146 billion, rising to R 192 billion in 2014. It is noted that what drives the demand for credit by the corporate sector is, amongst other things, the need for expansion, increase in floor plans, investments and alternative methods to produce energy. It is a well known microeconomic theory that expansion in production leads to an increase in labour demand, which translates into a reduction in unemployment in real economic terms. If we look closer into microeconomics, where firms function at an internal level, we realise that debt and equity finance is needed to finance projects of firms. This is a notion of capital structure that brings us closer to the literature of corporate finance. The trick in this literature is to strike a balance between equity and debt financing to avoid corporate failure. One cannot avoid debt altogether as it acts as a tax shield (Salubi, 2016).

In this paper, we argue that excessive debt increases firms' financial distress, hence the risk of default or bankruptcy. Corporate failure is costly to the economy since firms that fail negatively affect all their stakeholders. Tserng, Chen, Huang, Lei and Tran (2014) investigate corporate failure in the construction industry and emphasise the important role of the construction sector as the cornerstone of economic development. This sector connects other industries through the construction of buildings and roads, hence Tserng *et al.* (2014) investigated the financial health of construction companies using the method of logistic regression. According to Memic (2015), it is important to predict bankruptcy, as corporate bankruptcy involves a number of stakeholders of the company such as shareholders, credit risk managers, customers, suppliers and bankers. Bandyopadhyay (2006) argues that corporate borrowers who are in bad financial standing also have a negative impact on institutional lenders. He investigates the methods of assessing credit risk of corporate borrowers and evaluates the corporate loan portfolio of a lending institution using a logistic and Multiple Discriminant model to estimate default scores of the portfolio. Karminsky and Kostrov (2014) extensively examine the sustainability of banking institutions and assess the financial health of Russian banks by estimating the probability to default of the banks. Karminsky and Kostrov (2014) argue that failing banks have considerable potential to destabilise the economy, as observed in the 2007-2009 financial crisis. Canicio and Blessing (2014) investigate bank failure in Zimbabwe and concur that bank failure poses serious problems for depositors and the entire economic system, as banks are regarded as a hub for most financial activities.

In our study, we focus on bankruptcy prediction as opposed to corporate default. Corporate default is understood as the failure to meet financial obligations by institutions whereas corporate bankruptcy refers to a corporate that is under liquidation. However, the methods to predict corporate bankruptcy and corporate default/credit scoring are closely related, so much so that a model to predict corporate default or credit scoring can also be used to predict corporate bankruptcy. The difference is observed in the list of predictors that are used, however, the framework remains the same.

1.1 Literature on Corporate Bankruptcy

One of the earliest studies estimating corporate bankruptcy is that of Altman (1968), where the author bridges the gap between individual ratio analysis to assess the financial health of a company and statistical methods to predict financial health. Altman (1968) used Working Capital/Total Assets, Retained Earnings/Total Assets, EBIT/Total Assets, Market Value per Equity/Book Value of Total Debt and Sales/Total Assets as variables to predict corporate bankruptcy by means of a z-score. The ratio analysis method of discriminant analysis proved to be very efficient in predicting bankruptcy. In Altman's study, the ratio Market value per equity/Book value of total debt explanatory power proved to be more significant in predicting bankruptcy and was positively related to the z-score. According to Altman (1968), companies that are non-bankrupt will tend to have a relatively higher z-score. Ohlson (1980) collected the data of 105 companies, of which 18 were bankrupt, and used a number of accounting ratios as the predictors of bankruptcy. The main differentiator between Ohlson (1980) and Altman (1968) is that the former uses a probabilistic model as opposed to calculating a bankruptcy score. Beaver (1966) emphasises the importance of financial ratio analysis as the predictor of company failure but cautions that one needs to be careful as other financial ratios are weak in predicting failure. In his analysis, Beaver (1966) established that the cash-flow to debt ratio is efficient in predicting failure whereas liquidity ratios have weak classification power. Beaver (1966) analyses one ratio at a time and suggests that the model could be robust in a multi-ratio setting. Wilcox (1971) develops a theoretical model to explain the empirical findings of Beaver (1966) and concludes that the ratio of companies' wealth to net cash-flow is useful in predicting corporate bankruptcy.

Brîndescu-Olariu (2015) investigates 1,420 Romanian companies and concludes that debt-ratio has limited influence on default risk but cautions that the method used is within univariate settings and indicates that debt ratio may influence default risk to significant extent within a multivariate setting. Brîndescu-Olariu (2015) uses debt ratio as a single predictor of default rate and tests the performance of the model by means of calculating the area under the receiver operating characteristic curve, generally known as ROC curve. Rafiei, Manzari and Bostanian (2011) investigate the models to classify companies as bankrupt or non-bankrupt. In a multivariate discriminant setting where the z-score is calculated, Rafiei *et al.* (2011) control for Working Capital/Total Assets, Operating Income/Total Assets, Owners' Equity/Total Assets and Operating Income/Interest Expenses. The variable of interest is Owners' Equity/Total Assets, which measures the proportion of assets that are either financed by equity or debt. The coefficient of this variable is significant and shows that capital structure or leverage has an influence on predicting bankruptcy.

Clearly, the methods to predict corporate bankruptcy have evolved over the years since the work of Altman (1968), Beaver (1966) and Ohlson (1980). Recently there has been a proliferation of statistical and mathematical credit models which are virtually innumerable. Logistic regression and Multiple

Discriminant Analysis are very popular credit scoring models within the statistical framework (Lee , *Chiu, Chou & Lu* ., 2006). However, these methods are often criticised for their unrealistic assumption about the distribution of the data. Eisenbeis (1977) concurs with this and further suggests the use of quadratic functions instead of the current use of linear discriminant functions. Eisenbeis (1977) cautions that the *a priori* probabilities about the independent variables and the cost of classifications of Discriminant Analysis are questionable. Lee *et al.* (2006) note that Neural Network models are also becoming popular as they take into account the non-linearity features that are observed in the data. The disadvantage of this methodology is the need for large training data sets and interpretative difficulties Thomas and Edelman (2002) also classify Neural Networks, Linear Programming, Integer Programming, Genetic Algorithms and Expert Systems as non-statistical methods or mathematical methods. Hand & Henly 1997 also note that the credit markets experience increased competition as a result of more sophisticated mathematical and statistical methods having been developed. In their study. Lee et al (2006) draw our attention to Neural Networks and asset that the model is often critisised for its many local minima and maxima which pose a problem in optimization. To overcome the challenges that are associated with Neural Networks, Cortes & Vapnik (1995) introduce Support Vector Machines. Cortes & Vapnik (1995) introduce SVM. SVM intuitive idea is to optimize a strictly concave objective function to ensure the existence of global maxima. Cortes & Vapnik (1995) established that the accuracy of Support Vector Machines is better than that of other classical classification methods like Neural Networks and Decision Trees. Huang et al. (2004) investigates the determinants of corporate bond default by using a number of artificial intelligence techniques and also establish that Support Vector Machine does better than Logistic Regression and Neural Networks. Lee et al. (2006) suggest the use of Classification and Regression Tree (CART) and Multivariate Adaptive Regression Spline (MARS) as alternative methods. They assert that their models take into account complex relationship between the explanatory characteristics without making unrealistic assumptions. Moreover, the CART and MARS models do not require large training data sets.

Notable corporate failure studies recently carried out in South Africa include those of Kidane (2004) and Bruwer and Hamman (2006). Kidane (2004) focuses only on telecommunication companies and uses the z-score and Spring-gate score methods to predict failure. According to Kidane (2004), the z-score model is still relevant and easy to understand. However, the first shortcoming of the study is that it does not incorporate recent methods that have been discovered to improve on the Altman z-score model. Eisenbeis (1996) explores different methodologies of estimating corporate default and advises that researchers should strive to apply new methodologies in order to exploit opportunities that are inherent in such models.

The second shortcoming is the low performance rate of the methods used in the study as they failed to predict bankruptcy in the telecommunication sector of South Africa. Kidane (2004) attributes the dismal performance to the fact that the model was applied within the South African context of a different economic environment and industry. Bruwer and Hamman (2006) argue that it is better to use the variables that have proved to be significant in predicting failure as opposed to using the trial and error method of prediction. The cash-flow from operations to sales ratio and size proved to be significant in predicting bankruptcy.

Bruwer and Hamman (2006) use the CART model in estimating corporate bankruptcy and argue that this model performed reasonably well. The gap in the study is to claim that CART performs 'reasonably well' without having to compare it to more conventional models. Lee *et al.* (2006) compare CART to Neural Networks, Multiple Discriminant Analysis and Logistic Regression and establish that CART outperforms all the other models. However, Nie *et al.* (2011) compare two classification models, namely, Logistic Regression and the Decision Tree and arrive at the conclusion that Logistic Regression outperforms the Decision Tree. Kukuk and Ronnberg (2013) concur that the logit model is normally used to predict bankruptcy and credit default. As a result, we will use the classical Logistic Regression methodology as a method of reference and compare it to the CART and SVM model to establish if we can settle the argument of CART vs Logistic Regression vs SVM within a South African multi-industry setting. Eisenbeis (1977) maintains that the *a priori* assumptions made about the distribution of independent variables hamper the performance of the classical discriminant analysis. As a result, we propose using CART and SVM as these models are non-parametric in nature. Moreover, it is noted that the assumption of the linear classification function does not hold water in practice as classification functions are highly variable insofar as they maybe quadratic as opposed to linear (Eisenbeis 1977). SVM overcomes this problem by projecting data to an alternative dimension if an appropriate discriminating function cannot be found in the existing dimension. The study is the first of its kind to use SVM within the JSE setting.

Our study further aims to predict bankruptcy across a variety of sectors in the JSE, namely, mining, consumer goods, telecommunication, electronics and industrial. The reason why our study cuts across a variety of sectors is that we observed in data collection that failing companies have common attributes across all sectors. These characteristics are high book-to-market ratio, falling and low stock prices and persistent negative profits whereas thriving and continuing firms have the opposite characteristics. Since such attributes are common in both financially healthy and financially distressed firms regardless of the sector in which they operate, we assert that there is no need to limit the study to just one sector. Indeed, this is so as John, Hilscher and Szilagyi (2017) also note that firms which are more likely to default have experienced losses over time, have cash-flow difficulties and a high leverage ratio. It remains to be seen whether the models will make the correct predictions subject to the multi-sector data.

It is important to note that the distinctive contribution of our study is that it takes a multi-sector approach; some sectors, such as the financial, real estate or investment sectors, delist from the JSE as a result of scheme arrangements and not necessarily financial distress or failure. These sectors decide to delist without manifesting the typical characteristics of financially distressed firms. Therefore, classifying them as failed firms will hamper the classification ability of the models. However, Bruwer and Hamman (2006) define failure as the result of delisting on the stock exchange, which has a great potential to distort the discriminant ability of the model. Furthermore, the splitting rules of Bruwer and Hamman's (2006) model are not detailed insofar as the thresholds are simply presented as constants without theoretical rigour and there is no clear definition of the stop splitting rule of the terminal nodes.

In our analysis we are explicit and precise on how we define failure to minimise the error of misclassification. Eisenbeis (1977) clearly identifies the ambiguity in the definition of the classes as one of the pitfalls of discriminant analysis. We define failure as delisting on the JSE as a result of liquidation, not scheme arrangement. Moreover, the definition of splitting rules is clearly defined and backed up by financial concept. What further differentiates our study from those of Kidane (2004) and Bruwer and Hamman (2006) is the inclusion of the book-to-market (B/M) ratio and the leverage ratios as predictors of failure. The reason why the B/M ratio stands out in our study is that Fama and French (1995) explicitly use the B/M ratio as a proxy for financial distress and argue that firms with high B/M ratios are more likely to fail. Mathematically speaking, a high B/M ratio corresponds to significantly low stock price relative to book value. We deduce that a high B/M ratio also signals mispricing and how the company is perceived in the market by investors. Lyandres and Zhdanov (2013) begin by developing a theoretical model and establish that firms with high investment prospects are less likely to default. The high investment prospect is measured by the B/M ratio, and it is established that this ratio is significant (Lyandres & Zhdanov 2013). Kousenidis and Negakis (2000) concur with Fama and French (1995) and Lyandres and Zhdanov (2013) and use Athens Stock Exchange data to prove that a portfolio consisting of low B/M firms outperforms one with a high B/M ratio. This means that firms with low B/M have bright investment prospects as investors are inclined to purchase their stocks, which channels investments into such firms, thereby avoiding failure.

The reason why leverage ratio is also emphasised is that financially distressed firms are characterised by high leverage ratios as they are far away from their optimal debt ratios (Gilson 2009). Furthermore, the South African Insolvency Act, 1936, explains that agents are referred to as insolvent when liabilities exceed assets, hence the importance of leverage ratio as a determinant of financial distress.

The rest of this paper is divided into three sections. The first section provides the rationale for the variables used in the study and gives a tabled overview of how we arrived at our variables with respect to the literature. The second section describes the theoretical models that are used in the study while the final section details the empirical analysis and the challenges encountered during the data collection stage.

2. Variables Used in Studies on Financial Distress or Corporate Bankruptcy

There are countless variables that are used in bankruptcy studies. However, instead of picking any variable that we come across in the literature, we selected variables according to their level of significance in prior studies, which is a proxy for discriminating ability. Other studies suggest combining market data with accounting data (Charalambakis 2015). Tinoco and Wilson (2013) combine accounting ratios with market and economic variables and establish that accounting ratios are significant when supplemented by market and economic data. Table 1 below refers to the literature we consulted in selecting the variables.

Table 1: Variables from prior studies

Number of studies	Variables	Study	Level of Significant	Method Used
1	Profitability:Net Profit/ Average Assets, or		yes, but not stated	Back-Propagation Neural Network
	net profit/average share- holder's equity	Lee & Sung (2013)	yes, but not stated	
	Productivity: fixed assets/fixed liabilities		yes, but not stated	
	Liquidity: Liquid assets/total assets		yes, but not stated	
	Asset Quality: total loans/total assets		yes, but not stated	
2	**Book Leverage and Profitability**	Lyandres & Zhdanov (2013)	1%	Logistic regression
3	**Market value/Book-value of debt**	Altman (1968)	1%	Multivariate Discriminant Analysis
	Working capital/Total Assets		1%	
	Retained Earnings/Total Assets		1%	
	EBIT/Total Assets		1%	
4	Size	Ohlson (1980)	5%	Conditional Logit
	Total Liability/Total Assets		5%	
	Net Income/Total Assets		5%	
	Joint Current Liquidity- Working Capital/Total Assets		5%	
	Current Liabilities/Current Assets			
5	**Book-value**	Charalambakis (2016)	5%	Z-Score
	Profit before tax/Current liabilities		1%	
	Current Assets/Total Liabilities		1%	
6	Operating income/Total Liabilities	Tinoco & Wilson 2013	5%	Logit Regression
	Total Liabilities/Total Assets		5%	
	Retail Price Index		5%	
	Short term Bill rate		5%	
	Market Capitalisation		5%	
7	Size	Bruwer & Hamman (2006)	N/A	Recurssive Partitioning/CART
	Working Capital/Total Assets		N/A	
	Accounts Receivable/Total Revenue		N/A	
	Cashflow from operating activities		N/A	
	Cumulative cash flow from operating activities		N/A	

In our analysis, we use the variables of Altman (1968) and Ohlson (1980), namely, Current Assets/Current Liabilities, Net Profit/Total Assets and Total Liabilities/Total Assets. According to Berk and Demarzo (2013) a company need not default as long as the market value of its assets is greater than its total liability, hence the ratio of Total Liabilities to Total Assets should theoretically be less than 1 to keep the company within a safe zone. We augment these variables with a cash-flow variable, Cash-flow from Operating Activities/Total Assets. The cashflow variable is an important factor in predicting default probability because companies with reliable cashflow are less likely to default (Berk & Demarzo 2013). Moreover, market variables such as B/M ratio and share price are variables which we include to set our model apart.

3. Methods

In the following we assume that our data set is $\{(x_i, y_i)| \ i = 1,2, \dots \dots, n\}$ drawn from the population (X, Y), where the input covariate vector $X = (X_1, \dots, X_p)$ is p-dimensional and correspondingly $x_i = (x_{i1}, x_{i2}, \dots . x_{ip})$ is the i^{th} observed input covariate vector, $i = 1, 2, \dots, n$. Note that X_j denotes the j^{th} input variable.

3.1 Support Vector Machine

The support vector machine (SVM) is a method that uses linear decision boundaries for classification, leading to an optimal separating hyperplane between two classes, say class A and class B. Specifically, suppose our data set is:

$$\{(x_i, y_i) \mid i = 1, 2, \dots, n\}$$

Where x_i are the input covariate vectors and y_i are the categorical responses that take values 1 for say, class A and -1 for class B. An SVM estimator, or classifier, denoted be $\hat{C}$(x) with input x, is given by the sign of the linear function $f(x) = \beta_0 + \beta \acute{x}$ denotes the inner product of the vectors β and x. If $f(x) > 0, \hat{C}(x) = sign\,(f(x))$ takes the value 1 and predicts the response to be in class A; and $if\ f(x) < 0$ then $C\hat{}(x) = sign\,(f(x))$ takes the value -1 and predicts the response to be in class B. Here the hyperplane $f(x) = 0$ is called the separating hyperplane because it separates the responses into two classes, depending on which side of the hyperplane the input vector x falls into.

An estimator of (β_0, β) is usually obtained as the solution to one of many equivalent forms of an optimisation problem. One such optimisation form is to minimise the norm ||β|| with respect to (β_0, β) subject to the conditions

$$Y_i(\beta \acute{x}_I + \beta_0) \geq 1 - t_i, \varepsilon_i \geq 0, i = 1, 2, \dots, n \qquad (1.1)$$

$$\textstyle\sum_{i=1}^{n} \varepsilon_i \leq r_1 \,. \qquad (1.2)$$

Here the norm $||\beta||$ is defined through $||\beta||^{\,2} = \beta \acute{\beta}$. The parameter r_1 in (1.2) is a tuning parameter that controls the locations of so-called margins which are hyperplanes parallel to the optimal hyperplane $f(x) = 0$ at a distance of $||\beta|| - 1$. Only observations within the margins as well as misclassified observations contribute to the estimators $\hat{\beta}$ and $\widehat{\beta_0}$.The larger the tuning parameter r_1, the wider these margins will be.

To see why we take the classifier $\hat{C}(x) = sign\ (f(x))$, it is easier to look at an equivalent optimisation form. The above optimisation problem is equivalent to minimising

$$\sum_{i=1}^{n}[(1 - y_i)f(x_i)]^+ + r_2||\beta||^2 \quad (1.3)$$

with respect to (β_0, β). This is a penalised loss where the first part $\sum_{i=1}^{n}[1 - y_i f(x_i)]^+$ serves as a loss and the second part $r_2||\beta||^2$ serves as a penalty for larger values of ||β||. Here the notation $a^+ = max\ (0, a)$ takes the value 0 if $a < 0$ and is equal to a if $a \geq 0$. The parameter r_2 in (1.3) is a tuning parameter that plays a similar role as r_1 in (1.2). Let us examine the expected loss

$$E\ [(1 - y\,f(x))^+ \mid x] = (1 + f)^+ [1 - p(x)] + (1 - f)^+ p(x) \quad (1.4)$$

where $p(x) = P(y = 1|x)$. When $f(x) \geq 1$, the above is reduced to $[(1 + f(x)] [1 - p(x)]$ which is minimised when $f(x) \geq 1$ with minimum $2[1 - p(x)]$; when $f(x) \leq -1$, (1.4) is reduced to $[1 - f(x)]\, p(x)$ which is minimized when $f(x) = -1$ with minimum $2p(x)$. Finally when $-1 \leq f(x) \leq 1$, (1.4) is reduced to

$$[1 + f(x)][1 - p(x)] + [1 - f(x)]\, p(x) = 1 + f(x)\, [1 - 2p(x)],$$

$2[1 - p(x)]\ \ if\ \ p(x) > \frac{1}{2}$. In summary, the expected loss is minimised at $f(x) = 1$ when $p(x) = P\ (y = 1|x) > \frac{1}{2}$ in which case we should categorise the response into class A since its chance is higher than $\frac{1}{2}$, and when $p(x) < \frac{1}{2}$, we have $P(y = -1|x) = 1 - p(x) > \frac{1}{2}$ and the expected loss is minimised at $f(x) = -1$ which is when we should categorise the response into class B. Simply put, the sign of $f(x)$ indicates whether the chance for y = 1 is higher if $(f(x) > 0)$ or lower if $(f(x) < 0)$ than for $y = -1$.

Either tuning parameter, r_1 in (1.2) or r_2 in (1.3), can be selected using, for example, a cross-validation (CV) method. In a general SV method, the features (or input variables) x can be replaced by $h_j\ (x)$ for proper transformations h_j of the original variables x using different kernel functions. The number of transformed features can be much higher than the dimension p of the input covariate vector x. For more detail, see the example (Hastie, T., Tibsharani, R. & Friedman, J. , 2009).

3.2 Logistic Regression

The Logistic Regression, for two classes, models the logit (log of odds) as a linear function of the input covariates. In general, with K classes for the response variable, a logistic model assumes that

$$log\frac{P(y=k|x)}{P(Y=K|x)} = \beta_{k0} + \beta_k{}'x, \quad k = 1,2,\dots,K - 1. \qquad (1.5)$$

Here β_{k0} are scalars and β_k are p-dimensional vectors that match the dimension of the input covariate vector x. It is easy to see that under the above model assumption (1.5),

$$P(Y = k|x) = \frac{e^{\beta_{k0}+\beta'_k x}}{1+\sum_{J=0}^{K-1} e^{\beta_{j0} + \beta'_j x}}, \; k = 1,2,\dots,K - 1 \qquad (1.6)$$

And

$$P(Y = k|x) = \frac{1}{1+\sum_{J=0}^{K-1} e^{\beta_{j0} + \beta'_j x}} \qquad (1.7)$$

Typically the parameters $\theta = \{\beta_{k0}, \beta_k | k = 1,2,\dots,K\}$ are estimated by the maximum likelihood method. The log-likelihood function is given by

$$l\,(\theta) = \sum_{i=1}^{n} log(\prod_{k=1}^{K} P_k(x_i)\, I(y_i = k) \qquad (1.8)$$

where $p_k(x) = P(Y = k|x)$ are as given in (1.6) and (1.7). Here $I\;(y_i = k)$ is an indicator function which takes the value 1 if the condition is true, i.e. if $y_i = k$, and the value 0 otherwise. The maximum likelihood estimator (MLE) $\hat{\theta}$ maximises $l\,(\theta)$ in (1.8) and we can use it to estimate the probabilities $p_k(x)$ for $k = 1,2,\dots,K$. Once we have these estimators $\hat{p}_k(x)$, the prediction for the category that the response falls into is one that maximises $p\hat{}_k(x)$. Namely,

$$\hat{y} = \arg \underset{1\le k\le K}{\text{MAX}} \; \widehat{p_k}\;(x).$$

Asymptotic results from generalised linear models (GLM) can be used for statistical inferences on the MLE $\hat{\theta}$ as well as model or variable selection. For more detail see

3.3 CART

Another popular method in data analysis is the Classification and Regression Tree (CART). The basic idea of tree-based methods is to partition the feature space (i.e. the domain of the input covariate vector x) into boxes and make the average responses in these boxes.

Specifically, following Hastie *et al.* (2009), at any step of the tree-growing process, input variable in the input covariate vector $X = \{X_1, \ldots, X_p\}$. To find (X_j, s) at the initial step, let

$$R_1(j,s) = \{t = (t_1, \ldots, t_p) \in \mathbb{R}p \mid t_j \leq s\},$$

and

$$R_2(j,s) = \{t = (t_1, \ldots, t_p) \in \mathbb{R}p \mid t_j > s\}.$$

In other words, the space $\mathbb{R}p$ is divided into two big boxes, one with the jth component no more than s and the other larger. Let c_k be the average response of the observations falling inside the box R_k, k = 1, namely,

$$c_k = \frac{\sum_{i=1}^{n} Y_i \, I\,(X_i \in R_k(j,s))}{\sum_{i=1}^{n} Y_i \, I(X_i \in R_k\,(j,s))}, \quad k = 1, 2. \tag{1.9}$$

Note that the notation x_i in (1.9) refers to the input covariate vector of the i^{th} observation which is different from X_j we used above for the j^{th} input variable. We choose the pair (X_j, s) that minimises the sum of squared errors given by

$$SSE\,(X_j, s) = \sum_{i=1}^{n} \sum_{k=1}^{2} I\left(x_i \in R_k(j,s)\right)(y_i - c_k)^2, \tag{2}$$

over all choices. Here $I\,(x_i \in R_k(j,s))$ takes the value 1 if observation x_i falls into the k^{th} box $R_k(j,s)$, and the value 0 otherwise.

Once the pair (X_j, s), called a node, is found, we split the data into two branches according to the box they fall into. Then for each branch, we repeat this splitting process to grow the tree until, for example, a certain number of (terminal) nodes has been found. If we stop splitting a branch, this branch is called a terminal node, otherwise an internal node. This certain number, say m of terminal nodes of the grown tree, is our tuning parameter. If m is too small, the grown tree will not fit the data well. If, on the other hand, m is too large, it may lead to overfitting the data.

To choose the proper tree size, let T denote the grown tree and |T| its size, i.e. the number of its terminal nodes. Then we choose a tree that minimises a penalised SSE, given by

$$C_\alpha(T) = \sum_{k=1}^{|T|} \sum_{X_i \in R_k} (y_i - \hat{c}_k)^2 + \alpha|T| \tag{3}$$

where R_k is the box corresponding to the k^{th} terminal node of the tree and $\widehat{c_k}$ is the average response of those observations in RK. The parameter α determines the level of penalty we put on the tree size.

A higher value of α leads to a smaller tree and we can use a cross-validation method to select α. Note that SSE is not the only measure of errors we can use here. For classifications, we may use other measures such as the misclassification errors of deviances. For more detail, see Hastie *et al.* (2009).

Finally, once a tree is selected, we may categorise an observation by a majority vote. First, the box that the input covariate vector x belongs to must be found; then the largest group of responses in this box must be assigned to $\hat{y}$. For example, in a 3-class model, if 30% of the responses in the box is in class 1, then 10% is in class 2 and 60% in class 3, then set $\hat{y}$ = 3.

4. Empirical Analysis

The data of this model was collected from INET BFA. The data was then cleaned as it had commas and Excel recognised numbers as strings. Once the data was collected, we used the balance sheets and income statements of every company to compute accounting ratios. There were 79 continuing firms and 17 failed firms. The data collected ranged from 2007-2015 for both continuing and failed firms. We used the data between 2011 and 2015 as the testing data. Of the 17 failed firms, 7 were reserved for out-of-sample testing. Of the 79 companies that were continuing, 29 were used for testing.

4.1. Support Vector Machines

The coefficients of SVMs are estimated by explanatory variables, covariates and Lagrange multipliers; these are called support vectors. We used cross-validation to assist in choosing the model parameters, i.e. the parameter that allows us to choose the distance between the two separating hyperplanes and the misclassification magnitude.

Table 2: SVM Prediction Matrix for Training Dataset

Y	Predicted: Continuing	Predicted: Failed
Actual: Continuing	50	0
Actual: Failed	0	10

SVMs tend to fit the training data very well but fail to fit the testing data as detailed in Table 3. The results of the testing data set are detailed in Table 4.

Table 3: Prediction Matrix of the Testing Dataset

Y	Predicted: Continuing	Predicted: Failed
Actual: Continuing	28	1
Actual: Failed	6	1

The model failed on the testing dataset as it misclassified six failed firms as continuing firms whereas they had failed.

4.2 Logistic Regression

The variables that we used to predict the leverage impact on probability to default for the JSE-listed companies include Current Assets/Current Liabilities, Net Profit/Total Assets and Total Liabilities/Total Assets and Cash-flow from Operating Activities/Total Assets, where the dependent variable is the probability to default, or $P(y_i = 1|x_{ij})$ for company i . In estimating the logistic model, we ran a series of regressions with weak explanatory power until we got to the significant model that is detailed in Table 4.

Table 4: Logistic Regression Results

	Dependent Variable
	Probability of Failure
Variables	
D/E	0.071*
	(0.035)
LEVR	0.24
	(0.19)
B/M	0.041*
	(0.0244)
CFOA/TA	-0.413
	(0.43)
PRICE	-0.073***
	(0.014)

All the variables included in the model came out as predicted by economic theory. For instance, holding everything else constant, a percentage increase in the D/E ratio will lead to a 7.1% increase in the probability of failure. Likewise, all else equal, a percentage increase in the B/M ratio will lead to 4.1% increase in the probability of failure. This is consistent with the theory of Fama and French (1995) who use B/M ratio as a proxy for financial distress. Indeed, as per our model, we also observe a significant positive relationship between B/M ratio and probability of failure. The results of the cash-flow and price variables are not surprising as companies with good cash-flow and high stock prices are less likely to fail. These results are consistent with the assumption that we explicitly made at the initial stage of the study. The prediction matrix for the training data sets is detailed in Table 5 below.

Table 5: Prediction Results on Training Dataset

Y	Predicted: Continuing	Predicted: Failed
Actual: Continuing	48	2
Actual: Failed	2	8

According to the results in Table 5, logistic regression misclassifies four continuing firms as failed, and 1 failed firm as continuing. The results for the testing data set are summarised on Table 6.

Table 6: Prediction Results on Testing Dataset

Y	Predicted: Continuing	Predicted: Failed
Actual: Continuing	28	1
Actual: Failed	3	4

In the testing dataset, the model incorrectly classified one continuing firm as failed and three failed firms as continuing.

4.3 Classification Trees

The CART model gives five splitting variables, namely, CFOA/TA, Debt/Equity ratio (D/E), Price, Total Assets/Total Liabilities (TA/TL) and Current Liabilities/Current Assets (CL/CA). The model arrives at the splitting rule using a majority vote process. For instance, for the price variable, we have that (Price, 1.905), if the price is greater than 1.905 then the model will decide on assigning class 0 to a company based on the fact that the majority of companies that is continuing has a price that is greater than 1.905 threshold. The rest of the model is designed the same way using the variable that has a significant classification ability. The prediction matrix for CART is detailed in Table 7 below.

Table 7: Results of the Training Dataset

Y	Predicted: Continuing	Predicted: Failed
Actual: Continuing	49	1
Actual: Failed	1	9

CART performs fairly well on the training dataset as observed in Table 7. The prediction for the testing data set is summarised in Table 8.

Table 8: Results of the Testing Dataset

Y	Predicted: Continuing	Predicted: Failed
Actual: Continuing	29	0
Actual: Failed	3	4

In the testing dataset, the model incorrectly classified two continuing firms as failed and two failed firms as continuing. The tree of the CART model is illustrated in Figure 1.

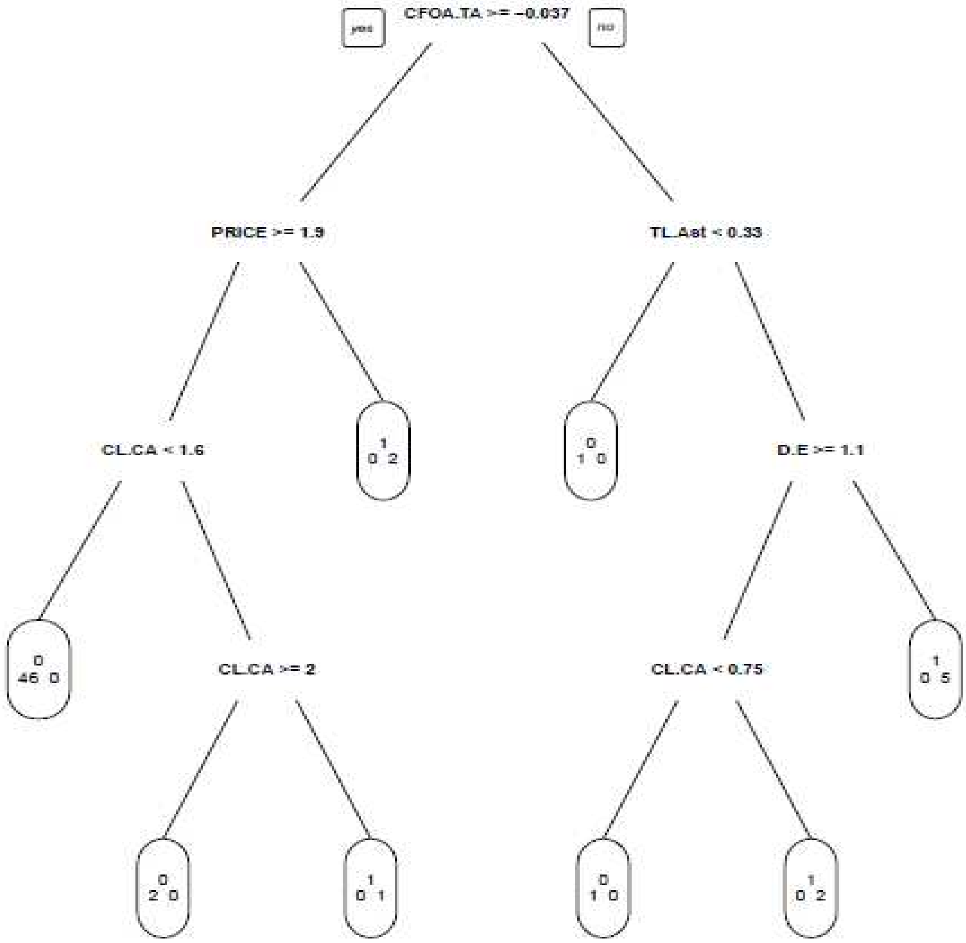

Figure 1: CART Model

In the figure above, 1 represents a failed firm and 0 represents a continuing firm. For instance, at the root, if $\frac{CFOA}{TA} \geq -0.037,$ then a company whose price is smaller than 1.9 will fail. Within the terminal node, we have 0 firms that belong to class 0, and 2 firms that belong to class 1. The rest of tree is interpreted the same way.

According to our observation all the models fit the training dataset very well and are challenged in out-of-sample testing. Table 9 below summarises how all the models compare in both the training and testing datasets.

Table 9: Summary of all Models' Performance

	Training Dataset		**Testing Dataset**	
	Accuracy		Accuracy	
	Continuing	Failed	Continuing	Failed
Model				
SVM	100%	100%	97%	14%
CART	100	100%	100%	57%
LOGISTIC	92%	86%	100%	57%

According to Table 9, SVM makes correct predictions for all the classes in the training dataset. However, in the testing dataset, SVM incorrectly classifies 86% of failed companies as continuing. The CART model and SVM achieve 100% accuracy on the training dataset whereas the logistic model scores the least on the training dataset. The CART and the logistic model outperform SVM on the testing dataset. Our results are consistent with those of Lee *et al.* (2004), as the authors predicted the CART model to be the best from among the Logistic Regression, MARS and MDA models. The results are also comparable with Bruwer and Hamman (2006), as their CART model also identified cash-flow variable to have a classification ability. As far as performance is concerned, CART performed better than the logistic model on the training dataset and was on par with the logistic model on the testing dataset. However, on the training dataset, our results are in conflict with those of Cortes & Vapnik (1995) who predicted that SVM would perform better than CART. Our study established that both CART and SVM make accurate predictions on the training dataset. Our study is the first of its kind to apply the SVM method on the South African dataset.

5. Conclusion

The objective of this paper is to identify variables that predict the probability of failure in South African JSE-listed companies. Our study is set apart insofar as it predicts failure in a multi-industrial setting as we are convinced that both failing and continuing firms have similar attributes, hence there is no need to limit the study to just one sector. Our hypothesis turned out to be valid because in the results we observe that all the classification models maintained their classification ability within the mult-industry frame-work. The other contribution of our paper is to include the B/M ratio that Fama and French (1995) used as a proxy for financial distress in the asset pricing theoretical model. Hence, the study created a platform for the B/M ratio to be empirically tested on its effect on the probability of default or financial distress. In achieving our results, we used different methods to predict failure. Amongst the methods that we used was the CART model, which proved to be the most effective as it performed well in out-of-sample testing data. Moreover, the SVM model managed to perform very well in the training dataset, and the model failed to perform in the testing dataset. Our study is the first to apply SVM on the South African listed companies. In the study, we also observed that Artificial Intelligence models are effective in carrying out economics research. For instance, the CART model searched for the variables that have a discriminant power and used these variables to solve the classification problem. Hence managers should always be on the look-out for cash-flow, stock price, liquidity and leverage ratio as these accounting variables need to be managed to ensure the going concern of a company in any industry. The limitation of our models is that we only used a small sample of failed companies, and one can be forgiving to the models that failed to perform in the testing dataset as there were no enough observations from which our models could learn the structure and attributes of failing companies. An area for future research is to incorporate macroeconomic variables to observe if these variables have an impact on predicting the probability of failure for companies.

References

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

Bandyopadhyay, A., 2006. Predicting Probability of Default of Indian Corporate Bonds: Logistic and Z-Score Model Approaches. *The Journal of Risk Finance*, 7(3), pp.255–272. Available at: ban.

Beaver, W.H., 1966. Financial Ratios As Predictors of Failure. *Journal of Accounting Research*, 4(1966), pp.71–111.

Berk, J. & Demarzo, P., 2013. *Corporate finance* Third Edit., New York: Pearson.

Brîndescu-olariu, D., 2015. The Potential of the Debt Ratio in the Prediction of Corporate Bankruptcy. *Journal of Public Administration, Finance and Law*, 2013(2), pp.37–46.

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

Canicio, D. & Blessing, K., 2014. Determinants of Bank Failures in Multiple-Currency Regime in Zimbabwe (2009-2012). *International Journal of Economics and Finance*, 6(8), pp.229–246.

Charalambakis, E.C., 2015. 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), pp.1–28.

Cortes, C. & Vapnik, V., 1995. Support-Vector Networks. *Machine Learning*, 297(973), pp.273–297.

Eisenbeis, R.A., 1977. Pitfalls in the Application of Discriminant Analysis in Business, Finance, and Economics. *The Journal of Finance*, 32(3), pp.875–900.

Eisenbeis, R.A., 1996. Recent developments in the application of credit-scoring techniques to the evaluation of commercial loansf. *IMA Journal of Mathematics Applied in Business and Industry*, (September 1995), pp.271–290.

Fama, Eugene F & French, K., 1995. Size and Book-to-Market Factors in Earnings and Returns. *The Journal of Finance*, 50(1), pp.131–155.

Gilson, S.C., 2009. Transactions Costs and Capital Structure Choice : Evidence from Financially Distressed Firms. *The Journal of Finance*, 52(1), pp.161–196.

Hand, D.J. & Henly, W.., 1997. Statistical Classification Methods in Consumer Credit Scoring : a Review. *Journal of Royal Statistical Society*, 160, pp.523–541.

Hastie, T., Tibsharani, R. & Friedman, J., 2009. *The Elements of Statistical Learning* Second., Springer-Verlag New York Inc.

John, Y., Hilscher, J.D. & Szilagyi, J., 2017. *Predicting Financial Distress and the Performance of Distressed Stocks*,

Karminsky, A.M. & Kostrov, A., 2014. The probability of default in Russian banking. *Eurasian Economic Review*, 4(1), pp.81–98.

Kidane, H.W., 2004. *Predicting financial distress in IT and services companies in South Africa*. University of the Free State.

Kousenidis, D. V & Negakis, C.I., 2000. Size and book-to-market factors in the relationship between average stock returns and average book returns : some evidence from an emerging market. *The European Accounting Review*, 9:2(1968), pp.225–243.

Kukuk, M. & Ronnberg, M., 2013. Corporate credit default models: A mixed logit approach. *Review of Quantitative Finance and Accounting*, 40(3), pp.467–483.

Lee, T.S. , Chiub . C, Chou, Y., & Lud C., 2006. Mining the customer credit using classification and regression tree and multivariate adaptive regression splines. *Computational Statistics and Data Analysis*, 50(4), pp.1113–1130.

Lyandres, E. & Zhdanov, A., 2013. Investment opportunities and bankruptcy prediction. *Journal of Financial Markets*, 16(3), pp.439–476.

Memic, D., 2015. Assessing Credit Default using Logistic Regression and Multiple Discriminant Analysis:Empirical Evidence from Bosnia and Herzegovina. *Interdisciplinary Description of Complex Systems*, 13(1), pp.128–153.

Nie, G., Wei, R, Lingling. Z., Yingjie. T., & Yong. S., 2011. Expert Systems with Applications Credit card churn forecasting by logistic regression and decision tree. *Expert Systems With Applications*, 38(12), pp.15273–15285.

Ohlson, J.A., 1980. Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research*, 18(No.1).

Rafiei, F.M., Manzari, S.M. & Bostanian, S., 2011. Expert Systems with Applications Financial health prediction models using artificial neural networks , genetic algorithm and multivariate discriminant analysis : Iranian evidence. *Expert Systems With Applications*, 38(8), pp.10210–10217.

Salubi, I.L., 2016. Corporate Borrowing and Tax Shield Among Listed Companies in Nigeria. *Journal of Academic Research in Economics*, 8(2), pp.239–252.

South African Reserve Bank, 2015. SARB Quartaly Bulletin. Available at: www.sarb.co.za

Tinoco, M.H. & Wilson, N., 2013. International Review of Financial Analysis Financial distress and bankruptcy prediction among listed companies using accounting , market and macroeconomic variables. *International Review of Financial Analysis*, 30, pp.394–419.

Tserng, H.P. , Chen, P. , Huang, W, Lei, M. C., & Tran, Q. H., 2014. Prediction of default probability for construction firms using the logit model. *Journal of Civil Engineering and Management*, 20(2), pp.247–255.

Wilcox, J.W., 1971. A Simple Theory of Financial Ratios as Predictors of Failure. *Journal of Accounting Research*, 9(2), pp.389–395.