Factors that Reduce the Probability of Loan Default by Survival Analysis of Zimbabwean SMEs

Tan Zhongming¹, Liberty Munashe Fungurai¹, Ding Guoping¹
¹School of Management, Jiangsu University, 301 Xuefu Road, Zhenjiang, Jiangsu, P.R. China
Corresponding Author: libertymunash1@gmail.com

ABSTRACT
This study focuses on modelling the probability of Small to Medium Enterprises (SMEs) defaulting payment in the event of them receiving loans. Under investigation were the factors that affect both default and survival of businesses since the default probability was based on survival analysis. Data drawn was analyzed using Cox Regression, a semi-parametric survival technique whose aspects of hazard and survival functions were a base for the analysis. The results interpreted on the basis of the fitted hazard ratios and overall Cox Proportional Hazards Model indicated that there is an inverse relationship between probability of survival and that of default. As reflected by the outcome of this study, businesses that have more time in operation are less likely to default payment. Entrepreneurs possessing self-employment, work experience and training qualifications, have a positive impact on survival of the businesses they are to start and run. In addition, the size of a loan and location of a business have no significant effect on the hazard (risk) associated with lending to SMEs. Implications of the study, limitations and future research directions are also discussed.

Key words: loan, default, survival, Zimbabwe, SME

INTRODUCTION
On a scholarly point of view, Small to Medium Enterprises (SME) are considered to be a group of people, all encompassed under a formally registered firm (regardless of size) operating as a profit making organization in any field of interest. On the other hand, SMEs can be defined as businesses whose personnel numbers fall below certain limits. The abbreviation “SME” is used in the European Union and by international organizations such as the World Bank, the United Nations and the World Trade Organization. However, mainly basing on the Zimbabwean economic state, with the issue of indigenization now being prominent, an SME as defined by the corporate sector is considered to be either an individual or group of individuals formally registered to operate independent from the formal sector of employment with the solemn aim of profit making. Basing on the latter definition of an SME, as indicated by the January 2013, Zimbabwean Monetary Policy statement, it is evident that approximately 90% of Zimbabweans are informally employed, 75% of this population being classified under SMEs, which clearly indicates that at present, SMEs are considered to be the backbone of the Zimbabwean economy. On the contrary, the attention channelled towards SMEs in Zimbabwe is very minimal which is rather distressing, regarding the fundamental role played by SMEs in the Zimbabwean economy. A recent survey by (FBC Security), as published in the Zimbabwean Independent,(2013) indicated that there are approximately 22 banks and roughly 100 Micro-finance institutions still operating in Zimbabwe. Of the entire Banking population, only about 30% of these Banks have considered or are still considering offering financial assistance to SMEs, which lands first as the major drawback in SMEs flourishing, lack of financing. This subsequently has a negative effect
on the growth of the Zimbabwean economy, considering that if SMEs perform well then high chances are the economy will likewise flourish. Due to simple structures of most SMEs which gives them flexibility, it is easier for SMEs to quickly respond to the constantly changing economic conditions so as to meet the local customer demands which would enable them to probably develop into bigger, better corporations. However, to a greater extent, the Banking and/or Micro-finance Institutions are justified for not extending their financial services to SMEs, which they branch under the Sub-Prime Market, a credit that is lent to people of questionable or limited credit histories as defined by Investopedia. The sub-prime market includes the business of sub-prime mortgages, sub-prime auto loans and sub-prime credit cards, as well as various securitization products that use sub-prime debt as collateral. Due to lack of collateral and mainly challenges in liquidity, from a Credit Risk perspective, SMEs are highly risky, which explains the forfeit in Banks and Micro-finance Institutions in lending to these. On the other hand, as Dietsch and Petey,( 2004) after analyzing German and French SMEs, concluded to say that, ‘SMEs are though riskier than larger businesses, have a lower asset correlation with each other.” Despite the risk associated with lending to SMEs, it is crucial that the lending practice be considered open to such markets as this might have a positive influence on the country’s economy, then the question that remains is, ‘how will return be maximized yet this practice is associated with very high risks”.

Decision making structures and processes, used by Zimbabwean Banks and MFIs to deal with the money lending process are mainly confined to the formal sector of employment, which constitutes a smaller percentage of the entire Zimbabwean population. However, money lending to the informal employment (sub-prime market) remains an area that very few banks and MFIs are willing to tap into as a result of high risk and uncertainty of retaining profits due to the lack of collateral and asset liquidity problems associated with this sector of operation. In as much as SMEs could be considered backbone of the Zimbabwean economy, most of them, regardless of promising viability end up either failing to grow and develop into bigger businesses or dying a natural death due to lack of financing. To the banking and Micro-finance Institutions, the sub-prime market can be a highly profitable one considering that they as the lenders can demand high interest rates but the problem now is that the high profits are dependent on the borrowers’ capability to repay their loans. Banks mainly determine viability of any SME judging from their business proposals or foresights as well as the physical check on the premises to be used and targeted market. However, this has proved problematic since the method of check lacks accuracy. The broad objective of this study is to model the probability of default of SMEs in Zimbabwe based on an analysis of their survival. However, the specific objectives are as follows;

1. Forecast the probability with which the SME/s will default payment in the event of lending using the survival analysis technique.
2. Identify the nature of the relationship between survival and default in the event of lending.
3. Make policy recommendations inclined on lowering risk of default based on the findings.

In pursuit of a new trajectory of accelerated economic growth and creation of wealth for the nation, the Zimbabwean Government in October 2013 formulated a plan known as the Zimbabwe Agenda for Sustainable Socio-Economic Transformation (Zim Asset). Moving with the notion, “Indeginize, Employ, Empower, Develop”, the Zim Asset scheme is to run till the year 2018. Considering that SMEs make up the bulk of the Zimbabwean economy, working towards their nourishment would be a giant leap towards full filling the vision of fighting unemployment through indeginization in Zimbabwe. However, the greatest challenge being faced by these SMEs is with the lack of financial aid which resultantly is due to the fear of Banking and Micro-finance institutions risking closing down due to failure linked to high levels of default associated with lending to SMEs. This study therefore seeks to bring an intervention to this setback, maintaining both counterparts, the Financial institutions as well as the SMEs running. This would considerably reduce the unemployment levels in Zimbabwe as many would be encorporated into the SME sector. Likewise the Zimbabwean economy would be affected positively through empowerment of SMEs leading to its development and with time stability.

2. LITERATURE REVIEW

Numerous factors have been identified in various studies as having an impact on credit management and loan repayment. Several factors such as interest rates, age, marital status, location and numbers of dependents are said to impact on the likelihood of default Lodha(2011). Some of these factors are as discussed below. The pioneering work of Stiglitz and Weiss (1981) cited by Godquin(2004) marked the beginning of attempts at explanations of credit
rationing in credit markets. They asserted that ‘... interest rates charged by a credit institution are seen as having a dual role of sorting potential borrowers (leading to adverse selection), and affecting the actions of borrowers (leading to the incentive effect)”. Weinberg (2006) advocated that interest charged and the amounts of debt are the two main factors affecting repayment obligations. Some banks use the interest rates that an individual is willing to pay as a screening device to identify borrowers with a high probability of repayment.

**Gender in Credit Management**

Studies endorse gender as a variable that could influence credit management practices among SMEs. Halkias (2008) pointed out that there is still a significant and systematic gap between gender in relation to business ownership and entrepreneurial involvement. Evans and Winston(2008) concurred with Halkias (ibid) that single, college educated women managed their credit more prudently than both men in general and married women, in a study conducted in Ghana. A number of important gender issues are recognized in terms of investigating successful SME development in Africa.

**Loan Size in Credit Management**

Godquin (2004) reported that both age and size of loans have an inverse relationship to repayment performance. This concept is related to a study done by Pang (1991) cited by Chong(2010) who pointed out that the main determinants of repayment obligations are the interest charged and the amount of debt. Furthermore, loans that are too big also lead to repayment problems, dissatisfaction and high dropouts Hietalahti and Linden(2006).

**Loan Period in Credit Management**

The loan period or term of a loan is usually classified as either short-term or long-term. A short-term loan in bank parlance is one that is repayable within a period of one year. A long-term loan on the other hand, is any loan with payment terms extending beyond one year. Although the relationship between loan maturity and borrower risk has been addressed in some theoretical models (Ortiz-Molina and Pe-nas, 2004), there is very little observed research that tests these theoretical models in the context of bank lending to small firms Berger and Frame(2005). Bragg (2010) asserted that "the short time frame reduces the risk of non-repayment to the bank, which can be reasonably certain that the business’s fortunes will not decline so far within such a short time period that it cannot repay the loan, while the bank will also be protected from long-term variations in the interest rate.

**Location in Loan Repayment**

Some studies consider various factors such as location as a determinant of business success and the performance of loan repayment Kang et al.(2005). McPherson (1995) cited by Rogerson, (2000) attested to this in a study conducted about key determinants of the survival rate of SMEs. The results indicated that businesses in commercial districts exhibit high success in comparison with the high failure rate experienced by home-based enterprises. In addition, soft information like distance between the borrower and the lender is important. A larger borrower-lender distance is associated with higher default risks because distance interferes with information collection.

**Age and family Circumstances of an Entrepreneur in Credit Management**

Cromie (1991), in a study of male and female owners of young firms, found that businesses managed by young people experience general management problems such as lack of people management and accounting skills. Age and the family circumstances of owners can negatively or positively affect the performance of the business. Small business owners with a supportive, experienced family structure tend to be able to cope with the pressure of running the business. Godquin (2004) reported that both age and size of loan have an inverse relationship to repayment performance. Athmer and De Vletter (2006) added that seventy per cent of defaulters in their study samples experienced a family problem such as death or health circumstances.

**Education and Training in Credit Management**

There is an indication of a positive link between flourishing SMEs in South Africa and education and training. The World Bank (1993) endorsed this concept by showing a direct correlation between sales and education in South African SMEs. The World Bank’s investigation concluded that entrepreneurs "who have achieved a Grade 10 of education have average turnover nearly twice that of those who have completed a Grade 8 level”. In an exploration of the determinants of success in a sample of emerging black-owned manufacturing SMEs in the Western Cape, Sawaya (1995) cited by Rogerson (2000) concluded that "the rate of success
was highly correlated with the level of education attained by the owner”.

**Sector of Business in Credit Management**

Mead and Liedholm (1998) pointed out that survival rates of small businesses vary by sector. The study concluded that enterprises in the service sector and manufacturing are less likely to close down than those in the wholesale and retail sector (ibid).

**Cash Flow Management in Credit Management**

Chong (2010) identified capacity (sufficient cash flow to service the obligations), colateral (assets to secure the debt), character (integrity), condition of the economy as well as capital (net worth) as needing to be included in the credit scoring model. The credit scoring model is a classification procedure in which data collected from application forms for new or extended credit line is used to assign credit applicants to "good" or "bad" credit risk classes, compared with enterprise start-ups Constantinescu, Badea, Cucui & Ceausu(2010).

**EMPIRICAL LITERATURE**

**Default Prediction Methodologies**

Literature pertaining to the techniques that can possibly be used to predict default is very substantial. A large number of authors studied and analyzed several possible methodologies that can be used to predict customers’ and/ or businesses’ default.

**Accounting-based Model**

Accounting ratios are widely used by banks in a bid to limit adverse selection and moral hazard problems in loan advancements. The methodology of the accounting-based approach is based on Multiple Discriminant Analysis (MDA) and logistic models that are the most useful in accounting based variables for classifying company default. Khorasgani (2009) argued that although there are numerous drawbacks to using accounting ratio-based models in predicting defaults, SMEs’ financial ratios derived from balance sheets and profit and loss accounts are regarded as good predictors of default. In addition, liquidity and activity are the most crucial factors in predicting an SME’s default, as well as the positive effect of age and size variables on an SME’s default prediction.

**Multivariate Discriminate Analysis**

This model is used when the researcher is already aware of the various classification groupings of the sample Fujikoshi et al.(2010). It serves to then classify the sample according to the classification rules. Mathematically the Multivariate Discriminate Analysis Model can be written as:

\[ Z = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_n X_n \]

Where: \( \beta_1, \beta_2, \beta_3 \) and \( \beta_n \) are the discriminant coefficient and \( X_1, X_2 \) up to \( X_n \) are the independent variables for analysis. Altman (1968) used a multiple discriminant analysis technique (MDA) to solve the inconsistency problem linked to the Beaver’s (1966) uni variate analysis and to assess a more complete financial profile of firms. His analysis drew on a matched sample containing 66 manufacturing firms (33 failed and 33 succeeded) that filed a bankruptcy petition during the period 1946-1965. Altman examined 22 potentially helpful financial ratios and ended up selecting five as providing in combination the best overall prediction of corporate bankruptcy. The variables were classified into five standard ratios categories, including liquidity, profitability, leverage, solvency and activity ratios. For many years thereafter, MDA was the most common statistical technique that was used in default prediction models. Quite a number of authors, the likes of (Deakin (1972), Edmister (1972), Blum (1974), Eisenbeis (1977), Taffler and Tis-shaw (1977), Altman et al. (1977), Bilderbeek (1979), Micha (1984), Gombola et al. (1987), Lussier (1995), Altman et al. (1995)) applied this technique too. However, in most of their studies, authors pointed out those two basic assumptions of MDA are often violated when applied to the default prediction problems. Moreover, in MDA models, the standardized coefficients cannot be interpreted like the slopes of a regression equation and hence do not indicate the relative importance of the different variables.

**Survival Analysis**

Unlike the above mentioned methodologies which can be implemented in determining the probability of corporate default, Survival Analysis is a statistical technique that more closely analyses the conditional probability of corporate default. The major difference between the methodologies explained earlier and this technique is that, the previously mentioned studies take not into consideration the continual existence of the firm/s in question. They made use of the inspection of other variables in determining whether or not the business will default in a given period of time. Survival is therefore the condition for determining default in this technique.
Survival-based Credit Scoring Model
Some banks and MFIs take the screening process prior to lending to another level by making use of the survival analysis method to measure response or time of an occurrence of an event. Luoma and Laitinen (1991) pointed out that the aim of the survival analysis method is to measure the link between illustrative variables and survival. Investigating the timing when customers are likely to go “bad” is important for effective credit management policies. The bank can manage and monitor profitability of clients to the bank over a customer’s lifetime. It has been shown previously by Narain (1992:109) and Banasik et al. (1999) that survival analysis can be useful to estimate default and repayment. Depending on the nature of factors influencing the time to an event, survival analysis can be non-parametric, semi-parametric or parametric. In this research, since the covariates (factors influencing the time to default of a firm) are both finite and infinite dimensional, then a semi-parametric survival analysis technique will be used to analyze the data collected in this research.

The Cox-Proportional Hazards Model
The Cox Proportional Hazards Model also known as the Cox Model is said to be the most widely used model in multivariate analysis of survival data. Just like it is with the Logit and Multiple Discriminate Analysis models, the Cox model applies the concepts of multiple regression to assess the relationship between independent variables influencing default and the to be predicted, non-metric dependant variable. However, unlike the major drawback of the Logit and MDA models in analyzing survival data, the cox model takes into account the aspect of survival time, a concept not covered by the prior mentioned models. This makes the Cox model preferable over the Logit and MDA models in the analysis of survival data, which is the centre of focus in this research. In addition, the cox model is also capable of predicting future loan performances as well as the likelihood of borrowers defaulting future loan payments as is done by the Credit scoring and markov models respectively. The only difference is that, the cox model goes a step further into not only predicting the performance of the loan but actually stamping a time frame to failure. In other words the cox model works towards predicting the period for which a loan will be performing which makes it more appropriate for the analysis of data in this research. In general, the Cox-Regression model can be represented as follows:

$$h(t, x) = h_0(t) \exp(x' \beta)$$

### Importance of the Cox regression Model
The CPH model is a popular survival analysis technique. Listed below are some of the reasons why this model became popular:

1. Robustness: The Cox model is a $0$-safe $0$ choice of a model in many situations.

2. Due to the form of the model:

$$h(t, x) = h_0(t)^e^{x' \beta}$$

the estimated hazards are always non-negative.

3. Even though $h_0(t)$ is unspecified we can estimate the $\beta$'s and hence compute the hazard ratio.

4. $h(t, X)$ and $S(t, X)$, the hazard and survival functions respectively can be estimated for a Cox model using a minimum of assumptions.

5. In survival analysis the Cox model is preferred to a logistic model since the latter takes not into account survival times as well as censoring information.

### METHODOLOGY

#### Data Collection
Secondary data is to be used for this research. Most of the data used in this study was obtained from the Zimstats Central Business Report which basically summarized everything as far as Zimbabwean businesses are concerned. Drawn for analysis was a mixed sample of 50 both survived and failed businesses from the period 2009-2014 is to be taken. The reason behind the period selected is that most SMEs have a short life span such that by the end of the two year period, most, if not all businesses would have failed. Most of the analysis of data is to be carried out using Statistical Package for Social Sciences (SPSS). One of the main reasons for selection of this statistical package is that it is one of the few that offers a platform to carry out analysis in Cox Regression which is to be used in this study. In addition, the package is considered to be user friendly. For the organization of outcomes in the form of graphs and tableaux's, Microsoft Excel is to be
used. One major advantage of Excel is that, it has an instantaneous graphing of data facility which makes it ideal for use in graphical presentation of results.

**Classification of Variables**
Dependant variable Time to default Independent variables Covariates, from X1 to X11. For conclusion purposes after analysis, the dependant variable is to be depicted by; (1 = survived; 0= failed). The independent variables are to be divided into 1. Business Specific Characteristics -Variables X1, X3, X4, X5, X10 and X11 2. Human Capital Characteristics Variables X7 and X8 3. Control variables Variables X2, X6 and X9

**General Description of Data Samples**
After classification of data, the samples are to be represented as follows Table 3.1 below presents the selected variables and the description of the categorical and continuous variables used in the analysis.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time to default</td>
<td>Business reported to have serviced loan until 2016 (failed to default = 0, defaulted = 1)</td>
</tr>
</tbody>
</table>

**Model Specification**
This research focused specifically on the semi-parametric type of survival analysis. This technique pulls out no assumption about the shape of hazard function, but makes assumptions on how covariates affect the hazard function, for example in Cox regression which will be used to surf data in this research. The Cox Regression procedure is useful for modeling the time to a specified event, based upon the values of given covariates.

In Cox Regression,
1. One or more covariates are used to predict a status (event).
2. The central statistical output is the hazard ratio.
3. Data contains censored and uncensored cases.

**Cox proportional hazard model Assumptions**
The following are underlying assumptions for the use of CPH model in this research:
1. The hazard rate is constant over time or similarly, the hazard for one individual is
proportional to the hazard for any other individual, where the proportionality constant is independent of time item Non-informative censoring

2. The design of the underlying study must ensure that the mechanisms giving rise to censoring of individual subjects are not related to the probability of an event occurring.

Model Formulation
The Cox proportional hazards model is also referred to as the Cox model, the Proportional hazards model or the Relative risk model. It incorporates a parametric modeling of relationship between the failure rate and specified covariates. By letting

\[ x_i = (x_{i1}, x_{i2}, \ldots, x_{ip}) \]

denote a collection of \( p \) explanatory variables that affect survival time, the hazard function for the Cox proportional hazards model is, to however have a positive hazard function without constraints on the coefficient values, the CPHM is given as; The model can be broken down into the following components;

- Hazard Ratio - The econometric measure of effect in the Cox regression involving only vector coefficients of \( \beta \).
- Maximum Likelihood Estimates - Estimates of the \( \beta \)s
- Baseline hazard function- represented by \( h_0(t) \)

\[ h_0(t) = h_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k) \]

where; \( h_0(t) \) is the baseline hazard.

\[ h(t, x) = h_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k) \]

\[ h(t) = h_0(t) \exp(\beta_1 x_1) \]

where; \( h_0(t) \) is the hazard ratio.

Empirically, the CPH model based on the standard proportional hazards model is given by;

\[ \ln H(t) = \ln H_0(t) + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k \]

The duration in the model is defined by: \( (t_0, t) \) and \( d \) where; \( t_0 \) - is the start date, \( t \) - The end of the observation window and \( d \) - Represents failure In this study, \( t_0 \) is the starting date for observing the funded businesses, \( t \) - time at which observation ends and \( d \) represents date at which firms are reported to have permanently exited the market. Since the incomplete nature of the observation occurs in the right tail of the time axis, such observations are said to be right censored. Our censoring indicator, represented by \( d \) is: 0, Right Censored, Business survival =1, Failure or market exit

Model Validation
Prior to fitting the model, it has to be ascertained that the drawn data set is indeed fit for analysis using the
selected model, Cox Proportional Hazards model in this case. In as much as the Cox Regression model is a survival analysis technique, it can also be considered as somehow a form of a multiple regression model. For this reason, apart from the assumptions of the model itself, it is necessary to test if the assumptions of regression analysis are met. Henceforth, the following assumptions will be tested for using SPSS. If all assumptions are met then indeed the CPHM is suitable for data analysis purposes but if not, then the converse is true.

1. The absence of heteroscedasticity amongst the residual values.
2. The absence of multicollinearity amongst the residuals.
3. Independence of residuals.
4. Normality of distributions of the residuals.

After all the stated procedures would have been undertaken, if all the assumptions have been met, the next step in this study would be to carry out the data analysis. However, given at least one of the assumptions is violated, suitable transformations are to be made to the data in order to make it suitable for analysis using Cox Regression.

Use of R-squared in Cox Regression
In survival analysis, measuring the predictive power of a model helps quantify the ability of prognostic factors to predict the time to an event, default in this case. The generalized R-squared is calculated as:

$$R^2 = 1 - \frac{X_{LR2}}{n}$$

Where;
- $X_{LR2}$ is the chi-square statistic for the likelihood ratio test for the overall model and
- $n$ = total number of businesses under investigation.

Although the generalized R-squared is recommended for Cox regression, its sensitivity to the proportion of censored values is not mentioned. However, the expected R-squared value decreases substantially as a function of the percentage censored values. For instance, with heavy censoring of say 50% the R-squared values can decrease by 20% or more.

DATA ANALYSIS AND RESULTS
The Cox Proportional Hazard Model

From the outcome of the data analysis, it was observed that some of the covariates whose effect on the time to default was being investigated did not have a significant effect on our variable of interest. This explains why the fitted CPH model shown below comprises of some, not all of the covariates included as part of the analysis. Basing on the CPH empirical formula provided by equation 3.4 the fitted model would be:

$$h(t, X) = h_0(t) \exp(-13.165X_1 - 0.3602X_2 + 0.523X_3 - 2.238X_4 - 1.149X_5 - 0.710X_6 + 1.134X_7 + 1.723X_8)$$

Where;
- $X_1$ is the variable for survived businesses.
- $X_2 = \text{business age}$
- $X_3 = \text{Education level}$
- $X_4 = \text{Work Experience}$
- $X_5 = \text{Entrepreneurial Training}$
- $X_6 = \text{Owner Contribution}$
- $X_7 = \text{Business Sector 1}$
- $X_8 = \text{Business Sector 2}$

Results from the test of Assumptions
Before the results of the CPHM Regression are presented, outcomes from the tested assumptions for use of this technique will be given. This stands as a justification for the use of Cox Regression, a survival analysis technique that was selected for analysis of data in this study. Cox Regression can be classified as multiple regression and hence before the technique can be used, a process of checking to ensure that the data can actually be analyzed using multiple Regression must be carried out. Below are the results on the test of all six assumptions necessary for use of Multiple Regression.

Test for Heteroscedasticity
In fig 4.1 below is a scatter plot diagram of time to default against all the other covariates that were selected for the CPH model.
It can be observed that the plots (error terms) are scattered and do not follow a particular pattern. This implies that the data exhibits homoscedasticity. Hence it is fit for analysis using multiple regressions.

**Test for Multicollinearity**

Provided in table 4.1 below, are the results for the tests carried out to check for multicollinearity in the data for this study. The interpretation of the results is also given. Under the column of collinearity statistics, it can be seen that all the tolerance values are well above 0.2. This is enough evidence to confirm that the data does not exhibit multicollinearity. In addition, all the VIF values are less than 10 for all covariates giving an average VIF value of 1.3 which is not substantially greater than 1. This is also evidence supporting that the data does not show traces of multicollinearity which makes it suitable for analysis using Cox regression.

**Test for Independence of observations (residuals)**

The outcome from the test of independence of residuals was given as follows

<table>
<thead>
<tr>
<th>Model</th>
<th>R</th>
<th>R Square</th>
<th>Adjusted R Square</th>
<th>Std. Error of the Estimate</th>
<th>Durbin-Watson</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.499</td>
<td>0.639</td>
<td>-0.036</td>
<td>16.771</td>
<td>2.358</td>
</tr>
</tbody>
</table>

The value of the Durbin Watson statistic of 2.358 as shown in table 4.2 above is greater than 2 but less than 4. Hence we fail to reject the null hypothesis that states that there is no autocorrelation. This means that there is no correlation between successive values of the same variables. We can therefore conclude that there is independence between the residuals which makes our data appropriate for analysis using CPH regression.

**Test for Normality**

In regard to the normality test on the residuals (errors), the outcome was as follows

![Figure 4.1: Test for heteroscedasticity](image1)

**Figure 4.1: Test for heteroscedasticity**

![Figure 4.2: Normality Test](image2)

**Figure 4.2: Normality Test**

**The Survival Function**

The graph in fig 4.2 below gives a general outline of the survival function. This graph was obtained by plotting all the covariates under investigation against survival, which in this case was considered to be the dependent variable.

![Figure 4.3: Survival Function](image3)

**Figure 4.3: Survival Function**

The survival function is a decreasing function in that, as the time frame increases, the probability of survival will be decreasing. Likewise, the probability of survival of a business decreases as the age of the business increases. In mathematical terms, the probability of survival is inversely proportional to a business’s age.

**The Relationship between survival and default**

As indicated in fig 4.3, the graph of the survival function outlines the relationship between business survival and default. The result is then explained below the graph. This graph was also plotted in the same way as the survival graph above, only difference being that for the current, analysis was carried out under the title of hazard functions.
From the results above, failed businesses have a lower survival curve than survived businesses. This means that businesses that fail have a shorter time to default when compared to those that survive. In other words, failed businesses have a higher risk of default in the event of loaning. This is highlighted by the hazard function curves in the fig below:

The hazard curve of survived businesses shown in fig 4.5 is lower than that of failed businesses which implies that failed businesses have a higher risk of default than survived businesses in the event of loaning. In general, if a business survives and continues to operate, it is less likely to default payment if it is to be given a loan. Mathematically, survival can be said to be inversely proportional to default.

**CPHM Regression Results**

At the end of this section, the Cox Proportional Hazards Model is estimated and the effects of the selected covariates on the outcome of interest (time to default) will be presented. However, focus will be given to business survival which was also selected as one of the covariates in estimating time to default. Firstly, the effects are to be explained in terms of the signs of the regression coefficients. A positive sign means that the hazard (risk of failure) is higher and hence the prognosis, that is probability of default in this case also higher for businesses with an increased proportion or higher value of that variable. Furthermore, the effects will be represented in the form of Hazard Ratios, given as $\text{Exp} (B)$ from the analysis. As stated by Box-steffensmier et al.(2004), when the CPH model estimates take the form of a hazard ratio (HR), interpretation of the results is such that: If the hazard ratio is less than one, an increase in the variable associated with the coefficient reduces the risk , thus resulting in longer survival time. In contrast, a hazard ratio greater than one portrays that, if the variable associated with the given coefficient estimate increases, the risk also increases leading to reduced survival time. Lastly, if the calculated hazard ratios are closer to one, the resulting implication is that the hazard rate is non-responsive to changes to the covariate, or the variable of interest has no influence on the increase or decrease of the hazard. Table 4.1 shows the coding for the categorical variables selected for estimation of the CPH model

**Interpretation of the Coefficient Estimates**

**Business Age**

In relation to business age, the negative regression coefficient indicates that an increase in the value of this variable will successively decrease the hazard, in this case risk of failure and default of the business. Collaborating it with the hazard ratio which is $\text{exp} (-3.60)$, this implies that a business that is one year older than the prior has a 0.697 less chance of defaulting payment in the event of loaning. This is probably due to the reason stability and consistency in performance exhibited by older, more established businesses as compared to the newly established ones. There is enough evidence supporting this claim at the 5% significant level.

**Loan Size**

With a hazard ratio of 0.741 which is closer to 1 and a 95% confidence interval that includes 1, this shows that loan size has no significant difference on the survival of a business. In other words, this covariate has no effect on the increase or decrease of the hazard and hence it can be concluded that the hazard rate is non-responsive to changes in the loan sizes. This could be because, regardless of a loan being big or small, depending on the performance of a business, default is equally likely to occur in both instances. Giving a small loan to a well performing business can decrease the hazard whereas the converse will increase the hazard hence the issue is not with the size of the loan but proportions of the size of loan to performance of the business.
However, there is no significant evidence supporting this hypothesis.

**Loan Tenure**

Regarding to loan tenure which is the repayment period of a loan, long-term loans were found to have a higher risk of default as compared to short-term loans. From the results, the hazard rate of 0.077 indicates that increasing the loan tenor, counter increase the risk of default 2.161 times. This can further be supported by positive regression coefficient of this variable. Related literature shows that short-term loans have a lower default record when being compared to long-term loans.

**Work Experience**

Businesses run by individuals possessing a past employment experience have a lesser risk of market exit as compared to those run by amateurs in business. Evidence to support this claim is highlighted by the resulted hazard ratio of 0.107 which implies that businesses owned by entrepreneurs with a higher work experience are more likely to survive by 10.7% as compared to those run by entrepreneurs with no or little past working experience. The argument behind this outcome may be due to the reason that entrepreneurs without employment experience struggle to keep up with the competition in business as they would be new to the strategies and techniques necessary for survival. However, this is in contradiction with Mannathoko’s (2011) view that if entrepreneurs possessing employment experience struggle in business they would not be motivated to stay since they have an option of moving to paid employment.

**Self employment**

Regarding self employment characteristics of business owners, the results indicate that entrepreneurs with prior-self employment have approximately a 49% chance of staying in business as compared to those who do not possess self-employment experience. This is supported by van Praag’s (2003) view that experience in the same industry as the business venture gives a better chance of survival. Unlike the line of argument put forward by Mannathoko (2011), the results of this study highlight that having experience in running some sort of business in any sector gives an individually considerable capability in running another business in any other sector, regardless of a change in environment.

**Entrepreneurial Training**

Generally, in relation to entrepreneurial training at start-up, businesses run by owners who went through entrepreneurial training at start-up of the businesses have a 31.7% chance of survival as compared to businesses run by untrained owners. This could be due to the reason that business owners who would have under-gone training have the knowledge as to how a business should be run. Apart from the determination required to run a business, knowledge on the various aspects of business is also required to stay afloat. Mannathoko (2011) however argues against this view saying that entrepreneurial training does not guarantee business success but rather there are various other critical success factors in running a thriving business.

**Owner Contribution**

The hazard rate of $\text{Exp}(-0.710)$ which was the outcome in regard to owner contribution to capital at start-up indicates that a business in which the owner has a made a contribution at start-up is 0.492 times more likely to survive than the converse. This argument could be due to the reason that if an entrepreneur contributes towards start-up of a business, he or she is more likely to want to protect their investment as much as they can. With such an aim, such an entrepreneur will pump in efforts to safeguard their investment through continual existence and operation of the business. There is significant evidence to support this claim at the 10% level.

**Education level**

In regards to the education level attained by an entrepreneur, the results of this study imply that businesses run by owners possessing a secondary level of education are 1.688 times more likely to survive as compared to those run by owners carrying a post-secondary level of education qualification. This could be based on the argument that furthering one’s studies does not make one a good entrepreneur like it is said that, “Entrepreneurs are born not made.” Advancement in studies only equips one with the theoretical aspect (knowledge) to owning and running a business, it does not specifically give them capability to act likewise. There is significant evidence to support this claim at the 5% level.

**Entrepreneur Age**

Some say age is nothing more than just a number but the results of this study do not discard the effects of entrepreneurial age on the business survival. According to the outcome of the data analysis, the regression coefficient of 1.382 suggests that a business run by an owner who is a year older than the
comparison is 3,983 times more likely to default payment in the event of loaning. This claim implies that businesses run by younger entrepreneurs are more likely to survive as compared to those run by older owners. With the world continually modernizing and technology continuously advancing, business owned by young entrepreneurs are more likely to thrive since the owners will be able to keep up with the competition brought up by these world changes. There is however no significant evidence to support this claim.

**Business Location**

The regression coefficient of -2.411 suggests that a business situated in the urban areas is more likely to survive as compared to one located in the rural areas. However, the hazard ratio of 0.90 which is quite close to 1 disqualifies the effect of location of a business on its chances of survival. What is suggested by this outcome of the hazard ratio is that changes in the variable of business location from rural to urban and vice versa has no significant effect on the hazard.

**Business Sector**

According to the outcome of the results, the effects of business sectors on survival vary depending on which sector of business it is. The regression coefficients of 1.134 and 1.723 respectively show that businesses in the manufacturing sector are 3.108 times likely to fail and those in the arts and entertainment sector 5.603 times likely to fail when compared to every other sector of business operation. An explanation to this could be because in the Zimbabwean economy other than acquiring raw materials for production of goods, most materials are now being imported. For this reason, to manufacture and hence sell becomes difficult as most customers would opt for imported goods which usually come at a lower price. In addition, the machinery for production purposes is now also imported into Zimbabwe which is yet another contributing factor to the high prices of our local products as compared to imported products. At the end of the day, businesses in the manufacturing sector will suffer losses and eventually fail completely.

Business Survival At the 5% level of significance, there is enough evidence to support the claim that business survival as a covariate on its own has an effect on a business’ time to default in the event of having received a loan. As indicated by the outcome of the analysis, businesses that survive have a lesser probability of defaulting loan repayment. This is supported by the hazard rate of 1.034 for businesses having failed before the end of the observation period. Theoretically, the hazard rate means that a failed business is 2.831 times more likely to default payment as compared to a survived business. However, in a more practical aspect, the risk of lending to a business exhibiting characteristics of failure as pre-explained in all the factors above, is 2.831 times more than that associated with lending to a business that is more likely to survive.

**CONCLUSIONS**

This study took into all the classes of factors influencing business survival as well as default in the event of lending. These included firm, entrepreneur and industry specific covariates. From the outcome of this study, in regard to businesses in the same sector, a business that has been running for a longer period has a higher chance of survival when compared to one that has been operating for a shorter period of time. In addition, the presence of records in self employment, entrepreneurial training and work experience prior to establishment of a business were said to lower the risk of businesses failing and consequently defaulting payment in the event of them receiving financial aid in the form of loans. Furthermore, having an entrepreneur to contribute towards the start-up capital of his or her business was aid to significantly lower the chances of such a business failing. Of all these conclusions, based on the results from the analysis, one of the main was that increasing the chances of a business surviving, in turn lowered the probability of the same business defaulting payment if it were to be given a loan, regardless of the loan size.

Situating one’s business in an environment where it is considered catchment area for most of the business’ clientele is considered as one of the common market strategies in business. Before one decides to setup his or her place of operation in business, they are to consider the sector of business they are operating in, the nature of the targeted clientele for that business and lastly the position that is most accessible to the bulk of their clients. On a generalized overview of all the business sectors that were considered in this study, situating one’s business in the urban areas poses as a better market strategy as compared to having it located in the rural areas. This could be because there is a lot more business activity in the urban areas as compared to the rural areas. However, the results of this study contradict this claim, the hazard rate for the location variable disqualifying the significant importance of business location on performance of a business.

Factoring in the issue of specificity of area of operation, the outcome of this study indicated that
businesses in the manufacturing and entertainment sectors are less likely to survive and flourish as compared to all the other sectors of operation. At present, the Zimbabwean economy does not fully support the production process due to scarcity and/or unavailability of raw materials for the process. As an alternative, most manufacturing companies end up resorting to importing the raw materials which then becomes more expensive unlike in the case where they are obtained locally. For this reason, purchasing readily made exported products becomes cheaper than buying locally produced items. These variations in prices have a negative impact on the performance of our local manufacturing businesses.

REFERENCES


[32]. Zimbabwean Monetary Policy Statement, January 2016