CREDIT RISK MEASUREMENT MODEL FOR SMALL AND MEDIUM ENTERPRISES: THE CASE OF ZIMBABWE

by

MARX DAMBAZA

submitted in accordance with the requirements

for the degree of

DOCTOR OF BUSINESS LEADERSHIP

at the

UNIVERSITY OF SOUTH AFRICA

SUPERVISOR: PROF JAN KRUGER

2020

DECLARATION

Student No: 72879971

I, Marx Dambaza, declare that **Credit Risk Measurement Model for Small and Medium Enterprises: Case of Zimbabwe** is my own work and that all the sources that I have used or quoted have been indicated and acknowledged by means of complete references.

I further declare that I submitted the thesis to originality checking software and that it falls within the accepted requirements for originality.

I further declare that I have not previously submitted this work, or part of it, for examination at Unisa for another qualification or at any other higher education institution.

______________________ ______________________

SIGNATURE DATE

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere gratitude to my supervisor, Professor Jan Kruger, for his careful and patient guidance during my research period. Whenever I had any questions relating to my research, he was always available and gave me invaluable suggestions and inputs. Professor Kruger has always inspired me to develop innovative thinking and encouraged me to face the challenges. I owe him great thanks for all his continued support through the period of study. I would also like to thank all Graduate School of Business Leadership (SBL) staff members who have helped me along the way. Moreover, I would like to thank every friend for sharing their experiences both of life and their studies during my study period at University of South Africa (UNISA). Finally, I would like to dedicate this thesis to my beloved parents with all my appreciation. I am deeply indebted to them for their unlimited love and unconditional support. Special thanks go to my dearest wife, Rorisang Dambaza, who took care of my family during my visits to Professor Kruger in South Africa. With her help and support, I was able to successfully complete Doctor of Business Leadership (DBL) research. Most importantly, I specially express my best gratitude to Dr Victor Gumbo, for his endless and sincere support for me. With his constant inspiration and encouragement, my heart has remained full of momentum and courage; therefore, this DBL thesis was finished completely within my designated schedule.

ABSTRACT

The advent of Basel II Capital Accord has revolutionised credit risk measurement (CRM) to the extent that the once “perceived riskier bank assets” are now accommodated for lending. The Small and Medium Enterprise (SME) sector has been traditionally perceived as a riskier and unprofitable asset for lending activity by Commercial Banks, in particular. But empirical studies on the implementation of the Basel II internal-ratings-based (IRB) framework have demonstrated that SME credit risk is measurable. Banks are still finding it difficult to forecast SME loan default and to provide credit to the sector that meet Basel’s capital requirements. The thesis proposes to construct an empirical credit risk measurement (CRM) model, specifically for SMEs, to ameliorate the adverse effects of SME credit inaccessibility due to high information asymmetry between financial institutions (FI) and SMEs in Zimbabwe. A well-performing and accurate CRM helps FIs to control their risk exposure through selective granting of credit based on a thorough statistical analysis of historical customer data. This thesis develops a CRM model, built on a statistically random sample, known-good-bad (KGB) sample, which is a better representation of the through-the-door (TTD) population of SME loan applicants. The KGB sample incorporates both accepted and rejected applications, through reject inference (RI). A model-based bound and collapse (BC) reject inference methodology was empirically used to correct selectivity bias inherent in CRM domain. The results have shown great improvement in the classification power and aggregate supply of credit supply to the SME portfolio of the case-studied bank, as evidenced by substantial decrease of bad rates across models developed; from the preliminary model to final model designed for the case-studied bank. The final model was validated using both bad rate, confusion matrix metrics and Area under Receiver Operating Characteristic (AUROC) curve to assess the classification power of the model within-sample and out-of-sample. The AUROC for the final model (weak model) was found to be 0.9782 whilst bad rate was found to be 14.69%. There was 28.76% improvement in the bad rate in the final model in comparison with the current CRM model being used by the case-studied bank.

Key words: Credit risk measurement (CRM) model, known-good-bad (KGB) sample, Reject inference (RI), Basel II IRB framework, Bound and Collapse (BC), ROC, bad rate, missing data, selectivity bias.

NGAMAFUPHI

Isivumelwano se*Basel II Capital Accord* sesishintshe indlela yokulinganisa ubungozi bokunikezana ngesikweletu *credit risk measurement (CRM)* kwaze kwafika ezingeni lapho izimpahla ezazithathwa njengamagugu anobungozi "*riskier bank assets*" sezimukelwa njengesibambiso sokuboleka imali. Umkhakha wezamaBhizinisi Amancane naSafufusayo, phecelezi, *Small and Medium Enterprise (SME)* kudala uqondakala njengomkhakha onobungozi obukhulu futhi njengomkhakha ongangenisi inzuzo, ikakhulu njengesibambiso sokubolekwa imali ngamabhange ahwebayo. Kodwa izifundo zocwaningo ezimayelana nokusetshenziswa nokusetshenziswa kwesakhiwo i*Basel II internal-ratings-based (IRB*) sezikhombisile ukuthi ubungozi bokunikeza isikweletu kumabhizinisi amancane nasafufusayo (*SME*) sebuyalinganiseka. Yize kunjalo, amabhange asathola ukuthi kusenzima ukubona ngaphambili inkinga yokungabhadeleki kahle kwezikweletu kanye nokunikeza isikweletu imikhakha enemigomo edingekayo yezimali kaBasel. Lolu cwaningo beluphakamisa ukwakha uhlelo imodeli ephathekayo yokulinganisa izinga lobungozi bokubolekisa ngemali (CRM) kwihlelo lokuxhasa ngezimali ama-*SME*, okuyihlelo elilawulwa yiziko lezimali ezweni laseZimbabwe. Imodeli ye-*CRM* esebenza kahle futhi eshaya khona inceda amaziko ezimali ukugwema ubungozi bokunikezana ngezikweletu ngokusebenzisa uhlelo lokunikeza isikweletu ababoleki abakhethekile, lokhu kususelwa ohlelweni oluhlaziya amanani edatha engumlando wekhasimende. Imodeli ye-*CRM* ephakanyisiwe yaqala yakhiwa ngohlelo lwamanani, phecelezi i*statistically random sample*, okuluphawu olungcono olumele uhlelo lwe *through-the-door (TTD) population* lokukhetha abafakizicelo zokubolekwa imali bama *SME*, kanti lokhu kuxuba zona zombili izicelo eziphumelele kanye nezingaphumelelanga. Indlela yokukhetha abafakizicelo, phecelezi *model-based bound-and-collapse (BC) reject-inference methodology* isetshenzisiwe ukulungisa indlela yokukhetha ngokukhetha ngendlela yokucwasa kwisizinda se*CRM*. Imiphumela iye yakhombisa intuthuko enkulu mayelana namandla okwehlukanisa kanye nokunikezwa kwezikweletu kuma *SME* okungamamabhange enziwe ucwaningo lotho., njengoba lokhu kufakazelwa ukuncipha okukhulu kwe-*bad rat*e kuwo wonke amamodeli athuthukisiwe. Imodeli yokuqala kanye neyokugcina zazidizayinelwe ibhange. Imodeli yokugcina yaqinisekiswa ngokusebenzisa zombili indlela isikweletu esingagculisi kanye negrafu ye-*Area under Receiver Operating Characteristic (AUROC)* ukulinganisa ukwehlukaniswa kwamandla emodeli engaphakathi kwesampuli nangaphandle kwesampuli. Uhlelo lwe-AUROC lwemodeli yokugcina (*weak model*) lwatholakala ukuthi luyi 0.9782, kanti i*bad rate* yatholakala ukuthi yenza i-14.69%. Kwaba khona ukuthuthuka nge-28.76% kwi-*bad rate* kwimodeli yokugcina uma iqhathaniswa nemodeli yamanje i*CRM* model ukuba isetshenziswe yibhange elithile.

Amagama asemqoka: imodeli yokulinganisa ubungozi bokunikezwa isikweletu, ukungalungeli ukunikezwa isikweletu (RI), isakhiwo seBasel II IRB, Uhlelo lwe*Bound and Collapse* (BC), ROC, uhlelo lwe-*bad rate*, idatha engabonakali, indlela yokukhetha ngokuthatha ingxenye

KAKARETSO

Basel II Capital Accord e fetotse tekanyo ya kotsi ya mokitlane (*credit risk measurement (CRM))* hoo “thepa e kotsi ya dibanka” ka moo e neng e bonwa ka teng, e seng e fuwa sebaka dikadimong. Lekala la Dikgwebo tse Nyane le tse Mahareng (SME) le bonwa ka tlwaelo jwalo ka lekala le kotsi e hodimo le senang ditswala bakeng sa ditshebetso tsa dikadimo haholo ke dibanka tsa kgwebo. Empa dipatlisiso tse thehilweng hodima se bonweng kapa se etsahetseng tsa tshebetso ya moralo wa Basel II internal-ratings-based (IRB) di supile hore kotsi ya mokitlane ya SME e kgona ho lekanngwa. Leha ho le jwalo, dibanka di ntse di thatafallwa ke ho bonelapele palo ya ditlholeho tsa ho lefa tsa diSME le ho fana ka mokitla lekaleng leo le kgotsofatsang ditlhoko tsa Basel tsa ditjhelete. Phuputso ena e ne sisinya ho etsa tekanyo ya se bonwang ho mmotlolo wa kotsi ya mokitlane (CRM) tshebetsong ya phano ya tjhelete ya diSME e etswang ke setsi sa ditjhelete (FI) ho la Zimbabwe. Mmotlolo o sebetsang hantle hape o fanang ka dipalo tse nepahetseng o dusa diFI hore di laole pepeso ya tsona ho kotsi ka phano e kgethang ya mokitlane, e thehilweng hodima manollo ya dipalopalo ya dintlha tsa histori ya bareki. Mmotlolo o sisingwang wa CRM o hlahisitswe ho tswa ho sampole e sa hlophiswang, e leng pontsho e betere ya setjhaba se ikenelang le monyako (TTD) ya batho bao e kang bakadimi ba tjhelete ho diSME, hobane e kenyelletsa bakopi ba amohetsweng le ba hannweng. Mokgwatshebetso wa *bound-and-collapse (BC) reject-inference* o kentswe tshebetsong ho nepahatsa tshekamelo ya kgetho e leng teng ho lekala la CRM. Diphetho tsena di bontshitse ntlafalo e kgolo ho matla a tlhophiso le palohare ya phano ya mokitlane ho diSME tsa banka eo ho ithutilweng ka yona, jwalo ka ha ho pakilwe ke ho phokotseho ya direite tse mpe ho pharalla le dimmotlolo tse hlahisitsweng. Mmotlolo wa ho qala le wa ho qetela e ile ya ralwa bakeng sa banka. Mmotlolo wa ho qetela o ile wa netefatswa ka tshebediso ya bobedi reite e mpe le mothinya wa *Area under Receiver Operating Characteristic (AUROC)* ho lekanya matla a kenyo mekgahlelong a mmotlolo kahare ho sampole le kantle ho yona. AUROC bakeng sa mmotlo wa ho qetela (mmotlolo o fokotseng) e fumanwe e le 0.9782, ha reite e mpe e fumanwe e le 14.69%. Ho bile le ntlafalo ya 28.76% ho reite e mpe bakeng sa mmotlolo wa ho qetela ha ho bapiswa le mmotlolo wa CRM ha o sebediswa bankeng yona eo.

Mantswe a bohlokwa: mmotlolo wa tekanyo ya kotsi ya mokitlane, *reject inference* (RI), moralo wa Basel II IRB, *Bound and Collapse* (BC), ROC, retie e mpe, datha a siko, tshekamelo ya kgetho

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

CHAPTER 1 ORIENTATION

1.1 BACKGROUND OF STUDY

The prevailing global economic environment is causing drastic changes in the lending function at all financial institutions (FIs), and in tandem with these changes many lending institutions are changing the way of doing business especially in how to manage the credit portfolio (Mileris, 2012; Anagnostopoulos, Skouloudis, Khan & Evangelinos, 2018). FIs must correspondingly respond to the realities of the market and the unabated declining margins on loans, due to disintermediation taking place in the face of the lending industry (Hawkins & Turner, 1999). With the fast-paced changes in the national and international financial environments, however, the profit margins that banks can acquire are no longer as good as before, and on the other hand, the degree of credit risk exposure also deepens (Beck, Demirguc-Kunt, Laeven. & Levine, 2008).

The 2007 financial crisis revealed that the financial sector was over-levered to the extent that the sector is not able to absorb the subsequent credit losses (Antão & Lacerda, 2011). Consequently, the sector has recently undergone major transformations in terms of stricter regulatory framework demands in order to arrest the banks' risks, credit risk in particular. According to Bruni, Beraldi and Iazzolino (2014), the subsequent aim of the modifications of the Basel Capital Accords was to preserve sustainability, to create a sound financial market and to develop new improved credit risk measurement (CRM) models to rescue financial institutions from over-leverage. Lending to industry, Small and Medium Enterprises (SMEs) as well as to consumers still constitutes, regardless of respective credit risk entailed, the core source of income of commercial banks and other FIs. Therefore, it is imperative that appropriate and well-performing models be developed for CRM to reduce and tame the credit risk that banks are facing today (Lai & Kuo, 2010).

On the other hand, banks and any FIs are specialists in credit risk management, their paramount challenge being able to differentiate between good and bad borrowers' precedent to the loan granting process (Ngwa, 2010; Bolton, Cecchetti, Danthine & Vives, 2019). Such differentiation is achievable if the lenders acquire expertise to both measure and quantify their credit risk exposure on their own behalf and on behalf of their clients (Collins, Morduch, Rutherford & Ruthven, 2009). By nature of their core business, commercial banks and FIs are credit risk intermediaries; they maintain an inventory of credit risk that must be measured carefully to ensure that risk does not threaten the intermediary's solvency (Lai & Kuo, 2010; Rehman, Muhammad, Sarwar & Raz, 2019; Khaled & Wahab, 2019).

The growing recognition of economic role played by SMEs has also drastically changed the lending market (Nyamwanza, 2014; Karadag, 2016;). SMEs are described as efficient and prolific job creators, seeds of big businesses, and fuel of national economic engines (Abor & Quartey, 2010; Wangmo, 2015). Even in developed economies SMEs absorb the largest share of employers of general workforce (Kritikos, 2014; Haga, 2017). Structural development of SMEs is topical in any policy debates in any economy regardless of the level of economic development; thereby governments at all levels find it imperative to undertake to promote growth of SMEs (Yoshino, 2013; Kritikos, 2014).

Sustainability in economic growth and job creation is achievable if and only if vibrancy in the national entrepreneurial activity is promoted, which is both a means and an end to the transition to market economy and creation of entrepreneurial culture (Bushe, 2019). This culture is a panacea that mitigates social costs that result from restructuring of SMEs and enhance the probability of a sustainable commitment to democracy and rule of law (Fotache, Fotache, Ocneanu & Bucşă, 2011). One important deterrent and challenging problem that SMEs, often face is access to capital (Ackar & Vuvor, 2011; Wangmo, 2015; Muriithi, 2017). The hardest hit are SMEs in emerging economies, where banks lack the capacity or are unwilling to give long-term loans to this sector of the economy (Wangmo, 2015; Aysa, 2019). They give preference to lend to large, established firms with well-developed balance sheets and traceable credit histories of additional assets for collateral requirement in traditional bank financing (Rahman, Rahman & Ključnikov, 2016; Al Baz, 2017; Kanapickiene & Spicas, 2019).

Substantial obstruction to adequate access to external financial resources plays a significant constraint on SME development and sustainability (Nkonge, 2013; Abiodun & Entebang, 2015; Coetzee & Buys, 2017; Sibanda, Hove & Shava, 2018) notwithstanding the recognition of the economic role of SMEs in the development and growth process in any country. For their expansionary and developmental objectives, SMEs are reliant on financial institutions, especially commercial banks, for external financing. Literature (Yeung, 2009; Fatoki & Van Aardt, 2011; Mlachila, Dykes, Zajc, Aithnard, Beck, Ncube & Nelvin, 2013; Creal, Schwaab, Koopman & Lucas, 2014) confirms that the main formal source of SME financing is the bank but to the contrary, banks prefer firms with proven track records and enough hard assets for collateral, which are difficult to obtain for most of the SMEs. The financing of SMEs, in most economies, is heavily dependent on informal sources of finance and on retained profits to complement their working capital (Ackar & Vuvor, 2011; Padachi, Howorth & Narasimhan, 2012; Wangmo, 2015). This paltry funding is not enough for their investment needs. Owing to the constraints to formal external financing, SMEs use informal external financing such as friends and family as a fall-back alternative (Nyankomo, 2014; Asad, Asad, Abdullah & Khalid, 2018).

Commercial banks and other FIs acknowledge the need to enter into this non-traditional lending market but are sceptical of credit risk entailed in the SME lending market. They are aware that lending activities are susceptible to credit risks therefore explicit understanding and quantification of such risks is paramount to bank credit management as well as wholesome stability of the economy at large (Al Baz, 2017; OECD, 2018; Tursoy, 2018). Kuritzkes *et al*. (2003) established that credit risk makes up about 50-60% of the total risk in a large bank. This is quite substantial.

Loans are any lending institution's major source of revenue, and conversely, loans also represent the greatest source of risk to the institution's safety and soundness and have been the ultimate cause of losses and bank failures, consequently the cause of worldwide financial crises (Li & Zou, 2014; Saeed & Zahid, 2016). The viability of each lending institution is in its loan portfolio and success of the institution depends on how well this portfolio is managed (Lagat, Mugo & Otuya, 2013; Magali, 2014). Bank loans are the prime source of business finance, even for SMEs (Veiga & McCahery, 2019). Therefore, banks as custodian sources of SME loans are in an economic dilemma since lending to the latter is stereotyped risky, untrustworthy and inflicted of great uncertainty in the way of lending business.

Banking is an attempt to manage multiple and seemingly opposing needs (Wolfe & Amidu, 2012; Amidu & Wolfe, 2013; Bogale, 2018). The conventional approach to credit evaluation that banks adopt mostly focus on the collateral and credit records of loan applicants and guarantors, and subjective means of judgements such as experience principles or credit ratings are used (Hagos, 2010; Lai & Kuo, 2010; Wang, 2013). Such subjective approaches are being fast outpaced by ever increasing changes in the financial market environment due to technological and regulatory advances. Innovations and objective evaluation of credit risk could be an answer to the calamities bedevilling the banking industry.

Predicting and mitigating default events is at the core of appropriate credit risk management and this can be greatly helped by employing suitable quantitative CRM models, without however precluding the reliance on human expert judgement (Fraser & Simkins, 2010; Wang, 2013). Owing to the complexity of risky lending environment in which FIs are operating, CRM has become increasingly reliant on statistical or quantitative methodology (Nkonge, 2013; Abiodun & Entebang, 2015; Coetzee & Buys, 2017; Sibanda *et al.*, 2018). Over the past decade, commercial banks have devoted resources to developing internal ratings-based (IRB) quantitative models to better quantify their credit risks and assign economic capital (Berger, 2004; Gakure, Ngugi, Ndwiga & Waithaka, 2012; Wang, 2013; Li & Zou, 2014).

Metric measurement of credit risk attracted the interest of academic and scientific researchers to help FIs improve credit quality of obligors as well as optimally determine risk-adjusted regulatory capital requirements (Horstedt & Linjamaa, 2015; Ma, 2016). CRM tools are currently relied upon by contemporary banking institutions to reject those deemed too risky to lend (Chen & Astebro, 2012; Mustafa & Perssons, 2017). The use of such vital tools has significantly reduced both loan processing costs and aggregate default costs, while radically expanding credit supply to SMEs (Chen & Astebro, 2012; Kennedy, 2013). Kennedy (2013) postulates that an accurate and well-performing CRM model offers lenders the control of their risk exposure through due diligence granting of credit based on the objective analysis of historical customer data. The advent of Basel Capital Accords (Basels I, II & III) has brought useful revolutionaries in the measurement of credit risk by financial institutions thereby facilitating accessibility to financing by non-traditional banking clientele, specifically SMEs (Veiga & McCahery, 2019). Unfortunately, the adoption of such prudence is not obligatory, SMEs continue to be shunned by non-Basel banks mostly in the emerging economies (Schmukler, de la Torre & Martinez, 2008; Wehinger, 2013; Dambaza & Kruger, 2018).

The Basels II and III direct the international credit systems to pay closer attention to measuring and to managing credit risk of all bank assets (Spuchľáková, Valašková & Adamko, 2015; Tursoy, 2018). This is a deliberate ploy to accommodate all credit exposures in the economy, specifically to help banks extend their business into the theoretically opaque assets like the SMEs (Griffith-Jones, 2003). Commercial banks dominate the financial systems, leaving little leeway for SMEs looking for alternative financing to bank loans (Ibe, Moemena, Alozie & Mbaeri, 2015). CRM can objectively play a major role in bringing the two (2) parties together by minimising scepticism from both sides.

Basel II, in particular, has revolutionised financial approach to measuring credit risk of SMEs. The failure, during the Basel I regime, to capture the intrinsic credit risk in the banking business, has aroused the need for quantitative/statistical approaches to measuring credit risk induced by lending SMEs (Muthinja, 2016). Basel II is designed to ensure that SMEs, like big firms, are not hindered from accessing credit (Giuseppe & Ughetto, 2007; Yoshino & Taghizadeh-Hesary, 2017). To transform the traditional financial marketing, emerging economies are fast adopting the Basel II framework.

1.2 PROBLEM STATEMENT

CRM practices is an issue of concern in today's banks and other FIs. There is growing demand to develop improved processes and systems to deliver better due diligence selection of who to grant or not loan at application stage in all bank assets. An erudition of the 2007 financial crisis was that the financial sector was over-levered and not able to absorb the subsequent credit losses (Antão & Lacerda, 2011). As a consequence, the banking industry has, of recent years, undergone drastic changes in terms of stricter regulatory framework demands in order to arrest the adverse effects of banks' credit risks (Bruni *et al.*, 2014.). At the same time, FIs have grown financial appetite to extend credit to SMEs due to the growth and return potential in this business segment and also due to thinning profit margins in the traditional lending market (corporate and retail) lending market. Yurdakul (2014), ratifying Lai and Kuo (2010), pointed out that the most important risk that banks are exposed to is credit risk, which involves loans that are not paid back. In fact, credit risk is the underlying cause of the devastating 2007 financial crisis, this is compelling for researchers and scholars to adopt rigorous empirical research so as to understand the different underlying factors that caused the financial crisis (Al Baz, 2017).

Guided by Basel II Capital Accord, several banks and FIs have designed CRM models for corporate and retail assets, while only a few banks have been developed CRM models specifically for SMEs. In Zimbabwe, the same trend is conspicuous as SMEs CRM is still considered a grey area that has not been tamed, as there is scarce literature on the subject. Notwithstanding the conspicuous potential for banks to garner a higher return on assets (ROA) by extending credit facilities to this large sector of the economy. The growth of SMEs is a critical issue for the economic development of any country. SMEs rely on bank loans for their external financing (Wangmo, 2015; Al Baz, 2017; Kanapickiene & Spicas, 2019), but they often have difficulties in obtaining sufficient funds from banks in a timely manner. Yoshino and Yamagami (2017) pointed out that SMEs lack public information on their economic activities compared to large enterprises. This gives birth to rampant asymmetry of information between them and banks, thereby making it difficult for the latter to measure the credit risk of the former. In order, to minimise the moral hazard effect of information asymmetry between these vital economic players, CRM has been suggested by scholars as the ultimate solution in place of the subjective relationship lending technology being used by small banks (Altman & Sabato, 2007; Wangmo, 2015).

In that endeavour, statistical CRM models are being developed yet some crucial statistical principles are being overlooked with great certainty in the CRM domain. There are controversies among researchers on the adverse effects of selectivity bias prevalent in the CRM domain. Most CRM models (Altman & Sabato, 2005; Al Baz, 2017; Kanapickiene & Spicas, 2019) are developed on conspicuously truncated, non-random KGB samples made up of only the selected "goods", which by far not indicative of the TTD population of credit applicants. This compromises the authenticity and robustness of the eventual model, since the model development, KGB, sample is not a proper representation of the TTD population. The resulting CRM created or modified from a truncated KGB sample is eventually measured on the TTD population, breeding statistical deficiency, selectivity bias or sample bias. Verstraeten and van den Poel (2005) pointed out that the consequence of the presence of selectivity bias may be an erroneously selected acquisition strategy to expand the banking portfolio and the resulting lower-than-expected profits or even significant losses. Again, Jacobson and Roszbach (1999), ratified that it is of great importance to resolve the selectivity bias in the KGB sample. This anomaly in CRM domain turns out to be a serious issue as demonstrated by Chandler and Coffman, 1977, and Avery, 1977 who spearheaded early discussions on this statistical discrepancy. Ditrich (2015), among others, also pointed out that using a model designed solely on a truncated KGB sample often generate biased and misleading results. As banks and other FIs operate with high volumes of clientele, therefore, even a slight improvement of the discriminatory and predictive abilities of these CRM models may generate significant additional gains (Kennedy, 2013). Reject Inference has been pencilled as the panacea to the selectivity bias problem, although some reject inference techniques have been found futile in literature (Banasik & Crook, 2006; Ditrich, 2015).

FIs in many parts of the world, including Zimbabwe and sub-Saharan Africa, have explored the SME credit, since that is where the majority of such entities are hit hard by debt inaccessibility. Regardless of such effort, FIs are still in struggle on how to penetrate the SME credit business, leaving questions like; how do banks and other FIs measure SME credit; what CRM models do banks and what loan pricing strategy should banks apply with the aim to nurture SMEs portfolios and make them a profitable business. However, Hanson, Hashem and Schuermann (2008) noted that a majority of loans issued by banks are granted to SMEs, which the banks know scarcely about compared to matured segment publicly traded corporate. Therefore, as a significant amount of effort is needed as, to conduct accurate CRM and as well make these SMEs have full knowledge about what aspects that they are evaluated on to access the much-needed credit for their investments.

Therefore, the problem statement is:

> *That SMEs credit risk has not been readily quantifiable for loan underwriting, due to predominance of relationship lending. As a result, SMEs have limited access to financing from Commercial banks due to perceived high credit riskiness, regardless of their crucial role in economic growth and development of any country, especially emerging economies.*

1.3 PURPOSE OF THE STUDY

This study had two (2) purposes: (a) to examine the extent at which banks are involved with SME financing in Zimbabwe and (b) to develop an internal rating-based SME CRM model for a bank in accordance with the Basel II Capital Accord framework.

The first purpose of the study is to assess the extent to which the banks are involved with SME financing in Zimbabwe. The proliferation of SMEs has necessitated the need to take stock of financial services sector’s response to the emergence of a non-traditional lending market, in the wake of dwindling of the corporate and retail traditional assets. In 2012, the World Bank carried a survey of Micro-Small and Medium Enterprises (MSMEs) in Zimbabwe. This was an MSME perspective survey which investigated the challenges and constraints the sector was facing. For the sake of the purpose of this thesis, a double perspective exploration was carried in order to corroborate views from both suppliers of credit (banks) and demanders of credit (SMEs). In that regard, two (2) surveys were carried out to gauge the extent of the interaction between the two (2) important economic players.

For the bank survey, the target respondents were bank managers, credit managers and loan officers of banks that were ex-ante thought to be serving SME sector and banks that were perceived ex-ante to have little or no intention to serve the emergent SME sector. The instrument of measurement was a questionnaire. At the time of the bank survey, there were 14 banks in Zimbabwe of different sizes and ownership. A sample of 14 banks was conveniently selected since participation was strictly voluntary and banks are generally secretive. Both the face-to-face interviews and self-administered methods of data collection were employed for this survey to extract as much views, opinions, attitudes towards SME financing, as possible.

For SME survey, a purposive sampling design was employed targeting SME owners and administrators. In Zimbabwe there is no sampling frame for SMEs but there is an association of SMEs called Small and Medium Enterprise Association of Zimbabwe (SMEAZ). This was used as a proxy to the population of SMEs in Zimbabwe. Using SMEAZ's website, a questionnaire was flighted querying membership on issues to do with sources of finance for their investments and growth of their enterprises as well as for financial constraints they face. Those who found survey interesting responded. A sample of 300 owners responded out of a membership of 705, responded. Surveys on both the supply (Bank survey) and demand (SME survey) sides were carried out to confirm and compare the current existing relationship between the two (2) sectors in Zimbabwe.

The second purpose of this thesis was to design an IRB SME CRM model for a bank in Zimbabwe to facilitate objective measurement of SME credit risk at loan application stage with the goal to ameliorate the adverse effects of information asymmetry between banks and SMEs. In fact, the purpose of phase II of the thesis at hand is to first explore how to design an All-Known-Good-Bad (AGB) sample using theoretically based reject inference technique, then eventually design a CRM model on the resulting representative sample. Currently, using judgemental method to select SME loan applicants for granting, ZimSME bank, a case-studied bank, assesses them individually on 26 characteristics variables. So, it was also the purpose this phase to explore which qualitative and quantitative criteria does ZimSME bank uses in its credit granting decision for SMEs. The author also aim to reduce the adverse effects of selectivity bias prevalent in CRM domain by applying a model-based BC methodology to first construct AGB sample before developing the CRM model This innovation breeds a better population indicative sample, unlike the KGB sample on which most CRM models, in literature, are built. A model built on AGB sample will be compared to model developed from KGB sample to assess if there is any improvement in the classification of SME loan applications into good and bad debtors before loan is granted.

Throughout phase II of the thesis, the author takes the perspective of the case-studied bank. In that regard, the author gathered data of the bank through bank loan register and interviews of credit personnel at ZimSME bank in Zimbabwe which is offering credits to SMEs. This purpose is quite relevant since SMEs have more problems to attract and generate capital than larger enterprises. Again, SMEs accounts for the main part of the Zimbabwean lending market, this is an important issue to further investigate.

The overall purpose of the study is to help commercial banks and FIs to build a profitable, well performing SME loan portfolio, through an effective and systematic application of internal quantitative CRM regardless of high informality, opacity and lack of trading history which characterise the SME sector. CRM modelling has the potential to change SME banking systems to the benefit of both parties and economy at large, leading to increased competition among SME credit providers, hence increased availability of credit destined to the sector.

1.4 RESEARCH QUESTIONS

To achieve the main goal of this study, the following research questions need to be addressed:

1. What are the factors that cause SME financing constraints from the perspectives of SMEs and the banks in Zimbabwe?
2. How do banks measure credit risk of SMEs in Zimbabwe?
    - 2.1 How do we minimise selectivity bias present in the KGB sample?
    - 2.2 What are the criteria used by banks when measuring credit risk of SMEs in Zimbabwe?

1.5 EXPECTED BENEFITS OF SME CRM MODELING

Development of a CRM model would benefit the participating banks, the SME community and/or broader society in the emerging economies and business in general in the following ways:

- Reduced subjectivity and increased objectivity in bank credit risk assessment.
- Increased pace and consistency of credit risk assessment process.
- Automation of SME credit risk assessment process is a substantial possibility.
- Substantial decrease of transaction costs of SME credit risk assessment.
- Economic and improved allocation of bank resources in credit risk assessment process.
- Objective and better determination of SME loan parameters: interest rate, loan duration, loan amount, capital requirement, etc.
- Improved SME credit risk management.
- Improvement in the way banks collect and recover their loans from SMEs.
- Possible dominance and gain of an edge in the non-traditional SME lending market.

1.5.1 Aim of the study

To develop a CRM model of a portfolio of SME loans in a bank in accordance to Basel II Regulatory Framework.

1.5.2 Specific objectives

The specific objectives of the study are to:

- explore the extent to which banks are involved with SME financing in Zimbabwe.
- apply Bound and Collapse Bayesian Reject Inference methodology to build ‘all-good-bad” sample for CRM model development.
- develop a logistic CRM model for SMEs.

1.6 RATIONALE OF THE STUDY

The study develops quantitative CRM mechanisms that allow (FIs) to evaluate SMEs loan applications as objectively and efficiently as possible prior to loan granting. For an overall national economic development of any economy, SMEs need to be financially supported by banks and at the same time banks, as financial intermediaries, need to safeguard the depositors' money as well as getting reward from their ultimate lending business. If an accurate and well-performing CRM modelling for SMEs in emerging financial markets is achieved, access to SME finance would be improved thereby enhancing the development of entrepreneurial activity in the national economy. CRM is important for bankers because it helps determine several features of a loan: interest rate, maturity, collateral, capital requirements and other contractual obligations a loan entail. If CRM modelling is not done adequately, default rates escalate to the extent that banks are eventually pushed into insolvency and bank failures and global financial crisis ensue.

For any statistical approach to correctly infer about the credit risk of an SME or any credit applicant there is need to go to the basics in terms of sampling theory on which any statistical inference is grounded. The issue of truncated development or non-random samples, a product of selectivity bias, on which most CRM models are built, appears with great certainty in credit risk forecasting domain (Lin, 2007; Chen & Astebro, 2012; Ditrich, 2015). Truncated samples engender biased estimates due to selectivity bias and non-randomness of the development sample consisting of only the pre-screened "good" loan applicants, leaving out the "bad" ones. This scenario engenders a direct and conspicuous mismatch between the model development, KGB, sample and the TTD population which is supposed to be credit risk measured in future. To counter this statistical deficiency caused by selectivity bias, the study employs theoretically supported reject inference methodology to develop first a representative random development sample, AGB) sample, on which the envisaged final CRM model would be built. The imputation model-based Bound and Collapse (BC) technique is adopted in this thesis for it is theoretically hinged on Bayesian inference analysis, in which reject inference is treated as a missing data problem. This is contrary to other reject inference techniques (re-weighting methods or extrapolation methods) which are hinged on the tenuous assumption that the distribution of good clients is the same as the bad clients. On the other hand, owing to the increase in off-balance sheet activities that create implicit contracts and obligations between prospective lenders and buyers, CRM has become more complex.

1.7 EXPLORATORY LITERATURE REVIEW

Emerging economies strive for a wholesome economic development with the aim to reduce poverty, improve in the standard of living of the majority poor; therefore, they see SMEs as the conduit to that destiny. SMEs are increasingly seen as playing an important role in economies of many countries, especially in emerging economies. Owing to their simple structure, SMEs can respond quickly to economic conditions and meet customers' needs, growing sometimes into large and powerful corporations or failing within a short time of the firm's inception (Altman & Sabato, 2005; Nyamwanza, 2014; Bushe, 2019). In general, emerging economies experience high levels of poverty and income inequality, in this regard; Fin Mark (2006) suggests that the best way to address these socio-economic ills is to improve SME development.

At their disposal, SMEs have two (2) primary sources of external finance, which are equity and debt. Owing to the rigorous requirements for equity accessibility, SMEs are left dependent on bank loans. Despite the SMEs dependence on debt finance, paradoxically access to debt finance is very limited to them, especially in emerging economies (Sharma & Gounder, 2012; Fatoki, 2014; Osano & Languitone, 2016; OECD, 2018). Still under this unsupportive background, SMEs, unlike large firms which have access to different sources of financing, rely on banks as source of external financing although commercial banks are hesitant to lend to them due to fear of credit risk and high transactional costs the business entails.

Fin Mark Trust (2006) provides evidence that only 2% of new SMEs in South Africa can access bank loans and that the use of suppliers' credit by SMEs is virtually non-existent. Mutezo (2015) reports that about 75% of applications for bank credit by new SMEs in South Africa are rejected. Balkenhol and Evans-Klock (2002) put the use of trade credit by new SMEs in South Africa at only 0.2%. As Smorfitt (2009) puts it, new SMEs in South Africa struggle to raise external finance. These statistics reveal the gravity of SME financing problems bedevilling emerging financial markets. This is a real dilemma in which emerging economies find themselves in. An in-depth research is required to nurture SMEs financially by involving the providers of external finance, banks, at the same time allaying their fear of credit risk and high transactional costs.

A recent surge of interest in default risk in the 1990s (Madeira, 2018) is a clear indication of how increasingly complex the credit risk management of financial institutions has become. According to the Basel II Capital Accord, default is the situation when the obligor is unlikely to pay its credit obligations, or the obligor is past due more than 90 days on any material credit obligation. Studies follow the second part of this definition and bias has been on corporate default, the traditional lending market for commercial banks.

Statistics demonstrate that SMEs are contributing significantly to economic growth of nations. It is confirmed (Fatoki, 2014; Banwo, Du & Onokala, 2017; Gherghina, Botezatu, Hosszu & Simionescu, 2020) that for Organisation for Economic Co-operation and Development (OECD) members, SMEs produce three-fourths of total jobs and often more than one-third of the country's GDP. Again, in OECD countries, SMEs represent almost 99% of the total number of firms, they are responsible for 78% of the job offer of the country, but around 80% of SMEs are shut down before one year of activity (Altman, Sabato & Wilson, 2009; OECD, 2018). In South Africa, SMEs contribute 66.4% of private sector employment and 34.8% to GDP (Statistics South Africa Labour Force Survey, 2006). These statistics substantiate the need for banks to reconsider their marketing and business strategies in the wake of the great need for SME sector development in pursuit of the much-needed profitability from their lending business.

On the other hand, competition intensity in the financial markets from both the supply side and the demand side has prompted some financial practitioners, academics and researchers to explore ways of making lending business to SMEs profitable through objective evaluation of their creditworthiness prior to loan granting. Basel II has drastically transformed the way banks do business with SMEs. The opacity, a consequence of information asymmetry, through the adoption of the Basel II IRB framework, is shading off between FIs and SMEs for the good of the economy in general (Yoshino & Taghizaden-Hesary, 2017). This has allayed the severity of poor relationship between these two (2) pillars of any emerging economy. The success stories in the developed countries (Schwaiger, 2002; Saurina & Trucharte, 2004; Altman & Sabato, 2005; Al Baz, 2017; Kanapickiene & Spicas, 2019) give an indication that SMEs in emerging economies can also access financing from commercial banks with ease if the latter adopt the Basel II framework, Internal Ratings-Based (IRB) approach to measuring credit risk.

In-depth analysis shows that defaults of SMEs are more weakly correlated than among corporate (Agrawal & Maheswari, 2019). In general, defaults among bigger companies are thought to be primarily caused by systematic risk factors, while defaults by smaller businesses are driven by idiosyncratic risk factor (Curcio, Gianfrancesco & Malinconico, 2011; Novales & Chamizo, 2019; Maiti, 2019; Agrawal & Maheswari, 2019). As there are no clear-cut calculation of SME capital requirement in the Basel II Capital Accord, SMEs are either treated as corporate or retail assets. Literature review on SME capital requirement determination has seen financial innovations which have given indications on how lending to SMEs should be treated, approaches viewed as a hybrid of corporate and retail lending (Dietsch & Petey, 2004). If some SMEs are considered as corporate, however, capital requirements are slightly greater than under the existing Basel I Capital Accord. Altman and Sabato (2005) suggest that the most eligible is that banks would use a blended approach that is considering some SMEs as retail and some as corporate customers. In same article, Altman and Sabato (2005) urged banks to use the Advanced Internal Ratings-Based (A-IRB) approach to measure and to manage SME credit risk on a pooled (portfolio) basis. This has led to the derivation of statistical-based CRM models, which predict an enterprise's probability of default (PD) and assign credit scores to SMEs in a portfolio.

It is in the emerging economies, where the external financing constraint is severe and at the same time the SME is seen as the seedbed for economic and social solution to a plethora of economic problems they face (OECD, 2017; Bakhtiari, Breunig, Magnani & Zhang, 2020). SME sector comprises the great majority of enterprises which are viewed as the potential and promising market for lending business. Banks and other FIs can effectively target this sector and manage the risks in lending without heavy reliance on collaterals (Mullineux & Murinde, 2014; Altman, 2018).

With hope to enhance socio-economic development, some governments have put in place policy framework to financially support and subsidise SME growth by even going to the extent of canvassing reluctant banks to bail out SMEs through political and uneconomic financing schemes. Government assistance strategies in both developed and emerging countries often try to achieve a combination of equity objectives (alleviating poverty and addressing social, ethnic and gender inequality) and efficiency objectives (raising the productivity and profitability of firms) (Wangmo, 2015; Taiwo & Falohun, 2016). All these efforts have not yielded tangible results. As Ojo (2003) argues, all these SME assistance programmes have failed to promote the development of SMEs. Oftentimes, the finance provided have been misdirected, gone to wrong persons or found to be inadequate to impact on the expected development of the assisted firms. Tumkella (2003) reiterates that these government programmes cannot achieve their expected desires due largely to abuses, poor project feasibility evaluations and monitoring as well as moral hazard involved in using public funds for promoting private sector enterprises. Such a background has aroused the interests of researchers (Altman & Sabato, 2005; Dainelli, Giunta, & Cipollini, 2013) and banking practitioners (Jimenez & Saurina, 2004; Berger, 2004; Aubier, 2007; Al Baz, 2017) to develop efficient and objective methodologies and introducing rigorous control practices in measuring and managing SME credit risk with an eventuality of facilitating SME access to financing from banks and other FIs without financial rationing.

1.8 SCOPE OF THE STUDY

The study takes a case of an emerging economy, Zimbabwe, as an important context for SME financing scholarship. Generally, SMEs in emerging markets, face less developed financial markets, feeble institutions for distribution of capital as well as fierce volatility in social and economic developments that limit availability of major resources and increase cost of such resources (Bamper, Fernandez-Stark, Gereffi & Guinn, 2014; Nyamwanza, 2014; Ross, Ligang & Cai, 2018). It is on this background that the study starts by looking at the financial services relationship with SMEs (banks-SMEs surveys). It has been noted that many emerging countries in Sub-Saharan Africa have large number of SMEs relative to the size of the economy but the majority of these are not formal. In the case of this thesis the focus is put on SMEs that are larger than micro companies and are part of the formal economy. To assess the degree of bank involvement in SME financing two (2) major surveys were carried out. The bank survey covered all banks in Zimbabwe (large, foreign, small and local). The survey on SMEs, targeted the population of all SMEs in different spheres of the Zimbabwean economy.

The second phase of the study includes an empirical development of a statistical CRM model from a real data of a SME loan portfolio of a case-studied commercial bank in Zimbabwe according to the dictates of Basel II Capital Accord. It is conducted by taking the perspective of the case-studied bank leaving out the views of regulators, SMEs and other stakeholders. This is in line with Basel II which is grounded in recognizing an individual credit risk through internal-rating-based (IRB) system; a bank's manager must correctly measure the credit risk and price it correctly (Basel II, Capital Accord, 2004). In that regard, the financial and historical loan repayment information of an SME portfolio included in the study is obtained from a commercial bank under study, pseudonym ZimSME on ethical and confidentiality grounds. Again, in tandem with the problem statement, the researcher will only consider the CRM modelling of SMEs only, no discussion about corporations will be included in the analysis. Only SME clients who sought credit from ZimSME were considered as the KGB sample. The case of Zimbabwe in the development of the CRM model excluded consideration countries with regulatory frameworks, market conditions and other country specific characteristics. This limiting scope was meant to come up with a homogenous group with similar conditions in terms of country specific that influence the CRM procedures of banks differ otherwise geographical extension may contribute with interesting findings.

1.9 LIMITATIONS OF THE STUDY

The first phase of the study aimed at extracting information from banks in Zimbabwe on their involvement with SME funding as well from the SME owners regarding their demand for external financing. The following were the anticipated limitations to the first phase of the study:

- Owing to secretive nature of the operations of banks, access to data and information is not as easy as accessing data in the public domain.
- Access to important bank policy documents regarding SME financing is an uphill task owing to fear of losing competitive advantage to rivalry institutions.
- Most respondents from both the supply and demand sides are reluctant to tell the truth due to fear of divulging confidential information of the bank and SME as well as due to fear of jeopardising bank position and future financing on the part of SME clients.
- Some banks do not allow loan information of their clientele for academic study purposes.
- Some respondents do not respond to questionnaire on time.

In anticipation of these limiting factors, which would have jeopardised the success of the project, the researcher had to establish good rapport and entered into signing of ethical codes of conduct with participating banks so as abide with confidentiality character of any bank information. The participating banks and SME owners were assured that the subsequent analysis of information collected during surveys would be done in aggregate not linking any result to any individual participant and that resultant report would be accessible to any participant at request.

The second phase of the study constitutes the empirical SME CRM model development process. The phase is heavily dependent on the availability of historical portfolio of SME loans from a case-studied bank in Zimbabwe. ZimSME bank in Zimbabwe agreed to provide the much-needed portfolio, which it did on condition that its name never be mentioned and that the privacy of its clientele be kept intact and never be exposed. The bank's SME clientele personal details are removed from the portfolio during the initial data cleaning procedure. Besides the difficulty encountered in getting the SME loan portfolio the following limiting factors to CRM modelling phase were acknowledged:

- CRM modelling may be developed on a non-random KGB sample of loan applicants who have been pre-screened and given loan only (Ditrich, 2015), excluding potential applicants who would have been rejected under the reigning bank acceptance policy. Therefore, accuracy of the resulting CRM systems for the rejected clientele is still an open question. Due to the fact that, the resultant CRM model may not perform satisfactorily when applied on the TTD population of SME loan applicants since the data used to construct it is not representative of the TTD population. Inaccuracy of the model may erode the benefits of cost savings through commission of type I errors (rejecting good client; bad given loan) and type II errors (accepting a bad client; good given bad). To overcome this limitation which appears with certainty in CRM modelling domain, a model-based BC, a Bayesian reject inference technique to impute the missing outcome values of the rejected applicants, to minimise the effect of selectivity bias induced by the accept/reject decisioning by banks.
- Change of patterns for instance population drift, is a limiting factor since the major assumption in CRM is that past loan performance can predict future performance (Lin 2007; Wood, 2012; Kritzinger & van Vuuren, 2017). This implies that the measurement characteristics of past loan applicants who are subsequently classified as "good" or "bad" debtors can be used to forecast the credit risk of a new and future applicants. But due to the tendency for the change of patterns of characteristics over time, this means there is need to constantly re-engineering of the CRM model to stay relevant.
- The use of CRM model as an underwriting tool engenders the possibility that banks, or other FIs become heavily reliant on the CRM model technology to the detriment need for prudent judgement and exercise of domain knowledge. For ultimate prudent decision-making, the credit analyst or bank manager must use his/her discretion and judgement in support or contra the model output.
- A developed CRM model will never forecast underwriters with certainty what the future credit risk of an individual SME loan applicant will be. There is possibility of committing type I error, where some loan applicants will be denied loan even though they would have repaid and type II error, where an applicant is granted loan but eventually would default. Both errors hurt the banks bottom-line. Therefore, there is a need to statistically diagnose, validate and stress test the final CRM model before adoption to minimise the commission of the ensuing errors.

- Default data are scarce and such limitation may encourage the use of various simplifying assumptions like the independence of the determinants of credit loss, model parameters are assumed to be stable and borrowers within pre-defined risk segments are considered to be statistically identical (BCBS, 1999). By virtue of similar economic characteristics of emerging financial markets, the study was limited to a commercial bank and SMEs in Zimbabwe. The envisaged model is developed from personal data of owner(s) and financial data from Zimbabwean financial market as representative of the emerging economy. The definitions of SME, default, probability of default and loan portfolio used in this thesis were spelt out to the understanding of ZimSME bank (case study) and in conjunction with the definition enunciated in Basel II Capital Accord. Therefore, the final CRM model must not be adopted generically but adjusted to suit the respective bank and country situations. In fact, the eventual, CRM model can be applied to many other emerging countries because the issue of SME lending is a global challenge (Al Baz, 2017).

1.10 CONTRIBUTION OF THE STUDY

The contributions of research are:

- This research focuses on fundamental statistical deficiencies of most current CRM technologies. Non-random, truncated development samples (KGB) samples) appear with certainty in the domain of CRM modelling (Smith & Elkan, 2004; Chen & Astebro, 2012; Ditrich, 2015). The inherent lenders accept/reject decisioning in lending process engenders selectivity bias (Ditrich, 2015; Horstedt & Linjamaa, 2015). The subsequent truncated development sample generates a potential bias whenever loan applicants are either selected or dropped out of the loan process, because rejected applications would not have been used to develop future CRM model but just used the accepted applications only (Chen & Astebro, 2012; Ditrich, 2015; Nguyen, 2016). The shortcoming of such truncated KGB samples is that they produce biased model estimates which do not fit to evaluate credit risk on the TTD population of loan applicants. Upon that background this thesis starts off by scrutinising whether the intricacies of basic sampling theories are delicately taken into consideration in the process of developing an authentic and objective SME CRM model. A random and representative model development, AGB, sample must be constructed first to subsequently develop a SME CRM model in the context of the Basel II Capital Accord for calculating credit risk-based capital requirements for loan issuing bank (Lin, 2007; Kennedy, 2013). In CRM modelling, both predictor data and outcome data are needed for all observations (Nguyen, 2016; Chen & Astebro, 2012). In that regard, reject inference is a special case of missing value problem, since the outcome values of rejected applicants are missing not at random (MNAR) and are clearly non-ignorable missing values (Sebastiani & Ramoni, 2000; Chen & Astebro, 2012; Nguyen, 2016; Kennedy, 2013). The non-ignorability of these missing outcomes is dependent on the type of person they are and the reigning rejection policy, based on data derived from the loan application forms. To resolve the reject inference problem when data are MNAR, a theoretically supported model-based BC, a Bayesian reject inference methodical approach was used to develop an AGB sample, a sample which obeys the basic principles of sampling theory on which the envisaged CRM model would be built.

- The thesis provides, in particular for Zimbabwe and emerging economies in general, insight into whether an objective and efficient CRM for SMEs can be developed aptly for large and small banks that are likely to adopt the A-IRB approach for CRM under New Basel II Accord. In fact, this research addresses pertinent SME financing problem for a bank in Zimbabwe, given that there is a clear issue with respect to financing SMEs in this emerging economy. This SME CRM model can also be applied to many other emerging countries because the issue of SME lending is a global challenge (Al Baz, 2017).
- The research provides a detailed overview of reject inference techniques involved in the construction of an AGB sample on which the envisaged final CRM model is built. Its primary objective is to illustrate how well theoretically supported model-based reject inference techniques fare in the imputation of the credit quality (default/no-default behaviour) of the rejected applicants against the ruling rejection policy of the bank. The research adopts a Bayesian based reject inference technique, model-based imputation methodical approach called BC technique (Chen & Astebro, 2012: Sebastian & Ramoni, 2000). In CRM modelling, the credit quality of rejected applicants is MNAR according to the missing data theory (Chen & Astebro, 2012; Districh, 2015; Dong & Peng, 2013).
- The research provides an empirical application of BC, (which imbeds a generalised beta distribution, Dirichlet probability distribution), through its two (2) major bounding and collapsing steps, imputes the missing credit quality of the rejected applicant. When an AGB

sample has been developed, the traditional logistic regression modelling is applied on the modified sample. This procedural development of a CRM is done to overcome several inherent problems encountered in measuring credit risk for any bank asset. It also provides insight into the incorporation of the missingness function in the inference of the missing payment behaviour (credit quality) of the rejected clients as well as choice of Dirichlet priors. The results of the thesis provide some illustrations on why missing data should not be ignored but incorporated in the subsequent inference of the loan process (refer Figure 3.1). When missing data mechanism is MNAR, it is non-ignorable that is it must not be ignored in the inference of the missing credit quality. Reject inference techniques which improperly ignore specifying the missingness function and built on tenuous assumptions have proven not valid (Newman, 2014; van der Meijs, 2018). For MNAR condition there is need to specify an approximately authentic and correct missingness function (Chen & Astebro, 2012; Chen & Haziza, 2018).

- The thesis has demonstrated that banks that aspire to fully benefit from the regulatory requirement for capital under the New Basel II Accord (BCBS, 2006) need to produce their own IRB CRM models based on their trading credit registers. The eventual CRM model will permit banks to behave in a prudent and conservative fashion when dealing with non-traditional SME market.
- The Basel II Accord promotes special treatment for the SMEs loans in cognisance of the fact that such an exposure derives from idiosyncratic risk and much less from systematic risk factor. This research has been driven by the Basel II Accord has shown that CRM model for SME in Zimbabwe using SME owner data, enterprise demographic and financial data as well as economic data can be engineered successfully to classify SMEs into defaulting and non-defaulting firms before loan is granted and to eventually forecast PD of any future SME loan application.
- This research critique different definitions of SMEs which literature has apportioned the blame of poor SME financing on imperfect definitions this vital sector receives from country to country. From private sector development path, an SME sits between microenterprises and the corporate, therefore a perfect definition of an SME must spell out its behavioural and functionality attributes that makes it different from its extremes. As a result, the following definition of SMEs has been suggested for future adoption to enhance

bank involvement with financing of this vital sector: ***An SME is a formal enterprise with annual turnover, U.S. dollar terms, of between 10 and 1000 times the mean per capita gross national income, at purchasing power parity, (GNI/PPP) of the country in which it operates"*** This is because this definition gives an unequivocal demarcation of SMEs from its extremes which do not carry the same characteristic with it. SMEs, unlike its extremes, is distinctively growth dynamic and is a developmental asset class, able to grow a very small entity to a large sized enterprise, unlike microenterprise which rarely grow out of their size category, even corporates. The steep 10 to 1000 range emphasises pivotal attribute of intense growth dynamism that characterise the sector.

- During variable selection process, the value of predictor variable transformation is examined by using fine and coarse classing; weight of evidence (WoE) and information value (IV) for improving models' predictive accuracy. Some researcher uses principal component analysis (Al Baz, 2017) to reduce variable dimensionality for parsimonious CRM model building.
- This research has demonstrated that reject inference is a vital preliminary step to CRM modelling for SMEs with eventual goal to expand credit supply to this sector, which has been traditionally starved of financing by banks. Improved SME CRM would lead to more accurate predictions and less capital reserves for benefit of economy and businesses in general. Again, accurate and well-performing SME CRM models ameliorate the adverse effects of information asymmetry prevalent between the immature SME segment and banks.
- This research extensively investigates the CRM modelling based on the model-based imputation Bound and Collapse (BC) methodical approach for predicting the SMEs PD. In the context of statistical models-building, the eventual SME CRM model is imperatively validated prior to adoption for final use for assigning CRM scores that is to ascertain the ability of the model to predict PDs of enterprises that apply for loans from a bank Therefore accuracy of the default predictive measure.
- To emulate real bank situations, the research applies different cut-off points on the different level of financial distress using this to validate the models and examine the banks' different lending policies and risk appetite. This eventuates in having two (2) different models: Weak and Strong selection models, where weak selection model emulates for a risk-taker

banks and strong for a risk-averse banks. This is done to vary the degree selectivity bias which is common in truncated non-random KGB samples. The two (2) resulting models are assessed on their degree of the commission of type I and type II errors by using the confusion matrix and for predictive strength by using the F1-score metrics. The ROC is used to validate the two (2) models. To select the eventual model between the two (2), AIC is used. The model with lower AIC is the final CRM model, the major outcome of this research. The weak selection model came out to be final CRM model (equation 99) for SMEs at ZimSME bank.

1.11 DISSEMINATION

The major output of the research is the SME CRM model, an underwriting tool which can be adopted by ZimSME bank in an effort to expand credit supply to a sector which has been traditionally starved of loans from banks. Generally, banks consider SME asset risky due to opacity, informality and information asymmetry as a result they find it difficult to underwrite them therefore ration or refuse to do business with them to safeguard the depositors' money. On the other hand, SMEs find the banks as indispensable sources of finance for their expansionary and developmental investment objectives. Therefore, the outcome of this study comes in as a solution for the seemingly divergent vital sectors of the economy. Banks see the SMEs as a lucrative and unsaturated lending market but are short of strategies to enter this potentially promising market. This has drawn in the academia in search of ways to facilitate smooth relationship between the two (2) sectors for the benefit of the economy of a country at large.

Therefore, the findings of this study must be communicated to the following audiences; banks, SMEs, academia and businesspeople though publications, oral presentations in conferences, poster presentation, abstract and web. The study involved some surveys where individual banks and SME owners participated for the success of the study. In recognition of their resilient participation, some communication of the findings of the study will take place through letting them see copies of the final report, giving executive summary of the report and sending them a note of thanks.

1.12 WHY AT BUSINESS SCHOOL

The research fits well to be done in a business school since the research findings promote the flourishment of business in the economy at large. SMEs are described as efficient and prolific job creators, seeds of big businesses, and fuel of national economic engines (Abor & Quartey, 2010; Atistain-Suarez, 2012; Babak & Xu, 2017; Rankhumise & Lestoalo, 2019). If adequate CRM for SMEs in emerging financial markets is achieved, access to SME finance would be improved thereby enhancing the development of entrepreneurial activity in any national economy. The research outcome of this study is a business facilitator in both sectors involved (banks and SMEs). If aggregate default is reduced due to objective and efficient screening of the bad SME clients, it eventuates in a boom of profitability from the SME lending market due to the flourishment of entrepreneurial activity in the economy. This benefits the financial service sector business resulting in cut-throat competition which dovetails into good service delivery to the SMEs which, in turn, develop into big businesses, as they are a developmental asset class (IMF, 2019). On the other hand, the CRM model achieved in this thesis facilitates improved SME credit supply which dovetails into heightened business in the economy. SMEs are a seedbed of a sustainable private sector development which is an engine for economic growth of any economy. This is an illustration of how well this work augurs to be done at a business school.

1.13 SIGNIFICANCE OF THE STUDY

Despite that this thesis has a focus on Zimbabwe given that there is a strong issue with respect to SME funding in this country, the eventual mechanics CRM modelling can also be applied to many other countries since the issue of SME financing is a global challenge. The study will contribute to the improvement of SME credit supply as well as diminish aggregate default of the SME loan portfolio to a bank. In dealing with SMEs, banks currently use relationship lending approaches which are majorly dependent on “soft” information of individual loan applications. Owing to great numbers of SMEs in any economy, it is no longer feasible to underwrite them on individual basis due to the subsequent cost involved. These approaches are normally associated with small community banks which used to dominate the SME lending market. Owing to the thinning out of profit margins in the traditional market for large banks (local and international), all banks are moving into non-traditional SME lending market therefore relationship banking is no longer feasible (phase I findings). Again, most FIs have become increasingly interested in granting credit to SMEs owing to the growth and return potential in the sector. Many banks and FIs are venturing into SME lending, but not clear on how to penetrate the SME credit risk, leaving questions on how lenders quantify SME credit risk and how to price it to grow the SME portfolio and make it profitable asset for the bank.

For successful penetration into this market, banks need to adopt arms-length technologies to isolate bad SME loan applicants from good ones in an objective and efficient way. This is where the product of this thesis will immensely contribute towards helping banks and any other FIs in making decision on which SMEs to grant loan with the aim to decrease the aggregate default in safeguarding depositors' money. In tandem with the Basel II Capital Accord IRB approach, each bank must develop its own CRM model, not generic models as banks are intrinsically different as well as their respective clientele.

Academically, the thesis aims to make several contributions:

- It empirically designs a representative random AGB sample, consistent with statistical principles, on which the eventual CRM model was constructed. Non-random samples, infested with selectivity bias, are common in CRM domain. Statistically non-random samples produce biased estimates thereby making eventual CRM models incompetent to forecast the creditworthiness of loan applicants prior to loan granting. A model-based approach BC methodology which hinged on Bayesian Statistics theory and missing data theory, was adopted in order to curb the undesirable effects of selectivity bias, which most researchers ignore. The end product of BC procedure is a database made up of both accepted and rejected loan applicants, AGB, which is indicative of the TTD population of SME loan applicants. It is on this AGB where the CRM model was built. The CRM model is constructed based on logistic regression method to forecast the PD of each of SME loan applicants in the TTD pool of SME loan applicants. The CRM model include significant qualitative and quantitative credit risk drivers for the Zimbabwe SME's sample portfolio which are distinct from most studies employing data from developed countries like US (Altman & Sabato, 2005) and Sweden (Mustafa & Persson, 2017).
- In the literature on SME CRM in Zimbabwe, to the author's knowledge this is the first study to develop a CRM model for SMEs. This contribution is empirical in that the eventual CRM model can be applied to measuring credit risk of SMEs in Sub-Saharan African and other emerging economies. Databases of banks' SME customers can further be improved and made more representative of the TTD population, on which internal CRM models are

created and modified by following steps outlined in this thesis. In conclusion, a powerfully PD forecasting CRM models will make credit risk management of the whole financial system to improve to the extent that lenders can accurately price the credit risk, which constitutes the majority of risks FIs are exposed to. In tandem with the law of numbers, as the database becomes large, the accuracy of the CRM scores tends to increase. Therefore, the construction of database is paramount since the resulting CRM model does not predict credit risk on a case-by-case basis but on loan portfolio basis.

- This thesis also offers practical contributions in respect to unexplored aspects of due diligence. The major product of this thesis is the SME CRM model, an empirical tool that can improve FIs' ability to objectively and efficiently approve loans to eligible SMEs at arms-length, thereby ensuring higher credit supply coupled with minimal credit losses. An accurate, well-performing and robust SME CRM model would go a long way in helping banks and other FIs meet the Basel Accord capital requirement adequacy (Ngwa, 2010; Antão & Lacerda, 2011; Rankhumise & Lestoalo, 2019).

For successful development of a CRM model the following outcomes must be observed; development of an AGB, a random sample built through the employment of a theoretically supported reject inference techniques (Bound and Collapse Method, Multiple Imputation or any), determination of the missingness function as well as Dirichlet prior probabilities (conjugacy analysis of incomplete samples), application of logistic regression on the subsequent AGB sample, and validation of the model as well as back-testing prior adopting the appropriate eventual model using AIC model selection criterion.

1.14 VALIDATION

The measurement instruments used respectively for the bank and SME surveys were comparable in several aspects to those used by World Bank and other researchers (de la Torre, A., Soledad Martínez Pería, M. & Schmukler, S.L, 2010), confirming the reliability and validity of the research design and data collection instruments used in this work.

After a final CRM model has been developed it requires adequate diagnosis and validation, before taken aboard or recommended for adoption by the case studied bank, ZimSME. The validation process was carried out to confirm whether the developed model would serve the purpose it was built for. Will the final model applicable to the TTD population of SME loan applicants for underwriting? The other purpose of validation was to check whether the resultant “final” model was not over-fitted. As in the model development stage, the 100% of the AGB sample constituted the development sample whilst the an arbitrary 80% of it constituted the holdout sample. The resulting holdout sample was used for validation process. To carry out the validation, some goodness of fit statistical measures was employed. For this work, the Receiver Operating Curve (ROC) was used, which is a plot of the true positive rate against false positive rate at different cut-off points. The area under the ROC is used to measure the CRM model’s classification power.

1.15 OUTLINE OF THE THESIS

The remainder of this work is organised as follows: Chapter 2 sets the stage for foundation studies of CRM (historical perspective), summarising the prediction studies, and common tools used for credit forecasting modelling. This section also alludes to why SME financing is a challenge to FIs by explaining the paradox in the definition of SMEs and the derivation of the Benchmark Risk Weight (BRW) function for SMEs. Impact of Basel II on SME financing was also elaborated in depth and concluded by elaborating on binary choice models as candidate models for CRM models.

Chapter 3 looks at conceptual issues of CRM, modelling issues and literature review of methodologies used in SME CRM: judgemental and statistical scoring methodologies. On statistical CRM modelling, much emphasis was put on reject inference which is instrumental deal with incomplete or truncated samples common in financial forecasting field. It further looks at issues to do incomplete samples due to missing data problems and the proliferation of reject inference techniques in trying to solve problems of selectivity bias prevalent in CRM domain. Bayesian analysis of incomplete samples and Bound and Collapse methodology were extensively discussed in detail. Conclusively reject inference was reduced to a missing data problem thereby qualifying the use of the Bayesian model-based reject inference technique adopted in this work. In fact, the technique proposes a framework that can be used to generate a random AGB sample on which the final model was eventually developed.

Chapter 4 details the designs and methodologies adopted for each of the two (2) phases entailed in this research work. The first phase entails an exploratory study where a survey-based research design was implemented, and a case-study research design was instituted to develop a CRM model for the case-studied bank; ZimSME bank.

Chapter 5 summaries the analysis of data obtained from the two (2) surveys (bank and SME) carried out in the first phase of the work. The output was displayed using different descriptive statistical diagrams to exposed inherent inter-connection and relationships between or among variables of interest to eke out the intricacies and coexistences between banks and SMEs.

Chapter 6 gives a summary of the development of a CRM model for the case-studied bank.

Chapter 7 discusses the findings from both phases thereby answering to objectives and research questions set for this work.

Chapter 8 summarises key contributions of this work and highlights opportunities for future research.

CHAPTER 2: THEORETICAL FRAMEWORK

2.1 INTRODUCTION

This chapter describes the evolution of CRM modelling in the lending industry, spelling out the approaches used to mitigate bank failures. In fact, the chapter gives a brief evolution of the concept of CRM and why it has become so topical over the years. The risks to which the FIs are exposed due to their lending activity are discussed and more emphasis is given to SME credit risk. This would lead to an overview of CRM under the Basel II Capital Accord frameworks. The Basel II frameworks' treatment of the SMEs would follow, revealing why banks consider SMEs differently from both retail and corporate exposures when it comes to lending. The chapter ends up by having a close look at problems of the implementing the New Basel Capital Accord to CRM and explains how this can be improved by using credit risk models.

2.2 EVOLUTION OF CREDIT RISK: HISTORICAL PERSPECTIVE

Default risk must be one of the oldest sources of financial risk that one can think about and yet there has been a surge of interest in the area only in the 1990's (Madeira, 2018). This is true, as one looks at the trajectory evolution of bank failures due to default by firms. In 1932, Fitzpatrick uses financial ratio analysis to measure risk drivers of firms to mitigate bankruptcy. He finds out that two (2) financial ratios, Net Worth to Debt and Net Profits to Net Worth are significant for ex-ante assessment of loan applicant firms. This work was followed by Smith and Winakor's work in 1935, Merwin in 1942, Chudson, in 1945 and Jackendoff, in 1962. All these studies were centred on financial ratios of firms to help banks to avoid lending to risky firms. These are termed bankruptcy prediction studies, which only focussed on single type of quantitative default drivers.

It is remarkable how little attention was paid to CRM in the past (Farrar, 2010). Only a slight improvement was initiated by Beaver (1966) who changed the tide of research in CRM when he suggested as future research that multiple financial ratio analysis could better the univariate predicti ve ability of bankruptcy. So, began the evolution of bankruptcy prediction models in the multivariate direction and alternative approaches like Bayesian Statistics, Data Envelopment Analysis (DEA), MachineLearning (ML), Support Vector Machine (SVM) and Genetic Algorithm (GA) were also employed in attempt to quantify credit risk of companies prior to loan granting.

Altman (1968) was the first researcher to take up the call of Beaver, when he used the multivariate discriminant analysis (MDA) to develop a five-factor model to predict bankruptcy of manufacturing firms. His model, referred to as the "Z-score Model", has a 95% predictive ability for initial sample one year before failure. Quite excellent results, but due to the dynamics of borrowers' environment (population drift) and that of the market and selectivity sample bias, the Z-score predictive power wanes with time as it goes down to as little as 29% predictive accuracy four (4) years before complete failure. This prompted researchers and academics to fill the void that a number and complex of bankruptcy prediction models have increased dramatically, since Altman's work.

The point of departure for most research is the clarity of the definition of 'failure' as is used in finance literature. There is a diverse set of definitions of failure used for prediction studies (Pretorius, 2008). This makes it difficult to compare the effectiveness of failure prediction models being constructed by different researchers. Others define failure as 'actual filing for bankruptcy or liquidation, whilst another school of thought, defines it as suffering financial stress or an inability to pay financial obligations. Most studies adopt the latter as well this thesis.

The bankruptcy prediction studies gave birth to what we now call credit risk forecasting/measurement concept, a leading topical issue in modern finance, of which the concept of default risk rather failure is pivotal. During the past 10-15 years, marked progress has been achieved in credit risk forecasting. This could be due to an upsurge of CRM at FIs which is increasingly becoming complex and reliant on quantitative solutions. In that respect, there is a correspondently remarkable surge of interest in default risk studies in the 1990's (Madeira, 2018).

In actual fact, analysis and development of SME CRM by banks has made much progress since the bad loan problems of the 1990s and global financial crisis. For instance, the implementation of the IRB system that ranks enterprises according to financial strength as well as their major financial indicators, such as the capital adequacy ratio (Nemoto, Yoshino, Okubo, Inaba & Yanagisawa, 2018). This also saw an upsurge of financial risk modelling since the 2000s, which became pervasive as noted by Financial Services Authority (2003). Statistical methodologies to estimating the probability of bankruptcy became handy by using PD to determine loan extension as well as loan rates spreads. To reduce cost involved in lending, CRM methodology does not manage risks on a case-by-case basis but manages the risks on loans throughout the portfolio control based on the law of large numbers (Nemoto *et al.*, 2018).

As found with the concept of 'failure' in the early studies of bankruptcy prediction, there are also different definitions of 'default', a crucial component of CRM. Bank for International Settlements (BIS) defines default as the situation when an obligor is unlikely to pay its credit obligations, or the obligor is past more than 90 days on any material credit obligations. Many academic researchers (Kanapickiene & Spicas, 2019; Agrawal & Maheswari, 2019) adopt the second part of this definition. A scrutiny of it reveals some deficiencies which may cause differences in CRM methods. As some obligors may pay off their obligations back even after 90 days, overdue payment which can be treated as a payment indiscipline rather than real lack of income to pay a loan.

The point is that default does not necessarily imply losses (Gonzalez-Watty, 2016). This clarification is necessary because default and loss are critical terms to CRM. This makes comparison of CRM models difficult for the measurement of default and loss differ from article to article. For instance, in the US, researchers define default as when and only when obligor is "re-organisation" bankrupt, in accordance to the US Bankruptcy Code, referred to as Chapter 11 Bankruptcy (Ba Zhang, 2009). Still this definition is deficient, because a firm can default on debt obligations but still not declare bankruptcy. Conclusively there is no universal or better definition of default, when measuring credit risk. In this thesis, the operational definition of default is the one adopted by the BIS.

CRM modelling has evolved dramatically over the past twenty years in response to several secular forces that have made its application more important than ever before. Among these forces have been:

- Worldwide structural increase in the number of bankruptcies,
- A trend towards disintermediation by the highest quality and largest borrowers,
- More competitive margins on loans,
- A declining value of real assets (and Therefore collateral) in many markets and
- A dramatic growth of off-balance sheet instruments with inherent default risk exposure, including credit risk derivatives (Altman & Saunders, 1998).

These secular forces have taken academics and practitioners to a new era, an era of developing innovative and more objective early warning mechanisms in the financial lending market. There is growing need to develop and adopt CRM tools that efficiently calibrate credit risk of a portfolio of loans rather than of individual loans, as the demand for loans is incessantly growing. Loans must be correctly priced and forecasted to avoid increases of bankruptcies and financial crises. Quantitative revolution in pricing and hedging credit risk has begun as banks are increasingly adopting and developing sophisticated quantitative CRM models to allocate capital internally and to convince regulators that capital adequacy standards are being met (Ishtiaq, 2015; Butt, 2017; Song, 2016).

All this points to the fact that reckless subprime lending can no longer be practised as the repercussions are disastrous to the worldwide economy. Predicting and mitigating default events is at the core of appropriate and efficient CRM and this can be greatly helped by employing suitable quantitative CRM models, without however precluding the reliance on human expert judgement (Kennedy, 2013; Konovalova, Kristovska & Kuduriska, 2016; Shen, Nguyen & Ojiako, 2013). This means the CRM stage of underwriting process must be quantitatively and objectively thorough to minimise credit risk.

2.3 CRM

In the following sub-section, subjective CRM is explained in details as most FIs use relationship lending, which fundamentally depends on 'soft' information in assessing creditworthiness of loan applicant. The subjective CRM assessment of lending relationships is mainly based on credit historical data and on the acquaintance with the loan applicant (Petersen & Rajan, 1994; Berger & Udell, 1995), The section gives a brief description of how CRM has transversed over the past 20 years in tandem with technological advances in data handling and statistics.

2.3.1 Subjective CRM

Credit risk arises from non-performance of a borrower, either as an inability or unwillingness to perform in the pre-committed contracted manner (Kipsang, 2014; Muritala & Taiwo, 2013). Therefore, ex-ante assessment of loan applicants is a crucial step in any lending activity, since that is when a crucial decision of granting or not is made to minimise credit risk. Originally methods used in pre-credit granting assessments were purely subjective and based on the view of the expert underwriter. In fact, 20 years ago, most financial institutions depended virtually on subjective analysis or the so-called expert systems to assess the credit risk on corporate loans (Altman, 2018; Gakure, 2012; Wang, 2013; Tsai, Li, Wu, Zheng & Wang, 2016).

During underwriting process, what mattered most were the 4C's of credit - character of borrower (reputation), capital (leverage), capacity (volatility of earnings) and collateral (Altman & Saunders, 1998) or the 5C's of credit - character, capital, collateral, capacity and condition (Akhavein, Frame & Lawrence, 2001; Mester, 1997). This constitutes relationship lending, contrast to transaction-based underwriting, which is being adopted by contemporary bankers. Owing to relationship lending's subjectivity; financial institutions are increasingly moving away from the expert systems over the past years towards systems that are more objectively based (Akelola, 2012; Wang, 2013; Ball, 2016), in fact adopting arms-length technologies to help decide whether to grant or not loan to an applicant.

In trying to move away from expert underwriting, the first stop was the univariate accounting-based CRM systems. Here the ex-ante assessment is based on comparing various key accounting ratios of potential borrowers with industry norms. After Beaver's (1966) suggestion of multivariate approach to financial ratio analysis, Altman (1968) combined key financial ratios and weighted them to produce either a CRM score or a PD measure (Altman, 2018). This was the birth of multivariate, quantitative CRM approach in the ultimate goal to minimise credit risk in lending business. This approach proved a thorough scrutiny of loan applicants for it could 95% accurately predict bankruptcy of a firm one year before complete failure fours later when is supposed to be re-engineered (Altman, 2018).

Up to so far, there are four (4) major methodological approaches to developing multivariate credit-risk measurement modelling systems, which have been tried and tested:

- The logit model (Ohlson, 1980; Altman & Sabato, 2005; Bolton, 2009; Memic, 2019; Ranganathan, Pramesh & Aggarwal, 2017).
- The probit model (Koh, 1987).
- The discriminant analysis model (MDA) (Altman, 1968; Memic, 2015).
- Artificial Neural Network (ANN)

- Machine Learning (Kennedy. 2013; Dada, Bassi, Chiroma, Abdulhamid, Adetunmbi & Ajibuwa,, 2019; Kraus, 2014; Yi, 2019; Roelofs, 2019; Vabalas, Gowen, Poliakoff & Casson, 2019).
- Support Vector Machine (Smith & Elkan, 2004; Goh & Lee, 2019)

According to the Journal of Banking and Finance publications, multivariate discriminant analysis (MDA) and logistic models top the list of most dominant methodologies in CRM modelling mostly for corporate and retail credit exposures. But in the late 1990's, the Genetic Algorithmic (GA) family of methods, Neural Network methodologies have become the most popular to CRM modelling.

Table 2.1: Model types commonly used in CRM modelling

	Discriminant Analysis (DA)	**Logit Analysis**	**Probit Analysis**	**Neural Networks**	**Others**
1960's	2	0	0	0	1
1970's	22	1	1	0	4
1980's	28	16	3	1	7
1990's	9	16	3	35	11
2000's	2	3	0	4	3
Overall	63	36	7	40	26

Note: "Other" methods include linear probability, judgemental, Cusp catastrophe and Cox proportional hazards models.

2.4 BANKS RISKS AND SME LENDING APPROACH

In the following sub-section, the concept of Small and Medium Enterprises (SMEs) is explained as it central to this thesis. Definitions of SMEs are different country by country. In some countries the criteria for the categorization is capital, in some countries it is based on the number of employees, and other countries use a mixed criteria like Japan where it varies in each business sector.

2.4.1 Definition of Small and Medium Enterprises

Like in any field of analysis it is of crucial importance to define the precise meaning of terms used and if no consensus can be reached on the universality of the meaning of the crucial terms of reference, operational definitions are adopted for clarity of analysis concerned. This is quite true about the definition of SMEs. The definition of this concept varies from country to country and even within a country, therefore not universal. In any economy, developed or developing, SMEs are acknowledged to be of great importance to economic growth and development, but regardless of that, there is no generally agreed definition of this indispensable sector of the economy. This is so because of numerous factors of socio-economic nature influence the definition of SMEs.

In the study of SMEs, it is a major task to give a precise definition of the sector since there are many definitions pronounced by different organisations and government departments in one country or different countries. The non-existence of uniform definitions of SMEs is thought to be directly affecting the sector in accessing external financing from banks and other FIs. Regardless of this misnomer, globally it is generally agreed that SMEs constitute major engines of growth and employment in any economy (Nyamwanza, 2014; Karadag, 2016).

For SME financing analysis, there is need for a unified operational definition for SMEs appropriate to developing economic structure and business environment. This would, in turn, play a contributory effect to consistency and efficiency of compiling data and therefore help researchers and policy makers with valuable statistics on credit extended by FIs to this crucial sector of the economy. In some instances, a small firm that performs exceptionally well in its credit material obligations, it is classified as a corporate firm. Consequently, confusion ensues on when to treat SMEs as retail or corporate thereby the sector does not receive the attention or support justified by its pivotal role in the economy of any nation.

From an international point of view, SMEs constitute a diverse and dynamic group of enterprises. The Organisation for Economic Co-operation and Development (OECD) has noted that the characteristics of an SME reflect the economic, the cultural as well as the social dimension of a particular country, therefore accepts the diversity of the definition of SME as countries are diverse. There is consensus on the criteria normally adopted in country to country definitions of SME; number of employees, invested capital, turnover and industry type. The OECD countries, for statistical analysis, use the number of employee criterion, thereby defining an SME as a firm with fewer than 500 employees. This is in contrast to the definition given in the Don Cruickshank Report (2002) which defines SMEs as those firms which are no longer treated as personal customers by providers of money transmission services and credit but are too small to have direct access to competitive capital markets, consisting mostly of enterprises with turnover of up to 10 million pounds, or employment of up to 250 people. The Standard Bank of South Africa, for credit management purposes, defines SMEs as firms with turnover of between R150 000 and R5 million per annum.

This divergence of definitions of the same concept may inhibit clarity of an eventuality of the intended analysis, therefore the need for an operational definition of SMEs given the economic, cultural and social status of a country or region. Some countries define SMEs differently in the manufacturing and services sectors of the same economy. For instances, in some countries, like Canada, they break down the SME definition: small business as one which has fewer than 100 employees if business is a goods-producing based or fewer than 50 employees if the business is a service-based. A firm employing more than these cut-offs but fewer than 500 employees is classified as a medium-sized business.

In South Africa, the National Small Business Act framework has been adopted to counter divergent view on SMEs when treating this vital economic sector. Therefore, South Africa defines SMEs as: small enterprise is a firm which employs up to 50 people and exhibit more complex business practices and medium enterprise employs up to 100 or 200 for the mining, electricity, manufacturing and construction sectors, characterised by decentralisation of power to an additional management. Under the Basel Capital Accords, an SME is understood to be a company where the reported sales for the consolidated group of which the firm is a part is less than 50 million euros (Lin, 2007; Serov, 2011; Vandenberg, Chantapacdepong & Yoshino, 2016).

2.5 CRITIQUE OF SME TRADITIONAL DEFINITIONS

In emerging economies, the official definitions of SMEs are flouted with a lot of inadequacies and inconsistencies of the sector which consequentially lead to distortions in the allocation of funding for private sector development. An appropriate and well-performing measure of the SME relative size is more needed than the conventional measurements by number of employees or value of assets. In that regard, volume of turnover would be the most appropriate measurement of SME relative size. Basel II Capita Accord suggests a single definition of SMEs for multiple countries in different stages of economic growth; this in turn leads to additional distortions in the way this sector is treated. Such distortions, in SME development policy circles, have led to question the rationale behind spending taxpayers or foundation moneys on SME initiatives, causing unclear SME development policy and funding problems by governments and non-governmental organisations (NGOs).

The working definition/concept of SMEs must be able to answer the following questions in respect to private sector development process in an economy:

- Where do large firms come from?
- How does a country diversify its economy?
- What is an SME? (Gibson & van der Vaart, 2008).

The term SME has been in existence for the past quarter century referring to a segment of businesses occupying the space between microenterprises and corporate, which means they present challenges and opportunities much different from the two (2) extreme groups (Gibson & van der Vaart, 2008, Nyamwanza, 2014; Kanapickiene & Spicas, 2019; Sitharam & Hoque, 2016). The current definitions of SMEs are not able to distinguish the SME sector from the two (2) extreme sectors that border it, thereby diluting the claim that SMEs are the backbone of any economy, a claim which has been made without supporting evidence and effort to understand what constitute an SME. How can the sector be a backbone when there is heterogeneity in its conception? Such a distortion does more harm than good considering the much-needed private sector development majorly in emerging economies (Wangmo, 2015). Gibson and van der Vaart (2008) acknowledge that there is still a gloomy picture in reaching an international consensus on the universality of an SME definition. Such disparities being observable among SME definitions have caused complexities in the international SME dialogue, thereby giving openings to mistrust and misdirection of financial resources much needed for the development of this sector. Therefore, there is great need for a less imperfect SME concept than the current mosaic of SME definitions in common use in different countries at different stages of economic development (Nyamwanza, 2014; Gibson & van der Vaart, 2008; Yoshino & Taghizaden-Hesary, 2016). In turn, the multiplicity of SME definitions makes SME funding, SME policy analysis and development an unsystematic and less perfect process. Are functionality attributes of the current definitions such as number of employees, asset and turnover relevant to a universally accepted more perfect SME definition?, ask researchers. In response to that interjection, Gibson and van der Vaart, 2008, suggest that an SME definition would be more meaningful if the functional and behavioural attributes are much more pronounced rather than the Procrustean quantifications of employees, assets and turnover. Justifiably functional characteristics answer to the reasons for which taxpayers' money is used to help private sector development through adequate public funding of the SME sector. To clearly expose the aptness of the prevailing SME definitions across the globe, it is worth scrutinizing each of the functionality attributes that embody them.

2.5.1 Critique of SME definition by employment

Since an SME is a sector that occupies the space between the microenterprises and corporate, therefore a good proxy SME definition would be the one which distinguishes the sector from the two (2) extremes in the private sector development trajectory. A profound look at the functionality attributes employment and asset sizes as determinants of an entity bounded by microenterprises and corporate firms have been seen as detrimental attributable measurements. At the same time, turnover provides an adequate measurement which proxies the functional and behavioural attributes of SMEs. In other circles, these two (2) components which constitute the concept of SMEs are disjointly defined. Some researchers (Ba Zhang, 2009) have pointed out that the separation of "small" and "medium" is also another source of distortion and detriment, considering that, Gibson and van der Vaart, 2008, recommend the merger of "small" and "medium" into single compact concept. This distinction dilutes the broadness and oneness of the concept into narrow spectrum thereby leading to loss of essence and status of being the backbone or seedbed of economic diversity and growth dynamism. It is on that background that US banks refer to this family of businesses as small businesses, a compact concept (Ba Zhang, 2009).

Why defining SME by the number of employees is misleading? This incorrectly points to the conjecture that the larger an enterprise is, the more employees it has, implying that for growth the enterprise must be measured by the number of people it employs, a notion which is not financially justifiable. The SME definition by use of number of employees is detrimental and risky in the sense that it promotes classifying businesses by their inefficiency and lack of value addition (Nyamwanza, 2014; Ackah Vuvor, 2011; Yoshino & Taghizaden-Hesary, 2016: Wangmo, 2015). The larger the number of employees does not mean efficiency and justification of existence of the enterprise. Judging the financial worthiness an enterprise by number of employees may send wrong signals to financiers and taxpayers which may result in distortions in private sector development process in an economy (Gibson and van der Vaart, 2008).

2.5.2 Critique of SME definition by assets

In most SME definitions, the asset criterion is constantly included as a measurement of business size although precision in their estimates of fixed assets normally underestimates them. Therefore, this criterion is irrelevant since SMEs rarely have assets in instances where asset taxes are required. In most economies there is no systematic count of what assets are used for when defining business size, some governments use fixed assets and land while others use only fixed assets, no consensus in that respect. This makes cross-country comparison of SMEs very difficult. In times of inflation or economic depression, the measure of business size using asset criterion is rather futile as there is glaring understatement of the "true value" of business assets which is never corrected even if the situation improves. In other situations, fixed asset value may diminish in contrast to increase of revenue and employment thereby causing distortion in the definition of business size. Therefore, in summary asset-based SME definition does not recognise capital efficient in as much as employment-based definition tend not recognise labour efficiency (Ackah & Vuvor, 2011; Yoshino & Taghizaden-Hesary, 2016; Gibson & van der Vaart, 2008; Muriithi, 2017). For banks and other FIs, it is illogical to appraise an enterprise for a loan based on the number of employees and net assets of the firms. Such decisional measures do not give the lenders the security that they would get their money back because they do not pinpoint the actual size of the enterprise being lent the depositors' money. This is the reason why SMEs currently find difficult in securing credit from banks and other FIs.

2.5.3 Critique of turnover-based SME definition

Logically, when entrepreneurs are asked about the size of their businesses, they do not answer basing on the number of employees and even based on net asset but based on sales they made the previous business cycle. For any SME that might have graduated to a corporate level is appraised for loan allocation by banks based on growth in turnover and market share. This implies that turnover proves to be a good measure of business size therefore an SME definition based on turnover looks quite objective and measurable in the determination of the size of business. The measurability of business size by turnover closely mirrors functional and behavioural attributes of an SME (Pratt & Virani, 2015; Kerr, Kerr & Xu, 2017). Gibson and van der Vaart (2008) further pointed out that the advantages of adopting a turnover-based SME definition are as follows; indexation to the international currency US dollar and cross-industry consistency. The dollar value of a firm's sales determines the position of a business on the business size spectrum in an economy. Whilst this is not possible to rank enterprises based on the number of employees as well as based on asset value because such criteria vary among government departments as well as organisations in the same economy and across economies. Therefore, indexation of sales to the US dollar looks universal and convenient to distinguish the SME sector from microenterprises and corporate extremes.

Owing to the current distortive and deficient conceptions of the SME sector, there is no cross-industry consistency in terms of number of employees as well as assets. Comparative analysis of business sizes across industries in the same economy or across economies would be possible if their turnovers are the same for the SMEs share several organisational attributes. The turnover-based definition of SME would help reject the notion that the manufacturing is the dominant standard and default proxy SME for all industries, a notion derived from the incumbent SME concept. This is a wrong perception definition as it excludes the agricultural sector which by far the largest sector in the emerging economies.

By the same analogy, the upcoming service industry is also on the side-line of the manufacturing sector. A turnover-based definition of the SME would help include all SME sectors and improve SME financing by commercial banks and other FIs, which scrutinise financial statements before any loan granting. In that respect turnover information majorly help banks extrapolate the repayment behaviour of the SME loan applicants, a thing not possible from employment and asset-based SME information.

2.5.3.1 Turnover definition of SME by formula

Owing to the current distortive deficiencies of the incumbent multitudes of conceptions of SMEs it is worthwhile come up with a consistent and functional definition of this vital sector of any economy. A concept which induces comparability of the SME sector across countries in the world, distinctively distinguishes it from micro-enterprises and corporate firms which border its space in private sector development trajectory, and unequivocally pinpoint its typical developmental dynamism that best describes it.

Currently some countries do not define SMEs as a single component but define 'small' and 'medium' enterprises as if they are different components of the sector. This is contrary to reality as it relates to issues of funding eligibility and other assessments of this SME sector because small businesses are regarded in the same manner as medium businesses; even scholarly studies of SMEs do not highlight that distinction. Instead of locating correctly and distinctively the SME sector as a broad sector in private sector development trajectory, the current SME definitions seem less broad due to the unwarranted division into two narrower distinct subgroups. This, in turn, dilutes the dynamic growth characteristic that SME sector must carry (Gibson and van der Vaart, 2008; Yoshino & Taghizaden-Hesary, 2016).

Therefore, the SME must be looked at as a single size-compact group that is distinctively sliced off from the microenterprises and large firms. In this regard, Gibson and van der Vaart, 2008, 2008, suggest that if economies are to benefit from dynamic growth characteristic of SMEs, it is time for policymakers to make the *de facto* merger of 'small and medium' a *de jure* recognition of SME as a developmental and dynamic asset class or a single size compact group.

From the preceding discussions, it is evidently visible that employment and asset-based definitions of SMEs are misleading insofar as highlighting the distinctions in the phases of private sector development. Analogically, using one absolute number of turnovers or any other firm size measure may also mislead by failing to capture the functionality and behavioural aspects of the sector as a growth dynamic and developmental asset class. The envisaged definition of SMEs must consider the degree of environmental economic development of a home country in consideration. In fact, it must be a turnover-based definition, since turnover is a good proxy measurement of firm size although turnover figures alone do not explicitly reveal consistency in the definition of businesses by relative size in their respective economic environments.

Economic environments of countries are comparable by using economic indicators like Gross National Income (GNI) and Gross National Income/Purchasing power parity (GNI/PPP). The former indicator effectively compares one country's economy to the other whilst the latter is a standard proxy measure for comparing standards of living between countries. In fact, the GNI/PPP best spells out the functional attributes of an absolute amount of income to the citizens of each respective economy whilst GNI just mirrors how much money is in the national income coffers. From a relevance perspective, GNI/PPP is more appropriate in measuring levels of wealth and poverty in a given economy of a country.

A definition of SMEs capable to distinguish enterprise size across countries from relative enterprise size within a country is needed, therefore the corresponding need to improve on the turnover-based definition by making relevant adjustments to capture the economical contextual environment of the country in question and including firm size measurements (Berisha & Pula, 2015). A satisfactory conclusive definition of an SME must consist of the following impeccable components:

- Replacement of a single nominal maximum cut-off in defining SMEs for all countries, with a formula that adjusts a single, readily available information to provide a relevant definition of SMEs taking into cognisance the economic contexts of individual country.
- Adoption of annual turnover as the best single measurement of business size.
- Designation of "SME" as single size compact group within a given range, no more division of the sector into 'small' and 'medium' and establishment of limiting criteria to exclude both micro-enterprises and corporate firms from this developmental asset class. (Gibson & van der Vaart, 2008, 2008; Reeg, 2013; Chin & Lim, 2018; Berisha & Pula, 2015).

In conclusion, Gibson and van der Vaart (2008), propose an SME definition using a turnover-based definition, which is open for further qualification in business, policy and academic circles:

"*An SME is a formal enterprise with annual turnover in U.S. dollar terms, of between 10 and 1000 times the mean per capita gross national income, at purchasing power parity, (GNI/PPP) of the country in which it operates*"

This definition conspicuously and precisely demarcates an SME from either micro-enterprise or corporate extremes that borders it in the private sector development path and spells clearly the growth dynamism which must characterise SMEs differently from microenterprises and corporate firms. This definition bases on readily available data as GNI data for countries on the globe are annually published by the World Bank, in either ranking by country or by country income groups. The steep 10 and 1000 range indicates the essential attribute of intense growth dynamism of SME as a developmental asset class. SME growth dynamism implies that an SME can tremendously grow from very small to a large sized enterprise unlike micro-enterprises which rarely grow out of their size category and large firms normally begin as large firms. In fact, SMEs start out larger than microenterprises, and with substantial initial investment have capacity to grow into 1000 times their initial size. Therefore, SME financing should be to facilitate dynamic growth of the sector from being small to large. The 10 to 1000 times GNI/PPP range spelt out in the formula definition of SME mirrors the explicit functionality of SMEs in the context of emerging economies.

The Gibson and van der Vaart definition of an SME is quite relevant to the bank financing of the sector, which must be directed to facilitate dynamic growth from small to large as figured by the wide range 10 to 1000 times of the GNI/PPP. The table below shows the relevant and consistent SME definitions for emerging economies juxtaposing the current Global SME definitions:

Table 2.2: Proposed SME Turnover-based vs World Bank and UNDP SME definitions for some Developing Economies

Country	Country PC/GNI	Country PC/GNI/ PPP	New definition Lower - Upper	World Bank Lower -Upper	UNDP range Lower - Upper
Morocco	1,730	4,360	43,600-4,360,000	100,000-15,000,000	67,000-10,000,000
Egypt	1,250	4,410	44,100-4,410,000	100,000-15,000,000	67,000-10,000,000
Ghana	450	2,370	23,700-2,370,000	100,000-15,000,000	67,000-10,000,000
Tanzania	340	730	7,300-730,000	100,000-15,000,000	67,000-10,000,000
Malawi	160	650	6,500-650,000	100,000-15,000,000	67,000-10,000,000
Zimbabwe (*)	1,020	2,190	21,900-2,190,000	100,000-15,000,000	67,000-10,000,000

Of the Table 2.2, column 3 provides PC/GNI/PPP for each of the country and column 4 gives SME formula definition which results when the PC/GNI/PPP is multiplied by our constants 10 and 1000. Columns 5 and 6 give SME definitions used by the World Bank and UNDP respectively.

The proposed new SME formula definition takes the local state of the economy into context and adjustable as the context changes unlike the one-size-fits-all SME definitions used by the World Bank and UNDP respectively as if economies of different countries are uniform, far from reality. The latter definitions do not capture the diversity that characterize different economies whilst the former keeps consistent approach across countries thereby giving correspondingly separate definition adjustable to consistently reflect the per capita size of its economy using broad turnover range defined by both minimums and maximums. In fact, SME formula definition is country specific dependable on PPP which captures significant differences in purchasing power parities and exchange rates among economies. It is evident that the World Bank and other international institutions do not use PPP to determine SME loan size limits. Accordingly, these global institutions, the standing slogan is a dollar is dollar without taking into cognisance that a dollar is less money where it buys less than where it buys more. This difference is well captured by PPP which provides us with an additional adjustment in favour of consistency with local economic context (Gibson & van der Vaart, 2008). The explicit scaling of SME definitions to local economic context gives weight to SMEs as they are viewed as vital players in the global marketplace. Therefore, SMEs' dynamic growth and developmental aspirations must be underpinned on the production and services of the domestic markets.

Home market-based SME success is sustainable due to relatively lower transport costs, reduced vulnerability due to currency and trade restrictions. This helps SMEs focus on attaining financing, management and technical assistance for their eventual growth and development. It is upon this background that the turnover-based SME definition proves its worth for it matches the SME sector with the realities of the home markets because it reflects more domestic economic conditions than a more homogeneous global marketplace as do other commonly used in the multiplicity of SME definitions.

Regardless of the forthcoming perfections of the incumbent SME definitions through the adoption of the turnover-based definition, this may require drastic changes in policies, agreements and in other documents. Great effort of change management may inhibit the much-needed improvement on the distortions caused by the current multitude, deficient and inconsistent SME definitions.

Improvement of access to long-term finance to SMEs is the only panacea to foster growth of a greater proportion of SMEs into new corporate firms in any given economy. Therefore, any SME development policy must work out on how improve that paramount accessibility to finance. This is where determination of appropriate SME size range comes handy in any SME development policy. Significant SME expansion in their early years of existence is crucial for such expansion would lead to significant growth in revenues, wages paid, taxes paid, import substitution and increase in exports, thereby propping up the national economy at large.

It is paramount that any SME development policy must seek to focus on expansion of financing systems. This path has been pursued by the World Bank and other multilateral and bilateral institutions which heavily invested in risk capital, venture capital and other forms of capital tailored specifically to improve SME financing. Owing to poor SME definitions such effort has not yielded desired eventuality because the SME scaling is not consistent with economic contextual conditions of a country in question.

The one-size-fits-all nature of the incumbent SME definitions of the World Bank and UNDP stifle SME development and dynamic growth which should stimulate the developmental vibrancy of the SME sector. Currently sizes of SMES in two (2) different economies are incomparable, for instance, saying the size of SMEs in Argentina and Zimbabwe are the same makes no sense. Therefore, the SME financing policy of the two (2) countries are never congruent because the respective local economic prevailing conditions are not identical as presupposed by incumbent SME definitions used by multilateral and bilateral institutions.

With regard to SME definitions, Gibson and van der Vaart, 2008, made the following summative observations:

- The degree of diversity and conflict among official SME definitions is currently so great that it borders upon, or surpasses, irresponsibility not to reconsider how they are derived and applied.
- Multi-country SME definitions cannot legitimately be said to be consistent among countries if they do not consider the differing levels of poverty among such countries and the differing levels of relative competition among private enterprises.
- Official national definitions vary too greatly in proportion to national economies for responsible use by international organisations.
- To avoid further distortions in the generation of SME policy and the resulting misapplication of funds, the major multilateral development institutions should take steps,

as a group, to introduce some coherence of rationale among their SME definitions and encourage the same to individual national governments.

- Microenterprises and SMEs are distinctly different, do not naturally elide in an unbroken continuum, and cannot be usefully discussed together. Definition by turnover has multiple advantages over definitions by either employment or assets, given that it is the most consistent across sectors.

2.6 BASEL CAPITAL ACCORDS

This section covers briefly the historical perspective of the New Basel Capital Accord, Basel II, which forms the fundamental basis for SME CRM modelling, especially for retail and corporate exposures to FIs. The impetus for Basel Capital Accords (Basel I, II and III) came into being as a cushion against credit losses, of the major banks, which had become dangerously low after persistent erosion through competition (Lall, 2009; Seliane & Sello, 2015; Ozili & Outa, 2017). The discussion emphasises on the change of the lending activity, adoption of risk-adjusted return approach to loans and the favourable treatment of the SME exposure has received in the wake of lack of financial support to this sector because of perceived low levels of credit worthiness. Therefore, the Basel framework gives the best background for SME CRM modelling.

2.6.1 Basel II framework

In June 1999, the Basel Committee released a proposal to replace Basel I Capital Accord with a more flexible and more risk-sensitive framework, on which more than 200 comments were received from bankers and supervisors worldwide (Fullenkamp & Rochon, 2014; Flanney & Bliss, 2018). This implies that the new Basel II Capital Accord framework is born out of wide consultation and the unfavourable conditions presented by the Basel I regime. Safety and soundness in today's dynamic and complex financial system can be attained by the combination of effective bank-level management, market discipline and supervision (Fullenkamp & Rochon, 2014; Flanney & Bliss, 2018; OECD, 2018). The framework's major emphasis is on bank's own internal control and management, the market discipline and supervisory review process.

In 2004 a New Capital Accord, Basel II was proclaimed as a replacement of the one-size-fits-all regulatory framework, Basel I. Basel II's main purpose is to further strengthen the soundness and stability of the international banking system through encouraging banks to improve their risk management practices in the allocation of capital as buffer against losses. This framework provides a spectrum of approaches from simple to advanced methodologies for the measurement of both credit risk and operational risk in determining capital levels. In this regard, subject to supervisory review process, each bank adopts approaches which best suits their level of sophistication and risk profile (OECD, 2018). This is a deliberate incentive for stronger and more objective and accurate CRM approach. In fact, it is designed by regulators to meet the remaining perceived regulatory challenges they and their banks face (Jones & Zeitz, 2018; Claessens & Kodres, 2014; Gottschalk, 2014; Hoskins & Labonte, 2015). Aubier (2007) views that this New Capital Accord, Basel II, is designed to ensure that SMEs, though theoretically riskier than big firms, are not hindered from accessing credit, among other risk-related objectives, because of its 'SME inclusivity' flare.

Its predecessor regime (Basel I) stereotyped this SME sector riskier to the extent that banks could hardly do business with SMEs, regardless of the sector's assumed socio-economic significance. Aubier (2007) further spells out the main purpose of Basel II as to further strengthen the soundness and stability of the international banking system through encouraging banks to improve their risk management practices by incorporating new financial risks into the allocation of capital to buffer bank losses due to the respective risks. Unlike Basel I, which just considered credit risk in allocating capital and this New Capital Accord is too prescriptive; therefore, it is seen as an improvement of its predecessor. It intends to provide approaches which are both more comprehensive and more sensitive to financial risks than Basel I while maintaining the overall level of regulatory capital (OECD, 2018).

2.6.2 Standardised approach to CRM

This is one of the tools put at the disposal of the FIs by the Basel II which is based on external rating systems. In fact, it is conceptually the same as the Basel I, but more risk sensitive (OECD, 2018). It brings with it, distinctive character from the Basel I when it comes to aligning regulatory capital requirements more closely with key components of banking risks by the introduction of a wider differentiation of risk weights and wider recognition of Credit Risk Mitigation (CRM) techniques. It is more risk sensitive as it produces capital ratios closely aligned to actual financial risks that banks are facing on their daily business.

In other words, its introduction has improved the incentives for banks to enhance their credit risk management and measurement capabilities and reduce incentives for regulatory capital arbitrage (Benzin, Truck & Rachev, 2003). To capture the broad risk diversity, different risks levels are assigned to different categories of assets. This variety of risk levels is determined by independent, objective and transparent external rating agencies chosen by the bank and recommended by respective supervisor. This is the simplest CRM instrument adoptable by many banks in the world, not necessarily located in low-income countries (Ishtiaq, 2015; Gonzalez-Watty, 2016).

Since the proclamation and adoption of the Basel II Capital Accord, Standardised Approach (STD) has received little research attention and has been adopted mostly in retail, securities and corporate sectors and relies on credit ratings of borrowers assigned by external credit assessment institutions (ECAIs) to compute banks' regulatory capital to credit risk (Ishtiaq, 2015; Gonzalez-Watty, 2016). The major role bestowed to ECAIs has evoked heated debate among supervisors and academic researchers. Supervisors are questioning the functionality of the credit rating agencies with subsequent goal to better understand the credit industry.

Credit ratings are a subjective the assessment of the counterparty's PD and as such differ across ECAIs due to difference in opinion, methodology and rating scale, thereby bringing about the difference in regulatory risk weights and subsequently different capital requirements. On that premise confusion ensues and as a result no uniformity is achievable in the eventual regulatory capital requirements calculations. In the first place, the choice of the identity and the number of ECAIs is to the discretion of the bank concerned. Although the recognition and validation of the ECAI's assessments are the responsibility of national supervisors, it is questionable whether the bank's choice discretion is the one which cause differences in the calculation of capital requirements, making the monitoring work of supervisors difficult since there is no benchmark for objective supervision has been instituted.

From an academic point of view, some studies have pointed to some potential peril of entwining regulatory risk weights to credit ratings, including a greater pro-cyclicality of capital requirements and greater volatility of capital requirements for banks located in emerging economies (Gambacorta & Karmakar, 2016; Lesle & Avramova, 2012). In that respect, studies are being carried to limit the number of external ratings a bank is supposed to use at a time to ensure higher degree of transparency and to reduce unwarranted reliance on a single, perhaps more favourable rating (Gonzalez-Watty, 2016). If this goes unchecked, wider differences in external ratings could affect bank's lending policies and hence the quality and efficiency of credit allocation as well as risky monitoring.

Regardless of wide Standard Approach (STD) adaptation by banks, Bartels (2005) makes an observation that for the majority of SMEs, a bank internal-rating applies, since external-rating agencies are not able to rate such firms. This has been attributed to the fact that SMEs are unfamiliar with formal rating processes and most SMEs are ill-prepared in terms of documentation on strategy, market position or potential, and the use of the instrument of controlling. Financial statements are generally geared towards a tax avoidance strategy rather than a business and investment strategy. This observation demonstrates that STD approach is not yet applicable to measuring SME credit risk.

2.6.3 Internal Ratings-Based (IRB) approach to CRM

The major pronouncement of the Basel II is that banks can employ their own internal estimates of borrower's creditworthiness to assess credit risk in their portfolios, subject to strict methodological and disclosure standards (Gonzalez-Watty, 2016). In fact, the Basel II approach to CRM represents a significant step forward in banking regulation because it combines practical applicability with a solid theoretical basis. Its methodical applicability is universally acceptable to banks of different sizes, business structure and risk profiles, therefore a common quantitative approach to modelling credit risk across for all types of banks, a new development for regulatory purposes and taming of credit risk.

The IRB methodology is based on a CRM model which establishes the likelihood of a borrowing company being unable to repay its debt, as determined by the difference between the value of its assets and the nominal value of its debt. The value of the firm's assets is modelled as a variable which changes over time due to impact of random shocks. As a result, default is assumed to occur when a firm's assets are insufficient to cover debt. The corresponding measure of credit risk within a certain time frame, normally a year, is the PD (Ishtiaq, 2015). In accordance with the IRB framework, the required minimum capital to cushion credit losses is based on the distribution of losses due to default in a portfolio of loans.

Banks now have the leeway to choose between two (2) types of methods when risk weighting loans to determine the corresponding capital requirements. In addition, in January 2004, the Basel Committee of Banking Supervision agreed that IRB capital requirements cover only the unexpected losses (UL), losses not covered by provisions. In fact, the IRB approach is based on measuring expected loss (EL) and unexpected loss (UL), but risk weighting function only includes unexpected loss (UL). This approach especially advanced IRB enables the banks to achieve substantial capital savings in return for establishing a system that will ultimately result in better risk selection (Aubier, 2007).

To calculate capital requirements for a loan's default risk under Basel II depends on six (6) risk components:

Probability of Default (PD): an estimate of the likelihood of the borrower defaulting on its obligation within a given time horizon, normally a year. All banks must provide an internal estimate of the PD associated with the borrowers in each borrower grade. Each estimate of PD must represent a conservative view of a long-run average PD of the grade in question and be grounded in historical experience and empirical evidence (Henneke & Truck, 2006).

Loss given default (LGD): loss on the loan following default on the part of the borrower, commonly expressed as a percentage of the debt's original nominal value, that is:

Equation 1: Loss-given default equation

$$LGD = \frac{Loss\,on\,Loan}{Debt's\,OriginalValue} \times 100\%$$

While the PD associated with a given borrower does not depend on the features of the transaction, LGD is facility specific. Losses are generally influenced by key transaction characteristics such as the presence of collateral and the degree of subordination (Henneke & Truck, 2006). To model LGD it is important to look at what happens after counterparty goes into default.

Exposure-at-Default (EAD): nominal value of the borrower's debt. It is also facility specific. As a firm tends towards default it will normally attempt to increase its leverage, since the reason for default is generally a liquidity problem. In summary, it can be seen as an estimation of the extent to which a bank may be exposed to counterparty in the event of, and at the time of, that counterparty's default.

Maturity of loan (M): Where maturity is treated as an explicit risk component, banks are expected to provide supervisors with effective contractual maturity of their exposures. Where there is no explicit adjustment for maturity, a standard supervisory approach is presented for linking effective contractual maturity to capital requirements (Henneke & Truck, 2006).

Correlation to systematic risk (R): estimate of the link between the joint defaults of two (2) separate borrowers.

In this respect, the IRB model relies on a single-factor asset value model to describe the co-movement of defaults in a portfolio. The unobservable common factor can be interpreted as a variable which represents the state of the economic cycle/business cycle. IRB correlations to the single systematic risk factor are a function of the firm's side and credit quality in accordance with Basel Committee on Banking Supervision (BCBS) framework.

Risk Weight Function: a function relating the loss forecast to minimum capital requirements (MCR); IRB risk weights are specified under the Basel Committee on Banking Supervision (BCBS) framework.

The IRB approach consists of the Foundation internal ratings-based (F-IRB) and Advanced Internal ratings-based (A-IRB) approaches. The major distinction from the Standardised Approach (STD) is that, to measure credit risk, banks use internal modelling techniques to calculate the risk components. Under the F-IRB, banks use their own credit risk models to determine PD, but other risk components like LGD, Exposure at Default (EAD) and loan maturity (M), are furnished by the regulatory authorities. In case of A-IRB, banks can determine, through modelling techniques and database, all the risk components, especially PD and LGD.

The advent of the Basel II was not at first a welcome to SMEs. The question was, "Does the Basel II discourage lending to SMEs?" Several economic commentators focused on the potentially negative effects the new capital requirements might have on the financing of SMEs. They thought the adoption of the IRB methodical approach to measuring credit would put more emphasis on financial ratios than on qualitative factors, resulting in more expensive and rationed credit to borrowers perceived as of higher risk (SMEs in particular) and cheaper credit to borrowers regarded as of lower risk.

Owing to information asymmetry, SMEs are obviously perceived as of higher risk than corporate, therefore would be penalised heavily under the New Capital Accord, Basel II regime. In addition, start-up firms would be highly disadvantaged, as they do not have a credit history to support their creditworthiness. Higher credit costs or decrease in the supply of credit would lead to drastic deterioration in SME financing by banks. If these negative developments were to materialise, the repercussions could have been disastrous, since SMEs are assuming an important growth and developmental role of any economy.

However, the final shape of the New Capital Accord framework dispelled these concerns, as SMEs are in fact expected to obtain rather more favourable treatment under Basel II than under the Basel I regime (Ishtiaq, 2015). In this regard, Bartels (2005) observes that the Basel II Accord on capital adequacy has triggered a fundamental change in the attitude of banks towards these non-traditional borrowers, especially SMEs. Banks are now reviewing their portfolios along the lines of the criteria laid down by the Basel II, adoption of the IRB methodology, which is designed to permit a more accurate assessment of the true risk of a loan.

This requires objective assessment of SME creditworthiness, since banks do always have not good insights into the economics of SMEs. Basel II, a revision of Basel I, is a global standard by which the final soundness of banks is assessed and is designed to create a more competitive, albeit safer banking world (Bartels, 2005). Basel II was designed to ensure that SMEs are not denied access to external financing regardless being presumed theoretically riskier than retail and corporate assets of a bank. However, the capital requirement figure itself is still derived from a supervisory formula provided by the Basel Committee that has been calibrated to reflect the risk of specific asset types and to ensure that overall capital levels in G10 countries remain broadly unchanged.

2.7 DERIVATION OF THE BENCHMARK RISK WEIGHT (BRW) FUNCTION

A default event is a discrete which follows a Bernoulli probability distribution.

Letting default, Y, a random variable which obeys the Bernoulli probability law. This implies Y is a discrete variable which takes values either 0 or 1, where Y =1 means a borrower defaulting otherwise not defaulting. Basing on the theory of Merton (1974) on default, the structural modelling theory of credit risk employs the value of a firm as an input variable for determining default probabilities. According to this theory, a firm default if the value of the total assets V falls below a determined threshold D, the contractual value of its obligation. In probability theory the PD can be expressed as follows:

Equation 2: Probability of default

$$P(Y = 1) = P(V < D)$$

Let therefore $Z_{i,t}$ be the asset change of company *i* within a time interval of length *t*. Using a one-factor model (Belkin *et al*, 1998), in which $Z_{i,t}$ is considered to have a standard normal distribution with mean 0 and variance 1, that is, $Z_{i,t} \sim N(0, 1)$. This variable is decomposable into two (2) risk components (systematic and idiosyncratic risks):

Equation 3: Decomposition of the asset change

$$Z_{i,t} = \sqrt{\rho} X_t + \sqrt{1-\rho}.\varepsilon_{i,t} \qquad with \;\; X_t \to N(0,1) \; and \; \varepsilon_{i,t} \to N(0,1)$$

This can be interpreted as the asset value of borrower *i* is a combination of a systematic risk factor X_t which affects all borrowers, and idiosyncratic risk factor $\varepsilon_{i,t}$ only affecting borrower *i.* The coefficient of the systematic risk factor is termed loading of systematic risk factor and is interpreted as the sensitivity of asset value to systematic risk. In mathematical terms it is the square root of the correlation coefficient of the asset value process with the systematic risk factor. On this premise, the PD is given as:

Equation 4: Unconditional PD for systematic risk

$$P(Y_{i,t} = 1) = P(Z_{i,t} < c_i) = \Phi(c_i).$$

This is unconditional PD. Suppose that the outcome of systematic risk is known, conditional PD is calculable as follows:

Equation 5: Conditional PD for systematic risk

$$
\begin{aligned}
P(Y_{i,t} = 1 | X_t = x) &= P(Z_{i,t} \le c_i | X_t = x) \\
&= P(\sqrt{\rho} X_t + \sqrt{1-\rho}.\varepsilon_{i,t} \le c_i | X_t = x) \\
&= P(\varepsilon_{i,t} < \frac{c_i - \sqrt{\rho} X_t}{\sqrt{1-\rho}} | X_t = x) \\
&= \Phi\left(\frac{c_i - \sqrt{\rho} X_t}{\sqrt{1-\rho}} \right)
\end{aligned}
$$

It must be noted that Φ denotes the cumulative standard normal distribution function. The process is the modelling of the PD for an individual loan, which can extend to portfolio modelling.

Let us consider a loan portfolio consisting of *n* loans to different borrowers where each borrower's PD is as above (Equation 5) and it is assumed that all borrowers have the same default threshold (cut-off point) $c_i = c$. From probability theory, sum of Bernoulli variable follows a binomial probability distribution, which is the probability of having *k* defaults in the portfolio, is

given as:

Equation 6: PD for a portfolio

$$
P\left(\sum_{i=1}^{n} Y_{i,t} = k | X_t = x \right) = \binom{n}{k} (p(x))^k (1 - p(x))^{n-k} \quad \textit{where} \quad p(x) = \Phi\left(\frac{c - \sqrt{\rho} x}{\sqrt{1-\rho}} \right)
$$

Basing on the laws of expectations, the probability of k defaults is the expected value of the conditional probability of *k* defaults:

Equation 7: Conditional PD for a portfolio

$$
\begin{aligned}
P(\sum_{i=1}^{n} Y_{i,t} = k) &= \int_{-\infty}^{\infty} P\left(\sum_{i=1}^{n} Y_{i,t} = k | X_t = x \right) \Phi(x) dx \\
&= \int_{-\infty}^{\infty} \binom{n}{k} (\Phi\left(\frac{c - \sqrt{\rho} x}{\sqrt{1-\rho}} \right))^k (1 - \Phi\left(\frac{c - \sqrt{\rho} x}{\sqrt{1-\rho}} \right))^{n-k} \Phi(x) dx
\end{aligned}
$$

The question is how this theoretical model of defaults in a portfolio relates to the Basel II Internal Ratings-Based (IRB) approach to CRM? IRB functions have their base in Value at Risk (VaR) measure. Given the probability of *k* defaults in a homogenous portfolio of size *n*, the cumulative loss distribution function of the portfolio is:

Equation 8: PD for a homogenenous portfolio of size n.

$$P\left(\sum_{i=1}^{n} Y_{i,t} \leq m\right) = \sum_{k=0}^{m} \int_{-\infty}^{\infty} \binom{n}{k} (p(x))^k (1-p(x))^{n-k} \quad where \quad p(x) = \Phi\left(\frac{c-\sqrt{\rho}x}{\sqrt{1-\rho}}\right)$$

From here, there is need to determine VaR (99.9%) which translates to computing $P^{-1}(0.999)$, not an easy task, for it would need the use of numerical methods. To circumvent the tedious numerical methodical work, Gordy (2002) demonstrates that the VaR can be approximated efficiently in one-factor models, which provide a portfolio-invariant rule for capital charges at the level of a single loan, giving a foundation of the Basel II IRB function.

Letting $\alpha_{0.999}$ denotes the adverse 99.9% quartile of the state of the economy X_t, implies that a worst outcome of the systematic risk factor has a probability of 0.001% of occurrence. Since $X_t \to N(0,1)$ and with small values of X_t, being unfavourable to a firm, $Var(99.9\%) = \Phi^{-1}(0.001)$. Therefore, conditional on this bad state of the economy, the PD for an individual loan is given as:

Equation 9: Conditional PD for an individual on the bad state of the economy

$$P\left(\sum Y_{i,t} = 1 \middle| X_t = \alpha_{0.999}\right) = \Phi\left(\frac{c_i - \sqrt{\rho}\,\Phi^{-1}(0.001)}{\sqrt{1-\rho}}\right)$$

and the expected loss (EL) on the loan is:

Equation 10: Expected Loss (EL) on a loan

$$E\left[L_i \middle| X_t = \alpha_{0.999}\right] = LGD \times \Phi\left(\frac{c_i - \sqrt{\rho}\Phi(0.001)}{\sqrt{1-\rho}}\right)$$

Gordy (2002) has demonstrated how the sum of these expected conditional losses approaches the true VaR (99.9%) of the whole loan portfolio. The threshold c_i can determined from the PD of the respective loan as follows:

Equation 11: PD of a loan

$$PD_i = P(Y_{i,t} = 1) = P(Z_{i,t} < c_i) \ \ with \ \ Z_{i,t} \to N(0,1)$$

we get:

Equation 12: Cumulative PD of Loans

- $PD_i = \Phi(c_i) \Leftrightarrow \Phi^{-1}(PD_i) = c_i$

This leads to the core of the Basel II Internal Ratings-Based (IRB) function to determine the regulatory capital charge on a single loan. Based on the symmetric characteristic of the standard normal distribution around the origin, we derive the core of the benchmark risk weight (BRW) function in the Basel II Capital Accord:

Equation 13: Benchmark Risk Weight (BRW) in Basel II Capital Accord

$$E\left[L_i \middle| X_t = \alpha_{0.999}\right] = LGD \times \Phi\left(\frac{\Phi^{-1}(PD) + \sqrt{\rho}\,\Phi^{-1}(0.999)}{\sqrt{1-\rho}}\right)$$

Having derived the BRW function, there is need to illustrate the calibration of this function with deliberate focus on its effect on capital requirements for SME asset. According to the second consultative document (CP2), the RWA should be calculated using the formula below (BIS, 2001):

Equation 14: Risk Weight of an Asset (RWA) equation

$$RWA = \min\left[\frac{LGD}{50} \times K(PD)(1 + b(PD)(M-3); 12.5 \times LGD\right]$$

It must be noted that the factor *b(PD)* is a maturity adjustment and *K(PD)* is the calibrated BRW function. The calibration of *b(PD)* according to the Basel Committee is a smooth functional relationship between PD and b(PD) according to:

Equation 15: Relationship between PD and $\boldsymbol{b(PD)}$

$$b(PD) = \frac{0.023 \times (1 - PD)}{PD^{0.44} + 0.047(1 - PD)}$$

and the calibrated BRW function is of the form:

Equation 16: Calibrated BRW equation

$$K(PD) = 976.5 \times \Phi\left(\frac{\Phi^{-1}(PD) + \sqrt{0.20}\,\Phi^{-1}(0.995)}{\sqrt{1-0.20}}\right)\left(1 + 0.047 \times \frac{1-PD}{PD^{0.44}}\right)$$

It is observable that the parameter for the asset correlation in the 2001 proposal was set at $\rho = 0.20$ for all loans while the quartile for VaR calculation was set to the 99.5% level. The term 976.5 was a constant scaling factor to calibrate the *K(PD)* to the 100% for PD = 0.7% and LGD = 50%. The middle factor

Equation 17: Middle factor of the Calibrated BRW equation

$$\Phi\left(\frac{\Phi^{-1}(PD) + \sqrt{0.20}\,\Phi(0.995)}{\sqrt{1-0.20}}\right) = \Phi\left(1.118 \times \Phi^{-1}(PD) + 1.288\right)$$

represented the sum of expected loss (EL) and unexpected loss (UL) and was associated with a hypothetical, infinitely granular portfolio of one-year loan having an LGD of 50%, based on the idea of the one factor model. The last factor $\left(1 + 0.047 \times \frac{1-PD}{PD^{0.44}}\right)$ was an adjustment since the benchmark risk weights for loans were calibrated to a three-year average maturity.

A close look at the eventual suggestions of the Internal Ratings-Based approach of the second consultative document (CP2) the following criticisms are raised:

- The desired incentive character of IRB approach for banks is questionable, since risk weights in many cases are higher for Internal Ratings-Based (IRB) approach than for Standardised (STD) approach because of the tendency in the Internal Ratings-Based approach of assigning lower risk weights to companies with very good ratings and much higher risk weights to such firms with ratings worse than BB-. This means that the SME sector is highly condemned because most of them belong to the speculative grade area of the credit rating.
- The assumed linear relationship between maturity and the assigned risk weights is questionable. This would imply that higher risk weights are assigned to exposures with longer maturities. As a result, many SMEs are afraid that due to higher risk weights for long term loans, banks could even refuse to make such contracts anymore.

Since the CP2 outcomes were vehemently criticised for being unfavourable to SME asset, changes were eminent. The treatment of SME was revised and several changes in the BRW function were undertaken to take cognisance of the economic significance SMEs entail and the crucial need for their adequate financing. Generally, under Basel II, SME borrowers are defined as firms with less than Euro 50 million in annual sales. In recognition of a spectrum of risks associated with SME borrowers, the IRB approach for corporate credits would be allowed to separately distinguish loans to SME borrowers from those to corporate.

In fact, banks that manage SME exposures in a manner similar to retail exposures will now be permitted to apply the less capital requiring retail IRB treatment to such exposures, provided that the exposure to a bank by an individual SME is less than Euro 1 million. Such exposures are treated the same way as credits to private customers. Presumably this special treatment of SME exposure would result in an average reduction in the IRB approach for corporate loans.

Furthermore, the final version of the IRB framework saw changes in the maturity adjustment term of K(PD). The assigned maturity adjustment in CP2, specifically the linear relationship between maturity horizon and the assigned risk weights for more risky loans was heavily criticised and changed. As a result, under Foundation Internal Ratings-Based (F-IRB), all exposures will now be assumed to have an average of 2.5 years instead of 3 years of maturity.

Under Advanced Internal Ratings-Based (A-IRB), supervisors have an option of exempting smaller domestic firms from the maturity framework. Consequently, smaller domestic firms are defined as those with consolidated sales and consolidated assets of less than Euro 500 million. If done so, these firms will be assumed to have an average maturity of 2.5 years, as under the Foundation Internal Ratings-Based (F-IRB) approach. For firms with sales greater than Euro 500 million in the Advanced Internal Ratings-Based (A-IRB) approach the maturity adjustment will be included according to the factor:

Equation 18: A-IRB Maturity Adjustment factor

$$\frac{(1+b(PD)).(M-2.5)}{1-1.5b(PD)}$$

With:

Equation 19: Maturity Adjustment equation

$$b(PD) = 0.011852 - 0.05478(\log(PD))^2$$

It must be noted that the denominator in the fraction of the factor is interpreted as an adjustment to the average maturity of 2.5 years while the numerator is the maturity adjustment based on the PD of the exposure and its maturity. As a result of this factor, exposures with higher PD the effect of maturity is more than for higher rated exposures, contrary to the outcome of the CP2. This new result is in line with the Basel Committee objective of avoiding high BRW and Therefore high capital requirements for exposures to SME exposures with longer maturities. All these changes and improvements of the derived BRW function, call for an ultimate refinement of the same function.

As a result of conceited consultations and debates, the Basel Committee eventually imposed a refined version of the BRW function. The refined version of the function distinctively relates PD, asset correlation, firm size and maturity to capital requirements, unlike the originally proposed in the CP2 document. The two (2) formulae differ in various ways, for instance, the CP2 BRW formula incorporates an implicit assumption that asset correlations for all exposures are equal to 0.20, whilst the refined formula assumes that the asset correlation declines with PD and decreases with the size of the firm according to the following relationship (for $5 \le S \le 50$).

Equation 20: Refined BRW equation.

$$\rho(PD;S) = 0.12\left(\frac{1-e^{-50*PD}}{1-e^{-50}}\right) + 0.24\left(1-\frac{1-e^{-50*PD}}{1-e^{-50}}\right) - \left(1-\frac{S-5}{45}\right)\times 0.04$$

For S > 50 the last term will assume value of 0, while for S < 5 it takes value of 0.04. Ignoring the adjustment for the size of the firm, for lowest *PD* value the asset correlation equals 0.24 and for the highest *PD* value it is equal to 0.12. Between 0 and 0.04 can be subtracted from the value of the asset correlation according to size of the firm. In summary for firms with turnover of Euro 5 million or less the assumed asset correlation is reduced by 0.04 while for companies with a turnover greater than Euro 50 million there is no need to reduce the assumed asset correlations at all. In between there is a linear relationship between size and asset correlation.

It is to the interest of the Basel Committee to reduce the BRW for exposures to SMEs with higher PDs. In that regard, the confidence level that was implicit in the formula has been increased from

0.995 (CP2) to 0.999 to cover some of the elements previously dealt with the 976.5, scaling factor. This value enters the main formula for the capital requirement and risk-weighted assets. With the improved confidence level and $\Phi^{-1}(0.999)$, the refined BRW formula is the following:

Equation 21: RWA equation

$$RWA_i = K(PD,\rho) * 12.5 * EAD$$

With:

Equation 22: Capital requirement formular

$$K(PD,\rho) = LGD\left[\Phi\left(\frac{\Phi^{-1}(PD) + \sqrt{\rho}\,\Phi^{-1}(0.999)}{\sqrt{1-\rho}}\right) - PD\right] . \frac{(1+b(PD)).(M-2.5)}{1-1.5b(PD)}$$

Eventually the following formulae package has been adopted for use for SMEs when the A-IRB approach is applied when exposure is treated as either retail or corporate. All the formulae follow the same calculations steps involving inputs for correlation (R), capital requirement (K) and risk weighted assets (RWA). The most important input variables, to be provided by the banks are three (3) (PDs, LGDs and EADs), while the asset correlation (R) is implicitly given by the Basel formula (Altman & Sabato, 2005). The list of formulae that follows below, is a list contained in the last version (June 2004) of the New Basel Accord, Basel II. It must be noted that when SMEs are classified as retail, the formulae are the ones for the “other retail exposures” whilst when classified as corporate, the formulae to be used are the ones for the corporate exposure, considering the size discount factor (BIS, 2004).Formulae when SME is treated retail asset

Equation 23: SME as retail Correlation (R) Equation

Correlation (R) =

$$0.03 * \left(\frac{1-e^{-35*PD}}{1-e^{-35}}\right) + 0.16 * \left(1 - \frac{1-e^{-35*PD}}{1-e^{-35}}\right).$$

Equation 24 : SME as retai Capital Requirement Equation

Capital Requirement (K) =

$$LGD * \Phi\left[(1-R)^{-0.5} * G(PD) + \left(\frac{R}{1-R}\right)^{0.5} * G(0.999)\right] - PD * LGD.$$

2.7.1 Formulae when SME as corporate assets

Equation 25: Correlation equation when SME considered corporate asset

Correlation (R) = $0.12\left(\frac{1-e^{-50*PD}}{1-e^{-50}}\right) + 0.24\left(1-\frac{1-e^{-50*PD}}{1-e^{-50*PD}}\right) - 0.04 * \left(1-\frac{S-5}{45}\right)$

Equation 26:Capital Requirement formular when SME (corporate asset)

Capital Requirement (K) =

$$LGD * \Phi\left[(1-R)^{-0.5} * G(PD) + \left(\frac{R}{1-R}\right)^{0.5} * G(0.999)\right] - PD * LGD * (1-1.5b(PD))$$

Equation 27: Maturity Adjustment equation for SME (corporate asset)

Maturity adjustment (b) =

$$(0.11852 - 0.05478 * \log(PD))^2$$

2.8 IMPACT OF BASEL II ON SME FINANCING

The inception of the Basel II was never palatable to SMEs and politicians as they had a notion that the envisaged New Capital Accord would further discriminate them from accessing external finance from banks, due to the Accord's heightened risk-sensitivity character it seemingly possesses. The major objective of this Accord is to magnify riskiness alignment of a loan to capital requirements. In that respect, loans to SMEs are generally risky due to various factors, which would imply that the more risk-sensitivity character of this Basel II would further entrench this sector into serious financial distress.

Therefore, politicians and SMEs expressed concern that high capital requirements in Basel II might reduce banks' lending to SMEs and make capital more expensive. This concern was echoed by Cardone-Riportella and Trujillo-Ponce (2007) who observe that because Basel II sets up capital requirements that are more sensitive to credit risk, it would therefore increase the risk premium that banks charge on SME loans and, as a result, it would exacerbate their very well-known financial accessibility difficulties.

In response to this concern, the Basel Committee allayed the fear by assuring them that the risk from lending to SMEs is lower in practice than indicated by share capital and tangible assets (Ekpu, 2015; Blundell-Wignall & Atkinson, 2010). The fact is that lending to many fairly small SMEs implies high degree of diversification in a bank's portfolio and this reduces the risk of major total losses to the bank. It is true that on an individual basis SMEs are riskier than large firms because the expected default probability is inversely related to firm size, an argument normally used by banks to justify their higher risk premium on SME loans. But a portfolio of SME loans is not necessarily riskier than a portfolio of loans to large firms (de la Torre *et al.*, 2010; Yoshino & Taghizaden-Hesary, 2017).

In that regard Dietsch, and Petey (2004) found that default correlations are lower within a group of SMEs than within a group of large firms, implying that lower default correlations can offset the higher individual default probabilities within a pool of credits. Chionsini, Marcucci and Quagliariello (2010) argue that the special treatment of SMEs by Basel II is because SMEs are typically more affected by idiosyncratic shocks than by systemic factors. SMEs are certainly riskier than larger firms, but their riskiness tends to be scarcely influenced by systemic drivers; in other words, asset correlation is lower, and therefore their defaults tend to be not correlated to each other. As a result, banks can diversify such a risk by pooling many claims on SMEs in their loan portfolios. Their argument falls in with Dietsch and Petey's point of view, who further justified his point by affirming that firm-specific risk is diversifiable as opposed to systemic risk, which characterise large firms.

SMEs show a great deal of flexibility in the transformation of their business when macroeconomic conditions deteriorate or improve, while large firms are often locked into existing organisational structures and technologies. Based on such arguments, it is justifiable for Basel Committee to assure the SME sector that the New Capital Accord has more favourable treatment to SMEs than previously perceived. Again, the Basel II allows small firms to make use of collateral, guarantees and credit derivatives, on-balance sheet netting to mitigate credit risk. It is, therefore, interesting for banks to know the impact of such techniques on their capital requirements (Cardone-Riportella, Trujillo-Ponce. & Briozzo, 2011).

The importance of SMEs for the Basel Committee on Banking Supervision is evident from the various modifications that were made with the object that Basel II should not turn out to be too detrimental for SMEs in terms of the capital required. The formulae for calculating the regulatory capital associated with SMEs were modified three times in the Basel II consultative documents of 2001, 2003 and 2004. In all these modifications, the definition of SMEs has been consistent, under the Basel Capital Accords; an SME is understood as a company where the reported sales for the consolidated group of which the firm is a part is less than 50 million Euros (Cardone-Riportella *et al.*, 2011).

As a result of these positive changes in the financial landscape, the changes must work in favour of SME financing. In addition to Basel Capital Accords, new information and communication technologies contribute to reducing information asymmetries between lenders and borrowers at lower cost, thereby making SME lending more attractive. Moreover, due to progress in information technology, new banking methods are being developed and implemented, for instance, banks can adopt new portfolio CRM models that allow them to allocate and price their resources more effectively (Yoshino & Taghizaden-Hesary, 2017).

Basel II, through the IRB approaches, encourages banks internal activities, allowing banks to focus on comparative-advantage activities, notably credit risk assessment, loan origination, and credit monitoring - all activities crucial for the provision of the finance to SMEs (Yoshino & Taghizaden-Hesary, 2017). In fact, in recognition of the idiosyncratic nature of risks associated with SMEs, banks, through the IRB framework can distinctively treat loans to SME borrowers from those to corporate borrowers.

Other banks that may treat SME exposures as retail may also be permitted to apply the less IRB capital requirement provided the total exposure of such banks to an individual SME is less than one million Euros. This implies that SMEs exposures can be treated as if they are exposures to private bank clientele. This has been illustrated in the benchmark risk weight (BRW) derivation in section 2.7 where there is clear show of reduced regulatory capital exposures to this SME sector.

2.9 TRADITIONAL APPROACHES TO SME FINANCING AND BASEL IMPACT

Traditionally, in comparison to corporate, SMEs are often more opaque and information asymmetric, thereby making them less transparent. This situation makes SME financing a challenge as information asymmetry breeds adverse selection and moral hazard problems. The repercussion of such a situation is credit rationing (Lesle & Avramova, 2012; Lin, 2007), implying that SMEs do not get as much credit as they want although they are willing to meet the conditions set by the lenders on equivalent credit contracts. Therefore, SMEs worldwide complain about financing problems and the behaviour of their bankers. Reflecting on these challenges, SMEs often have no other choice than to rely on bank relationships for their external financing while large firms turn to banks as well as to capital markets (Yoshino & Taghizaden-Hesary, 2017; Simpasaand & Pla, 2016).

Long relationship with creditors is one way of reducing asymmetric information; relationship banking as opposed to transaction banking is conducive if three (3) conditions are met:

- The intermediary gathers information beyond readily available public information.
- Information gathering takes place over time through multiple interactions with the borrower, often through the provision of multiple financial services.
- The information remains confidential (Yoshino & Taghizaden-Hesary, 2017).

In summary, relationship banking may create value as it can stimulate the channelling of information on the borrower to the lender. It allows inter-temporal smoothing of financing costs (Arnoud & Matej, 2008; Giovannini, Lacopeta & Minneti, 2013). For instance, a bank may subsidise a firm at the beginning of a product cycle and receive compensation for initially accepting a low interest rate when product has matured.

2.10 BINARY CHOICE MODELS - CANDIDATES FOR CRM MODELS

Credit risk of bank client is a binary variable characterised by two (2) outcomes default or no default and as such it can be treated as a Bernoulli random variable. On the other hand, binary choice models are extension of the Bernoulli random variable distribution. Therefore, binary choice modelling approach provides the best candidature for CRM modelling. Therefore, it suffices to consider credit risk as a Bernoulli random variable denoted by Y that takes on two (2) values 0 (non-default) and 1 (default) with the following probabilities:

Equation 28: Probabilities for non-default and default

$P(Y = 1) = p$ and $P(Y = 0) = 1 - p$

The corresponding probability mass function of Y given p is given as follows:

Equation 29: Conditional Probability of a portfolio

$$f((y|p) = p^y(1-p)^{1-y},$$

The respective expectation and variance of the random variable Y are given as follows:

Equation 30: Expectation of a portfolio

$$E(Y) = p.1 + (1-p).0 = p$$

Equation 31: Variance of a portfolio

$$Var(Y) = E\left[\left(Y - E(Y)\right)^2\right] = p(1-p)^2 + (1-p)(0-p)^2 = p(1-p)$$

The property that $E(Y) = p$ becomes crucial when we extend Bernoulli to binary choice models where we would use the maximum likelihood estimation (MLE) methodology to estimate the unknown parameter p.

Let $y_1, y_2, y_3, \dots \dots \dots \dots, y_n$ be a random sample from a Bernoulli distribution with probability of success given by p. Then the likelihood function for p is given as follows:

Equation 32: Likelihood of a function portfolio

$$L(p|y_1, y_2, \dots \dots .. y_n) = f((y_1, y_2, \dots \dots, y_n|p) = \prod_{i=1}^{n} f(y_i|p) = \prod_{i=1}^{n} p^{y_i}(1-p)^{1-y_i}$$

The log-likelihood function is

Equation 33: Log-likelihood function of a portfolio

$$l(p|y_1, y_2, \dots \dots \dots .. y_n) = lnL(p|y_1, y_2, \dots \dots \dots .. y_n)$$

$$= \sum_{i=1}^{n} \ln[p^{y_i}(1-p)^{1-y_i}]$$

$$= \sum_{i=1}^{n} [y_i lnp + (1-y_i)\ln(1-p)]$$

To find the maximum likelihood estimator for p we find the partial derivative of the log-likelihood function and equate it to zero:

Equation 34: Maximum likelihood Estimation of PD of a portfolio

$$\frac{\partial}{\partial p} l(p|y_1, y_2, \dots \dots \dots . . y_n) = \frac{\partial}{\partial p} \sum_{i=1}^{n} [y_i lnp + (1 - y_i) \ln(1 - p)]$$

$$= \sum_{i=1}^{n} \left[y_i \frac{1}{p} - (1 - y_i) \frac{1}{1 - p} \right]$$

$$= \frac{1}{p} \sum_{i=1}^{n} y_i - \frac{1}{1 - p} \sum_{i=1}^{n} (1 - y_i) = 0$$

The result would be that:

Equation 35: Estimated Pd of a portfolio

$$\hat{p} = \frac{1}{p} \sum_{i=1}^{n} y_i$$

2.10.1 Binary choice models

When dealing with binary choice models the interest is in the estimation of the conditional probability, $P(Y = 1|x_i)$ the probability that an agent i chooses $y_i = 1$ conditional on independent variable*s* x_i. In CRM modelling, the interest is in the probability that a credit applicant defaults given credit characteristics information such as age, purpose of loan, salary and so forth.

From Bernoulli distribution, consideration is in the case in which p varies across individuals. Therefore, it is assumed that p_i is the probability that individual i have event $Y = 1$ and p_i is a function of the independent variable x_i:

Equation 36: Conditional Expected PD of a portfolio

$$p_i = F(x_i) = P(y_i = 1|x_i) = E[y_i |x_i]$$

This means that different choices of the functional form $F(.)$ in equation (36) provide us different binary choice models. The three (3) choices of $F(.)$ could either be linear probability model (LPM), probit model or logit model. The choice of the functional form is dependent with the modeller for different form provides different model respectively.

2.10.2 Linear Probability Model (LPM)

Considering a linear regression model given as:

Equation 37: Linear regression model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + e = \boldsymbol{x\beta} + \boldsymbol{e}$$

where $\boldsymbol{\beta}$ is (k+1) vector of parameters, $\boldsymbol{x}$ is $n\ x\ k$ design matrix of explanatory variables, and $\boldsymbol{e}$ is a residual vector. In this instance the functional form $F(.)$ is a linear function, that is,

Equation 38: Functional form of thr regression model

$$F(x_i) = \boldsymbol{x_i'\beta}$$

such that

Equation 39: Linear probability Model

$$P(y_i = 1|x_i) = \boldsymbol{x_i'\beta}$$

To estimate the model parameter, β we use the least square estimation (LSE) method to obtain the least square estimator (LSE) vector, $\widehat{\boldsymbol{\beta}}$. The fitted value $\widehat{y_i}$ of y_i is actually the predicted probability of the event that $y_i = 1$ conditional on independent variable x_i:

Equation 40: Estimated LPM

$$P(y_i = 1|x_i) = \boldsymbol{x_i'\widehat{\beta}}$$

The problem with the linear probability model is that the predicted probability $\widehat{y_i}$ may be less than 0 or greater than 1, thereby flouting the laws of axiomatic probability. It is on this basis econometricians rarely employ the linear probability models for binary choice modelling in CRM. In fact, for this work, there is complete abandonment of the LPM and the corresponding Ordinary Least Squares (OLS) approach to estimating binary CRM models. Therefore, the need to explore the other two (2) functional forms in CRM modelling.

2.10.3 Probit and Logit Models

Let us consider a class of binary response models of the form:

Equation 41: Conditional LPM

$$P(Y = 1|x) = G(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k) = G(\boldsymbol{x\beta})$$

where G is a function taking on values strictly between 0 and 1: $0 < G(z) < 1$, for real numbers z. The above model is normally called index model because $P(Y = 1|\boldsymbol{x})$ is a function of the vector $\boldsymbol{x}$ only through the index:

Equation 42: LPM

$$\boldsymbol{x\beta} = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$$

This is simply a scalar quantity. The condition that $0 < G(\boldsymbol{x\beta}) < 1$ ensures that the estimated response probabilities are strictly between 0 and 1. G is a cumulative density function, monotonically increasing in the index z ($z = \boldsymbol{x\beta}$) with:

Equation 43: Probabilities of Default and Non-default

$$P(Y = 1|x) \rightarrow 1 \ as \ \boldsymbol{x\beta} \rightarrow \infty$$

$$P(Y = 1|x) \rightarrow 0 \ as \ \boldsymbol{x\beta} \rightarrow -\infty$$

It is explicitly clear that G must be a non-linear function which cannot be solved by Ordinary Least Squares method. It has been suggested several such functions but the commonest are the logistic distribution and the standard normal distribution which yield the logit model and probit models respectively. Both the logit and probit functions are monotonically increasing in $\boldsymbol{x\beta}$ and increase relatively swiftly at $\boldsymbol{x\beta} = \boldsymbol{0}$ while the effect on G at the extreme values of $\boldsymbol{x\beta}$ tends to be zero. The implication of the latter condition is that the partial effects of changes in the explanatory variables are not constant, unlike in the LPM.

The standard normal cumulative distribution function (cdf) has a similar shape as that of the logistic, implying that choice between the two (2) models for analysis is not crucial. There is an alternate way of explaining how probit and logit models resolve some shortcomings of linear probability model. This involves not using a response to a functional form problem. Tradition has viewed probit and logit models as models suitable for estimating parameters when the dependent variable is not fully observable. This is quite relevant to CRM of loan applicants where credit quality of the rejected is not readily observable.

Let y_i^* be a continuous random variable that we do not observe, that is to say it is a latent variable determined by the following model:

Equation 44: Latent Variable

$$y_i^* = x_i'\beta - \varepsilon_i$$

where ε_i is a residual assumed uncorrelated with x_i, that is, x_i is not endogenous. The latent variable y_i^* is unobservable but only observable is the discrete choice made by the individual according to the following choice rule:

Equation 45: Possible values of the outcome variable

$$y_i = \begin{cases} 1 \; if \; y_i^* > 0 \\ 0 \; if \; y_i^* \leq 0 \end{cases}$$

In the CRM view, y_i^* can be interpreted as risk of defaulting of payment that is defined as:

Equation 46: Latent outcome variable

$$\begin{aligned} y_i^* &= x_i'\beta - \varepsilon_i \\ &= \beta_0 + \beta_1 age_i + \beta_2 income_i + \beta_3 ratio_i - \varepsilon_i \end{aligned}$$

and loan applicant chooses:

Equation 47: Loan outcome variable

$$y_i = \begin{cases} 1 \; default \; if \; \beta_0 + \beta_1 age_i + \beta_2 income_i + \beta_3 ratio_i - \varepsilon_i > o \\ 0 \; nondefault \; if \; \beta_0 + \beta_1 age_i + \beta_2 income_i + \beta_3 ratio_i - \varepsilon_i \leq 0 \end{cases}$$

Assuming the residual ε_i in the model (46 and 47) has a cumulative distribution function $F(.)$ then the following conditional probabilities are defined:

Equation 48: Conditional Probabities of the loan outcome variable

$$\begin{cases} y_i = 1 \underset{by\ choice\ rule}{\Longleftrightarrow} y_i^* > 0 \underset{y_i^* = x_i'\beta - \varepsilon_i}{\Longleftrightarrow} x_i'\beta - \varepsilon_i > 0 \underset{y_i^* = x_i'\beta - \varepsilon_i}{\Longleftrightarrow} x_i'\beta > 0 \underset{y_i^* = x_i'\beta - \varepsilon_i}{\Longleftrightarrow} P(y_i = 1|x_i) = F(x_i'\beta) \\ y_i = 0 \underset{by\ choice\ rule}{\Longleftrightarrow} y_i^* \leq 0 \underset{y_i^* = x_i'\beta - \varepsilon_i}{\Longleftrightarrow} x_i'\beta - \varepsilon_i \leq 0 \underset{y_i^* = x_i'\beta - \varepsilon_i}{\Longleftrightarrow} x_i'\beta \leq 0 \underset{y_i^* = x_i'\beta - \varepsilon_i}{\Longleftrightarrow} P(y_i = 0|x_i) = 1 - F(x_i'\beta) \end{cases}$$

This implies that different choices of the cumulative distribution function $F(.)$ provide different binary choice models for CRM: probit and logit models.

2.10.3.1 *Probit model*

The probit model assumes the error term ε_i in the model equation (46) is distributed as standard normal and F(.) in equation (48) is a cumulative distribution function of the standard normal distribution such that:

Equation 49: Cumulative Normal Distribution function

$$F(z) = \phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} exp\left(-\frac{1}{2}u^2\right) du$$

Then the intended conditional probabilities will be given as follows:

Equation 50: Conditional Normal probabilities

$$\begin{cases} P(y_i = 1|x_i) = \phi(x_i'\beta) = \displaystyle\int_{-\infty}^{x_i'\beta} \frac{1}{\sqrt{2\pi}} exp\left(-\frac{1}{2}u^2\right) du \\ P(y_i = 0|x_i) = 1 - \phi(x_i'\beta) = 1 - \displaystyle\int_{-\infty}^{x_i'\beta} \frac{1}{\sqrt{2\pi}} exp\left(-\frac{1}{2}u^2\right) du \end{cases}$$

On the other hand, the logit model assumes the error term in the model equation (46] is distributed as logistic distribution such that:

Equation 51: Logistic distribution function

$$F(z) = \wedge(z) = \frac{\exp(z)}{1 + \exp(z)}$$

Using the analogy of probit model the conditional probabilities of interest are given as follows for the logit model:

Equation 52: Conditional logistic probablilities

$$\begin{cases} P(y_i = 1|x_i) = \wedge(x_i'\beta) = \dfrac{\exp(x_i'\beta)}{1 + \exp(x_i'\beta)} \\ P(y_i = 0|x_i) = 1 - \wedge(x_i'\beta) = 1 - \dfrac{\exp(x_i'\beta)}{1 + \exp(x_i'\beta)} = \dfrac{1}{1 + \exp(x_i'\beta)} \end{cases}$$

2.10.4 Estimation of probit and logit models

It has been demonstrated that the LPM is estimated by employing the Ordinary least squares method due to linear nature of the model. On the contrary, probit and logit models have a non-linear nature therefore maximum likelihood estimation is appropriate for this type of the latter models. Considering that the samples are made up of identically independently distributed (iid) random variables; $\{y_i, x_i\}_{i=1}^{N}$ in which the outcomes y_i are Bernoulli distributed. The Bernoulli conditional density function is given by:

Equation 53: Conditional Bernoulli density function

$$f(y_i|x_i;\beta) = p_i^{y_i}(1-p_i)^{1-y_i}$$

where p_i is the probability of $y_i = 1$ and $1 - p_i$ is the probability of $y_i = 0$. Applying the F(.) function form and setting up

Equation 54: Cumulative probability of a portfolio

$$p_i = F(x_i'\beta) = \begin{cases} \phi(x_i'\beta)\ if\ we\ use\ probit\ model \\ \wedge(x_i'\beta)\ if\ we\ use\ logit\ model \end{cases}$$

Therefore, the conditional density function becomes:

Equation 55: Conditional probability of the portfolio

$$f(y_i|x_i;\beta) = F(x_i'\beta)^{y_i}\big(1 - F(x_i'\beta)\big)^{1-y_i}$$

A likelihood function of the sample is a joint conditional density function and since the constituent random variables are identically independent distributed (iid) then the likelihood function for the sample is:

Equation 56: Likelihood function of the sample

$$L_n(\beta) = \prod_{i=1}^{N} f(y_i|x_i;\beta) = \prod_{i=1}^{N} F(x_i'\beta)^{y_i}\left(1 - F(x_i'\beta)\right)^{1-y_i}$$

Then taking log we obtain the following log-likelihood function:

Equation 57: Log-likelihood function

$$l_n(\beta) = ln\prod_{i=1}^{N} f(y_i|x_i;\beta) = \sum_{i=1}^{N} lnf(y_i|x_i;\beta) \ = \sum_{i=1}^{N} ln\left\{F(x_i'\beta)^{y_i}\big(1-F(x_i'\beta)\big)^{1-y_i}\right\}$$

$$= \sum_{i=1}^{N}\{y_i lnF(x_i'\beta) + (1-y_i)\ln\,[1-F(x_i'\beta)]\}$$

To estimate the parameters, we find the partial derivates log-likelihood function with respect to β and equate to zero, that is:

Equation 58: Partial derivative of the log-likelihood function

$$\frac{\partial}{\partial\beta} l_n(\beta) = \frac{\partial}{\partial\beta}\sum_{i=1}^{N}\{y_i lnF(x_i'\beta) + (1-y_i)\ln[1-F(x_i'\beta)]\}$$

$$= \sum_{i=1}^{N}\left\{y_i\frac{f(x_i'\beta)x_i}{F(x_i'\beta)} + (1-y_i)\frac{-f(x_i'\beta)x_i}{1-F(x_i'\beta)}\right\}$$

$$= \sum_{i=1}^{N}\left\{\frac{y_i}{F(x_i'\beta)} - \frac{1-y_i}{1-F(x_i'\beta)}\right\} f(x_i'\beta)x_i$$

$$= \sum_{i=1}^{N}\left\{\frac{y_i\big(1-F(x_i'\beta)\big) - (1-y_i)F(x_i'\beta)}{F(x_i'\beta)\left(1-F(x_i'\beta)\right)}\right\} f(x_i'\beta)x_i$$

$$= \sum_{i=1}^{N}\left\{\frac{y_i - F(x_i'\beta)}{F(x_i'\beta)\left(1-F(x_i'\beta)\right)}\right\} f(x_i'\beta)x_i = \mathbf{0}_{\boldsymbol{k}\times\mathbf{1}}$$

Note that:

Equation 59: Partial derivative of the linear model

$$\frac{\partial}{\partial\beta}F(x_i'\beta) = f(x_i'\beta)x_i$$

The solution of the equation (58) gives the maximum likelihood estimator of the parameter. Again, the equation has no analytical solution for $\widehat{\boldsymbol{\beta}}_{\boldsymbol{ML}}$ for $\boldsymbol{\beta}$ unless numerical methods are employed.

In probit model we have

Equation 60: Standard Nomal distribution of linear model

$$\begin{cases} F(x_i'\beta) = \phi(x_i'\beta)\ cdf\ of\ stadard\ normal\ distribution \\ f(x_i'\beta) = \phi(x_i'\beta)\ pdf\ of\ standard\ normal\ distribution \end{cases}$$

Therefore, the original equation becomes:

Equation 61: Transformed linear model

$$\sum_{i=1}^{N} \left\{ \frac{y_i - \boldsymbol{\phi}(x_i'\beta)}{\boldsymbol{\phi}(x_i'\beta)\left(1 - \boldsymbol{\phi}(x_i'\beta)\right)} \right\} \phi(x_i'\beta) x_i = \mathbf{0}_{\boldsymbol{k}\times\mathbf{1}}$$

In the logit model the following ensues:

Equation 62: Distribution and density function of the logistic distribution

$$\begin{cases} F(x_i'\beta) = \Lambda(x_i'\beta)\ cdf\ of\ logistic\ distribution \\ f(x_i'\beta) = \lambda(x_i'\beta)\ pdf\ of\ logistic\ distribution \end{cases}$$

Therefore, the equation for the solution of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{ML}}$becomes:

Equation 63: Logistic parameter estimator

$$\sum_{i=1}^{N} \left\{ \frac{y_i - \Lambda(x_i'\beta)}{\Lambda(x_i'\beta)\left(1 - \Lambda(x_i'\beta)\right)} \right\} \lambda(x_i'\beta) x_i = \mathbf{0}_{\boldsymbol{k}\times\mathbf{1}}$$

where the cdf and pdf of logistic distribution is as follows:

Equation 64: Cumulative distribution function of the logistic distribution

$$\lambda(z) = \frac{\partial}{\partial\beta}\Lambda(z) = \frac{\partial}{\partial\beta}\left\{\frac{exp(z)}{1 + exp(z)}\right\} = \left\{\frac{exp(z)\left(1 + exp(z)\right) - exp(z).exp(z)}{\left(1 + exp(z)\right)^2}\right\}$$

$$= \frac{exp(z)}{\left(1 + exp(z)\right)^2} = \frac{exp(z)}{1 + exp(z)}.\frac{1}{1 + exp(z)} = \Lambda(z)\left(1 - \Lambda(z)\right)$$

Therefore:

Equation 65: Probability density function(pdf) of the logistic distribution

$$\lambda(x_i'\beta) = \Lambda(x_i'\beta)\left(1 - \Lambda(x_i'\beta)\right)$$

which simplifies equation above as follows:

Equation 66: Simplification of the pdf of the logistic distribution

$$\sum_{i=1}^{N}\left\{\frac{y_i - \Lambda(x_i'\beta)}{\Lambda(x_i'\beta)\left(1-\Lambda(x_i'\beta)\right)}\right\}\lambda(x_i'\beta)x_i$$

$$= \sum_{i=1}^{N}\left\{\frac{y_i - \Lambda(x_i'\beta)}{\Lambda(x_i'\beta)\left(1-\Lambda(x_i'\beta)\right)}\right\}\Lambda(x_i'\beta)\big(1-\Lambda(x_i'\beta)\big)x_i$$

$$= \sum_{i=1}^{N}\{y_i - \Lambda(x_i'\beta)\}x_i = \mathbf{0}_{k\times 1}$$

2.11 SUMMARY

In this chapter an outline of the evolution of credit risk from historical perspective and the methods banks used to underwrite loan applications was explained to give basis for further research into classification problems involved with CRM. A close look was given to subjective CRM methodologies which are prevalent in the lending industry. Owing to unprecedented upsurge of default risk and some other secular forces in the banking industry, the need for CRM has increased to tame the risks banks are exposed to as a result of their core business of lending. The advent of the Basel Capital Accord Internal ratings-based (IRB) CRM approaches led to the introduction of quantitative methodical approaches to CRM in the form of the MDA (Z-score), the O- score by Altman (2018) and Ohlson (1980) and many others. These arms-length technologies to measure credit risk of loan applications assisted banks with objective, consistent and efficient underwriting.

The multiplicity of the definitions of SMEs was critically scrutinised to correctly located the SME sector on the private sector development trajectory. A perfect concept of SME must distinctively demarcate the sector from the microenterprise sector and corporate extremes and must describe the sector as a dynamic and developmental class. An SME is normally defined in terms of employment size, asset size and turnover, which sometimes misrepresent the dynamic and functional character of the SME sector thereby defeating motive of banks to finance the sector. A conclusive and universal definition of SMEs was adopted to facilitate bank financing and comparison of SMEs across economies and sectors.

Subsequently the chapter looks at the advent of the Basel Capital Accords to reign stability and soundness in the banking industry by cushioning the credit losses after persistent erosion induced through intensive competition. This culminated in the setting up of minimum risk-based capital requirements for banks, Therefore the establishment of the Cooke ratio, which compelled banks to hold regulatory capital to at least 8% of their total loan commitments. The Basel I, because of its one-size-fit-all character did not achieve the initial intentional objective of the Basel Committee, therefore was replaced by Basel II, which facilitates the derivation of the benchmark risk weights of assets, thereby dawning for SMEs not to be hindered from accessing credit, though theoretically risker than retail and corporate assets.

This culminated in the advent of IRB methodologies to CRM. The New Basel II Accord (2004) has proposed CRM approaches that took into consideration the individualised, quantitative framework adopted by banks to meet their specific requirements. Research has evidenced that the Basel II based new CRM technologies have made a stronger contribution to progress in the intricacies of banking business. This made it possible and much easier for FIs to increase their prudence on credit exposure from any asset and to instil flexibility in the extension of credit to even non-traditional assets because of better information systems and technologies.

Conclusively the chapter, demonstrates the derivation of the benchmark risk weight (BRW) function of the retail and corporate assets, as the SME asset can either be considered as a retail or corporate asset in the calculation of respective capital requirements. The impact of Basel II Capital Accord was pronounced and considered the basis on which the envisaged model is built, as an improvement of the traditional technologies which discriminate against the financing of the SME sector.

Given the dichotomous nature of credit quality of loan applicants, binary choice models place good candidature for the modelling of measurement of the credit risk banks. Credit risk of bank client is a binary variable characterised by two (2) outcomes default or no default and as such it can be treated as a Bernoulli random variable. The three (3) binary choices models could either be linear probability model (LPM), probit model or logit model. Conclusively probit or logit model provides the ultimate candidacy to credit risk modelling because they obey the axioms of probability.

CHAPTER 3 LITERATURE REVIEW: CONCEPTUAL CREDIT RISK MEASUREMENT MODELS

3.1 INTRODUCTION

This chapter looks at conceptual issues regarding the CRM and the modelling of the CRM in general. It looks in detail on technical terms, instruments which have been used in CRM and as well as past work and current work on CRM modelling with special emphasis on SME CRM. Tools like credit scoring models and rating models for both retail and corporate exposures are analysed with the aim to find headway to simulate the same tools for SME exposure. The risk components in the form of PD, LGD and EAD are revised in detail for better understanding of the envisaged eventual CRM modelling. But much emphasis is put on the PD, a measure of creditworthiness of SME borrowers in particular. Statistical mechanics to construct a random model development sample are illustrated to expatiate on the basic inference theory from random sample. The inherent selectivity bias in CRM modelling is decisively dealt with using reject inference technique, Bound and Collapse Bayesian inference-based methodology for eventual measurement or the quantification of credit risk banks are exposed to for lending depositors' money to borrowers.

3.2 FINANCIAL RISKS

Lending is a risky business, in the sense that there are two (2) mutually exclusive outcomes: a possibility of loss or an opportunity for gain. In that respect, risk can be defined as the volatility of unexpected outcomes. As for lending activity by FIs, credit risk arises when obligors are unwilling or unable to fulfil their contractual obligations due to various reasons. There is a risk of involuntary default, when the counterparty may not have enough money to pay for the principal loan and interest or a strategic default, when the borrower may just simply refuse to pay up (Schoen, 2017; Baily, Litan & Johnson, 2008). The effect of such default is quantifiable by the cost incurred in replacing the cash flows, thereby missing the original performance target of a portfolio. Therefore, the real credit risk is the deviation of portfolio performance from its expected value (Schoen, 2017; Baily, Litan & Johnson, 2008). In short, credit risk is the risk that counterparty will fail to meet his/her payment obligation, resulting in a loss on the part of the lender. This type of risk is historically considered the main risk for banks.

Under the Basel II Capital Accord, a bank has onus to internally measure its credit risk and retain capital for it. Owing to the devastating effect of credit risk to the banking industry, in particular, its measurement has long been viewed as the best way to tame it by understanding the various ways in which banks manipulate and mitigate such effects. In that respect, credit risk prediction is of paramount importance, which brings in statistical prediction modelling from financial statements, customer transactions, repayment records and any credit risk pointing events. It is therefore prudent to take bank perspective in CRM modelling.

Prediction of credit risk entails PD, the major credit risk component in the lending business. To forecast PD is a major challenge in banking industry, especially to financial exposures with inadequate informational input, like the SME loans. In that respect, statistical modelling methods have long been used to model the quantification of PD, to facilitate lenders distinguish between the “good” and “bad” before deciding to grant or not a loan to a loan applicant, a crucial decision in any lending activity.

In modelling PD, the major stumbling block is that there is no universally accepted definition of default. In that regards, regulators as well as rating agencies define it as any of the following occurrences: bankruptcy, write-off, 90 days past due loan on internal non-accrual list. On the other hand, the Basel Committee, to ensure uniformity in the banking industry, has defined default as involving one or more of the following criteria. It is determined that the obligor is unlikely to pay its debt obligations (principal, interest, or fees) in full:

- There is a charge-off.
- The obligor is overdue more than 90 days on any credit obligation.
- The obligor has filed for bankruptcy or similar protection from creditors.

This, still, can cause confusion among banking practitioners, so the operational definition of default has been reduced to only two (2) criteria: more than 90 days overdue, and unlikely to pay in full. Based on this definition, the modelling of the PD as well as other credit risk components like LGD and EAD is uniform across the banking industry. These credit risk components are input variables to Internal-Ratings Based (IRB) methods, which allow banks to set the capital requirements for different exposures. In fact, the best estimate of exposure to the borrower is dependent on the credit risk components, so the need for detailed approach to their estimation process.

3.3 PROBABILITY OF DEFAULT (PD)

PD is defined as the likelihood that a loan will not be repaid and fall into default (Spuchľáková, *et al.*, 2015). From statistical point of view, PD is known as expected default frequency, which falls into the former definition. It is a parameter used in the calculation of economic capital or regulatory capital under Basel II for a banking institution. It is calculated for each client who has a loan that is for wholesale banking or for a portfolio of clients with similar attributes especially for retail or corporate banking. Therefore, it is an attribute of any bank's client. For its calculation, credit history of the counterparty or portfolio and the nature of the exposure are central. Since PD of a client or portfolio is a variable, time framework is important to incorporate into its definition. Therefore, PD can be defined as a measure of credit risk that is assigned internally or externally to a client or portfolio with the aim of estimating the PD within a year. This is achievable through CRM tools which have developed and proposed in literature.

Under Basel II, banks have, to their disposal, alternative methodologies and procedures for estimating the PD. It may be estimated from a historical database of actual defaults using modern statistical tools and techniques like multivariate discriminant analysis (MDA), logistic regression (logit), and neural network. They can also be estimated from observable prices of credit swaps, bonds, and options on common stock. The former approach is the one which rating agencies adopt.

Up to so far, no external agencies have ventured into SME PD estimation business, therefore banks must internally estimate PDs for their SME clientele. Under Basel II IRB approach, SME PD estimation, logistic regression has been found to be popular basing on historical database of defaults as well as basing on the dichotomous nature of credit risk. In that regard, some procedural approach to PD modelling has been suggested:

- Analyse the credit risk aspects of the counterparty/portfolio (Scoring)
- Map the counterparty to an internal risk grade which has an associated PD (Rating)
- Determine the facility specific PD (Rating)

3.4 CRM MODELLING

To measure credit risk, PD is a major input variable in the capital ratio, benchmark risk weight (BRW) function and correlation coefficient formulae. It means CRM anchors on the precise determination of this major variable. In fact, PD is an essential part of business intelligence and customer relation management systems in the FIs (Radmehr & Bazmara, 2017). Underestimating this important component and the LGD, might threaten the stability and smooth running of financial markets.

From a CRM point of view, the result of predictive accuracy of the estimated PD is more valuable than the standard binary classification: credible or non-credible customers. In that regard, the Basel II Accord recognises the methods of reducing credit risk and PD and LGD as important components of A-IRB (Abdulrahman, 2018). The process of assessing the PD of counterparty or portfolio is referred to as CRM. In fact, CRM is defined as a set of decision models and their underlying techniques that aid lenders in the granting of credits. It can also be defined as a method of evaluating the credit worthiness of borrowers by using a formula or set of rules.

In the lending activity, it is a crucial decision to determine who gets credit, how much credit and what operational strategies will enhance the probability of the borrowers paying back to lenders. Therefore, CRM modelling plays a fundamental role in the risk management practices at most lending institutions. For instance, commercial banks' major business is to do with extending loans to borrowers, generating loans and credit assets for a profit. This business is risky therefore the need to assess the quality of the banks' assets in line with the maximum risk a bank is willing to accept. This is to protect itself against incidents that may have adverse repercussions on its profitability and its capital base or share price, therefore referred to as bank's risk appetite. In that regard, banks need to quantify credit risk with the aim to manage it, therefore most appropriate and advanced tools are crucially needed for a successful future of banking system. CRM models have been proved a powerful toolkit to quantify risk at counterparty or transaction level and are dependent on the nature of the counterparty. This means CRM models for corporate counterparty are different from SMEs or retail counterparty.

Credit cycle constitutes three (3) phases: application phase, behavioural phase and collection phase. CRM models cannot be uniformly applied along the credit cycle. This explains how dynamic CRM modelling is. Accordingly, care must be exercised when using this tool, key steps must be analysed during CRM model's lifecycle, which constitute assessment, implementation and validation, highlighting the main requirements imposed by Basel II. For lenders, under Basel

II, to calculate their capital requirements need to increase their attention and consideration of CRM models for stability and soundness in their business.

CRM assumes that past experiences can be used as a guide in forecasting credit worthiness of a certain category of borrowers (Ditrich, 2015; Kanapickiene & Spicas, 2019; Al Baz, 2017). Credit risk prediction is of great importance. It involves analytical processes and prediction models whose purpose is to use financial statements, customer transaction, repayment records and other risk drivers to predict credit quality and to reduce the uncertainty and default (Gonzalez-Watty, 2016). Credit assessment process consists of comparing risk predictive characteristics of incumbent credit applicants with other earlier period borrowers whose repayment performance is known. If the prospective borrower's characteristics are a carbon copy of previous debtors who defaulted, then the applicant is automatically rejected for loan.

Contrary if the loan applicant characteristics are congruent to those who did not default, and then the loan is granted. Such crucial decisions in lending business require quantitative instrumental and objective weaponry to help credit analysts and underwriters make prudent decisions. It is on this regard that, Crook (1996) suggests, two (2) types of CRM models have been identified; Judgemental CRM and Statistical CRM models.

3.4.1 Judgemental CRM modelling

Credit assessment approach is classified judgemental if it consists of individual assessment of credit applicant by a credit analyst, implying that the success or failure of the judgemental evaluation is wholly dependent on the expertise, experience and common sense of the credit analysts (Rommer, 2005; Asrat, 2018). This CRM modelling approach is based on the traditional standards of credit analysis, whereby factors such as payment history, bank and trade references, credit agency ratings and financial ratios are scored and weighted to produce an overall CRM score of the counterparty (Chen & Astebro, 2012; Goh & Lee, 2019; Mamo, 2011).

Consequently, the major setback of such an approach includes subjectivity, inconsistency and individual preferences. The credit evaluation process cannot be automated if there are many credit applicants to be considered. Here expertise and experience of the credit executive play a major role as the choice of factors to use and their weighting in the CRM modelling process is dependent on the executive's past experience in the lending business. Factors that reflect the individual characteristics and policies of their own firm make the judgemental models more predictive and straightforward. The major strength derived from using method is that qualitative characteristics of the debtor and wealth of experiences and expertise of the credit analyst are taken aboard during the CRM process. The only problem with this type of CRM modelling is that is too subjective and when the number of borrowers increases it becomes less effective.

3.4.2 Statistical credit risk measurment modelling

This type of CRM modelling functions the same as the judgemental models except that the choice of factors to be scored and weighted is statistically done rather than based on the experience and judgement of the credit executive (Altman, 2018; Dong & Peng, 2013; Ball, 2016). Statistical methodology considers many factors at once, a process which calculates, and analyses multivariate correlation thereby identifying relevant trade-offs among credit drivers and assigns statistically derived weights. This type of modelling is quite objective and easy for automation, making it appropriate for use for prompt credit decision-making for a host of applicant borrowers. This in turn improves customer service, retention of good profitable customers and facilitates low cost in credit supply. In fact, by employing statistically extracted cut-off credit scores, acceptable and unacceptable loan applicants are easily and objectively segregated. This approach is anchored in the addition and subtraction of statistically derived number of points relating to the applicant's CRM score given to the risk-pointing characteristics.

Credit driving factors are normally captured from credit agency reports and the credit files of the clients. In literature, these models are described as a scorecard, pooled scorecard or a custom scorecard. In this regard, a scorecard uses data from one firm, whilst a pooled scorecard uses data from various firms. Lastly a custom scorecard blends a statistical model with some factors used in the judgemental model (Wood, 2012) The major handicap with CRM modelling is statistical in nature. In CRM modelling, in general only those who are accepted for credit will be followed up to find if they really do turn to be good or bad credit risk in accordance to default definition adopted. This engenders truncated or biased or non-random samples which compromise the basis of statistical inference of the TTD population of applications due to selectivity bias. If a sample consisting only accepted applicants is used to construct a new CRM model, then selectivity bias is introduced. Owing to this inherent selectivity bias misleading CRM scores result threatening banking industry and the purpose for which it is built.

Over the last 50 years, several statistical methodologies have been used to build CRM models. It was in the late 1950s, the simplistic univariate credit scoring was rife and subsequently overtaken by the advent of multivariate approaches to CRM modelling. The predominant statistical methodology introduced by academic researchers (Dong & Peng, 2013; Schouten, Lugtig & Vink, 2018) was the multivariate discriminant analysis (MDA). As the CRM became topical in the finance domain, a variety of statistical CRM modelling techniques was introduced to outperform previous suggested methodologies. For instance, Ohlson (1980), instead of the MDA, applied for the first time, conditional logit to credit risk prediction. Thereafter, several other statistical tools were put to test with ultimate goal to improving prediction power of CRM models given the high frequency of financial crises due to subprime lending and the contagious and disastrous nature of such crises.

The following techniques have been applied to CRM modelling; linear regression, probit analysis, logit modelling (binary choice models), Bayesian methods, neural network just to mention a few. The advent of the Basel II has brought an incentive for proper evaluation and measurement of credit risk at bank internal level. In fact, CRM modelling is gaining new importance with the Basel II, which focuses on techniques that allow banks and supervisors to evaluate properly the various risks that banks face (Tursory, 2018; Lesle & Avromova, 2012; Ferreiro, 2016). Under the New Capital Accord, Basel II, CRM modelling works in conjunction with the Internal-Ratings-Based (IRB) approach, where it broadly contributes towards internal risk assessment process of a financial institution. Therefore, various new techniques for statistical CRM modelling are being introduced to estimate real credit risk given the adverse effect it may have on FIs if it goes unchecked.

3.5 CRM MODEL LIFECYCLE

Under Basel II Capital Accord, commercial banks strive to implement the most advanced approach to calculate their minimum capital requirements. To achieve uniformity, strict and common rules regarding how banks should build their internal models along the CRM model lifecycle. This lifecycle constitutes model assessment phase, implementation phase and the validation phase. Therefore, all banks implementing A-IRB approach to CRM must follow the designed standard model lifecycle and regulators have published specific requirements for each one of the phases (Altman & Sabato, 2005; Reserve Bank of Zimbabwe (RBZ), 2011).

3.5.1 Model assessment phase

The basis for CRM modelling is on the assumption that the future will be similar to the past, so the objective is to risk rank new or existing borrowers on that basis. It is logical that if an applicant or existing client had certain historical credit behaviour, it is likely that a new applicant or existing client with the same characteristics will show the same credit behaviour. From that perspective, to develop a CRM model we need a random sample of past loan applicants or clients' data to the same product as the one we want to use our measuring for. For banks, the CRM modelling requires availability of appropriate historical data is of paramount importance to developing an empirical model (Rehman *et al.*, 2019; Hao, Alam & Carhy, 2010).

From the random sample covering the time horizon appropriate for adequate statistical analysis, a good estimate of the true payment behaviour of banks' clientele is observable. Great care must be exercised at this stage in terms of proper definition of dependent and independent variables for easy development of the desired model. In the case of the conditional logit analysis introduced by Ohlson (1980), the dependent variable is the credit quality/loan performance of the clients, defined as either default or non-default event associated with each client in the sample, Therefore a binary outcome variable. The characteristics of the clients at the beginning of the selected period are the predictors/measurement characteristics.

3.5.2 Model implementation phase

This phase consists of the automation of the decision system to manage a large number of applicants because it is practically impossible to manually refer to credit analysts for final decision over a multitude of clients. To achieve this task through a CRM model, the cut-off or decision threshold of CRM model must be clearly defined; otherwise it may be difficult to separate with acceptable level of confidence between expected 'good' clients and expected 'bad' ones. To achieve the best out of this tool, the optimal cut-off must be incorporated bearing in mind the misclassification costs related to Type I (bad given good) and Type II (good given bad) error rates as Siddiqui, (2005) and Chen and Astebro (2012) point out. This again must be considered, exercising careful consideration of each particular bank peculiarities, which include among others; risk appetite, profit-loss objectives, recovery process costs, efficiency and marketing strategies. For instances, modern advanced banks use profitability criterion to set cut-offs at account level.

This is quite possible due to the availability of information technology, which enables banks to follow the lifecycle of any client, from application to the end of banking relationship; with monthly updated CRM scores calculated by CRM models related to the phase of the credit cycle where the client is located; origination, account maintenance, collection, and write-off (Altman & Sabato, 2005).

3.5.3 Model validation phase

Sophisticated banks even in emerging countries, are willing to adopt the prudential Basel II IRB-Advanced approach for efficient CRM. In doing so, these banks are required to put in place a regular cycle of model validation that should include at least monitoring of the model performance and stability, review of the model relationships and testing of the model against outcomes, which is called back-testing (BCBS, 2005, 2006). Stability and performance are very important information about the quality of the CRM models.

3.6 MULTIVARIATE DISCRIMINANT ANALYSIS

It was Altman in 1968 who first applied the multiple discriminant analysis technique (MDA) in developing multivariate CRM model. The subsequent model could effectively predict business failures using a set of financial ratios. Its performance was seen to be better than the univariate CRM models, like the one developed by Beaver (1966). Altman's model is referred to as the Z-score model and uses five (5) financial ratios: working capital/total assets; retained earnings/total assets; EBIT/total assets; market value equity/BV of total debt and sales/total assets. The model is based on two (2) restrictive statistical assumptions:

- The independent variables included in the model are multivariate normally distributed;
- The group dispersion matrices or variance-covariance matrices are equal across the "bad/defaulting" and "good/non-defaulting" group.

After its introduction into statistical CRM, MDA was the mostly used statistical technique for default forecasting (Edminster, 1972; Altman, 2018). In most of these studies the basic assumptions of MDA are often violated when applied to CRM problems. It has also been discovered that in MDA models, the standardised coefficient cannot be interpreted like slopes of a regression equation and hence do not indicate the relative importance of the different variables involved in the model (Sabato, 2005). It is upon these statistical and technical challenges that other researchers like Ohlson (1980), introduced conditional logit model to CRM problems. On the other hand, Zmijewski (1984) was the pioneer in applying probit analysis to predict default as explained in Chapter 2, under Binary Choice models.

Empirical evidence in the default prediction study has shown that until now logit analysis has given better results. Regardless of these breakthroughs into bettering the estimation of PD, several statistical techniques are being tested to improve the prediction accuracy of CRM models. These include linear regression, logit and probit (binary choice models), Bayesian methods, Genetic algorithms (GA), Data envelopment analysis (DEA), machine learning and neural networks just to mention a few and no empirical evidence has shown significant improvements in default forecasting problem (Sabato, 2005).

3.7 DEFAULT PREDICTION LITERATURE

During the past 40 years, academics and banking practitioners have done extensive studies in default prediction methodologies with the ultimate goal to objectively and efficiently forecast the customer's default behaviour or business bankruptcy before the actual granting of loan. As reported previously, seminal works in this field are of Beaver (1967) as well as of Altman (1968), who developed default prediction models using univariate and multivariate statistical approaches respectively. They both used matched sampling technique of default and non-default observations for constructing the model development sample prior to eventual modelling.

Through the univariate analysis Beaver (1967) used a dichotomous classification test to determine the error rates of a potential creditor would experience if a company is classified as failed or non-failed (Roggi, 2015; Roggi & Altman, 2013). This classification was based on 14 financial ratios. The problem with his approach was inconsistency, a loophole Altman (1968) used to develop a multivariate approach to default prediction modelling. He adopted the multivariate discriminant analysis (MDA) to circumvent the inconsistencies presented in Beaver's univariate approach. The MDA proved a better default classification technology which several researchers (Edmister, 1972; Gombola, Haskins, Ketz & Williams, 1987; Altman, 2018) adopted for different firms' default prediction modelling. Despite this widespread use of this multivariate statistical technology, it was found to be deficient in its application; in that its two (2) basic assumptions are often violated when it is applied to default prediction problems. As a result, the standardised coefficients cannot be interpreted as slopes of regression, therefore do not indicate the relative significance of different independent variables considered in the model (Altman & Sabato, 2006).

Karels and Prakash (1987) revealed the MDA application deficiencies; a ground Ohlson (1980) used to develop a conditional logit methodology which is free of the incumbent restrictive assumptions for the same purpose and allowing to be working with disproportional sub-samples. This statistical approach fits very well the characteristics of the default prediction problem, where the dependent variable is binary (default/non-default) and with the subsamples being discrete, non-overlapping and identifiable (Altman & Sabato, 2006). Like the Altman (1968) Z-score model to CRM problems, Ohlson (1980) introduced the O-score which claimed superiority over the former as it outdoes the fundamental challenges of the MDA in its applications. The O-score model is the first probabilistic default prediction model ever in the literature of corporate failure forecasting.

Ohlson (1980) used 9 explanatory variables to generate a default prediction model but he never represented any theoretical justification for these chosen variables (du Jardin, 2009; Rezende, da Silva Montezano, de Oliviera & de Jesus Lameira, 2017). He asserted that the predictive power of default prediction models is majorly reliant on the choice of the threshold or cut-off point and the significance of each of the resulting type of the errors (type I or type II) for specific model. He also suggested that choosing a larger development sample of estimation and adding more default indicators to increase the predictive power of the subsequent default forecasting model.

Realising the utility value of the work of Ohlson (1980), most of the academic researchers started to use logistic regression to modelling default prediction of corporate and other bank assets (Wang, 2013; Wu, Gaunt & Gray, 2010; Lawrence, Pongsatat & Lawrence, 2015; Yi, 2017; Wang, 2011; Yi, 2019; Belyaeva, 2014; Mamo, 2011). Thomas (2000) confirmed logistic regression model as the most popular technique in CRM modelling, since it is easy to implement, to understand, and to interpret. The coefficients are not interpreted as is in multiple regression; they are interpreted as log of odds of being non-defaulter or defaulter and the odds ratios can be used where characteristics with odd ratios greater than 1 are considered significant risk factors.

The problem which most researchers overlook is the problem of incompleteness of the model development samples. Most previous default forecasting models are built on sample data that is an inaccurate representation of the through-the-door (TTD) population of loan applicants. There is a conspicuous mismatch between the model development sample and the future credit risk scored population, a bias that is so critical in default prediction modelling (Shen, Nguyen & Ojiake, 2013; Nguyen, 2016; Ditrich, 2015).

When a representative random development sample has been established, it is when logistic regression modelling is applied. CRM models are continuously being built and updated upon the availability of bank's customer's historical performance from where key predictive characteristics are statistically identified to help in describing the future performance of new customers.

From a statistical viewpoint, the obvious assumption modellers make is that the development sample used to come out with a future CRM model is representative of the overall population and is random, assumptions overlooked in most cases of CRM modelling. Most model development samples for CRM modelling are datasets of the likely "good" customers not inclusive of the likely "bad" ones, making it conspicuously and inherently a truncated sample. This engenders selectivity bias which could explain why the Z-score, or the O-score could wane with lapse of time since the development sample becomes more unrepresentative of the through-the-door population it is supposed to measure. This is a clear violation of the basic principles of sampling theory, a problem that needs to be resolved first prior to any eventual CRM modelling.

3.8 SMES DEFAULT PREDICTION LITERATURE

Prior to Basel II, most default prediction models were designed for corporate entities whose financial default indicators were readily available. Berger (2004) observes that after the Basel II is widely implemented in many countries, some academic studies now pay their attention to default forecasting models of SMEs segment of the economy to analyse the effects of the Basel II rules on this type of firms. Kolari and Shin (2004) found that normally the SMEs are riskier than their larger counterparts, but they comprise the large portion of the economy, so the issue of number can be very beneficial for banks in terms of lending. Studies (Schwaiger, 2002; Saurina & Tricharte, 2004; Udell, 2004; Berger, 2006; Kolari & Shin, 2004; Altman & Sabato, 2005, 2006; Lin, 2007; Chen & Astebro, 2012; Kennedy, 2013; Nguyen, 2016; Al Baz, 2017; Kanapickiene & Spicas, 2019) have demonstrated that the use of CRM systems for SMEs clients is a significant strategic and competitive issue for banking organisations in order to achieve internal efficiency and maximise profits linked to the SME business (Yi, 2017; Ahn, 2016; Wagner, 2008).

In any economy, SMEs constitute the larger portion of business. The indispensable source of external funding of these enterprises is basically the commercial banks, as confirmed by literature (Udell, 2004; Berger, 2006; Lin, 2007; Ackah & Vuvor, 2011; Kennedy, 2013; Abiodun & Entebang, 2015; Wangmo, 2015; Osano & Languitone, 2016; Al Baz, 2017;), whose main source of income is bank credit. Banks are in a dilemma when it comes to lending to SMEs, which have been stereotyped riskiest under the Basel I Regulatory Framework regime. Under Basel I, SMEs are an ungraded bank asset; with benchmark risk weighted (BRW) of 100%, the riskiest asset for any banking institution. Upon this background, banks are still being extremely cautious about SME lending thereby imposing the toughest requirements to ration the much-needed external funding to this sector.

Academic studies as well as the Basel II framework are contrary to imposition of prohibitive and illogical bank requirements of very high level of capital for SMEs credit allocation. If the unavailability of credit for SME financing perpetuates, credit crisis of SMEs is eminent and would negatively affect the national economy of any country. Unanimously Basel II Capital Accord is urging banks, especially commercial banks to develop unique, specialised CRM models for SMEs to gain the best possible outcome. It encourages banks and FIs to internally correctly measure the credit riskiness they are exposed to by their different clientele by aligning capital requirements to the inherent risk in credit. Correct, efficient and objective CRM of SMEs would go a long way in reducing significantly loan processing costs as well as aggregate default costs and at the same time radically expanding credit supply (Kennedy, 2013; Al Baz, 2017; Kanapickiene & Spicas, 2019) to the credit stranded enterprises.

Empirical evidence in the default prediction study has shown that until now logit analysis has given better results (Memic, 2015; Agrawal & Maheswari, 2019; Tserng, Chen, Huang, Tran, Lei & Zhang, 2014). Regardless of these breakthroughs into bettering the estimation of PD, several statistical techniques are being tested to improve the prediction accuracy of CRM models. These include linear regression, Bayesian methods, Machine learning (ML), support vector machine (SVM) and neural networks just to mention a few and no empirical evidence has shown significant improvements in default forecasting problems (Altman & Sabato, 2006).

The banks' dilemma on lending to SMEs also arise in that SMEs have been included by banks in the retail segment only recently, often forced by national regulators and to follow the Basel II

Capital Accord rules (Jones, 2014; Chen & Astebro, 2012; Ma, 2016; Gonzalez-Watty, 2016;). The implication here is that SME portfolios in some banks have a small history that the available data may not warrant development of an authentic CRM model due to few or zero defaults. In such a scenario, we can consider that portfolio of SMEs as a Low-Default Portfolio (LDP). In fact, insufficient defaults do not warrant a credible CRM modelling tool, thereby creating a missing data problem. SME LDPs constitute a challenge to both academics and practitioners. The traditional logistic regression methodology cannot be applied directly but the missing defaults need to be estimated first (Gelman, Jakulin, Pittau & Su, 2008; Ranganathan *et al.*, 2017; Bolton, 2009).

In their study, Löffler, Posch, and Schöne (2005) proposed a Bayesian methodology for it is an appropriate approach to work with small datasets as well as missing data to achieve a credible CRM model. Altman and Sabato (2005) proposed the shock methodology to estimate the missing defaults then apply the logistic regression over an imputed development sample. This current research work seeks not to find solution of the SME LDPs, but the other problem commonly found in CRM modelling; which is the mismatch between model development sample (KGB sample) and the TTD population to be scored in future.

The problem researchers overlook in CRM modelling is the treatment of incomplete development samples, datasets consisting only of pre-screened loan applicants, the likely "good" excluding the likely "bad" thereby engendering non-random, truncated samples. Chen and Astebro (2012) confirm that non-random sampling appears with certainty in CRM modelling where complete data is typically only available for those that have gone through a pre-screening process and have been accepted for credit granting, therefore giving birth to incomplete samples. These types of samples challenge the analyst because they destroy the very foundations of modern statistics (Copas & Li, 1977) since the missing data affect the randomness of a sample, thereby removing the grounds for the use of any statistical inference. A model built on an incomplete development sample engenders biased estimates thereby producing a decision tool unfit to do the envisaged task (Chen & Astrbro, 2012; Nguyen, 2016; Dong & Peng, 2013; Barakova, Glennon & Palvia, 2013).

Such tools in CRM modelling have far reaching consequences in the commission of type I and type II errors, whose commission must be minimised. This has been the trend in previous studies where CRM models for corporate, retail and SME assets have been developed without prior considering of disastrous effects of selectivity bias contained in the model development samples.

The same observations are ratified through studies (Smith & Elkan, 2004; Nguyen; 2016) which point out that, if a default prediction is based on non-random samples, the accuracy results of the subsequent model cannot be generalised.

To minimise this selection bias, Chen and Astebro (2012) suggested reject inference as an interim process of estimating the missing credit quality of the rejected applicants. In fact, the problem of incompleteness in the development sample must treated as a selectivity problem solved guided by the principles of missing data theory (Dong & Peng, 2013; Nguyen, 2016).

Statistical models for predicting defaults in SME credit industry and elsewhere suffer from the non-availability of default information for customer who were denied credit in the first place (Lin, 2007; Moradi & Rafiei, 2019; Wehinger, 2013; Gambacorta, Huang, Qiu & Wang, 2019; Hasumi & Hirata, 2010). This known as the reject inference problem, it affects the estimation of the CRM model parameters (Nguyen, 2016, Chen & Astebro, 2012; Ditrich, 2015). In fact, reject inference is a statistical methodology for inferring how rejected credit applicants would have behaved had credit been granted to them. In other words, it is the prediction of what repayment behaviour the rejected clients would have shown, had they been accepted.

Credit quality data on rejected applicants are usually missing not at random (MNAR) (Chen & Astebro, 2012; Schouten, Lugtig & Vink, 2018, Dong & Peng, 2013), an assumption adopted for CRM development sample construction. Therefore, the goal of the current study is to start developing a complete development sample on which the envisaged CRM model can be built or improved in such a way that it would be eventually used to measure the credit risk of the bank's future TTD population of loan applicants. It is upon this sample that any traditional statistical method (MDA, logit, probit or Linear Regression) can be used to develop the eventual CRM model.

3.9 LOAN PROCESS

Statistical modelling has gained widespread acceptance in CRM modelling arena; therefore, sampling principles must be adequately scrutinised for correct eventual inference of the population of potential loan applications (TTD). In credit measurement modelling, researchers struggle to determine a random sample of the target through-the-door (TTD) population. A model development sample for designing a CRM model is supposed to take all potential loan applicants into account, to satisfy the randomness character of a statistically random and representative sample. In contrast, usually portfolio managers and credit analysts consider only the accepted applications whose credit quality/loan performance is known in developing tools they would apply on the target TTD population.

To iron out this frequent and inherent error in measuring credit risk of borrowers of a bank or any financial institution, it is logical to take a scrutiny of the loan process itself to determine the mechanisms involved and strategize how to overcome the non-randomness problem that always ensues the model development samples. According to Mok (2009) through an illustration in Fig 3.1, demonstrates a loan process, highlighting the inherent mechanisms and decision nodes that constitute the process. The process starts when a client applies for a loan from a creditor. This is followed by accept/reject decision by lender based on the characteristics of the applicant picked from the loan application form and other reference sources. From that point, only the accepted application is given the loan.

After some specified lapse of time, the loan performance of the accepted application is classified as either good or bad. Contrary, the loan performance of the rejected application is not observable for similar classification. From the illustration, the loan process is made up of two (2) major mechanisms viz: selection and outcome. The selection mechanism constitutes the accept/reject decision node and determines whether the application is accepted or rejected whilst the outcome mechanism node determines whether the loan performance of only the accepted application is either good or bad. The loan performance of the rejected application is not observable and therefore non-classifiable.

To measure credit risk of the accepted application, we model the outcome mechanism. This implies that it is easy to determine the probability that an accepted application will be bad based on the loan performance of the accepted application. On the other hand, the probability that the rejected application will be good is not estimable since the loan performance of the rejected is not observable. This constitutes the major problem in CRM modelling, because if we develop an acceptance policy based on data on only accepted applications leads to incorrect and biased results due to selectivity bias and non-randomness of the source sample of statistical inference - the truncated sample. Treating the accepts as if they were a random sample for the TTD population of loan applications is not in any justifiable assumption if the original model had any predictive validity (Smith & Elkan, 2004; Nguyen, 2016; Lin, 2007; Ditrich, 2015).

The loan process engenders incomplete, missing or partial data samples. The impact of such data on eventual statistical modelling may have adverse influence on the eventual inference. To correct this prevalent anomaly in CRM, reject inference has been suggested in literature as the solution to problems arising from incomplete, missing or partial data (Smith and Elkan, 2004; Nguyen, 2016; Dong & Peng, 2013).

As earlier on explained, reject inference is a process of estimating credit risk of loan applicants that are rejected under the current acceptance policy. By virtue of its nature, reject inference is a missing value problem. The impact of incomplete or missing data on the randomness of a sample is dependent on the mechanisms which has caused the missingness of the values (Chen & Astebro, 2012, Dong & Peng, 2013: Nguyen, 2016; Lin, 2007; Ditrich, 2015). Missing data theory has been extensively worked on by Rubin (1976); Little & Rubin (1987); and Gelman, John, Hal and Rubin (1995).

In figure 3.1 illustrates the loan process, which starts off with a potential borrower applying for a loan from a lender, the lender using current acceptance policy, accepts or rejects the application (selection mechanism). The selected applications are granted loan and the rejected are discarded straight away. After an elapse of time, those selected and given loans are either classified as non-defaulters or defaulters, a classification the rejected will not have (outcome mechanism).

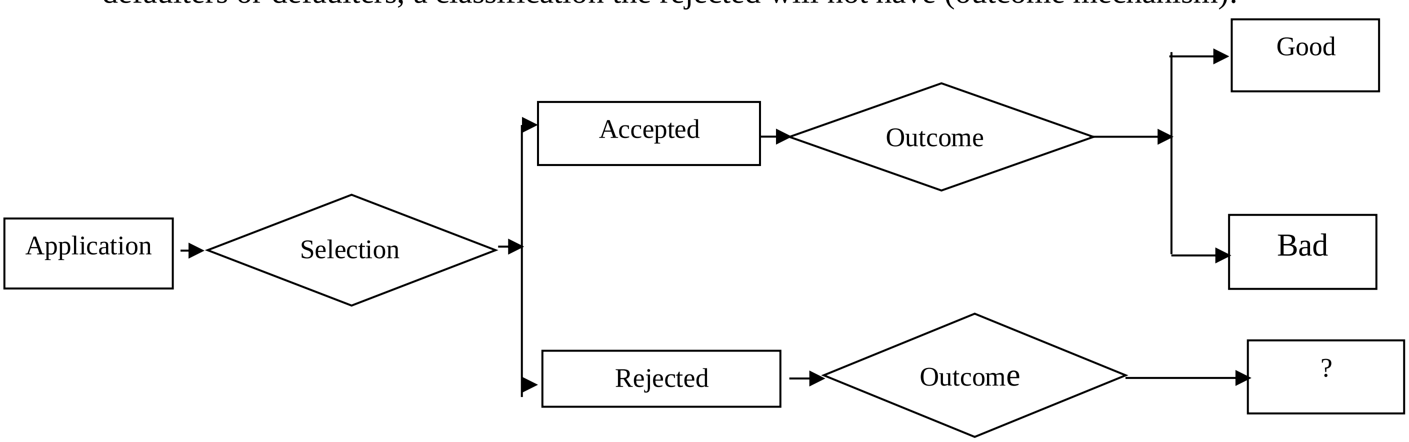


Figure 3.1: Loan process

Source: *Mok (2009)*

3.10 REJECT INFERENCE

The CRM system is intended to be applied to the entire TTD population that enters a bank requesting for a loan, not only to those approved by the previous bank selection system (Smith & Elkan, 2004; Nguyen, 2016; Chen & Astebro, 2012). If a CRM system is constructed using only accepted applicants, then it will contain the effect of the previously used methods that are to be replaced (Smith & Elkan, 2004, Nguyen, 2016; Dong & Peng, 2013). When that happens, the resultant sample is non-random and not representative of the target TTD population of all loan applications due to selection bias. Therefore, the great need for reject inference techniques to reduce the magnitude of the effect of that selection bias otherwise it is not practical in banking industry to offer loans to the TTD population of credit applicants.

Lewis (1992) suggests making an inference of the future performance of the rejected applicants (a process of augmentation) which enables addition of the inferred "goods" and "bads" to the AGB sample. Therefore, the current inherent selectivity bias in most CRM methodologies, namely, that data on good and bad risks belongs mostly to debtors whose applications were once accepted is a cause of concern. This means that only the probability of an accepted debtor going bad can be estimated, but the inferred probability of that a rejected applicant was in fact good, is biased (Nguyen, 2016; Ditrich, 2015; Barakova *et al.*, 2013).

Reject inference technologies are employed to construct a representative development sample on which CRM model, capable of assessing creditworthiness of any applicant from the through-the-door (TTD) population, is built. In most cases, data used to develop CRM models is an inaccurate representation of the population of all future applicants. This bias, resulting from a mismatch between the model development sample and the future measured TTD population, becomes a critical issue (Smith & Elkan, 2004; Dong & Peng, 2013; Nguyen, 2016; Ditrich, 2015; Barakova *et al.*, 2013).

Therefore, reject inference technologies come in as way to account for and correct the inherent selection bias by inferring the good or bad loan status of the rejected applicants. If a sample consisting only accepts is used to construct the new CRM scorecard then bias may be introduced (Nguyen, 2016; Ditrich, 2015). Therefore, if the sampling bias error is minimized, it would be possible to build a CRM model that is more representative of the future through-the-door CRM rates low (Nguyen, 2016; Ditrich, 2015).

Reject inference comes in to cover the methodological gap in CRM modelling that has not received enough attention in default prediction literature, to construct a representative credit applicants' sample. If this is achieved, it would help banks and FIs to objectively and automatically identify creditworthy loan applicants that they are currently rejecting, that is, leaving money on the table, thereby increasing the approval rates in future while holding bad. Such a representative development sample of the credit application pool aid model developers deduce the likely payment behaviour of the rejected based on the preliminary CRM system. This is only achieved through objective and theoretically based reject inference techniques, a process of trying to infer the true credit status of the rejects, using their characteristic vectors (Nguyen, 2016; Ditrich, 2015).

Henley (1995) gave reasons why reject inference is greatly needed as potential selectivity bias is introduced whenever a sample consisting only of accepted applicants is used to construct a new CRM model and to procure an accurate estimate of the portion of potential goods (from the full application pool) being rejected by existing CRM model. Eisenbeis (1977) confirmed that truncated sample of accepted applicants can frequently generate misleading results. This was also confirmed by Smith & Elkan 2004; Nguyen (2016); Ditrich (2015) and Hand (1998) who also highlighted the possible selectivity bias that can result from using an incomplete sample of credit applicants to build a CRM model tool with which to assess the TTD population. To prove the point, Henley (1995) conducted a survey of reject inference methods proposed in literature, compares and contrasts their performances including methods that use supplementary information (referencing) in the form of distributional assumptions. He came with a conclusion that has caused a lot of dilemma to CRM model developers.

The threat of potential selectivity bias which may result from not including the rejects in the development sample is well documented and understood throughout the credit industry. But Henley (1995) concluded that reliable reject inference is not possible unless extra information is incorporated in some way. There is no additional data can be used to improve reliability of reject inference as CRM model developers continue to use ad hoc methods of reject inference with full understanding of its potential impact on the eventual CRM model implementation (Nguyen, 2016; Ditrich, 2015).

CRM model developers are in dilemma for most of the proposed reject inference techniques are not yielding promising results so the need for continual search for techniques which are hinged on firm theoretical and tested assumptions. Literature (Kanapickiene & Spicas, 2019; Al Baz, 2017, Nemoto *et al.*, 2018; Lin, 2007; Kennedy, 2013; Nguyen, 2016) has confirmed discouraging outcomes of the currently used reject inference techniques in CRM model development (Chen & Astebro, 2012; Banasik, Crook & Thomas, 1999; Chen & Haziza, 2018; van der Meijs, 2018; Newman, 2014; Ditrich, 2015; Barakova *et al.*, 2013). It is on this background that theoretically supported rejected inferences are being developed to fill this knowledge gap.

3.10.1 Illustrative example of reject inference

Reject inference is based on the principle that just as some of the accepted debtors have defaulted (false negatives) so too, would many of the rejected debtors have paid as agreed, had they been accepted (false positives). So, reject inference consists of the problem predicting what the rejected applicants would have done, had they been accepted. To illustrate reject inference in CRM modelling, we give two (2) scenarios where a bank does construct a CRM model without applying reject inference and the other scenario when it is applied first.

3.10.1.1 Scenario 1: No reject inference applied

Suppose a ‘through-the-door’ population of 10,000 (N = 10,000) applies for loan from a bank. The bank has a standing credit selection policy of accepting 75% of the total applicants and grants them the loan. The bank does not practise reject inference. In accordance of the incumbent bank selection policy, 7,500 applicants are accepted and 2,500 are rejected. Of the 7,500 accepted, 6,000 debtors pay back as per loan contract, referring them as, ‘goods’ and the 1,500 end up defaulting, referring them as, ‘bads. In the pool of the 2,500 rejected applicants, the distribution of the goods and the bads is not known in the same way we have done from the accepted pool. In that instance, for CRM model development, we use a sample of the distribution of the accepted clients for forecasting the PD of future applicants with the same characteristics as the current ones. Given the non-randomness of the sample on which the CRM model has been built, the resultant PD prediction model’s parameters are biased due to sample selectivity. This bias can only be corrected if the loan performance behaviour of the 2,500 rejected pool is inferred.

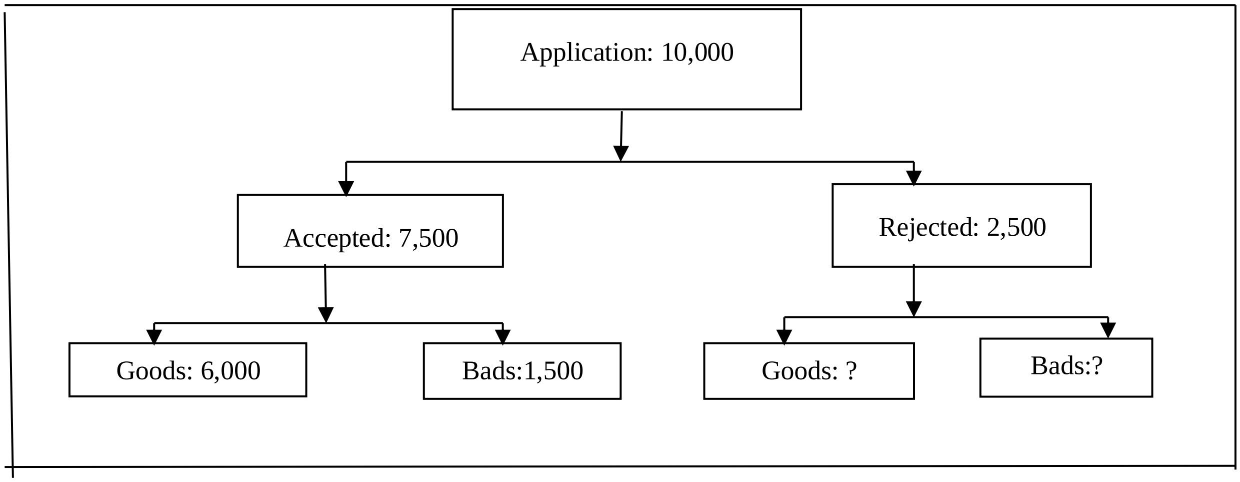


Figure 3.2: Known status of applications when no reject inference applied

When no reject inference is applied, there are 4000 bad debtors (2500 + 1500) which means the bad rate is 0.4.

3.10.1.2 Scenario 2: Reject inference applied

The second scenario is when the same bank realises the anomaly caused by not using reject inference. The same number of applicants, N = 10,000, apply for loan, that is a through-the-door population, the same selection policy stands such that 7,500 applicants are selected and given loan and 2,500 are rejected. The same scenario repeats itself in the 7,500 accepted pool where 6,000 proved goods and 1,500 turned bad. Applying any appropriate reject inference technique on the 2,500 pool of rejected, say 1,000 applicants are inferred good customers had they been accepted, and the remaining 1,500 applicants are correctly inferred as “bad”. This means that the total good applicants after reject inference is 7,000 (6,000 + 1,000) and the total bad applicants is 3,000 (1,500 + 1,500). It is observable that the net result from reject inference for the bank is an increase of 1,000 goods and a loss of 1,500 bads. This implies more income to the bank and no costs, Therefore more profit. Again, by reject inference it is possible to accept some formerly rejected, but good clients and to reject some formerly accepted bad debtors.

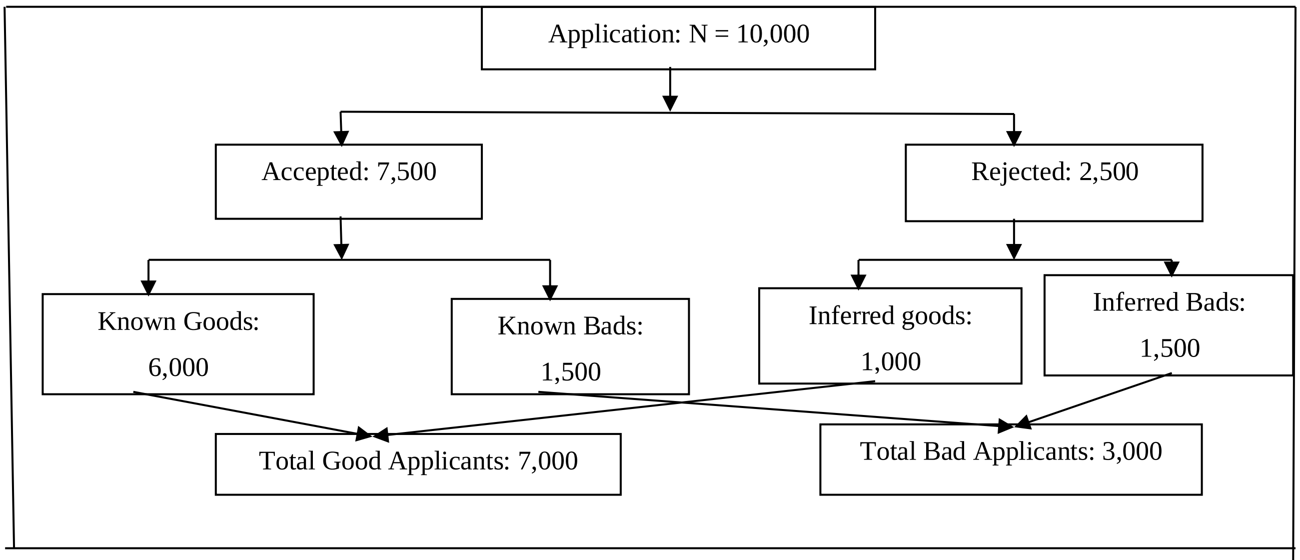


Figure 3.3: Known and inferred application status with reject inference

When the reject inference is applied, there are 3000 debtors (1500 + 1500) and the bad rate is 0.3 which is smaller than when no reject inference is applied - better classification of debtors.

3.10.2 Categories of Reject Inferences Techniques

Reject inference techniques can be classified into two (2) broad categories given which problem a technique is trying to solve. Since the CRM model built must be applied to the whole through-the-door population not only to those who would have been approved by a previous CRM system. This gives rise to two (2) problems: the first problem is that the number of observations on applicants available for analysis is reduced and the second problem is that the development sample is not randomly selected, that is referred to as sample selectivity problem. Accordingly reject inference techniques are classified as follows: The first category of reject inference techniques is a set of methods capable of handling sample selectivity bias and such methods include: Maximum Likelihood estimation (Chen & Astebro, 2012) and the Heckman two-stage estimator (Mok, 2009). The second category is a class of methods used to overcome a missing data problem on credit quality of rejected clients. It is when the reject inference problem is treated as a missing data problem (Chen & Astebro, 2012; Ditrich, 2015; Barakova *et al.*, 2013; Nguyen, 2016). That is a class of techniques that can be used to handle reject inference as a missing data problem such as Hot Deck imputation, List wise deletion, EM algorithm, Bound and Collapse (BC), MCMC simulation and multiple imputation. This thesis adopts reject inference as missing data problem using BC methodology, a model-based technique that can be used to overcome the problem of missing data in the financial behaviour by imputing the missing credit quality of the rejected loan applicants.

3.11 MISSING DATA MECHANISMS

The current taxonomy of missing data analysis is based on the work of Rubin (1976); Little & Rubin (1987) and Gelman *et al.* (1995). To conceptualise the missing data analysis, consider a sample classified according to the values of two (2) categorical vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$, where $\boldsymbol{X}$ is an independent vector which is always observable and $\boldsymbol{Y}$ is the dependent vector subject to missing values. We denote the missing value of Y by $Y =?$ and the resulting incomplete sample by $(X = i, Y =?)$ where the always observed vector $\boldsymbol{X}$ affects the outcome vector $\boldsymbol{Y}$ subject to missingness. The classification of the missing data mechanism is dependent on whether the probability of $Y =?$ depends on the state of $\boldsymbol{X}$ and /or $\boldsymbol{Y}$ (Rubin, 1976; Little & Rubin, 1987; Gelman *et al.*, 1995). As a result, there are three (3) types of missing data patterns: (1) missing completely at random (MCAR) when the probability that the $Y =?$ is missing neither depends on the status of $\boldsymbol{X}$ nor on the status of $\boldsymbol{Y}$; (2) missing at random (MAR) when the probability of $Y =?$ depends on the observable $\boldsymbol{X}$ only and (3) missing not at random (MNAR) when the probability of $Y =?$ depends on the states of both $\boldsymbol{X}$ and $\boldsymbol{Y}$.

The distinction among missing data mechanisms is of great importance since reliable reject inference techniques on handling missing data is dependent on the assumptions of randomness of the missing data. In CRM domain, credit quality data of rejected applications are usually MNAR (Chen & Astebro, 2012; Sebastiani & Ramoni, 2000; Barakova *et al.*, 2013; Dong & Peng, 2013) the assumption pivotal to eventual CRM modelling for this thesis. Pertaining to CRM theory, the observable characteristics of an applicant are based on what has been filled on the application form as well as on the applicant's credit history obtainable from central credit bureau. Such characteristics are observable for every loan applicant. Based on these, an applicant is either accepted or rejected, and this constitutes a selection mechanism (from loan process). The accepted applicants are given loans that are later assessed, at a lapse of outcome window, as either "good" or "bad" accounts in terms of loan payment behaviour. This is not the same with the rejected application who miss the outcome classification in loan payment behaviour. This process constitutes the outcome mechanism, not observable for all applicants but missing for rejected applicants.

To explain the mechanics of loan process in conjunction with missing data theory, let us denote the outcome mechanism (financial behaviour/credit quality) by class label $y \in \{0, 1\}$ and the selection mechanism by an auxiliary variable a $\in$ {0, 1}. From the selection mechanism, this implies that if a loan application is accepted then $a = 1$, and it is rejected then $a = 0$. From the outcome mechanism, the credit quality, y is observable when $a = 1$, that is the accepted applicant has good loan performance otherwise if bad not observable (Mok, 2009, Sebastiani & Ramoni, 2000, Chen & Astebro, 2012; Barakova *et al.*, 2013; Nguyen, 2016).

The indicator variable, a is defined to indicate what is known and what is missing, and this indicator variable is referred to as the missingness variable (Schafer & Graham, 2002). Impact of missing data on the randomness of a sample depends on the process responsible for the disappearance of some of the sample (Sebastiani & Ramoni, 2000; Chen & Astebro, 2012, Dong & Peng, 2013) which is called missing data mechanism (MDM), a process which concerns the relationship between missingness and the values of variables in the data matrix (Chen & Astebro, 2012; Smith & Elkan, 2004; Nguyen, 2016; Ditrich, 2015).

Looking at the missing data theory from a probability theory point of view and in conjunction with a bank situation we may have the following scenarios: Assuming a bank randomly accepts applications then we say the class label *y* is missing completely at random (MCAR), implying that the probability that an application will be accepted neither depends on the selection of the applicant nor on the loan performance. In probability terms this can be modelled as follows:

Equation 67: Probability definition of MCAR

$$P(a = 1|\boldsymbol{x}, \boldsymbol{y}) = P(a = 1)\ MCAR$$

Under this scenario, no reject inference problem is presented as analysis of accepted applications will engender unbiased model estimates. But this is not feasible for a bank to randomly accepted applications since it would increase the default risk which result in high costs to the banks (Mok, 2009).

Traditionally banks use standing rejection policy, that is, the selection mechanism model to make accept/reject decisions on loan applications. If selection mechanism is based on the observable characteristics of the applicants only, not on the loan performance, then the credit quality of rejected applications is said to be missing at random (MAR). The implication is that the probability that application is accepted depends on the observable applicant characteristics only. In other words, MAR means that the PD, given all exogenous variables of the model, is the same whether an applicant is granted a credit or not, that is:

Equation 68: Probability definition of MAR

$$P(a = 1|x, y) = P(a = 1|x)\ MAR$$

This means the probability of a good loan performance depends also on observable applicant's characteristics and not on the selection model as formulated below:

Equation 69: Alternative definition of MAR

$$P(y = 1|x, a = 1) = P(y = 1|x, a = 0) = P(y = 1|x)$$

Equation (69) implies that the loan performance of the rejected applications has the same distribution as the one of the accepted applications for fixed values of $\boldsymbol{x}$ (Mok, 2009). This scenario presents no selectivity bias therefore no reject inference problem, because the observed values of $\boldsymbol{Y}$ are not representative of the overall through-the-door population, only when considered within the categories of the selected only, X. When the credit quality data is MAR or MCAR then the missing data mechanism is pronounced as ignorable, implying that the inference of the credit quality does not depend on the missing credit quality data.

In some circumstances, the applicant selection process is also based on the impressive performance of the applicant as well as other qualitative unobservable characteristics; the credit quality of rejected applicants is said to be missing not at random (MNAR). Probabilistically that would imply that the chance that the application is accepted depends on the credit quality given the characteristics of the applicant, that is:

Equation 70: Probability definition of MNAR

$$P(a = 1|x, y) \neq P(a = 1|x)\ MNAR$$

This implies that the probability of a good credit quality depends also on the selection mechanism when conditioned on the observable quality of the applicant, that is:

Equation 71: Assumption for MNAR

$$P(y = 1|x, a = 1) \neq P(y = 1|x, a = 0)$$

The implication conveyed here is that the loan payment behaviour of the accepted applications has a different distribution from the one of the rejected applications for any fixed values of $\boldsymbol{x}$ (Mok, 2009). This is the scenario which constitutes a selectivity bias problem and requires a reject inference solution. When credit quality data is MNAR, the missing data mechanism is said to be non-ignorable, since the probability of the missing credit quality is dependent on both the observable characteristics and the missing loan performance itself. This shows that the resultant incomplete sample is conspicuously not representative of the overall population (TTD) to be measured under any respect (Sebastiani & Ramoni, 2000).

3.12 STANDARD MISSING DATA APPROACHES TO REJECT INFERENCE

It was until the late 1970s that some techniques to deal with missing data problems were developed as a direct consequence of advances in computer technology. Parallel to technical development, statistical procedures are also being developed specifically for missing data. Dempster, Laird and Rubin (1977) formalise the Expectation-Maximisation algorithm, a computational method for efficient estimation from incomplete data. It proves a useful computational device that signals a fundamental shift in the way statisticians view missing data. Before then, missing data was viewed as a nuisance to be gotten rid of by either case deletion or mean imputation (Fogarty, 2000). From that shift, missing data is now viewed as source of variability to be averaged out and that from incomplete dataset, the observed values provide indirect evidence about the likely value of the unobserved ones. If such evidence is entwined with certain assumptions, a predicative probability distribution is born of the missing values that should be averaged over in the statistical analysis (Barakova *et al.*, 2013; Nguyen, 2016). This turning point has led to the development of missing data techniques which are categorised into four (4) procedures (Ditrich, 2015; Nguyen, 2016; Dong & Peng, 2013):

3.12.1 Complete-case procedures

When variables are not recorded for some of the units, the method is to discard them and analyse only the units with complete data (Nie, Hull, Jenkins, Steinbrenner & Brent, 1975). This procedure is quite easy to carry out and may bring satisfactory results when there is small amount of missing data. The most common technique under this procedure is list-wise deletion missing data technique, most frequently used by researchers (Gilley & Leone, 1991; Fogarty & Blake, 2002). Its core task is to eliminate all classes with any missing units from the eventual analysis. Its application to reject inference problem is analogous to working with accepts only in credit scoring thereby engenders selection bias.

3.12.2 Available-case procedures

Pair-wise deletion is a missing data technique of this category, which works by deleting information only from those statistics that need information. It engenders the disadvantage that the sample base varies from variable to variable according to the pattern of missing data. It has the advantage, over list-wise deletion, in terms of statistical power, although it may be problematic if data are not missing completely at random (MCAR) (Little & Rubin, 1987).

3.12.3 Weighting procedures

These procedures are common in sample survey application. Randomisation inferences from data with non-responses are commonly based on design weights, which are inversely proportional to the probability of selection. The weighting technique modifies the weights to adjust for missingness, quite similar to mean imputation technique.

3.12.4 Imputation-based procedures:

This constitutes data imputation techniques, a group of methods that impute the value of the items that are missing. Roth (1994) spells out the advantages of imputation-based procedures which include among others the following:

- Imputation strategies save a great deal in terms of information over list-wise deletion since an individual is not deleted from the analysis as a result of missing a small amount of information.

- Imputation strategies save even more data than does pair-wise deletion.
- Imputed data preserve deviation from the mean and the shape of the distribution.

The methods under this category of procedures tend to produce less biased estimates than simple missing data techniques (MDTs) such as list-wise of pair-wise deletion methods. Once missing values have been filled in, standard complete data analysis methods can be used on the entire dataset. In many cases the imputation can be created just once by the database administrator who may have much better information about and understanding of the process that creates missing data than the analyst.

3.12.5 Types of Imputation-based procedures

In the following sub-sections , some types of imputation-based procedures are briefly explained and sometimes exemplified to how useful missing data is for eventual analysis.

3.12.5.1 Non-model imputation procedures

These include the mean imputation techniques where the means from sets of recorded values are substituted thereby allowing the researcher to use mean value of a variable in place of missing values for the same variable. This is a very popular missing data technique (MDT) used in current commercial activities and described by Chen and Haziza, (2018). The mean imputation preserves data and is easy to use and it also tends to attenuate variance estimates in statistical procedures (Chen & Haziza, 2018). Dong and Peng (2013) emphasise that using the mean imputation may make the analysts believe that they have more degrees of freedom than is warranted since substitute means are not independent from other observations in the data.

3.12.5.2 Model-based imputation procedures

This type of procedure constitutes a set of more flexible and less ad hoc than other missing data procedures (Henley, 1995; Chen & Astebro, 2012). Techniques in this category are used as starting point for investigation of possible methods of reject inference (Henley, 1995; Nguyen 2016). Model-based procedures can be broken down into two (2) categories:

3.12.5.3 Implicit model-based imputation approaches

This category consists of hot deck and cold deck imputation techniques. Hot deck approach's main idea is that researchers should replace the missing values with an actual datum from a similar case in the dataset. The implication is the resulting dataset with imputed values is termed "hot" since is current in use in the computer. Hot deck imputation approach performs better than list-wise, pairwise; mean imputation techniques in terms of accuracy because missing values are replaced by realistic values (van der Meijs, 2018; Newman, 2014). On the other hand, cold deck imputation consists of using a constant value external to the current dataset, such as a value from previous realisation of the same sample in place of missing data. The imputed value is termed "cold" because it is a datum from another data source that is not currently in use by computer.

3.12.5.4 Explicit model-based imputation approaches

These are a broad class of procedures, a resultant of defining a model for partially missing data and basing inference of the likelihood under that model. The common procedures under this class include the following;

- Regression imputation procedures in which missing values are filled in by estimates of regression coefficient estimates based upon available data. The problem presented when using regression procedures in general is that the resultant predictions are heteroscedastic due to fact that their variances are based on the independent X variables.
- Stochastic regression imputation approaches where the missing values are filled in by a datum predicted by regression imputation plus the residual term, a reflection of uncertainty in the predicted value. This is also constrained by the heteroscedasticity condition.
- Maximum likelihood (ML) procedures estimate parameters based on available data, then estimate the missing values based on the previously estimated parameters (Fogarty, 2000; Chen & Astebro, 2012).
- Expectation-Maximisation (EM) algorithm is an iterative approach which goes through a process of estimating missing data then estimating parameters (Henley, 1995; Mok, 2009).
- Composite methods combine ideas from different imputation procedures such as hot deck and regression combined by computing predicted means from regression and then adding

a residual randomly chosen from the empirical residuals to the predicted values when forming values of imputation (Fogarty, 2000).

- Multiple Imputation methods entail a procedure where more than one value is filled in for a missing datum. It is a method that allows valid estimates of variances of estimates to be computed using standard complete data procedures (Dong & Peng, 2013; Schouten, Lugtig & Vink, 2018).

In sum, there are some tangible advantages derived from explicit model-based imputation approaches which include the following among others; flexibility, the avoidance of ad hoc methods and the availability of large sample estimates of variance based on second derivative of the log likelihood which considers incompleteness in the data in those models (Chen & Astebro, 2012, Smith & Elkan, 2004; Schouten, Lugtig & Vink, 2018). Additionally, the assumptions underlying the resultant model-based methods can be evaluated (Rubin, 1976; Dong & Peng, 2013).

3.13 MAPPING REJECT INFERENCE TO MISSING DATA MECHANISM

To facilitate explain the interconnection between reject inference and missing data, we define the missing data of **Y** by Y = ? and define also the a vector of independent variables $X_1, X_2, X_3, \dots \dots \dots \dots .. X_p \ by \ \boldsymbol{X}$ of which the latter is entirely observable for each score range. By applying the selection mechanism, we assign each credit applicant *k (observation)* a CRM score as $S_k = f(X_k)$ using the incumbent loan acceptance policy. Using the existing threshold/cut-off point say δ such that $S_k \leq \delta$, credit is granted to the credit applicant *k*, otherwise no credit will be granted.

Table 3.1: CRM scores

$X_1, X_2, \dots \dots \dots \dots \dots \dots \dots \dots \dots\ X_p$	$Score, S$	Y	a
1	S_{min}	.	1
2	.	.	1
3	.	.	1
.	.	.	.
.	.	.	.
.	δ		1
.	$\delta + 1$	?	0
.	.	?	0
.	.	?	.
.	.	?	.
.	.	?	.
N	S_{max}	?	0

Source: *Ramoni and Sebastiani (2000)*

Using the auxiliary variable $a = 1$ if credit applicant k is offered credit, 0 otherwise, showing when credit quality is observable or not, for instance credit quality (default/non-default behaviour) is observable if $a = 1$ and missing when $a = 0$. In the same vein, we denote the observed credit risk as Y, implying $Y = 1$ if default, 0 otherwise. Setting the CRM scores bounded by $S_{min} \leq S_k \leq S_{min}$, the dataset can be set in the matrix form as in Table 3.1. It is observable that missing data is defined as a univariate pattern where missing values, Y = ?, happen on an applicant with CRM score greater than δ, but a set of $(p + 1)$ other items $X_1, X_2, X_3, \dots \dots \dots, X_p$ and S is completely observed. The auxiliary indicator variable a identifies what value of Y is known and what is missing, so is referred to as an auxiliary variable for the missingness (Feelders, 2000), whose distribution may need to be specified (Schafer & Graham, 2002).

3.14 BAYESIAN ANALYSIS OF INCOMPLETE SAMPLES

From Table 3.1, we can define our development sample as (S, Y), on which we intend to infer the envisaged CRM model. When the sample (S, Y) is complete, Bayesian conjugate analysis can be used to estimate the conditional distribution of $P(Y = j|S = i)$ and the marginal distribution of $P(Y = j)$. Assuming (S, Y), the sampling model has a multinomial distribution with parameter vector $\boldsymbol{\theta}$ and $\theta_{ij} = P(S = i, Y = j \,|\boldsymbol{\theta})$, where a standard conjugate prior for $\boldsymbol{\theta}$ is a Dirichlet distribution and the posterior distribution of $\boldsymbol{\theta}$ is still Dirichlet (Lindley, 1964). Given that the sample is complete, from the posterior distribution of $\boldsymbol{\theta}$, the posterior distributions of $P(Y|S)$ and of $P(Y)$ can be easily deduced. This is not the case when the sample is incomplete because the information about Y and S is unbalanced. But if the missing data is MCAR and the frequencies of complete are large, the incomplete cases can be ignored, and Bayesian conjugate analysis is carried out in the same way as complete sample.

When the missing data mechanism is MAR, the posterior distribution of $P(Y|S)$ is still conjugate and estimates of $P(S, Y)$ and $P(Y)$ can be easily computed but posterior distributions of $P(Y|S)$ and the marginal of $P(Y)$ cannot be expressed in simple terms as was done for complete and when missing data pattern was MCAR. The situation even becomes complex if the missing data mechanism is MNAR because the observed data do not carry information about missing data and exogenous information about the missingness is needed to be used for imputing the missing data. To solve this problem, the Markov Chain Monte Carlo (MCMC) methods such as the Gibbs sampling which treats the missing entries as unknown quantities of interest are currently being resorted to.

The EM algorithm has been suggested to help learn from conditional probabilities but the problem with these current presumed powerful methods is that they are based on assumption that the missing data mechanism is MAR. This means that they lose power as that assumption is violated. It is upon this background Sebastiani and Ramoni (2000) developed a new reject inference methodological framework called Bound and Collapse (BC) which is versatile for both MAR and MNAR missing data mechanisms. This new technique incorporates the impact of data source by imputing missing data of the dependent variable (credit quality) based on the estimated probabilities of missingness. It is a malleable technique that allows for outside supplementary information about the rejected region to be incorporated in the model. Unlike the other reject inference technique, BC is a model-based technique, firmly supported by theoretical Bayesian assumptions, a character that other reject inference techniques do not possess.

In the same way we can use set theory to describe SME bank loan portfolio by denoting it a sample of the form: $Y = Y_o \cup Y_m$ where the sample with complete observations is represented by Y_o and the sample with missing observations is denoted by Y_m. Supposing that we probabilistically complete Y_o by Y_k such that $Y_k = Y_o \cup Y_{dk}$ where Y_{dk} is a probability distribution of possible completion of the unreported cases in Y_m.

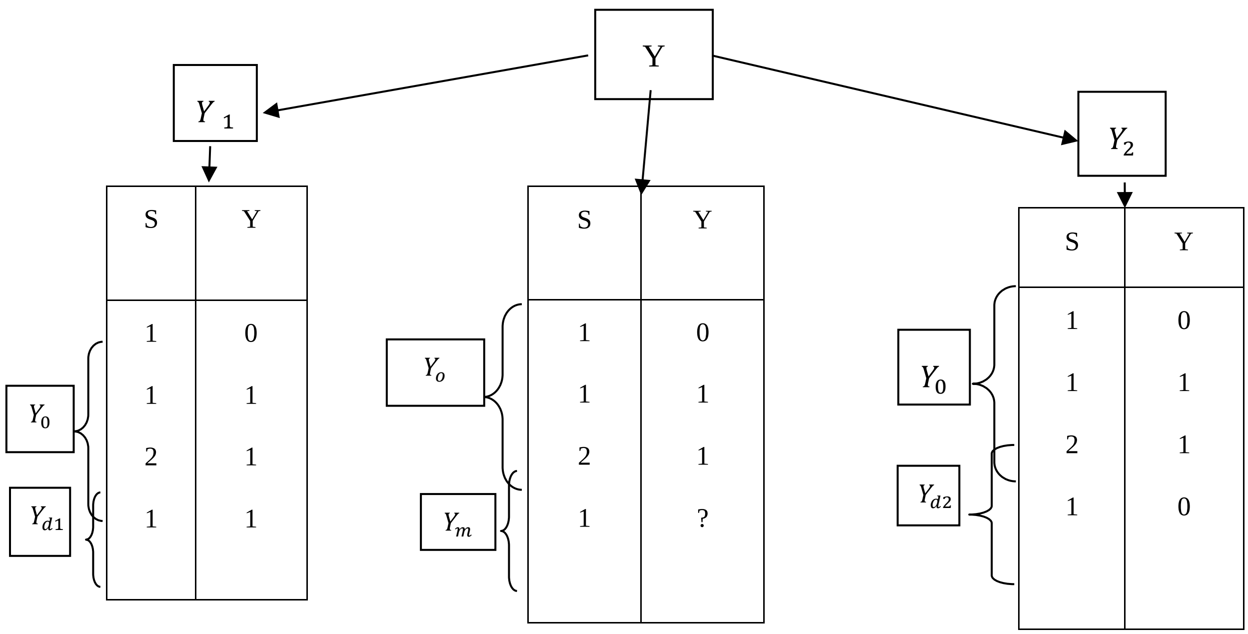


Figure 3.4: Possible Completions of the incomplete sample

Figure 3.4 shows possible completions of the incomplete sample$(S = i, Y = ?)$. In the incomplete sample let us denote n_{ij} the frequency of complete cases $(X = i, Y = j)$ and let m_i be the frequency of cases $(S = i, Y = ?)$. Consequently $n = \sum_{ij} n_{ij}$ is the number of cases completely observed, $m = \sum_i m_i$ is the total of cases partially observed and $n + m$ is the sample size of Y, the completed sample.

Using the argument of Ramoni and Sebastiani (2000) based on the Law of Total Probability (LTP), the exact posterior of $\mathbf{\Theta}$ is a mixture of Dirichlet distributions and the probability of the possible completions of Y, where the latter is computed if the missing data mechanism information is known. Therefore, the posterior distribution of $\mathbf{\Theta}$ is given as:

Equation 72: Posterior Distribution of the parameter space (Θ)

$$p(\mathbf{\Theta}|Y) = \sum_{d_k} p(\mathbf{\Theta}|Y_{d_k})p(A_{d_k}|Y)$$

with $p(Y_{d_k}|Y) \propto p(Y_{d_k}|Y_0)$ information computable when information about the pattern of the missing data is estimated. The information may lead to the formulation of a probability model of the missing data mechanism which looks as follows:

Equation 73: Probability model of missing data mechanism (MDM)

$$p(S = i, Y = j|Y = ?, \theta, \varphi) = \theta_{i+}\varphi_{j|i}$$

where $\varphi_{j|i} = p(Y = j|S = i, Y = ?, \varphi, \theta)$ which describes explicitly the embedded missing data mechanism.

Equation 74: Probabilistic model of missing data pattern

$$\varphi_{j|i} = p(Y = j|S = i, Y = ?, \theta, \varphi)$$

$$= \frac{p(S=i,Y=j|\theta)}{p(S = i|\theta)} \frac{p(Y=?|S=i,Y=j,\varphi,\theta)}{\sum_j p(Y=?|S=1.Y=j,\theta,\varphi).p(Y=j|S=i,\theta)}$$

$$\propto \theta_{j|i} p(Y = ?|S = i, Y = j, \varphi, \theta)$$

$$\cong \theta_{j|i} k_{ij}$$

where $k_{ij} \propto p(Y = ?|S = i, Y = j, \varphi, \theta)$, which explicitly describes the mechanism that generates the missing data. As soon as the probabilistic model of the missing data pattern is defined the mixture weights of the exact posterior distribution are computed as:

Equation 75: Conditional Probability of the MDM

$$p(Y_{d_k}|Y_0) = \int p(Y_{d_k}|Y_0, \varphi, \theta) p(\varphi, \theta|Y_0) d\varphi d\theta = \int p(Y_{d_k}|Y_0, \varphi, \theta) p(\theta|Y_0) p(\varphi|Y_0, \theta) d\varphi d\theta.$$

From equation (75) the effect of the missing data mechanism is visible in the shape it gives to the probability $p(Y_{d_k}|Y_0)$ through φ. For instance, if the missing data mechanism is assumed MAR, the probability of $Y = ?$ does not depend on Y but may depend on S, otherwise if not dependent on S neither, the missing data pattern is MCAR. This implies that the probability $p(Y = ?|S = i, Y = j, \varphi, \theta) = p(Y = ?|S = i, \varphi, \theta)$ and $k_{ij} = 1$ and therefore $\varphi_{j|i} = \theta_{j|i}$ from equation (75). This illustrates the ignorable character of the MAR and MCAR missing data mechanisms (Little & Rubin, 1987). Therefore, this implies that the mixture weights of the exact posterior distribution become:

Equation 76: Mixture weights of posterior distribution

$$p(Y_{d_k}|Y_0) = \int p(Y_{d_k}|\theta)p(\theta|Y_0)d\theta$$

The situation is drastically different if missing data mechanism is MNAR, where the probability of the missing on Y is dependent on both S and Y. Here the computations in equations (75) and (76) are only possible if only a prior distribution on φ is explicitly specified. As a result of the fact, under MNAR, that the required posterior conditional distribution is mixture on all possible completions of the sample in response to available information. Owing to this scenario the simplicity due to conjugacy is lost out, unlike under ignorable (MCAR and MAR) conditions. Under the non-ignorable (MNAR) conditions, some simplifying prior independence mechanisms can be used to compensate for the loss of conjugacy in analysis. Forster and Smith (1998) find that the posterior distribution of the marginal $\boldsymbol{\theta_I}$ and conditional $\boldsymbol{\theta_{J|I}}$ are independent and this simplifying condition is given in Theorem 1 that follows:

Theorem 1

Let Y be an incomplete sample in which n_{ij} is the frequency of observed cases $(S = i, Y = j)$, and m_i is the frequency of $(S = i, Y = ?)$. If $\boldsymbol{\Theta} \sim D(\boldsymbol{\alpha})$, the posterior of $\boldsymbol{\theta_I}$ is $D(\alpha_{1+} + n_{1+} + m_1, \ldots \ldots \ldots \ldots .., \alpha_{r+} + n_{r+} + m_r) \equiv D(\boldsymbol{\alpha_I} + \boldsymbol{n_I} + \boldsymbol{m})$, and $\boldsymbol{\theta_I}$ and $\boldsymbol{\theta_{J|I}}$ are independent.

(Sebastiani & Ramoni, 2000).The most fundamental and far reaching consequence of Theorem 1 is that in the posterior distribution of $\boldsymbol{\Theta}$, the marginal independence of $\boldsymbol{\theta_I}$ and $\boldsymbol{\theta_{J|I}}$ are not in function of the incompleteness of the sample either under MAR or MNAR assumptions. This implies that under both assumptions the Bayesian estimate of $\theta_{\boldsymbol{i}+}$ is the same given by:

Equation 77: Bayesian Estimator of the marginal distribution

$$\widehat{\theta}_{i+} = \frac{\alpha_{i+} + n_{i+} + m_i}{\alpha + n + m}$$

For the conditional probability distribution $\boldsymbol{\theta_{J|I}}$, under the MAR

assumption, is simplified because the incomplete cases are ignored in the inference about the conditional probabilities of $(Y = j|S = i)$. This means that for $\boldsymbol{\theta_{J|I}}$, the missing data mechanism is MAR, a result confirmed by Spiegelhalter and Lauritzen (1990) is in a theorem that follows:

Theorem 2

Suppose that the missing data mechanism is MAR. Then the distribution of $\boldsymbol{\theta_{J|I}}$ factorises into a product of independent Dirichlet distributions $D\left(\boldsymbol{\alpha_{J|i}} + \boldsymbol{n_{J|i}}\right)$. (Sebastiani & Ramoni, 2000).

The obvious consequence of the Theorem 2 is that the MAR assumption implies that missing data or incomplete cases in the sample are ignorable for the Bayesian estimates of $\theta_{j|i}$ calculated by:

Equation 78: Bayesian Estimator of the conditional of missingness

$$\hat{\theta}_{j|i} = \frac{\alpha_{ij} + n_{ij}}{\alpha_{i+} + n_{i+}}$$

A conspicuous exclusion from Theorems 1 and 2 is the type of the distribution of joint posterior probability of (S, Y) under the MAR assumption. The Dirichlet distributions of $\boldsymbol{\theta_I}$ and $\boldsymbol{\theta_{J|I}}$ have been respectively pronounced in the Theorems 1 and 2. In the joint posterior distribution of (S, Y), the lack of corresponding conjugate analysis is attributable to the unbalance report of information about S and Y in the sample. The loss of conjugacy is only recoverable if the missing data mechanism is MCAR and when there are high complete cases in the sample. Under such a scenario the Bayesian estimate of $\theta_{\boldsymbol{i}+}$ is given as:

Equation 79: Bayesian estimator of $\boldsymbol{\theta}_{i+}$ under MAR

$$\hat{\theta}_{i+} = \frac{\alpha_{i+} + n_{i+} + m_i}{\alpha + n + m} = \frac{\alpha_{i+} + n_{i+}}{\alpha + n}$$

under the assumption that $\frac{m_i}{m} = \frac{\alpha_{i+}+n_{i+}}{\alpha+n}$ and the incomplete cases are ignored and do not add value in the estimation of θ_{i+}. Such a simplification is never possible under MAR assumption because the condition $\boldsymbol{\frac{m_i}{m} = \frac{\alpha_{i+}+n_{i+}}{\alpha+n}}$ does not hold in the estimation of $\boldsymbol{\theta_{i+}}$. The only possible computations under this condition is the calculation of $E\left(\theta_{ij}\right)$ and $Var\left(\theta_{ij}\right)$ as well as $E\left(\theta_{+j}\right)$ and $Var\left(\theta_{+j}\right)$ in closed form such that it would be possible to use the following expressions for the posterior distributions of θ_{ij} and θ_{+j} respectively: $\boldsymbol{\theta_{ij}} = \theta_{i+}\theta_{j|i}$ and $\theta_{+j} = \sum_{i=1}^{r} \theta_{i+}\theta_{j|i}$. Based on Theorems 2 and 3 and the arguments presented by Wilks (1963) it is confirmable that $\theta_{ij} = \theta_{i+} - \theta_{j|i}$ where $\theta_{i+} \sim D(\alpha_{i+} + n_{i+} + m_i, \alpha + n + m - n_{i+} - m_i)$ and $\theta_{j|i} \sim D\left(\alpha_{ij} + n_{ij}, \alpha_{i+} + n_{i+} - \alpha_{ij} - n_{ij}\right)$. This leads to demonstrate that:

Equation 80: Bayesian posterior estimator of the model parameter.

$$\hat{\theta}_{ij} = E(\theta_{ij}|Y) = \frac{\alpha_{ij} + n_{ij} + m_i \dfrac{\alpha_{ij} + n_{ij}}{\alpha_{i+} + n_{i+}}}{\alpha + n + m}$$

This proves the point that incomplete cases are non-ignorable for the estimation of $\boldsymbol{\theta_{ij}}$ and these incomplete cases are distributed across categories $(Y = j|S = i)$ in accordance to the distribution of $\theta_{j|i}$. It must also be noted that the expression for $\hat{\theta}_{ij}$ above is a generalisation of the Maximum Likelihood estimation illustrated by Little and Rubin (1987) by including the flattening constant α_{ij}. Employing the same argument as above, it can also be shown that:

Equation 81: Bayesian estimator of $\boldsymbol{\theta_{+j}}$

$$\hat{\theta}_{+j} = E(\theta_{+j}|Y) = \sum_{i=1}^{r} \frac{\alpha_{i+} + n_{i+} + m_i}{\alpha + n + m} \frac{\alpha_{ij} + n_{ij}}{\alpha_{i+} + n_{i+}}$$

This reduces to $E(\theta_{+j}|Y) = \frac{\alpha_{+j}+n_{+j}}{\alpha+n}$ when the missing data mechanism is MCAR and the condition $\frac{m_i}{m} = \frac{\alpha_{i+}+n_{i+}}{\alpha+n}$ holds.

Subsequently if we let $Var(\theta_{ij}|Y) = E(\theta_{i+}^2|Y)E(\theta_{j|i}^2|Y) - E(\theta_{i+}|Y)^2E(\theta_{j|i}|Y)^2$,

where;

Equation 82: Expectation of the conditional $(\boldsymbol{\theta_{i+}^2}|\boldsymbol{Y})$

$$E(\theta_{i+}^2|Y) = \frac{\hat{\theta}_{i+}(1 - \hat{\theta}_{i+})}{\alpha + n + m} + \hat{\theta}_{i+}^2$$

And:

Equation 83: Expected conditional $(\boldsymbol{\theta_{j|i}}|\boldsymbol{Y})^2$

$$E(\theta_{j|i}|Y)^2 = \frac{\hat{\theta}_{j|i+}(1-\hat{\theta}_{j|i})}{\alpha+n+m} + \hat{\theta}_{j|i}^2,$$

then the posterior variance of θ_{ij} can be expressed in function of the Bayesian estimates of θ_{i+} and $\theta_{j|i}$

In conclusion, under MAR assumption, there are no simple expressions for the joint posterior probability distribution of $\boldsymbol{\Theta}$, the probability of $(S = i, Y = j)$ and for the posterior conditional distribution $\Theta_{+\mathrm{J}}$, the marginal probability of $(Y = j)$. Therefore the corresponding inference under MAR are based on approximating methodologies like Markov Chain Monte Carlo (MCMC) simulations in the form of the Gibbs sampling approach and also the Expectation-Maximisation (EM) algorithms. For the MNAR assumption, the situation is even worse for no such simplifications have been developed other than the encoding of the probabilistic model of the missing data mechanism within the categories of S.

The scenario of the MNAR condition can be explained as follows to develop the much-searched simplifications as done under MAR assumption:

Let the parameter vector:

$$\boldsymbol{\varphi} = (\boldsymbol{\varphi}_{\boldsymbol{J}|\mathbf{1}}, \boldsymbol{\varphi}_{\boldsymbol{J}|2}, \dots \dots \dots \dots, \boldsymbol{\varphi}_{\boldsymbol{J}|r}),$$

where

$$\boldsymbol{\varphi}_{\boldsymbol{J}|i} = (\varphi_{1|i}, \varphi_{2|i}, \varphi_{3|i}, \dots \dots \dots ., \varphi_{c|i})$$

such that $\boldsymbol{\varphi}_{\boldsymbol{J}|i}$ parameterises the encoded probability model of the missingness of dependent variable. Subsequently the posterior conditional distribution of $\theta_{J|i}$ is a mixture of Dirichlet distributions. The preceding theoretical framework is a firm foundation on which the flexibility of the Bayesian reject inference methodology can be utilised to co-opt the missing data mechanism assumption into the learning of the posterior probabilities for the eventual inference based on incomplete samples. It is upon this firm theoretical foundation that a versatile model-based Bound and Collapse methodology is developed.

3.15 BOUND AND COLLAPSE MODEL-BASED IMPUTATION TECHNIQUE:

In the following sub-sections, the Bayesian model-based reject inference technique, BC, is explained in full detail, for it is central in building the much-needed AGB sample for designing the SME CRD model for the case-studied bank in line with Basel II Accord. The background of the methodology is explained to demonstrate the theoretical base on which it was built, unlike other reject inference techniques built on tenuous assumptions. For appropriate application of the technique, the mechanics of the technique will also be detailed therein.

3.15.1 Background

Literature (Jacobson & Roszbach, 1999; Chen & Astebro, 2012; Banasik, Crook & Thomas, 2003) has confirmed that the majority of reject inference techniques which have been used in CRM modelling have not yielded positive results because the majority of them are based on tenuous, questionable assumptions and not theoretically supported. It is generally assumed that the probability distribution over the accepted sample is equal to that of the rejected, a tenuous assumption. Some estimation techniques have been built on the view that reject inference is a sample bias selectivity problem; have also been tried in CRM domain. Widely used methods are the Maximum Likelihood Estimation (MLE) and Heckman-two stage bi-variate estimation but have not also yielded promising results in this domain and this is because they are based on very tenuous and ad-hoc assumptions and lack the fundament theoretical basis.

At one time, reject inference has been viewed as a missing data problem and correspondingly some tried methods have shown some promising results especially methods which embeds the uncertainty of the missing data mechanism into estimation process. But still non-model-based reject inference techniques in this category have shown little promise (Ditrich, 2015; Nguyen, 2016). This is basically because in non-model-based reject inference techniques, the underlying missing data mechanism is not random therefore no concrete theoretical support props them up and also consider estimated missing values as real values. Learning from incomplete dataset cannot be based on the same assumptions as for complete case.

The rationale behind treating reject inference as a missing data phenomenon has led to development of powerful and promising techniques soundly supported by principled statistical assumptions. This means reject inference techniques must be based on missing data assumptions to learn effectively from incomplete samples. On the other hand, Bayesian inference provides a strong theory to statistically deal with incomplete dataset or missing data problems. Therefore, a reject inference technique built on such solid, theoretically supported basis will go a long way in successfully tackling problems of inferring from conditional distributions from incomplete datasets, a phenomenon common in CRM problem domain. Current methods being used to tackle this problem of estimating conditional probabilities from incomplete samples embed the missing at random (MAR) assumption, which is not universal to all missing data problem domains. The missing at random (MAR) assumption implies that the eventual inference process will ignore the missing data mechanisms, allowing non-model techniques like list-wise deletion to be implemented, without incorporating the uncertainty embedded in the missingness.

Stochastic and iterative techniques in the form of EM algorithm, sequential updating and Markov Chain Monte Carlo simulation such as Gibbs sampling constitute the current weaponry in learning conditional probabilities from incomplete samples. The handicap of these powerful techniques is that they base on the assumption that the missing data mechanism is MAR. Therefore, it is futile to adopt these techniques to tackle problems in the CRM problem domain, where the missing data mechanism is non-ignorable because the missingness in this domain is considered a case described by data missing not at random (MNAR).

It is on this background that a deterministic model-based imputation Bayesian reject inference technique was developed by Ramoni and Sebastiani (2000) to tackle the problem of conditional probabilities from incomplete samples, a technique which does not assume any pattern on the missing data mechanism. This non-stochastic and non-iterative technique is called Bound and Collapse (BC) due to the two (2) basic strategies in its embodiment. It is an imputation technique which embeds the impact of missing data mechanism as well as exogenous supplementary information about the incomplete sample to be malleably incorporated into the learning from incomplete samples.

Unlike other reject inference techniques, prevalent in literature, the BC has a theoretically firm base on which it is developed, an edge over other techniques. The fundamental idea behind the BC is that it is possible to extract valuable information from incomplete dataset to set probability intervals which contain possible estimates of the missing data, without knowledge of the missing data pattern, Therefore the bounding stage. Point estimation of the missing datum is only achieved when knowledge of the pattern of missing data is available, encoded in a probability model, is then used to collapse the previously determined probability interval into a point value of missing data, Therefore the collapsing stage. In fact, the BC yields a randomly imputed point estimate of the missing datum, eventually achieving a completed sample representative of the population being studied. The embedded strategies of this methodology show that BC is a model-based data imputation technique of completing incomplete samples.

3.15.2 The mechanics of bound and collapse in CRM

For applying the BC in building an AGB development sample for effective CRM modelling we make the following assumptions that the joint probability distribution of (S, Y) has a multinomial distribution with probabilities:

Equation 84: Probability of the loan outcome parameter

$$\theta_{ij} = P(S = i, Y = j|\boldsymbol{\theta})$$

where $S - credit\ score$, $Y - credit\ risk$ and

$$\boldsymbol{\theta} = (\theta_{10}, \theta_{11}, \theta_{20}, \theta_{21}, \dots\dots\dots\ \theta_{r0}, \theta_{r1}) = (\theta_{ij}), (\theta_{ij} \geq 0, \forall i, j\ and\ \sum_{ij} \theta_{ij} = 1)$$

This parameterises the joint probability distribution of (S, Y). Looking at the distribution of $\boldsymbol{\theta}$ we find that it has a standard conjugate prior which follows a Dirichlet distribution $D(\boldsymbol{\alpha})$ with $\boldsymbol{\alpha} = (\alpha_{10}, \alpha_{11}, \alpha_{20}, \alpha_{21}, \dots\dots\dots\dots\dots\dots\ \alpha_{r0}, \alpha_{r1})$ whose probability density function (pdf) is:

Equation 85: Dirichlet probability density function

$$p(\theta) = \prod_{i=1}^{r} \prod_{j=0,1} \frac{\Gamma(\alpha)}{\Gamma(\alpha_{ij})} \theta_{ij}^{\alpha_{ij}-1}$$

with $\alpha_{ij} \geq 0$ for all i and j, and $\alpha = \sum_{ij} \alpha_{ij}$.

The prior expectation $E(\theta_{ij}) = p(S = i, Y = j) = \frac{\alpha_{ij}}{\alpha}$ and the prior variance;

$Var(\theta_{ij}) = \frac{E(\theta_{ij})\{1-E(\theta_{ij})\}}{\alpha+1}$, which is a measure of the uncertainty about $p(S = i, Y = j)$. Since $Var(\theta_{ij})$ is monotonic decreasing function in α, for fixed $E(\theta_{ij})$. Therefore, the parameter α is called precision. Since the joint probability distribution of (S, Y) is parameterised, then the marginal probability distributions of S and Y as well as the r and c conditional probability distributions of $(Y|s = i)$ and $(S|y = j)$ are also parameterised. For the resultant Dirichlet distribution is a conjugate prior distribution. For this distribution, there are two (2) possible, non-informative Dirichlet prior distributions ($\alpha_{ij} = 0$ and $\alpha_{ij} = 1$). To develop the BC theory, the following theorem is applied without proof:

Theorem 3

Let $\boldsymbol{\theta} = (\theta_{10}, \theta_{11}, \theta_{20}, \theta_{21}, \dots\dots\dots, \theta_{r0}, \theta_{r1}) \sim D(\boldsymbol{\alpha}), \boldsymbol{\alpha} = (\alpha_{10}, \alpha_{11}, \alpha_{20}, \alpha_{21}, \dots, \alpha_{r0}, \alpha_{r1})$, and for $i = 1, 2, 3 \dots, r$ and $j = 0, 1$ define

$p(S) = \theta_{i+} = \sum_j \theta_{ij}, p(Y) = \theta_{+j} = \sum_i \theta_{ij}, p(Y = j|S = i) = \theta_{j|i} = \frac{\theta_{ij}}{\theta_{i+}}$,

$p(S = i|Y = j) = \theta_{i|j} = \frac{\theta_{ij}}{\theta_{+j}}, \alpha_{i+} = \sum_j \alpha_{ij},\ and\ \alpha_{+j} = \sum_i \alpha_{ij}$.

Then

$$\boldsymbol{\theta_I} = (\theta_{1+}, \theta_{2+}, \dots\dots\dots, \theta_{r+}) \sim D(\boldsymbol{\alpha_I}), \boldsymbol{\alpha_I} = (\alpha_{1+}, \alpha_{2+}, \dots\dots, \alpha_{r+})$$

$$\boldsymbol{\theta_J} = (\theta_{+0}, \theta_{+1}) \sim D\left(\boldsymbol{\alpha_J}\right), \boldsymbol{\alpha_J} = (\alpha_{+0}, \alpha_{+1})$$

$$\boldsymbol{\theta_{I|j}} = \left(\theta_{1|j}, \theta_{2|j}, \theta_{3|j}, \dots., \theta_{r|j}\right) \sim D\left(\boldsymbol{\alpha_{I|j}}\right), \boldsymbol{\alpha_{I|j}} = (\alpha_{1|j}, \alpha_{2|j}, \dots\dots., \alpha_{r|j})$$

$$\boldsymbol{\theta_{J|i}} = \left(\theta_{0|i}, \theta_{1|i}\right) \sim D\left(\boldsymbol{\alpha_{J|i}}\right), \boldsymbol{\alpha_{J|i}} = (\alpha_{0|i}, \alpha_{1|i})$$ (Sebastian & Ramon, 2000)

If our original development sample is complete, learning from conditional probabilities using Bayesian analysis is easy. However, but the difficult comes when other values of the credit quality are missing because the corresponding credit risk score is less than the cut-off value. It is when the reject inference is needed to impute the missing values.

Using set theory, the missingness mechanism can be easily illustrated as follows: Let C be the complete sample made up of two (2) sub-samples C_{obs} and C_{mis} where C_{obs} is a two (2) variable sub-sample with complete observations and C_{mis} has missing values of the credit quality such that $C = C_{obs} \cup C_{mis}$. Let the probability completion of C be denoted by C_p such that $C_p = C_{obs} \cup C_{misp}$, where C_{misp} is a probability distribution of the missing values in C_{mis}. To determine the size of the complete sample C, we denote the size of sub-sample C_{obs} by n_{ij}, the frequency of complete cases and also denote by m_i, the frequency of incomplete cases. Letting $n = \sum_{ij} n_{ij}$ be total number of observable cases in Y_{obs} whilst $m = \sum_i m_i$ be the total of partially observable cases in C_{mis}. Therefore, the sample size of C will be $n + m$.

In credit risk modelling the known-good-bad (KGB) sample is certainly incomplete. To characterize it we represent it by an $r \times (c + 1)$ contingency table, where $c = 2$ and the $(c + 1)th$

column represents the frequency of unknown cases for each category of S (score range) (Little & Rubin, 1987) as shown in Table 3.2

Table 3.2: Contingency table of incomplete sample

	Y		**Y**
Score ranges	0	1	?
1	n_{10}	n_{11}	m_1
2	n_{20}	n_{21}	m_2
3	n_{30}	n_{31}	m_3
.	.	.	.
δ	$n_{\delta 0}$	$n_{\delta 1}$	m_δ
.	. .		.
r	n_{r0}	n_{r1}	m_r

The posterior distribution of **Θ** is a mixture of Dirichlet distributions weighted by the probabilities of possible completion of C, can be derived if information about the missing data mechanism is availed (Sebastiani & Ramoni, 2000). If the missing data mechanism is non-ignorable, simple conjugate Bayesian Analysis is no longer possible. Using theory behind the Theorem of Total Probability (TTP), Sebastiani and Ramoni (2000) demonstrated that the incomplete sample can induce bounds on the possible estimates of missing data consistent with the available information, in the absence of any exogenous information about the missing data mechanism. If information about the missingness is availed, such information, encoded in probability, can be used to randomly select a single estimate within the set of possible ones. The possible estimates of $P(Y = j|S = i|C)$ bounded as:

Equation 86: Bound of $\boldsymbol{P(Y = j|S = i|C)}$

$$p_{\circ}(j|i) = \frac{\alpha_{ij} + n_{ij}}{\alpha_{i+} + n_{i+} + m_i} \leq \ P(Y = j|S = i|C) \ \leq \frac{\alpha_{ij} + n_{ij} + m_i}{\alpha_{i+} + n_{i+} + m_i} = p^{\circ}(j|i)$$

where $\alpha_{i+} = \sum_j \alpha_{ij}$ and $n_{i+} = \sum_j n_{ij}$.

For each credit risk measure score range, incomplete cases induce a set of $c = 2$ extreme distributions corresponding to the most extreme situations in which data are systematically missing on one category of Y. Any assumption about the pattern of missing data will induce a distribution of $\theta_{J|i}|S$ within these extreme distributions. Information about the missing data mechanism can therefore be used to identify a single distribution within these bounds, a key idea of the collapse step of BC.

We assume that some exogenous or external information on the missing data is available, from which a probabilistic model for missingness can be derived:

Equation 87: Probabilistic Model for missingness

$$P(Y = j|Y = ?, X = i, \varphi, \boldsymbol{\theta}) = \varphi_{j|i}, \text{where } \sum_j \varphi_{j|i} = 1 \text{ for all } i$$

Assuming also the most likely probability for $P(Y = j\big|S = i\big|C) = p^{\circ}(j|i)$ is $\varphi_{j|i}$ and correspondingly the most likely probability of $P(Y = j|S = i|C) = p_{\circ}(j|i)$ is $1 - \varphi_{j|i}$ a point estimate $\widehat{p_{J|\iota}}$ within the probability interval $[p_{\circ}(j|i), p^{\circ}(j|i)]$ via a convex combination of the extreme probabilities is estimated as:

Equation 88: Point Estimator of $\boldsymbol{P(Y = j|S = i|C)}$

$$\widehat{p_{J|\iota}} = \varphi_{(j|i)} p^{\circ}(j|i) + \left(1 - \varphi_{(j|i)}\right) p_{\circ}(j|i)$$

$$= \frac{\alpha_{ij} + n_{ij} + \varphi_{j|i} m_i}{\alpha_{i+} + n_{i+} + m_i}$$

For effective determination of such estimates from (88), there must be explicit specification of the non-ignorability of the missing data mechanism of equation (87). If the expression (87) is specified, equation (88) is used to impute datum for missing datum per respective score range. This estimate incorporates the CRM score into the missing data imputation. Therefore, great effort is exercised to estimate the missingness; otherwise the value for each missing datum based on equation (88) would not correctly represent the expected datum. When all missing data have been imputed we have a complete AGB sample for future CRM model development.

3.15.3 Estimation of the Missingness Probabilities

The issue of how missing data mechanisms should be estimated needs thorough discussion for that has some fundamental bearing on the application of the Bound and Collapse methodology. In the presence of the MNAR missingness, there are two (2) modelling approaches to determine the missing data pattern, these are selection and pattern mixture models. For this work, pattern mixture modelling has been adopted regardless of the appealing selection model. Probabilistically, the pattern mixture model is:

Equation 89: Pattern Mixture Model

$$P(Y_{miss}, Y_{obs}) = P(Y_{obs}|Y_{miss})P(Y_{miss})$$

The outright implication is that we classify missing data by their missingness and describe data within each missing group. Therefore, without information about Y_{miss}, we would not have known any characteristic of its distribution, resulting in a model undefined. This exhibits the fact that for a pattern mixture model we require some identifying constraints from information on missing data. As for the case of CRM, the probability of being bad is used as a proxy for the missing data mechanism (MDM) because probability of being bad is equal to probability that credit application is rejected. This implies that the original CRM score is equal to probability of missing. It is considered an external source of information for estimating the missing data mechanism (MDM).

The bad rate of the current accepts (preliminary model) can also be used to infer the missing data mechanism (MDM), internal source of information. The inference for missing uses linear model regressions in the form of linear, logarithmic or exponential extrapolations. For best results, the missing data mechanism (MDM) is best approximated by considering a weighted average of external and internal information. In this work, the missingness was modelled through simple linear regression model, where interval bad rate was regressed over mid-class CRM score. The missingness model was derived from the "accepted subsample" where the credit screening was applied to provide information of bad rates. This constitutes the internal information, which is extrapolated to estimate bad rate distribution in the corresponding "rejected subsample".

3.15.4 Selection of Priors

Since the BC model uses a multivariate generalisation of the Beta distribution, the Dirichlet distribution, as the conjugate prior distribution, there is great need to select appropriate priors for its application. Several non-informative Dirichlet priors offers strong candidature for the selection. For simplicity, model developers settle to choose between $\alpha_{ij} = 0$ and $\alpha_{ij} = 1$ priors. If one desires to achieve a uniform density such that the Dirichlet density function assigns equal weight to any vector $\boldsymbol{\theta}$ compliant to the constraint that $\sum_{ij} \theta_{ij} = 1$, one would settle for $\alpha_{ij} = 1$ for all i and j. On the other hand, if one settles for $\alpha_{ij} = 0$ for all j one would get an improper prior distribution, that would be uniform in the log (θ_{ij})'s but the resultant posterior is proper if there is at least one observation in each score range. Fortunately, if the concerned sample were relatively large, then the difference in results between these two (2) prior densities would not be that big. On that basis we, for this work, selected the non-informative Dirichlet prior $\alpha_{ij} = 0$ and based on simplicity in the application of the Bound and Collapse reject inference technology.

3.16 IMPLICATIONS FROM THE RESEARCH LITERATURE

The findings in the research literature herein summarised reveal that there is no overall best statistical approach used in building CRM model, implying that best technique for all circumstances does not exist. Developing a CRM model is a process that entwine different areas of statistics, in fact application of CRM methodologies have been extended to include various disciplines with subsequent goal to help credit decision-makers in the banking industry, in particular SME banking, to forecast their client's credit quality. The phenomenon of borrowing and lending has long history associated with human behaviour (Stepanyan, 2018; Astrom, 2015; Haldar & Stiglitz, 2016), but the history of CRM modelling, for SMEs in particular, is very short and literature is very limited.

The major deficiency of current CRM modelling methodologies is the overlook of the basic grounds on which statistical modelling is built. The representativity character of a sample is of paramount importance for any statistical estimation of parameters of the target population of interest. In fact, there should be no mismatch between model development sample and the target population to be measured in future, in the case of CRM statistical modelling. Therefore literature (Smith & Elkan, 2004; Ditrich, 2015; Nguyen, 2016; Schouten, Lugtig & Vink, 2018) has revealed that traditional statistical methodologies cannot be directly employed on truncated, incomplete samples, a case common in CRM statistical modelling.

The prerequisite sample must also embrace rejected applicants, a dilemma researcher's face, since banks do not record full profile of such clients. This can be resolved through reject inference approaches. Literature has demonstrated that traditional reject inference techniques have not yielded satisfactory results due to the fact they do not incorporate the missingness of the data and that they are not supported by theoretical assumptions. There are suggestions to solve the reject inference problem using missing data techniques. In that regard, data imputation approaches have proved the worthy to complete samples before the traditional logistic regression can be applied to CRM statistical modelling. This has led to the choice of the Bound and Collapse Bayesian inference approach, a deterministic model-based approach to missing data imputation, which incorporates the uncertainty caused by the missing data mechanisms into the subsequent estimation of the missing data.

3.17 SUMMARY

This chapter has provided broad reviews on financial risks, PD and credit risk models studies. Common CRM models are either judgemental or statistical. The choice of the model to measure credit risk of a bank is depended on the bank information technology and on the circumstances data availability. From this literature review, there is no single CRM model that has been classified as the best model that would be suitable for all banks. Internal ratings-based approach of the Basel II enunciate that a variety of factors determine the best model for the individual bank. In that regard, individual banks need quantitative expertise for internal CRM model development and for prudent selection from the available methodologies.

To estimate default probabilities for individual borrowers, statistical CRM models have been more frequently applied than market-based models. As a result, the model lifecycle has been defined to make it procedural approach to guide credit risk analysts. A series of statistical CRM models have been considered using personal, enterprise and accounting data to predict SME default. These include Multiple Discriminant Analysis (MDA), Logistic regression models, Data Envelopment Analysis (DEA), Neutral Network (NN), and Genetic Algorithm (GA) methods. CRM models play an important role in producing estimated PD for individual borrowers, typically using borrower demographic characteristics, financial ratios and other characteristics as risk predictive variables. The major advantages of the CRM models are that they are objective, efficient and consistent. In addition, these models are comprehensive borne out of robust statistical principles.

Various studies have so far been applied to measure the credit risk of SMEs since the number of these enterprises that seek financing from banks in any economy has exponentially grown. This has necessitated automation of the decision process, which requires processing of large volume of data, enables the banks to cut cost through speeding up the approval process henceforth improve on efficient service delivery to their clientele. Subsequently the scarcity of reliable database for

SME statistical CRM modelling, would be overcome as automation of decision process builds up large and reliable databases within a relatively short period. Statistical prediction of SME PD would not work properly if data is limited in quantity. Literature has confirmed that Berger and Udell (1995); Lin (2007); Nguyen (2016); and Saurina and Trucharte (2004) found out that a very large proportion of small loans are granted to SMEs, but researchers are still CRM modelling on experimental basis implying that a lot of research effort in the SMEs CRM are still open to incorporate innovative approaches.

Furthermore, a whole range of modelling techniques has been developed to analyse portfolio credit risk, but few studies have been done specifically for SME portfolio. The majority of them seek to offer alternative approaches to measuring the credit risk of an SME loan or a portfolio of SME loans. Such models are increasingly based on reject inference considered as a missing data problem to solve inherent problems of selectivity bias in CRM domain.

Theoretically model-based reject inference techniques are currently used in different segments in banking industry. From above literature, models which are being used traditionally to counter selectivity bias are built on tenuous assumptions and lack theoretical support, therefore have not yielded substantial results. Therefore, many criticisms have been raised by governments and SME associations that high capital charge for SMEs despite the importance of these firms are to the economy. Therefore, there is a need to demonstrate that banks should develop credit models specifically addressed to SMEs to minimise their expected and unexpected losses. Many banks and consulting companies already follow the practice of separating large corporate from small and medium sized companies when modelling their credit risk. In the academic literature, however, a study that demonstrates the significant benefits of such a choice is lacking.

CHAPTER 4 RESEARCH METHODOLOGY

4.1 INTRODUCTION

In this chapter a thorough discussions on data collection methods and methodology used to deliver answers to the problem statement and research questions pronounced in Chapter 1 were done. In line with the objectives set for this work, appropriate research designs, data collection tools, sampling procedures, data analysis methodology techniques have been justifiably selected and discussed in detail. It is therefore prudent to initially recapitulate the problem statement, research objectives as well as research questions. These are the determinants of the shape and form of this chapter.

4.2 PROBLEM STATEMENT AND RESEARCH QUESTIONS

CRM for SMEs in Zimbabwe is not readily quantifiable as compared to retail and corporate credit risk; as a result, Commercial banks have found it insurmountably challenging to enter into the promising but risky SME lending market. Since 2000, the economic situation in Zimbabwe has drastically changed to the extent that the majority of economic and business activities are being handled by SMEs (Maseko & Manyani, 2011; Gwangwava, Faitira, Gutu, Chinoda & Frank 2014; Nyamwanza, 2014). The SME sector's experiences have provided a scholarly laboratory to illustrate that an economy can be sustained and anchored on the dynamism of the SME sector and that social inequalities can be shaded off if the growth and development of the sector exponentially increase in the national economy.

From the supervisory perspective, Basel II Capital Accord principles are grounded in recognizing an individual credit risk through internal rating-based (IRB) systems, which implore banks' managers to correctly measure credit risk of any asset and price it correctly. The SME asset's credit risk has not been readily quantifiable to the detriment of SME financing; which translates to incorrect pricing of loans thereby stifling credit supply to the sector. Therefore, the overall goal of this research work is to develop a CRM model specifically for SMEs to facilitate banks, in Zimbabwe, in particular, and in emerging economies at large, to enter into the purported risky but promisingly lucrative SME lending market. To achieve this goal the following research questions must be answered:

1. What are the factors that cause SME financing constraints from the perspectives of SMEs and the banks in Zimbabwe?
2. How do banks measure credit risk of SMEs in Zimbabwe?
    2.1 How do we minimise selectivity bias present in the KGB sample?
    2.2 What are the criteria used by banks when measuring credit risk of SMEs in Zimbabwe?

4.3 PHASE 1: BANK INVOLVEMENT IN SME LENDING IN ZIMBABWE

The following sub-sections explain the methological procedures for Phase I of the thesis whose aim is to determine the extent to which banks in Zimbabwe are involved with SME lending. Bank and SME survey procedures are explained in order to answer the first research questions.

4.3.1 Data: Bank survey

At time of the study, Zimbabwe had a population of 14 Commercial banks, including both banks that were ex-ante thought to be serving SME sector and banks that were perceived ex-ante to have little or no intention to serve the emergent SME sector. Since official statistics concerning SME banking are scarce in Zimbabwe, a survey on banks was instituted with the aim to assess banks' involvement with SME financing and better understand their business models for serving the SME sector. This constituted a supply side survey. To explore and confirm the reigning situation with regard to SME financing in Zimbabwe, we adopted a non-probability sampling design since bank participation was voluntary. The sampling was non-probabilistic because banks are generally conservative and secretive that they hardly participate in such surveys for fear of exposure of their confidential and security information about their internal activities and clientele.

On that backdrop a convenience sample design was adopted. The sample represented the spectrum of domestic financial service sector of Zimbabwe. To facilitate data collection two (2) methods of data collection were used which were self-administered questionnaire and face-to-face personal interviews with people directly responsible for the SME banking issues (bank managers, credit managers and loan officers). To complement bank survey, some personal interviews were also carried out with alternative sources for SME financing such as Small Enterprises Development

Corporation (SEDCO), Infrastructure Development Bank of Zimbabwe (IDBZ) and some government agencies.

The questionnaires were distributed to all the 19 banks. The participating banks were assured that the information they provided would remain anonymous and analysis results would be presented in aggregated fashion. In the selection of banks, it was made sure that the selection process fitted the Zimbabwe lending infrastructure, which included the information environment, the legal, judicial and bankruptcy environment, the social, and the tax and regulatory environment (Berger & Udell, 2006). The country constituted a divergent lending infrastructure which epitomizes the lending infrastructure of any emerging economy at large. Therefore, given that background, the question was; how are the banks in Zimbabwe being involved with SME financing?

Literature (Kanapickiene & Spicas, 2019; Al Baz, 2017; Horstedt & Linjamaa, 2015; Nemoto *et al.*, 2018; Kennedy, 2013; Lin, 2007) confirms that commercial banks are the predominant sources of external funding of SME sector. This bank survey was comparable to the ones the World Bank and other authors carried out across countries (de la Torre *et al.*, 2010; Rocha, Farazi, Khouri & Pearce, 2011; Ackah & Vuvor, 2011; Berg & Fuchs, 2013). Most of such surveys were done across emerging economies. Zimbabwe being an emerging economy, the measurement instrument used in this bank survey was comparable to those used by World Bank and other researchers, confirming the reliability and validity of the research design and data collection instruments used in this survey.

During bank survey data collection process, ethical considerations were the guiding principles as any bank data is regarded highly confidential. Therefore interviews, questionnaires and data processing procedures were done anonymously maintaining maximum confidentiality to avoid damaging the reputation of the sampled banks. The data collection tool consisted of a tailored questionnaire with a maximum of 114 questions designed to extract information on eight (8) crucial thematic areas which would help assess Zimbabwean banks' side involvement with SME financing. The areas included among other things:

Table 4. 1: Bank survey questionnaire structure

Section	Theme	Number of Questions
A	Institutional Assessment	10
B	Government Policy on SME finance	6
C	Training of Personnel on SME lending	11
D	Competitive Environment for SME lending	10
E	Government Programs and Actions Affecting SME lending	16
F	FA: Bank Policies and Procedures to finance SMEs and Reduce associated costs: Bank's SME Business Models	25
	FB: Credit Risk Management Processes	20
	FC: Bad Loan Recovery Strategies	16
TOTAL		**114**

The questionnaire, in summary, queried how banks in Zimbabwe perceived the emergent SME segment vis-à-vis government programs in support of SME financing as well as the Basel regulatory and supervisory framework being adopted by the Reserve Bank of Zimbabwe (RBZ). It also queried on whether banks had been proactive enough to adopt business models that specifically suit SME financing. SME as a non-traditional bank asset is very distinctive of the retail and corporate assets, therefore the questionnaire examined what instruments banks were using to approve or evaluate SME loan applications prior to granting a loan. The issue of collateral is very crucial to any bank lending transaction; in contrast SMEs normally do not have collateralizable assets, so the questionnaire queried what type of collateral the bank used to secure loans extended to SMEs. Lastly the instrument, queried the banks' credit risk management as well as how to recover bad SME loans. In fact, the questionnaire tried to gauge the degree, type and pricing strategies of banking to SMEs in distinction with other traditional bank assets.

For statistical and comparison purposes, the heterogeneity of the SME definition was synchronized between banks and country. Country definition, through the Ministry of Small and Medium Enterprises and Cooperatives, is normally based on the number of employees, annual turnover and assets. On the other hand, banks generally define SMEs basing on average annual sales, a readily observable indicator of business activity and more useful to evaluating SME loans. Again, the threshold of annual sales used by banks varied in accordance with the size of the respective bank's portfolio and structure of its corporate sector (de la Torre *et al.*, 2010).

To achieve the overall goal of this work, the discrepancies caused by heterogeneity in the definitions of SMEs was toned down by only considering the surveyed bank's own SME definition to avoid comparison challenges as well. The questionnaire included both quantitative and qualitative questions to capture detailed business processes as well as operational challenges entailed in SME financing and identify factors that might facilitate the successful bank involvement with SME financing in Zimbabwe, in particular, and in any emerging market economy.

The strategic approach to understanding the bank-SME financing gap used in this work was to directly target the supply side (banks) and demand side (SMEs) players that confronted it every day. On the supply side the key personnel involved in loan underwriting included: SME Banking Managers, Credit Risk Managers, Credit analysts, loan officers, so were the targeted informers regarding banks involvement in SME lending business. To counter the prevalence of non-responding to questionnaires, in-depth interviews with key personnel of the banks were done on-site to all sampled banks within the scope of the research to illuminate the financing gap from supply side perspectives. Such interviews pinpointed the nuanced challenges that banks encounter in financing SME sector and specific links in the value chain. Banks as well as government efforts in filling the gap were also probed in the self-administered questionnaires as well as in the face-to-face interviews done at collection of self-administered questionnaires.

4.3.2 Data: SME survey

For the SME survey, the sampling design used was purposive because the sampling frame of SMEs in Zimbabwe was non-existent due to divergent definitions of SME. Fortunately, in Zimbabwe there is an Association of Small and Medium Enterprises (SMEAZ) which is constituted by a sizeable number of entrepreneurs from different towns and cities of Zimbabwe. The memberships worked in unison to market their products and to reach out for financing from banks and government agencies. This association has a role to entice commercial banks to be involved with

SME financing, given the new role SMEs have assumed owing to the unprecedented closure of big firms due to economic downturn, since 2000. The Zimbabwean economy has not been conducive for big businesses to the extent that the majority of large corporations folded into liquidation or scaled down operations. As a result, SMEs are emerging as the major providers of goods and services, but banks still regard them the riskiest asset ever in banking sector. This creates the need to fill up the financing gap that has engulfed the Zimbabwean economy between the suppliers of finance (banks) and the emerging demanders of finance (SMEs).

To corroborate information provided by banks, an SME survey was also carried out to investigate the extent of bank involvement with SME financing from a demand side perspective. SMEs as recipients of the much-needed financing from banks, the custodian suppliers of loans, have their expectations and needs which must be accounted for and understood to reduce information asymmetry between the two (2) sides. Since the sampling frame of SMEs in Zimbabwe is non-existent, the sampling frame of SMEAZ is used as its proxy. Like banks, SMEs fear to divulge information related to banking activities for fear of taxation as well as to strain relations with their bankers.

The association has as a website where members interact and share business strategies and information. The website facilitated internet survey, where a tailored questionnaire was flight and volunteers who responded constituted the purposive sample of 300 entrepreneurs from different economic activity areas out of 705 memberships at the time of survey. The questionnaire aimed to elicit SMEs particular financing constraints along their business development trajectory. The questionnaire comprised of 4 thematic sections soliciting different aspects: demographic data of the SME owners, general information on the type and situation of the firm, financing of the firm data and future, growth and obstacles data. The structure of the questionnaire is shown in Table 4.2

Table 4.2: SME survey Questionnaire Structure

Section	Theme	Number of Questions
A	Demographic part: General characteristics of the SME owner.	22
B	General Information on the type and situation of the enterprise	4
C	Financing of the enterprise	11
D	Future, Growth and Obstacles	7
Total		**44**

In summary the SME survey instrument of measurement collected demographic information of the SME owner(s) and general information about the respective enterprise. It also elicited information pertaining to the number of banking relations a given enterprise had and why the enterprise had either increased or reduced its operations with some certain banks. The questionnaire queried about the type and number of banking products the SME was using as well as how enterprises financed foreign trading, that is, whether they used internet banking or any form of banking for foreign business transactions. Lastly the questionnaire queried on the obstacles and challenges SMEs were constantly facing as well as their growth prospects in the face of constraint funding from banks.

4.3.3 Survey data analysis methodology

For a profound assessment of Zimbabwean banks involvement in SME lending, both quantitative and qualitative data were collected using a questionnaire similar to a World Bank designed questionnaire, which was used for cross country survey in Argentina, Colombia, Chile and Serbia (2006-2007). The objective pursued by this methodology was to find out whether there was any transformation in the way banks in Zimbabwe interact with SMEs from tradition, whether they now perceive the enterprises as core and strategic business partners given that the traditional lending market for the majority of banks is fast collapsing due to economic demise in Zimbabwe since the year 2000 (Maseko & Manyani, 2011; Nyamwanza, 2014; Gwangwava *et al.*, 2014). The degree of banks' involvement with SMEs would justify the need for conceited effort to develop specialised SME CRM technologies. The SMEs on the demand side of the equation were probed of their survival reaction in the face of constraint financing and how they strategically planned for growth under such unfavourable economic regime.

The bank and SME surveys allowed balanced analysis of bank engagement with SMEs basing on exhaustive viewpoints of both the supply and demand sides. During bank survey, it was noted that some of the quantitative information requested through the questionnaire or face-to-face interviews was not available and had to be estimated by average data for some banks, thereby making the extracted average data not representative of the whole banking sector in Zimbabwe. Given the nature of such datasets derived from surveys, econometric analyses of determining factors affecting banks involvement with SME financing would give biased and misleading results.

To circumvent such subsequent erroneous statistical inferences of respective target populations, it was resolved to confine discussions of the surveys' results within the bounds of descriptive statistics and observations from the two (2) surveys. For the banks' involvement with SME financing, data obtained was quantitatively analysed using descriptive statistics techniques; bar charts, pie charts as well as histograms. The analysis of questionnaires was done by using SPSS 13 software to derive useful information out of responses given.

Therefore, the evidence presented in Chapter 5 should be interpreted with caution as must not be interpreted as conclusive about the respective target populations studied through surveys. In fact, the ultimate goal of both surveys was to highlight the intrinsic correlations and observations which warrant further scrutiny or further research. For instance, the second phase of study, dovetailed from the discovered correlations and observations from the surveys' analyses.

4.4 PHASE II: CRM MODEL BUILDING

For this part of the work, the research design employed was a case-study, which is line with Basel II IRB Framework, which is grounded in recognizing an individual bank asset's CRM through IRB system; where the bank's managers must correctly measure credit risk and price it correctly. Any CRM model must be accurate on average across the range of the assets or facilities to which a bank is exposed and there must be no known material biases (OECD, 2015). For the case-study, a bank in Zimbabwe was conveniently selected, for fear of confidentiality and competition; the majority of banks were reluctant to offer their credit information for research purposes regardless of ethical assurance given by researchers. The case-studied bank offered credit register of its SME loan portfolio in the understanding that the ethical code of conduct signed between the bank and the researchers would be observed in principle during data collection and analysis. Unlike for the first phase, the data was secondary and historical covering a two-year period from 2010 to 2012.

In terms of default information, very few banks in emerging countries and in Zimbabwe, in particular, were beginning to collect such information to help them measure credit risk parameters for SME loans. This is unlike for retail and corporate loans where official statistics are readily available making it easier to measure any respective credit risk component parameter thereby exhibiting the U-shape in banks' financing patterns. This justified the research design used for the first phase to better understand the less known SME bank asset for effective measurement of its credit risk in order to expand SME credit supply.

4.4.1 Credit Data

To protect the reputation of the case-study bank, a pseudonym was given as; ZimSME Bank. The bank harbours the ambition to increase its market share and gain competitive advantage in SME financing but sceptical about how to sift the bad SMEs from the good ones. Its loan portfolio was currently made up of 354 observations (SME clients), which were operational under the current ownership since 2010 up to 2012. The enterprises were formal SMEs and fitted the bank definition of SME which was based on sales turnover. Since the size of the sample was small and the bank was cautiously entering into the SME lending market, no data observations were dropped at data exploration stage. The characteristics of each client and respective credit quality were recorded.

Using judgemental CRM systems, of the 354 observations, 281 were classified as good and 73 were bad during the loan tenure giving a bad rate of 20.62%. To qualify to get a loan, the SME loan applicants were assessed on 26 risk indicative characteristics by means of a judgmental credit risk evaluation tool which was based on the experiences of the credit analysts and discretion of the credit manager and loan officers of the bank. An enterprise was deemed delinquent if it disregarded its loan material obligation up to 90 days, a definition derived from Basel II Capital Accord. This original SME portfolio constituted what is referred to as known-good-bad (KGB) sample which was composed of SME loan applicants who were pre-screened good and given loans of different sizes. The rejected loan applicants were discarded at selection stage.

The loan performance window was 24 months (1 January 2010 to 31 December 2012). Owing to the sensitivity character of any bank information, the eventual portfolio had no names and account numbers of clients attached to it. Clients were coded by counting numbers thereby destroying any bank client identifiable link and such information had no statistical bearing on the envisaged model. The data collected included the owner of business's profile, SME business activities and financial data which were sourced from the loan application forms and bank records provided from the business's historical repayment behaviour of clients. There were positive indications that in the near future, larger datasets (portfolios) would be possible (evidence from bank survey). This placed relevance to development of CRM tools that would help banks correctly make accept/reject decisions on SME loan applications, therefore the envisaged CRM model may facilitate sample size enhancement in future.

4.4.1.1 CRM model development

The second part consisted of development of CRM modelling for SMEs based on the Basel II IRB principles. ZimSME Bank needed to measure credit risk of its SME clientele, not adopt generic or judgemental methods of credit risk assessment. As revealed in the literature review, several methodologies have so far been employed to measure credit risk of different bank assets which can be divided into two (2) broad categories: empirical and market-based (also known as structural or reduced form) models (Lin, 2007). The former models use historical default rates associated with each CRM score to identify the characteristics of the defaulting counterparties, while the latter models use counterparty market data to infer the likelihood of default (de Noni, Lorenzon & Orsi, 2007; Lin. 2007).

Empirical methodology was prescribed for this thesis, as market-based data for SMEs is a scarce commodity in the Zimbabwean financial service sector, the Zimbabwean Stock Exchange bourse does not accommodate the SME sector. The empirical approach uses historical repayment behaviour (credit quality) data to characterise borrowers that default. In this thesis, empirical CRM modelling was adopted in line with Zimbabwean financial services sector on SMEs where the risk level of a new credit customer can be predicted based on the similar credit customers in the past, whose subsequent results have been known (Dong & Peng, 2013, Horstedt & Linjamaa, 2015; Al Baz, 2017). Otherwise reduced methodologies used by other financial markets are not adoptable in this market and in other emerging markets at large (Lin, 2007).

The process flow-chart diagram, Figure 4.1 shows the workflow that was used to come up with the final CRM model for SMEs at ZimSME bank.

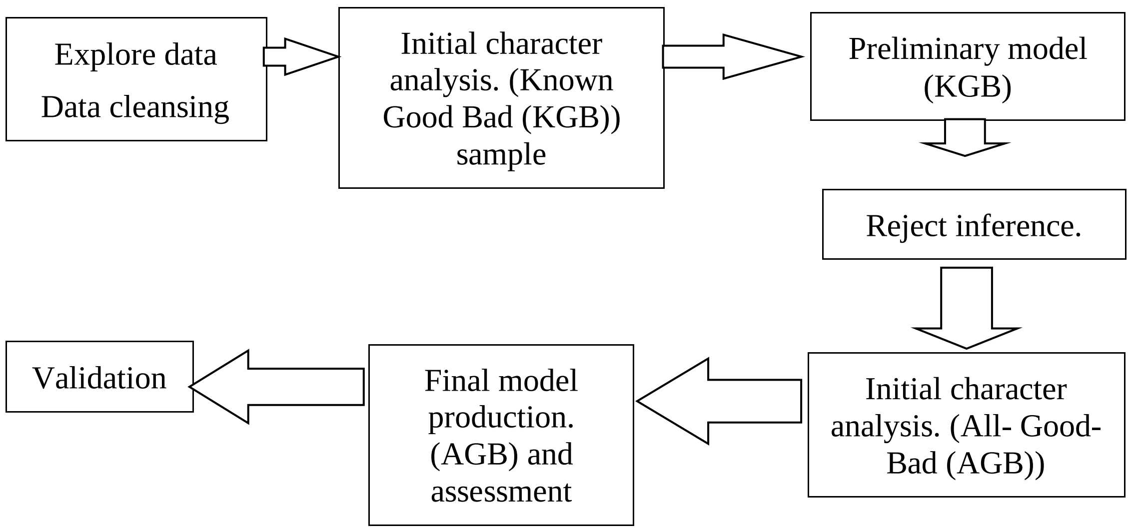


Figure 4.1: Model development steps

Source: *Siddiqi (2005)*

The reject inference step is the most crucial procedures which transforms the KGB sample to AGB sample. It is on the AGB sample where the eventual CRM model for the SME is developed. Traditionally, CRM models are built on KGB sample on the assumption that it is representative of the TTD population who apply for loan at a bank and minimal selectivity bias is assumed.

4.4.2 Data cleaning and data discretisation

Prior to building a CRM model, a common step is to perform variable screening and exploratory data analysis. This is the step where the data is known in detail and weeded out of the variables that are either ill-conditioned or simply contain no information. In that regard, data quality is paramount to the success of any CRM modelling for the process is susceptible to irrelevant or noisy and unreliable data. Therefore, for substantial knowledge discovery during CRM model development, data exploration and cleaning are crucial preliminary processes prior to modelling. For CRM modelling, data cleaning is done at record-level and at attribute-level. At record-level, some applications were excluded at the point of application due to some reasons which could compromise the selection process. The loan applications removed at this level from the portfolio are referred to as application exclusions. These include SME loan applications of staff members of the lender institution or applications which had already been approved (redundancy). The other types of applications which are excluded at record-level are decision exclusions, which consisted of applications that were discarded because their decision outcomes were unknown (rejected applicants).

At attribute-level, some characteristic variables are excluded from the model development process because they are removed on basis that they did not add any value and efficacy to the envisaged model. Such variables included the SME customer code or application number as well as the client account number. Again, predictive variables with 50% or more of missing characteristic values are removed from the list of predictive characteristic variables.

4.4.3 Variable selection for CRM model

Model candidate variables emanate from different sources used to record owner's profile, the SME business activity, financial position and macro-economic environment characteristics. Before the modelling process started, the data were cleaned, reduced and pre-processed to minimise background noisy and redundancy, since the quality of data is fundamental especially during the modelling stage. Originally the number of variables was large therefore some variable selection and reduction techniques were employed to identify variables that were predictive of the credit quality of SME loan applicants. This was where variables were statistically selected, in addition to using logical trend and business considerations to pinpoint the few predictive characteristics that best explain credit risk of past SME borrowers in a parsimonious way.

Literature has shown that a robust CRM model uses between 10 and 20 independent variables (Thomas, 2009) while Mays (2004) recommends 8 to 15 variables (Siddiqi, 2006). The reasons why the number of variables needed to be kept minimal were to do with cost reduction, identification of predictive variables that assist in providing clearer understanding the eventual model and application of the principles of Occam's razor, which state that a simple model with optimal predictive accuracy is more effective than a complex one, that is the principle of parsimony (Kennedy, 2013; Al Baz, 2017; Kanapickiene & Spicas, 2019; Lin, 2007).

Curse dimensionality was also important, which was a situation where many redundant and irrelevant variables were not enough to describe adequately the target population (Loughrey & Cunningham, 2005; Dada *et al.*, 2019; Ranganathan *et al.*, 2017; Kennedy 2013). This could eventuate into over-fitting whereby an induced model could classify accurately clients in the model development sample but performed poorly when applied to clients in an unseen (holdout) sample.

Adequate and robust characteristic variable selection were analysed through the factors that affect them. These were cost, legal, business logic and statistical analysis. Cost included both computational and financial costs involved in acquiring a certain input variable. The choice for candidature into final model was reached after a compromise between cost and explanatory use of the variable in question. The legal perspective of the variable was also considered. Any variable that posed legal, ethical and regulatory concerns was not selected for modelling.

In some circumstances modellers used business logic discretion to justify the inclusion and exclusion of variables, provided the reliability of the variable was accurately judged. This referred to as "forcing" a non-statistically significant variable into the final envisaged model. The statistical analysis objectively identified highly correlated variables which were supposed to be eliminated, as well as estimated the true contribution of each variable in assessing the credit quality of the SME client. The correlation-based variable selection was employed as well as backward logistic regression in R programming.

4.4.3.1 Data classing/binning

After some variable reduction analysis had been done, the number of variables drastically reduced to manageable levels. The remaining variables were pre-processed to put the data in a form appropriate CRM modelling. This was achieved by simplifying the structure of the data in way appropriate to the CRM modelling (Sudhakar & Reddy, 2016; Dimitras, Papadakis & Garefakakis, 2017). Binning/coarse classification is a common pre-process in the CRM domain used to code values of continuous characteristic into a number of small categories and similarly for both categorical and ordinal variables (Kennedy, 2013).

The disaggregation into small categories allows each small category of each characteristic variable to be treated as a dummy variable having its own weight in standard logistic regression model (Ranganathan *et al.*, 2017). This normally increased a CRM model's robustness by drastically reducing the likelihood of over-fitting and creating categories with adequate number of good and bad observations (Kennedy, 2013; Roelofs, 2019; Katuka & Dzingirai, 2015).

In this case it afforded non-linear relationship by creating many separate categories, each of which had its own weight in the standard logistic regression model. The other advantage of binning was its ability to treat the missing values as a separate category (Kennedy, 2013; Nguyen, 2016). One of the most commonly used binning statistics is weight of evidence (WoE) which is calculated as follows:

Equation 90: Weight of Evidence (WoE) formular.

$$WoE = ln\left(\frac{Distribution\ of\ goods}{Distribution\ of\ bads}\right) = \ln\left(\frac{P(value = good)}{P(value = bad)}\right)$$

For the WoE the range bin is converted into standardized good/bad odd ratio. The statistic shows that observations falling into the range bin have somewhat neutral credit risk if $WoE \approx 0$, indicates better credit risk if $WoE > 0$ and worse risk if $WoE < 0$. This interpretation helped further characteristic variable reduction.

The other famous binning statistic is information value, calculated as follows:

Equation 91: Information Value (IV) formular.

$$IV = \sum (Dist(Good) - Dist(Bad)) * WoE$$

For instance, the IV which is sometimes referred to as Kullback divergence measure was used to measure the difference between two (2) distributions (Anderson, 2007). If a characteristic with IV > 0.3, implies that the characteristic is highly predictive (Mays, 2004). Siddiqi (2006) proposes, using the rule of thumb, that if $IV > 0.5$ implies that the corresponding characteristic variable is too predictive and needs further investigation for inclusion into the eventual model. If $IV < 0.1$ implies that the corresponding characteristic variable has weak predictive contribution in the model therefore is a candidate for exclusion from the model development (Novianti, Jong, Roes & Eijkemans, 2015; Heinze, Wallisch & Dunkler, 2017).

For this thesis, predictive power of decision variables was determined by means coarse classing, whereby all the 26 characteristic variables currently used for SME applicants' selection at ZimSME bank were ranked by information value (IV) and weight of evidence (WoE). The WoE was used as a measure of the strength of each attribute or grouped attributes in separating *good* and *bad* accounts; in fact, it was a measure of the difference of the distributions of *goods* and of *bads* in each interval attribute.

On the other hand, the Information Value (IV) was also a useful concept for variable reduction during model development process. We calculated both the IV and WoE for each of the 26 original characteristic variables. The WoEs of each interval attribute of a characteristic variable were added up to give IV for the characteristic variable concerned. In fact, IV of a characteristic variable is related to the sum of the values for WoE over all groups or interval attributes. Therefore, the IV expressed the amount of diagnostic information of a characteristics variable for separating the *goods* from the *bads*.

Rules of thumb suggested by Siddiqi (2006) for variable selection made it convenient and handy in variable reduction and variable pre-processing, crucial transformations prior to CRM model development process.

Table 4.3: Information value scale

Information Value	**Predictive Power**
< 0.02	Useless for prediction
0.02 to 0.1	Weak predictor
0.1 to 0.3	Medium predictor
0.3 to 0.5	Strong predictor
> 0.5	Suspicious or too good to be true

Source: *Siddiqi (2006)*

From literature, variables with medium and strong predictive powers are selected for model development but other researchers advocate for just variables with medium IVs for a broad-based model development. When a characteristic variable is strong, it means it may be over-predicting, meaning that it is in some way trivially related to the good/bad information, which calls for a close check for that variable before including it in the model building process. For the current thesis, characteristic variables with at least weak IVs were selected into the initial model for further variable selection through backward logistic regression. In addition, some variables could be "forced" into envisaged on the grounds of business and logical trend considerations.

4.4.4 Development and holdout sample designs

Data quantity is of great importance in CRM modelling as the quality and robustness of the envisaged model was largely dependent on how large the original loan portfolio was. In normal situations, the amount of data to be used for any CRM modelling depends on the objectives pursued and the properties of data, whether the available data in tandem with the intended modelling objectives. Owing to the scarcity of credit data due to perceived confidentiality, data quantity was mostly dependent on availability.

Literature suggests at least 1 500 entries as the traditional data quantity suitable for accepted loan applications (Abdulsaleh; 2016: Dynan & Sheiner, 2018; Mills & McCarthy, 2016) and the other 1,500 additional entries to be created through reject inference. The current study used data availability criterion to determine eventual data quantity for the model owing to fact that ZimSME bank had just entered into the emergent SME lending market and aimed to increase its market share, so did not have large quantity of data for robust CRM modelling.

When the data cleaning and variable selection were satisfactorily done, the next step was to generate development and holdout samples from the working sample provided the sample size was 1500 or so (Kennedy 2013; Vabalas *et al.*, 2019). In preparation for the eventual modelling, the original sample was traditionally supposed to be split into two (2) subsamples: training/development and testing/holdout samples. The training sample was supposed to be used to develop the CRM model while the holdout sample for model validation (Leung, 2008). The sampling design used was stratified to ensure that the samples were random and to give a reflection of the population in some distinct characteristics.

Using stratified random sampling, the dataset was supposed to be split into an 80% development sample and a 20% holdout sample (Leung, 2008) while Siddiqi (2006) suggests a 70:30 split if only if the dataset was sufficiently large. From the loan process, it must be noted that these generated subsamples are drawn from only accepted applicants whose loan performance is known, excluding the denied/rejected loan applicants. If data is sufficiently small no splitting is exercised but straight use of statistical methods like cross-validation (Vabalas *et al.*, 2019) or bootstrapping

(Japkowicz & Shah, 2011; Xu & Goodacre, 2018) to estimate the model parameters are applied without loss of information (Thomas, 2009).

This was the case with ZimSME whose portfolio has 354 observations, far much below the traditional development sample sizes suggested in literature (Lewis, 1992; Siddiqi, 2006). The development sample was 100% of the original ZimSME bank SME loan portfolio. The holdout sample was randomly extracted 70% of the original to validate the subsequent model.

4.4.5 Sampling period

The other issue of great importance prior to modelling was the definition of the sampling period due to fact that, in CRM modelling domain, historic loan performance of a portfolio is believed to be a reliable pointer of the future loan repayment performance. With the aim to create a reliable dataset for CRM modelling, Martens, van Gestel, De Bdcker, Haesen, Vanthienen and Baesens (2010), suggests two (2) instances of every loan client taken at two (2) distinct points along the loan process trajectory. The first instance was taken at time when the SME client initiated the loan application process when the characteristics were recorded and the second instance at a lapse of time until when the client was labelled either good or bad. The duration between the two (2) instances was referred to as the outcome window. The length of the outcome window was dependent on the business objectives of the envisaged CRM model. Its determination was done with great caution, lest either chances of misclassification of a client or underestimation of the default would ensue.

A lengthy or too wide outcome window would give rise to the possibility of mismatch between the development sample and the TTD population, a yet unseen future population. This mismatch may be a result of population drift, dynamism in the macro-economic conditions, enterprise strategy shift and client personal circumstances (Hoadley, 2001; Kennedy, 2013; Kritzinger & van Vuuren, 2017; Sousa, Gama & Brandao, 2016). On the other hand, a too narrow or too short outcome window would give rise to loss of valuable data necessary for the eventual CRM modelling.

A compromise was exercised to decide for the optimal outcome window. It was a prerogative of the modellers/researchers to decide on the size of the outcome window prior to developing the envisaged model, provided the expected classification power of the model was not compromised.

Given the circumstances surrounding ZimSME bank, a 24-month outcome window was used to generate the initial SME loan portfolio data. The initial bank SME loan portfolio was made up of 354 observations (clients) of whom 281 were classified as good and 73 as bad after a two-year loan tenor giving a bad rate of 20.62%.

To qualify to get a loan, the SME applicants were assessed on 26 risk indicative characteristics which were grouped and classified as follows: owners characteristics (age, qualifications, experience, length of relationship and account with other banks), enterprise characteristics (income of directors, number of directors, sector, new firm, loan purpose, tenor of loan, technology level, asset size, number of employees, export and local trade) and financial characteristics (gearing, liquidity, stock turnover, annual turnover, net profit, loan amount, collateral, Interest rate, creditors' and debtors' days). To decide for a bad account, dependent on the current model and bank's view of success or failure (McNab & Wynn, 2000), the Basel II definition of default was adopted in this thesis. According to Basel II Capital Accord (BCBS, 2005), a debtor is declared bad if he/she is past due more than 90 days on any material credit obligation to the creditor.

4.4.6 Preliminary logistic CRM model

Owing to the binary character of the response variable (default/no-default); logistic regression technique was adopted for the eventual development of a CRM model for ZimSME bank. In fact, the preliminary model was developed by the application of the logistic regression on the statistically cleaned 100% development sample of the known-good-bad (KGB) sample. The logistic regression modelling suited well the situation because the predicted credit quality variable which is generally categorical and dichotomous. This generalised linear model was used in its multivariate form since the binary outcome; good/bad, depended on a number of predictor variables as deduced from initial characteristic analysis of the known-good-bad (KGB) sample. The logistic regression, like other generalised linear models, uses a set of predictor variables to forecast the likelihood of a specific target response variable. The logit transformation, the logarithm of the odds was used to linearize posterior probability and to limit the outcome of estimated probabilities in the generalised linear model to between 0 and 1. The transformation was defined as follows:

Equation 92: Logit transformation link

$$logit\ link = log\left(\frac{p(bad\ loan)}{p(good\ loan)}\right) = log\left(\frac{p(Y=1)}{p(Y=0)}\right) = e^{\beta_0 + \sum_{i=1}^{n}\beta_i x_i + \ random\ error}$$

The parameters were estimated by using the maximum likelihood estimation (MLE) method and appropriate interpretation of these estimated parameters. To determine attribute scores, the regression would be done against WOE created in the initial characteristic analysis. This would be done as an alternative bypass to grouped variable credit risk modelling. The normal approach was to regress the credit quality against a set of predictor variables, which constitute numeric and created dummy variables for categorical data. In the process of selecting a good fit model various regression model building techniques were employed these include forward selection, backward elimination and stepwise in R. This was additional to the initial characteristic analysis, where p-values were used to assess the suitability of each and every suspected predictor variable.

4.4.7 Designing an initial CRM model

The designing of a statistical model required the use of statistical tools to decide which independent variables best explained the variability observable in the outcome variable. These included statistical measures such as the Chi-square, R-square, p-values and other relevant statistical model adequacy tools. In addition to the decision of the eventual initial model, business goals like risk appetite, market share strategy among other things were considered in the model construction process. Given the importance of variable selection and reduction, a risk profile was first developed through initial character analysis. This was built using a variety of suspecting predictive variables which included demographics, financial data, repayment patterns, and time-related data except for credit bureau inquiries.

The credit bureau facility was non-existent in Zimbabwean financial service sector, a limitation to development of a robust model. The envisaged model was expected to be coherent with the decision support system of ZimSME bank. Of course, this intended model was supposed to be a sole arbiter, an epitome of an experienced credit analyst, thereby making the construction of a comprehensive credit risk profile was a must. The risk profile for the ZimSME bank included characteristics of the owner, the business as well as financial data, except for external characterization due to the unavailability of the credit bureau information.

4.4.7.1 Credit risk profile

The original characteristic variables were 26 and had to be strategically and statistically reduced to build a risk profile for the ZimSME bank which would be used to develop the preliminary CRM model on the initial know-good-bad (KGB) ZimSME sample.

Table 4.4: ZimSME Bank original characteristic variable profile

Owner characteristics	Enterprise characteristics	Financial characteristics
Age	Sector	Liquidity ratio
Qualifications	Local trade	Gearing ratio
Experience	Export trade	Stock Turnover
Income	New-firm	Debtors days
Number of Directors	Length of relationship	Creditors days
Income	Annual Turnover	Net Profit margin
	Number of employees	Other loans
	Collateral	Tenure of loan
	Technology	Interest rate
	Asset size	Loan amount
		Purpose of loan

Based on the eventual risk profile, a single regression approach would be adopted. When it was run, characteristics were placed in the regression equation in order based on information type as well as strength. Information type was ranked weaker to stronger accordingly and again attributes within each information type, characteristic variables were also ordered from weakest to strongest by means of IV. This was the sequence in which the single regression considered each characteristic. As an alternative, the predictor characteristic variables were IV ranked from highest to lowest regardless of information type.

The eventual accepted single regression model accounted for known performance of screened applicants. In fact, the preliminary model was built from a KGB sample was meant for measuring credit risk for those who have been screened to be good borrowers, the "cherry-picked" ones. This was in contrast to whole purpose of developing a decision to help bankers to classify good/bad applicants for loans from the TTD population. The adaptation of the developed preliminary model would bring in a lot of inaccuracies and misclassification due to selectivity bias induced by the non-randomness and truncation of the known-good-bad sample on which it was constructed. In fact, the known-good-bad sample represented the population of accepted applicants only and did not account for the rejected applicants. Therefore, if the preliminary model were applied to the TTD SME loan applicant population, a great deal of misclassification would ensue.

There was need to find a method to account for the missing credit quality of the rejected since all other characteristic variables were available. The method was found in the form of the reject inference techniques, to impute the credit quality of the rejected applicants such that an AGB sample was developed prior to the development of the envisaged CRM model able to classify into good or bad of the TTD SME loan applicants. There are various techniques for reject inference, but majority fall short substantive results as demonstrated by other researchers (Lin 2007; Nguyen, 2016; Kennedy; 2013; Al Baz, 2017). Therefore, for this thesis, a model-based reject inference methodology grounded on the theory of missing data and Bayesian inference analysis, built on theoretically supported assumptions was adopted.

The method incorporated the impact of the incomplete sample by imputing missing data of the response variable based on the estimated probabilities of missingness. It was a flexible approach which as well was able to incorporate supplementary information about the rejected into modelling process. In fact, it is grounded on a firm theoretical support, unlike other reject inference techniques which are grounded on tenuous assumptions, thereby making it have an edge over other reject inference techniques (Chen & Astebro, 2012; Nguyen, 2016; Ditrich, 2015; Kennedy, 2013). The method was fully described in Chapter 3, is called Bound and Collapse (BC) technique which uses Bayesian procedure to construct a model basing on the theory of missing data developed by Rubin (1976), a model-based imputation methodology.

4.4.8 Reject inference

Owing to the fact that credit quality of rejected loan applicants was not observable in ZimSME bank loan register, which contained complete data of only the presumed ‘good’ borrowers. This engendered a non-random sample due to selectivity bias which is quite rife in CRM modelling (Chen & Astebro, 2012; Smith & Elkan, 2004; Nguyen, 2016; Lin, 2007; Kraus, 2014). This background augured well for an effective implementation of the Bound and collapse (BC) methodology, to impute the credit quality of the rejected applicants thereby, eventually, developing an AGB development sample representative of the through-the-door population to be scored in future.

The preliminary model was applied on the known-good-bad (KGB) sample to create two (2) regions: the “accepted” and the “rejected”. To estimate the credit quality of the rejected clients, we emulated some banking acceptance policy making process. We defined $y = 1$ be the credit quality of an SME that has been in arrears on at least one material obligation within the two-year outcome window. This meant enterprises with high credit measurement scores were likely to default, unlike those with low credit scores. If a bank were risk-averse, it would set a low cut-off point whilst if it were a risk-taker, the cut-off score would be high, implying divergent rejection policies based on risk appetite. The bank risk appetite translates into “strong” and “weak” rejection policies defining their respective market share strategy.

These two (2) rejection policies were simulated by setting logical thresholds to eliminate selectivity bias. We first instituted a “weak” rejection policy by picking up a threshold of a high CRM cut-off score. Subsequently “accepted” and the “rejected” region were created. The same was done for the “strong” rejection policy, where a low threshold CRM score was picked. All this was possible after rank-ordering the KGB sample by CRM score. This gave birth to two (2) samples with different sizes of the respective “accepted” and “rejected” regions. To create an AGB random development sample, it was needed to impute the credit quality of rejected in the respective rejected regions of the two (2) samples created through weak and strong selection procedures. We use equation (83) to estimate the respective missing credit quality. The simulated AGB samples were labelled “weak selection” and “strong selection” respectively.

After successful variable selection and pre-processing, a preliminary CRM model was constructed from the non-random known-good-bad (KGB) sample. Reject inference was instrumental in providing complete random sample information indicative of the TTD applicant pool, on which the subsequent CRM model was estimated. To infer the credit risk for the rejected loan applicants, we apply the preliminary model the original KGB sample. From loan process, it is logical to suggest the probability of having a bad credit quality as a good proxy for the missing data mechanism (MDM) (Chen & Astebro, 2012; Kraus, 2014; Ditrich, 2015; Nguyen, 2016). This implied that the original credit risk score provides the much-needed information that paved way for the estimation of the missingness mechanism. For estimating the missing data mechanism, we used linear extrapolation of bad rates versus original score. The estimated probability for the rejected loan application's being bad is considered as the probability of missingness. The simple linear regression model (SLRM) is of the form:

Equation 93: SLRM for the probablity of missingness

$$Bad\ Rate = \beta_0 + \mathrm{\beta_1 Credit - score + C}$$

This SLRM model helps lenders observe bad rate - credit score relationship over time thereby influencing the change of credit acceptance policy by either tightening or loosening the cut-off or threshold levels. Credit policy change in turn has bearing to the probability of missingness, a framework crucial for the eventual application of the proposed Bayesian reject-inference method: Bound and Collapse.

Since the sample was small to guarantee sample split into development and testing samples, we used 100% of the original sample as development sample (refer 4.3.2.6). To simulate credit granting policies of ZimSME bank, the preliminary CRM model developed on the KGB sample was applied to the development sample. To generate missing data, we simulated two (2) rejection policies: weak selection when the cut-off point was high and strong selection when cut-off point was low. By strong selection we meant that the bank was risk averse by being strict in offering loans to its credit applicants and by setting a lower threshold meant rejecting many applicants. On the other hand, by weak selection, the bank loosened its lending policy by setting a higher threshold, that is, the bank is a risk-taker. Using the weak selection model, more applicants are selected for loaning. After defining the missingness function, we therefore applied a Bayesian theoretically based BC reject inference technique to impute the credit quality of the rejected applicants

4.4.8.1 Estimation of the missingness probabilities

The issue of how missing data mechanisms should be estimated needed thorough discussion for that had some bearing on the application of the BC methodology. In the presence of the MNAR missingness, there are two (2) approaches to determine the missing data pattern, which are selection models and pattern mixture models. For this work, pattern mixture modelling was adopted because the BC reject inference model was seen closely related to it (Little, 1993). Probabilistically, the pattern mixture model is defined as:

$$P(Y_{miss}, Y_{obs}) = P(Y_{obs}|Y_{miss})P(Y_{miss}).$$

The implication of this model is that missing data are classified by their respective missingness and describe the observed data within each missing group. Therefore, without the knowledge of the missingness of the missing data the pattern mixture model would be undefined. This pointed to the fact that for a pattern mixture model some identifying constraints were required from information on missing data.

Using the pattern mixture model to BC reject inference, observation with missing values were grouped in different missing patterns through an underlying CRM score, S. From loan process, it is logical to suggest the probability of having a bad credit quality as a good proxy for the missing data mechanism (MDM) (Chen & Astebro, 2012). This implied that the original credit score provides the much-needed information that paved way for the estimation of the missingness mechanism. For estimating the missing data mechanism, we used linear extrapolation of bad rates versus original score. The estimated probability for the rejected loan application's being bad is considered as the probability of missingness.

This general linear model helped lenders to observe bad rate - credit score relationship over time thereby influencing the change of credit acceptance policy by either tightening or loosening the cut-off or threshold levels. Credit policy change in turn has bearing to the probability of missingness, a framework crucial for the eventual application of the proposed Bayesian reject-inference method: BC. Once estimated missingness was computed, the simulated value of the missing datum was calculated using equation (83). Iteratively, we imputed all the values of the missing credit quality of rejected applicants thereby generating a complete AGB sample, a good representation of the TTD population to be scored in future.

4.4.8.2 Selection of priors

Since the BC model uses a multivariate generalisation of the Beta distribution, the Dirichlet distribution, as the conjugate prior distribution, there was great need to select appropriate priors for its application. Several non-informative Dirichlet priors offered strong candidature for the selection. For simplicity, it was settled to choose between $\alpha_{ij} = 0$ and $\alpha_{ij} = 1$ priors. If we desired to achieve a uniform density such that the Dirichlet density function assigns equal weight to any vector θ compliant to the constraint that $\sum_{ij} \theta_{ij} = 1$, it would be proper to set for $\alpha_{ij} = 1$ for all i and j. On the other hand, if it were to set for $\alpha_{ij} = 0$ for all j that would get an improper prior distribution, that would be uniform in the log (θ_{ij})'s but the resultant posterior is proper if there is at least one observation in each score range. Fortunately, if the concerned sample were relatively large, then the difference in results between these two (2) prior densities would not be that big. On that basis we, for this work, selected the non-informative Dirichlet prior $\alpha_{ij} = 0$ and based on simplicity in the application of the Bound and Collapse reject inference technology.

4.4.8.3 Verification

As soon as reject inference (Refer to 4.3.2.11) was successfully done, some simple verifications procedures were done. Firstly, some comparison of bad rates or odds of the post-inferred, "all-good-bad" and the "known-good-bad" samples were carried out to find out whether lending industry rules were not vilified. Reject inference techniques could be used to the satisfaction of industry norms which are usually based on the approval rate and the level of confidence of the preliminary model used for credit granting decisions. In this instance, if the preliminary model were good and the approval rate subsequently could be high, resulting consequentially in that the inferred rejects should have a bad rate at least three (3) times that of the approved.

There was also the need to carry out some comparisons of the bad rates of the grouped attributes for the KGB and the AGB samples. Some grouped attributes which are characterized by low acceptance rates and high WoE should display through the distributions of their WoE consistently with business considerations or in a way explainable by business experiences. It is also through this verification that the reject inference employed, and the corresponding estimated parameters are tested by means of what are called "fake rejects". To do this test, the accepted/approved subsamples was split into arbitrary accepts and rejects in the ratio 70% to 30% and the final model developed from the AGB sample. The classification of the 30% split was already known, therefore any observable misclassification due to the application of the final model was used to gauge the performance of the model developed.

Once thorough verification has been done, the combined sample of the approved and the inferred rejects was created to form an AGB sample, a random sample on which the final model would be built. This was the sample on which selectivity bias has been resolved, restoring the expected randomness character of a sample, a basis for any statistical inference of reality phenomenon. The resultant sample was assumed a better representation of the through-the-door population of the SME loan applicants. Using the same procedure as for the preliminary modelling, some characterization of the AGB sample was done prior to final model development.

4.4.8.4 Initial characterisation analysis on all-known-good-bad (AGB) sample

After the missing credit quality for the rejected SME loan applicants were imputed, initial characteristic analysis and statistical modelling procedures were carried out to generate final set of characteristics for the final CRM model. Post-inferred dataset constituted the AGB sample representative of the through-the-door SME loan application population on which the desired and final model was constructed. There was no limit to characteristics selected in the preliminary characteristics' analysis, as some characteristics became weaker and some stronger after imputation of missing credit quality. This implied that variable selection was repeated exploring the post-inferred development dataset. Unlike the preliminary model development, which was constructed on the KGB development sample, the final model was derived after performing initial characteristic analysis and running the logistic regression onto the AGB sample. The resultant logistic regression model, parameter estimates, and model performance statistics were the major outcomes from the post-inferred sample. The scaling of the scores, validation of point allocation, misclassification and strength of the model were addressed as soon as final model was derived.

4.4.9 Validation

When the final model was built, the next procedure was validation. The validation process was carried out to confirm whether the developed model was serving the purpose it was built for. Was the final model applicable to the through-the-door population TTD of SME loan applicants? The other purpose of validation was to check whether the resultant "final" model was not over-fitted.

As in the model development stage, the 100% of the working sample constituted the development sample whilst the arbitrary 80% constituted the holdout sample. The holdout sample was used for validation process (Siddiqui, 2005). This was a process whereby the distributions of measured goods and bads across the development and holdout samples were compared. To carry out the comparisons, some goodness of fit statistical measures were employed. For this work, the Receiver Operating Curve (ROC) was used. It is a plot of the true positive rate against false positive rate at different cut-off points. The area under the ROC is used to measure the CRM model's classification power.

The misclassification statistics were also used to assess the predictive prowess of the final model. For operational use of this statistic, a minimum level of acceptable bad rate was chosen as a "cut-off." Loan applicants whose CRM scores were below the set "cut-off" point were declined loan and tagged as potential defaulters. This process could lead to the commission of type I and type II errors, where an actual good is wrongly classified as bad and consequentially declined loan services and vice versa. To ensure suitability of a final model, it was measured in such a way that both errors were minimized, that is, such that the level of misclassification was at minimal. There were several measures for misclassification which were based on the confusion matrix. These include accuracy, error rate, sensitivity and specificity.

Table 4.5: Confusion matrix

Actual	**Predicted**	
	Good	**Bad**
Good	True positive (good decision)	False Negative (**Type I error**)
Bad	False Positive (**type II error**)	True Negative (good decision)

A good final credit risk measurement model would be one where the "true" cases are maximised, and "false" cases are minimised. The measures are defined as follows:

Equation 94: Useful Measures from a Confusion matrix

$$Accuracy = \frac{True\ positives + True\ negatives}{Total\ Cases}$$

$$Error\ rate = \frac{False\ positives + False\ negatives}{Total\ Cases}$$

$$Sensitivity = \frac{True\ positives}{Total\ Actual\ positives}$$

$$Specificity = \frac{True\ negatives}{Total\ Actual\ negatives}$$

Based on these measures, a bank may decide to strategically maximize the rejection of bads, with the aim to reduce losses (risk averse) therefore would choose a CRM model that maximizes specificity. If the bank, like the ZimSME bank, is a risk-taker in order to get a higher market share by even approving some bads for credit services. Such a bank would minimize the rejection of goods by developing a final model that maximizes sensitivity. The specificity and sensitivity statistic measures are being used as weaponry to achieve business goals of a bank through the development of a corresponding model.

4.4.9.1 F1-Score

The F1-score methodology was used also as a standard measure of comparison between different models and methodologies when the outcome is a binary value. This statistic measured classification power of a CRM model and it included precision and recall statistics of the test (Sasaki, 2007). The precision is the number of true positive observations divided by total classified as positive and it refers to the percentage of the true observations that were correctly predicted.

Equation 95: Precision formular

$$Precision = \frac{TP}{TP + FP}$$

On the other hand, the recall is the number of the true positive observations divided the total good observations and refers to the percentage of real good observations that were correctly predicted.

Equation 96: Recall formular

$$Recall = \frac{TP}{TP + FN}$$

After calculating the values for the precision and recall statistics, F1-score statistic can be calculated as in equation (94). It takes values between zero and one, one being indicative of the best score and zero the worst.

Equation 97: F1-score formular

$$F1 - Score = \frac{2 * Recall * Precision}{Recall + Precision}$$

4.5 ANALYSIS

For CRM, traditionally different parametric models were used for classifying input into one of two (2) groups, which is the main objective of statistical inference on credit-scoring problem. In case of SMEs most of the input variables are categorical (Bensic, Sarlija &, Zekic-Susac, 2005; Wang, 2013; Goh & Lee, 2019). In this thesis, Logistic regression methodology was chosen for data analysis because of nature of data collected as well as its suitability to deal with binary response variables. To improve on selection bias, some model-based reject inference technique in the form of Bound and Collapse methodology was adopted for this work.

4.6 LIMITATIONS

The first phase of this research is somehow like a social research and as such encountered some challenges of information asymmetry between the supply side (banks) and demand side (SMEs). In particular banks in Zimbabwe are so secretive that they hardly divulge information they deem confidential and security and argue that it is improper to inquire into their clientele payment behaviour. Regardless of the fact the research would help banks design appropriate business models to deal with the non-traditional SME lending market, banks still are reluctant to volunteer information. During data collection, both targeted participants were given assurance of how the data would be processed to preserve anonymity and non-identifiability of data source for confidentiality and security. Despite of that assurance, the two (2) forms of non-responses arose during data collection: sample non-responses where targeted participants chose not to answer the whole questionnaire on the reason that they did not offer information for any research and item non-response where other participants chose not to answer other questions of the questionnaire they deemed sensitive and unethical. That was great challenge.

The data analysis on the second phase was never exhaustive because it was limited lack of input variables from the referencing credit bureau, for the institution is non-existent in Zimbabwe. At times financial input variables for the SME performance were not available or not reliably authentic. Due to the confidential nature of credit default data, the researcher could hardly validate some of the input data. Even from the questionnaire some respondents answered some of the questions superficially to some questions which they thought sensitive and more confidential.

4.7 ETHICAL CONSIDERATIONS

In every research work, ethical considerations must be exercised for value addition of the final end product. Research ethics means the way a researcher treats both participants and information they provide, honestly and with dignity and respect. The same applied to the researcher of this work; bank information is highly confidential therefore everything was done to ensure anonymity and confidentiality. To build trust with participants in the project, the researcher signed Code of ethics with the data informants. The code of ethics gave the participants a candid explanation of the intentions and goals of the researcher and assurance that anonymity and confidentiality of their participation would be upheld in the final report. The participants were encouraged to feel ownership of the final output and assured that they may benefit in one way or the other. In accordance with the New Capital Accord, Basel II, Internal-based ratings approach was based on a bank's clientele, so banks which provided data for model building would direct benefit if they are satisfied with the final CRM model. The findings of the research would find channel into the public domain and shared worldwide for knowledge sharing.

4.8 SUMMARY

The aim of the study has been to investigate the methodologies that might be employed in further research into CRM modelling of SMEs in an emerging economy. Two (2) phases have been considered in this work: bank involvement with SME financing in Zimbabwe and CRM model development. The first phase is exploratory research where survey research design was used to explain the current interrelations between SMEs and banks. The lending equation of SMEs has two (2) vital components: the supply side (banks) and demand side (SMEs). To ascertain bank involvement with SME financing, two (2) surveys were designed for both sides for corroboration to get anecdotal evidence of the real relations between the two (2) economic players. Non-probability sample designs were used to extract sample surveys from respective target populations. The major instrument of inquiry for respective surveys was the questionnaire. For knowledge discovery in this phase of the research, descriptive statistics was the major tool for data analysis. The discovery of any intrinsic relationships between any components from either side will be instrumental in the second phase of study, which requires a lot of real data for modelling and simulations of the CRM in the second phase.

The second phase is about model building where data is of great importance. In accordance to the Basel II IRB methodology of CRM modelling, portfolio approach to estimating SME probabilities of default is appropriately employed. Internal ratings-based approach recommends banks to use internal data for eventual calculation of the capital requirement in tandem with the estimated credit risk exposure. The sample size is given by the size of the bank SME loan portfolio. This is the initial sample composed of the pre-screened good loan applicants and constitute what is called a know-good-bad (KGB) sample, which is non-random, and truncated, a case prevalent in CRM domain. To solve for this statistical deficiency, reject inference technology was applied to estimate the missing credit quality of the rejected applicants such that an AGB sample is constructed prior modelling of CRM.

Overall, the work in this Chapter has achieved the aim of minimising the effect of selectivity bias by applying theoretically based reject inference technologies not those based on tenuous assumptions and disregard the effect of missing data on the eventual inference. In subsequent chapter modelling will be based on an inclusive larger sample containing accepted and rejected SME loan applicants. This resulting sample is referred to as AGB sample which is random and representative of the through-the-door (TTD) target population of potential SME loan applicants. In fact, the chapter has also highlighted some of the crucial issues faced by researchers in the area of CRM modelling for SMEs. For subsequent the CRM modelling a larger range of predictive variables will be considered initially. Using data classing and other statistical transformational tools, the number of independent variables was drastically reduced to between 8 to 15 would allow for better and precise analysis of the future data.

For model building, maximum likelihood estimation was applied through backward logistic regression modelling. In emulating real bank situations, two (2) models will be built; weak selection and strong selection to vary the degree selectivity bias. Both models will be statistically diagnosed for goodness of fit and estimatability. The final model between the resulting two (2) will be selected using AIC, specificity ratio as well as F1-score. The final model will be back tested for robustness and applicability to serve the purpose it would have built for.

CHAPTER 5: DATA ANALYSIS: SURVEYS

5.1 INTRODUCTION

This chapter presents the results of the analysis of data collected from surveys of both the demand and supply sides of SME loan financing process. The data were collected in accordance to the problem statement pronounced in chapter 1 of this thesis (refer section 1.3). Data collection process was driven by two (2) fundamental goals, which also directed the subsequent data analysis, answering to research questions that were posed (refer section 1.4). The outstanding goals of this work were to assess the degree at which banks in Zimbabwe were involved with SME financing, as the sector is generally viewed as the vehicle to the indigenisation of the economy, and to develop a CRM model that would help banks reduce loan processing and diminish aggregate default costs (refer section 1.5). Through data analysis, the goals were accomplished as procedurally demonstrated herein. The findings presented in this chapter put to the fore the potentiality of SMEs as a vehicle to the economic prosperity of Zimbabwe on one hand and on the other hand banks find the next gold mine in SME lending provided appropriate lending technologies are developed and adopted.

5.2 BACKGROUND ISSUES

The ultimate goal of this research has been to contribute to the effort of generating solutions to the problem of SMEs financing problems especially in the Zimbabwean financial service sector as a case study of emerging market economies. Conventional wisdom generally apportions the blame of the failure to establish a smooth bank-SME relationship on banks, which are thought to be reluctant to do business with this non-traditional sector. To garner comprehensive anecdotal evidence, the research has been designed into two (2) interconnected phases: bank involvement with SME financing in Zimbabwe and SME CRM modelling. In pursuit of the aforementioned goals, the specific objectives (refer sub-section 1. 4. 2) guided the study.

In search of answers to the research questions pertaining to the first objective, measurement instruments in the form of questionnaires were framed and designed. The respondents were bank credit managers, credit managers, credit analysts and loan officers on one hand and SME owners and SME owners and administrators on the other hand. They were given opportunity to air their opinions and views over the economic interaction between the two (2) sides. Exploring opinions and views from both sides (supply and demand) of the SME financing equation would give robust and far-reaching insights into the issues under scrutiny. As explained in the methodology Chapter 4, the questionnaire for the bank survey consists of eight (8) thematic aspects concerning SME financing and for SME survey, four (4) themes were instrumental in soliciting opinions and views of these key respondents from both sides. Both questionnaires consist of both closed and open-ended questions to elicit robust responses. This would give a firmer foundation for the second phase of the research which developed an empirical CRM model based on a profound understanding, by banks, of the SME lending market, a market in which most banks in Zimbabwe are not familiar with.

5.2.1 Sample designs

Non-probability sampling design was the only way to extract respective samples for both bank and SME surveys, since no standing sampling frames for both target populations could be identified and that the participation in both surveys was voluntary. For the bank survey, a convenience sample of 8 banks was carved out of the target population of 14 commercial banks, according to register of banks at the Reserve Bank of Zimbabwe (RBZ). From the surveyed banks, bank managers and credit managers were targeted at for eliciting pertinent issues to do with SME financing from the perspective of the bank policy on SME funding. Therefore, for the bank survey, the eventual sample size was 8 banks.

For the demand side survey, SME survey, there is no existing sampling frame for the SMEs in Zimbabwe due to heterogeneity and non-universality of what constitutes an SME. In that scenario, a convenience sampling design was adopted for the SME survey. To synchronise the activities of the SME sector in Zimbabwe, the ministry responsible for the sector, has come out with some working definitions of what constitutes an SME. These definitions are based on annual turnover, number of employees and asset sizes. Regardless of these definitions, to the contrary banks use annual turnover as the basis of an SME concept. Given the economic significance of the sector, SMEs in Zimbabwe have formed an association of voluntary membership and therefore the list of member SMEs cannot be considered the sampling frame of all the SMEs in Zimbabwe. The association formed is called Small and Medium Enterprises Association of Zimbabwe (SMEAZ). This association provided a suitable platform to design a purposive sample for the SME survey. The association has a website where memberships interact and share information with each other.

Using the association's website, the demand side questionnaire was flight requesting membership interested to participate voluntarily and were assured that anonymity and confidentiality were upheld during and after the survey. The targeted respondents were SME owners. The internet-based survey managed to garner a sample size of 300 SME owners who electronically responded to the survey call and filled the questionnaire as per required. Therefore, for the SME survey, the sample size was 300 SME owners who were not distinguished according to age, gender, marital status or any other demographic characteristic. Notable was that the majority of the respondents were generally adults with reasonable academic qualification status as well as substantial business experience.

5.3 RESPONSE RATE OF THE BANK AND SME SURVEYS

In this study, two (2) major surveys were carried out to investigate the reality of both the supply side (banks) and the demand side (SMEs) of the SME financing equation. Logically the investigation started by querying the supply side, which is sometimes apportioned the blame of deliberately sabotaging the SME financing in emerging markets. To authenticate the investigation, the response rates of the bank and SME surveys were looked at simultaneously.

For the Bank Survey, the target population included both banks that were ex-ante thought to be lending to SMEs and banks that were ex-ante thought to have no or little relation with SMEs. The bank selection into the sample was done in such a way that approximated the true representation of the Zimbabwean financial market spectrum in the SME lending market. In that regard, 14 commercial banks operating in Zimbabwe, at the time of the survey, were targeted to cover a large part of the bank population. The instrument of data collection was a questionnaire which was distributed to all the banks that fell into definition of the target population, and 8 banks responded to the questionnaire. This rendered a response rate of 57% and sample non-response rate of 43%. Is this a satisfactory response rate to guide us to achieve unquestionable insights of the target population and valid survey estimates?

According to the Division of Instructional Innovation Assessment (DIIA) at the University of Texas, 2008, a response rate is of greater significance when the purpose of the survey is to measure effects or to generalise to a target population and less significant if the purpose is to gain insight. Furthermore, Gillham, 2000 agrees with DIIA and concludes that if the response rate is less than

30%, the value and validity of the design and estimate results are in question. Therefore, conclusively a reliable response rate must be at least 30%. With great confidence, it was concluded that the bank survey satisfactorily met the reliability test with a response rate of 57%. The response rate confirmed that the intended purpose of the study was to gain insight of bank involvement with SME financing in Zimbabwe, therefore such response rate would give a good picture of how the supply side behaves towards financing the SMEs.

For the SME survey, the association on which a sample of 300 SME owners was taken from had a membership of 705 at the time of the survey. This rendered a response rate of 42.55% and sample non-response of 57.45%. This response rate was good for the set survey objectives. According to Gillham, (2000), it was concluded that the SME survey moderately exceeded the threshold by 12.55% margin. Confidently substantial evidence was extracted on the demand side insofar as the reality of the bank-SME relationship. SME owners' views and opinions over this relationship were pertinent to its smoothening for eventual benefit to the economy at large.

5.3.1 Quantitative analysis of bank survey

To extract answers to our research questions as well as attain the objectives of the research, the bank and SME surveys' results were quantitatively analysed and graphically presented in accordance to the evidence derived from them (Appendix A(bank) and Appendix B(SMEs)). In any quantitative data analysis of survey data, the preliminary task was to check on the frequency distribution of each variable of interest, embodied in each question of the questionnaire, to acquire the correspondingly numerical value which in turn indicated the total number of responses for the respective question (variable). In that respect, for the research, the frequency distributions were derived through the analysis of the bank and SME surveys questionnaire findings.

In the data analysis, assertions or confirmations were made using substantive evidence from survey data. This implied that confirmatory data analysis was the driving approach in the statistical analysis of the bank and SME surveys' results. For the quantitative data analysis to be exhaustive, some relationships were extracted between variables through pictorial representative approach. In fact, variables were explored to regularise the common process by making inferences about relationships between them. Therefore, for the two (2) surveys, both confirmatory and exploratory data analyses were conducted to deduce the extent at which commercial banks are involved in

SME lending in Zimbabwe. In accordance with the conclusion of Fink (1995), those results of statistical analysis were descriptions, relationships, comparisons and predictions, so were the majority analyses in this part of the research. Therefore, results of surveys are presented using both graphs and tables accompanied with relevant discussions.

5.3.2 Bank survey data analysis

Taking a closer look at the results of the bank survey data, it was clear that SME banking was no longer confined to small banks as previously claimed by economists and policymakers. Large and foreign banks have aggressively entered the SME lending market which was once purported to be a preserve of small or niche banks. From the surveyed banks, SME sector had become a fulcrum of most bank lending businesses. In fact, bank involvement with SME lending has become a brisk business, a consequence which ratifies the findings articulated by Berger and Udel (2006) and by de la Torre *et al.* (2010). Results showed that banks in Zimbabwe have realised that SME lending market is fast becoming competitive but far from saturation.

Traditionally SMEs were stereotyped risky and unprofitable lending asset by most banks. Evidence showed that the traditional stereotype no longer holds in practice as every bank surveyed was in one way or the other is involved with SME business at some scale. Regardless of the economic downturn prevailing in Zimbabwe, the institutional and macroeconomic environments seemed not to deter banks pursuing involvement with SME funding. The only observable limitation as a result of the current economic environment was the range of products and services they were offering to SMEs. Currently banks mainly offer short-term loans and overdrafts **(Figure 5.1).**

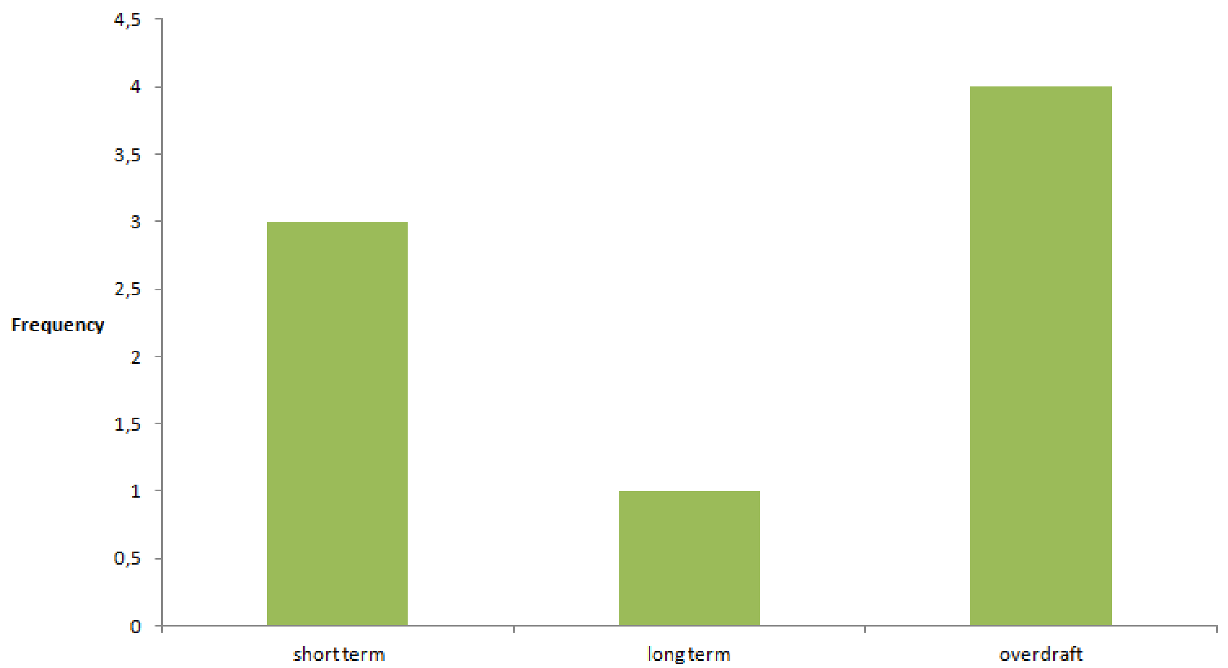


Figure 5.1: Type of financial services banks are offering to SMEs

The government's role as the facilitator of bank involvement with SMEs seemed to be paying dividends in greater pursuit of SMEs by many banks in Zimbabwe. To facilitate SME financing the government had established a ministry responsible for SME sector. On the question of whether the assessment of the impact of government SME policy on the sector had improved or not? 38% of the surveyed banks said it improved a lot whilst around 13% say it got worse. Government positive SME policy had helped contain divergence in the way banks look at SMEs due to the heterogeneity of what is called an SME.

There is no universality in the definition of SME which may compromise bank involvement with sector. Normally banks use turnover criterion to define SME but with diverging thresh holds, thereby making it difficult to compare one bank involvement with SME with another bank. To synchronise the relation between banks and SMEs, the ministry responsible for SMEs and Cooperatives had come up with homogenous SME definition adoptable by banks interested in doing business with the sector. This went a long way in making comparisons between banks' lending practice to earn consistence. This was consistent with the findings of Beck *et al.* (2008; 2010) and de la Torre *et al.* (2010) who downplayed heterogeneity of SME definition by considering whatever classification a surveyed bank adopted.

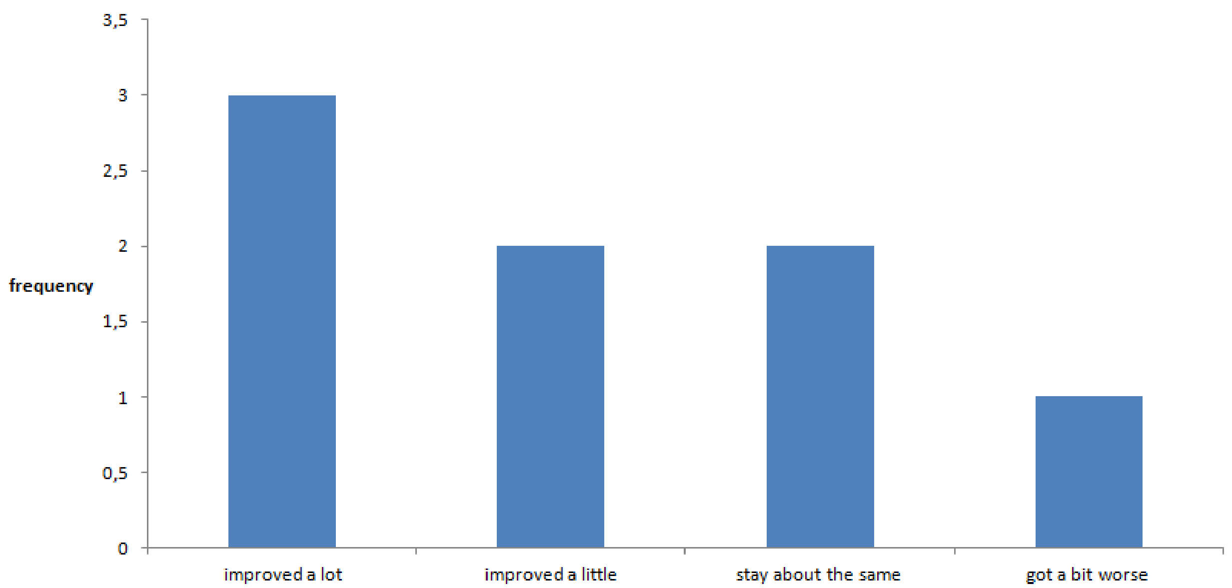


Figure 5.2: Government Involvement in SME Financing

Despite 38% of surveyed banks affirmed that government involvement into SME lending market vibrancy, the government's role still needed to be improved such that an enabling environment for Bank-SME relationship was achieved. Interviewed banks cited the non-existence of a credit reference bureau with adequate database that would help banks make informed credit decisions, thereby enhancing access to financing by SMEs.

Some banks suggested that the government ought to play a significant role in improving the judicial process in Zimbabwe, especially in enforcing the lender's rights when security lodged needed to be liquidated. Government was also called upon to make loans to SME cheaper by subsidising the banks involved in SME lending business. A vibrant SME sector was the panacea to a plethora of economic problems bedevilling Zimbabwe, so government was supposed to take a leading role. The establishment of a ministry responsible for SMEs was loudly applauded although adequate capitalisation was still a mirage.

On the issue of drivers of bank involvement with SMEs, banks in Zimbabwe seemed to be driven into SME business by various factors. From the bank survey, the factors which had been found driving banks' desire to get involved with the sector was the pursuance of profitability envisaged from the 'virgin' sector. Profits in the SME market are promisingly attractive. This upsurge seemed to have been exacerbated by the current downturn of the Zimbabwean economy which had seen the folding up of business by large firms and the rampant downsizing of operations. The implication was that banks, which were the traditional funders of large firms, are experiencing thinning of profit margins due to the fast dwindling of their traditional market. Therefore, spreading their tentacles into the non-traditional SME market has proven the logical thing to do, pursuing the fulcrum of their lending business. In the retail market cut-throat competition was the order of the day as there was an influx of other players in the form of departmental stores, loan sharks and others.

Lending to government had also hit a snug due to stringent fiscal policies and the ability of government to source funding from international capital markets. By re-positioning into SME lending market, banks are developing better technologies to assess this type of clientele, minimising the aggregate cost of lending which would eventually reduce interest rates to affordable levels to SMEs. There are also other factors besides perceived profitability which also drove banks to engage with SMEs **(Figure 5.3).**

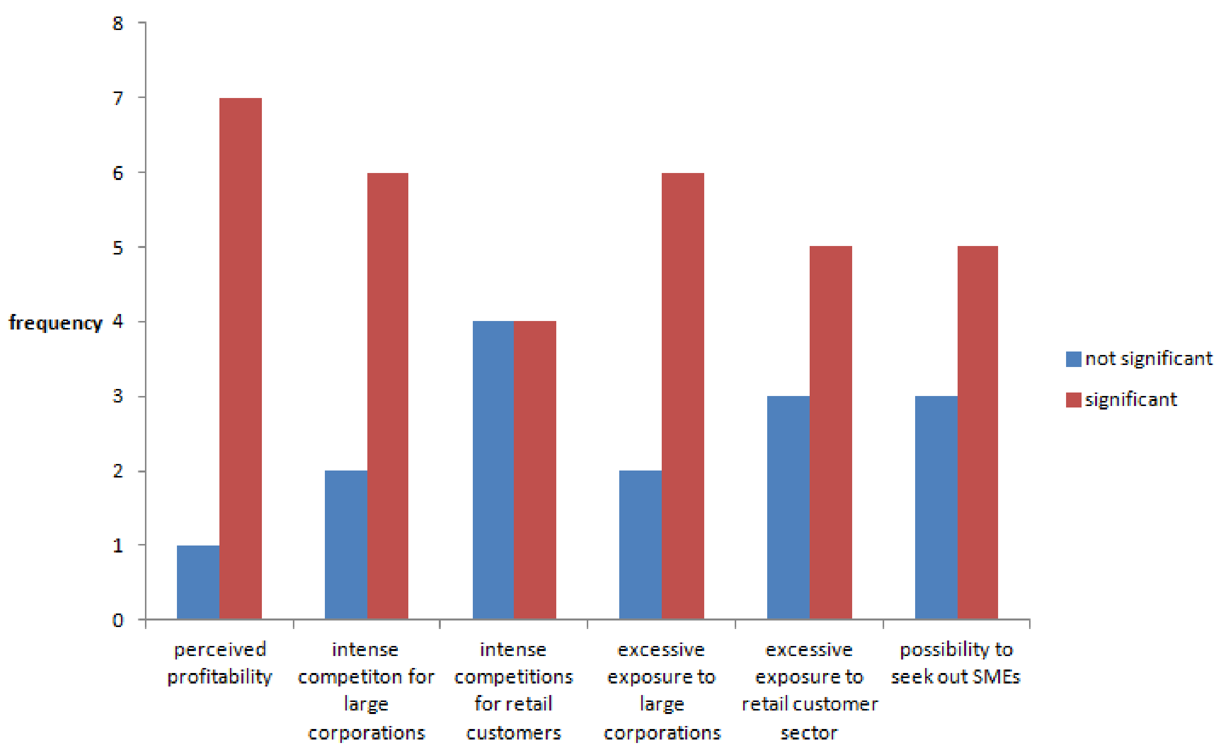


Figure 5.3: Drivers of Bank Involvement with SMEs.

On the other hand, factors holding banks from involvement with SME, the survey revealed that SME-specific factors, which include poor quality of financial statements, poor debtor information systems, inability of SMEs to manage risk and informality, are the most deterrent factors to bank-

SME engagement. Legal and contractual environment was also mentioned as an obstacle to bank involvement with SMEs. Interviewed banks have called upon the judicial service sector to regularise the bank-SME engagement for the benefit of both parties.

On the SME-specific factors, some banks suggested that they ought to play a pivotal role in nurturing SMEs adopt good business principles such that they can reach them out using arms-length technologies like credit scoring which require 'hard' information. In that regard, interviewed banks applauded the Small and Medium Enterprise Association of Zimbabwe (SMEAZ) for constant holding of seminars on how to do business professionally. Such practice would go a long way to alleviate the effect of SME-specific factors on bank involvement with SMEs.

Common wisdom could have been persuaded to cite macroeconomic factors as the major hindrance to SME lending given that Zimbabwe has been in the doldrums economically since 2000, which may have been- apportioned to poor bank-SME relation. Figure 5.4 shows how banks in Zimbabwe rate the major obstacles to banks involvement with SME lending.

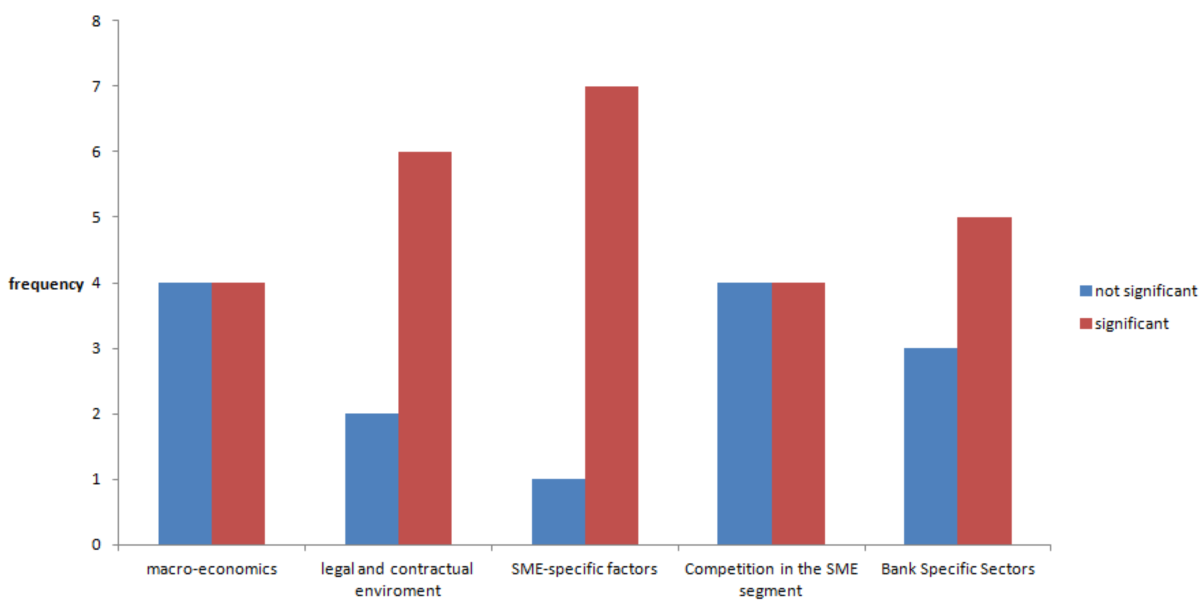


Figure 5.4: Obstacles to exposure Bank-SME Engagement.

Evidence from the survey was that the SME lending market was large and with significant prospects. As result banks were adopting bullish strategies given the prospective outlook of this market. 62.5% of the banks in the sample said that the SME lending market was large, and prospects were good although some other banks were still pessimistic about the prospects of the SME lending mark*et al*though most banks acknowledged the vastness of the market. On the competitiveness of the SME lending market, 87.5% of the surveyed banks said the market was competitive but not saturated **(Tables: 5.1 and 5.2)**. The evidence from this survey corroborated the World Bank/ Finscope 2012 MSME Survey in respect to the dominance of the Zimbabwean lending market by SMEs where about 5.7 million people were found to be involved. This was attributed to the unabated closures of large firms in Zimbabwe.

Table 5.1: Size and Prospects for SME market

Variable	Category	Frequency	Proportion (%)
What is your view on the size and prospects for SME market in general?	The SME market is small, and prospects are good	2	25.0
	The SME market is large, and prospects are bleak	1	12.5
	The SME market is large, and prospects are good	5	62.5

Table 5.2: Competitive nature of the SME lending Market

Variable	Category	Frequency	Proportion (%)
How competitive is the market for SME lending?	The market is not competitive and entry costs are low	1	12.5
	The market is competitive but not saturated	7	87.5

On banks' practices in SME lending, aspects to do with business models banks in Zimbabwe were adopting in this market, types of products they were offering, risk management (collateral requirements for SMEs, monitoring and credit analysis) were looked at. To deal with SME market, the bank organisational setup was to be arranged to suit the new lending market environment. About 38% of the interviewed banks mentioned that they have set up separated dedicated units to deal exclusively with bank-SME relations and the other banks were anticipating doing the same in the near future. This new model of doing business was a preserve of large domestic banks and foreign banks. This was a clear show that the SME lending asset was different from the traditional lending assets; retail and corporate. The major reason of adopting the separate dedicated units was to offer SMEs a wide variety of tailored products and services. Other banks which had not yet adopted the new organisational setup, offered standardised products to their SME clientele.

In as far as reaching out their SME clientele, banks used either headquarters or branches or both. The headquarters approach was designed based on which SMEs the bank would target and what products would be offered. On the other hand, branches attend to basic needs of the SMEs. There were activities done at both branches and headquarters. From the survey, banks were decentralising major activities by allowing them to be done at both headquarters and branches for instance for the activity: risk management; 12.5% of the banks said this activity was done only at the headquarters, 25% said it was primarily done at headquarters and 62.5 % allowed for both. In terms of type of products, banks surveyed offered standardised products (62.5%), equal number of standardised and tailored products (25%) and tailored (12.5%) **(Table 5.3)**. The significance of tailoring products rose with the size of the enterprise and banks design such products to a set of SMEs with similar needs.

Table 5.3: Standardisation of SME products

Variable	Category	Frequency	Proportion (%)
Indicate the most relevant statement regarding the standardisation of your SME products	Standardised	5	62.5
	Standardised/tailored	2	25.0
	Tailored	1	12.5

In SME lending, banks traditionally demanded collateral from SME borrowers which was normally higher than retail and corporate borrowers because credit risk of SMEs was difficult to evaluate. As a result, SME were generally considered riskier due to their informality and opacity character. To safeguard depositor’s money, banks in Zimbabwe require security in the form of collateral against any loan extended to SME business. Since SMEs were more vulnerable to both political and economic shocks and that their information was hard to assess for credit risk, the collateral demand was normally substantial.

Table 5.4: Collateral requirement for SMEs

Variable	Category	Frequency	Proportion (%)
Are collateral requirements higher for SMEs than for large corporate?	SMEs are more unstable	1	12.5
	SMEs are more informal	2	25.0
	SMEs have worse management	2	25.0
	SMEs are harder to evaluate	2	25.0
	SMEs collateral more difficult to seize in case of default	1	12.5

To secure their lent money to SMEs, banks normally use preventative triggers to monitor. These included debts outstanding, repayment frequency and reports by bank staff after regular visits. The monitoring was either done daily or weekly or monthly. From the survey, 50% of the banks did the monitoring daily while 25% monitored weekly and monthly respectively **(Table 5.5)**.

Table 5.5: Monitoring Frequency

Variable	Category	Frequency	Proportion (%)
How often are exposures relative to limits monitored?	Daily	4	50.0
	Weekly	2	25.0
	Monthly	2	25.0

5.4 SME SURVEY ANALYSIS

In any relation there are two (2) sides. In a bank-SME relation the two (2) sides are supply and demand sides in the sense that banks supply financial products and services which the SMEs consume for their growth and development. Therefore, to better understand the mechanics of this relationship, there was a need to closely analyse viewpoints from each side of the financing equation. These viewpoints from each side were collected through respective surveys with the aim to get anecdotal evidence to verify any hypothesis made over this relation. The bank survey had been analysed so to complement the bank side information, it was logical to also extract evidence from SME surveys, the demand side. The data was independently collected in such a way that it divulged different but complementary and corroborative viewpoints of the relation between SMEs and banks.

Although the data was collected from different sources, the SME survey data corroborated many of the findings obtained from bank survey because SME relate with banks as consumers of an assortment of bank products and services. Prior to the study common wisdom asserted that SMEs procured finance via relationship loans due to purported inherent opacity and informality but in practice, they also access financing products and services that were not accessed by processing “soft” information of the firm and the owner.

Evidence from the SME survey revealed the demographic details of the majority of SME owners were at least college and university leavers (33.3%), indicating that they were people with prerequisite literacy for business and trainable to business norms if need be, people with passion and moderate know-how of the trade they were into. Minority (8.3%) were owners with primary school education indicating basic literacy. The educational background profile of SME-owners in Zimbabwe is exhibited in Figure 5.5.

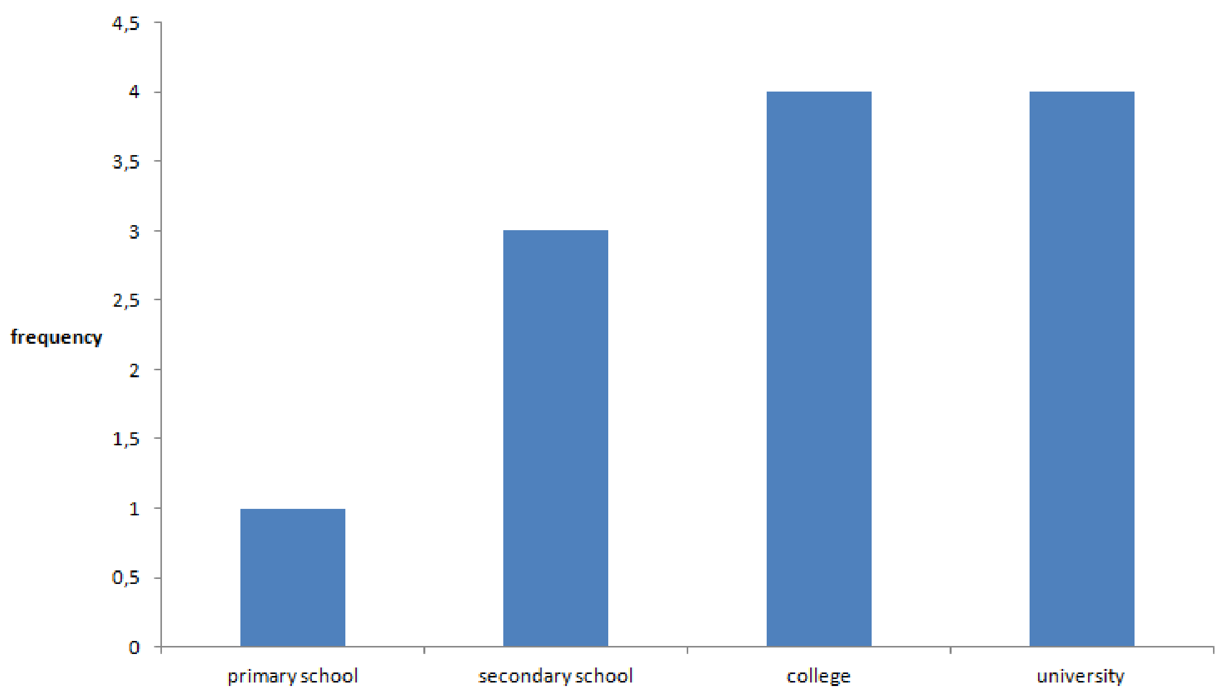


Figure 5.5: Educational background of SME owners in Zimbabwe

With regard to the biggest obstacles to the growth of their enterprises, 33.3% SME-owners cited obtaining additional financing and instability of demand products and services as the prime hurdles to SME growth and development in Zimbabwe. This revelation was in tandem with research output by other researchers like de la Torre *et al.* (2010). This was contrary to the assertion that government regulations inhibited SME growth and development as a paltry 8.3% confirmed that.

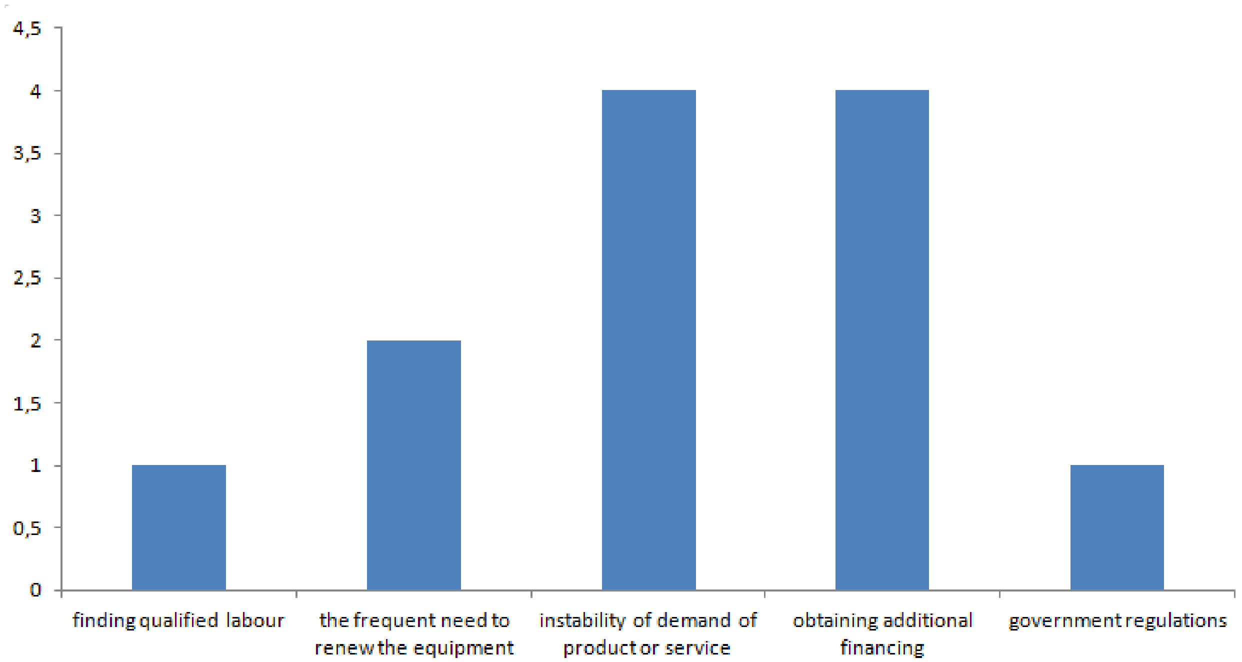


Figure 5.6: Obstacles to SME Growth

With regard to main activity in which the majority of SMEs in Zimbabwe are involved in, 25% are involved in both wholesale or retail and agriculture followed by mining and transport (16.7%) both. This is in line with the major economic activity spectrum of Zimbabwe.

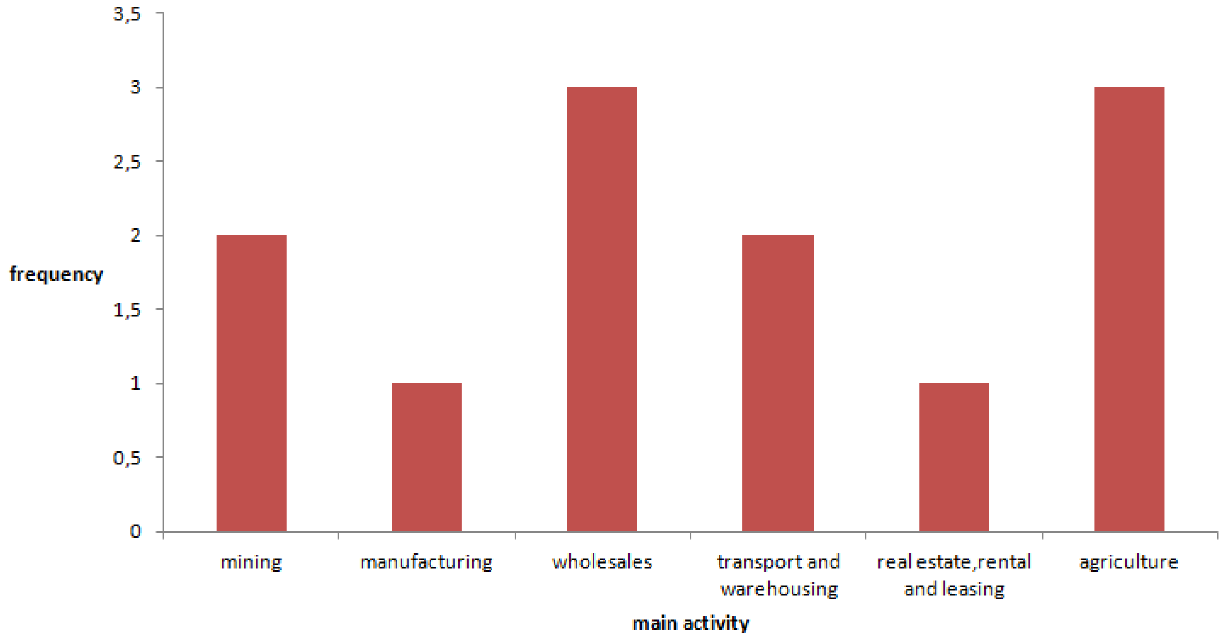


Figure 5.7: Main Sectors of the Economy in which SMEs are involved

5.5 SUMMARY

The evidence in this chapter queries the hypothesis that SMEs are neglected by most banks in Zimbabwe because of their chronic opacity and information asymmetry which made them illegible for large banks financing. They are said to be substantially dependent on relationship lending, a niche for small banks which have a natural comparative advantage in this type of banking. To the contrary, the new evidence unravelled by this research suggests, in reality, that all kinds of banks in Zimbabwe view SMEs as a strategic sector for their prime business owing to closures or scaling down of operations by large companies. In fact, large banks are extending their operations to this SME sector with great aggression in pursuit of profitability, as profit margins are thinning out in the traditional market (retail and corporate). As a result, the SME lending market is growing increasingly competitive and promisingly profitable if relevant business models are developed to deal with intrinsic and specific characteristics of the sector. It has also been noted that the sector is far from being saturated.

SME survey has demonstrated that SMEs in Zimbabwe no longer exclusively procure financing through relationship lending but also access financing products that depend on arms-length technologies which do not need the bank processing soft information on the firm. This implies that

SMEs which were ever a niche market for small banks business are no longer the case. Bank specific and SME specific factors seem to be at play in the formation dynamic patterns that have engulfed the non-cyclical phenomenal bank-SME relations. In response to the turning of events, banks swiftly climatizing to prevailing scenario by developing appropriate new business models, arms-length technologies, and CRM systems to serve this emerging lending market. Despite lending to SMEs, banks are moving into offering holistically a wide range of products and services, with fee-based products have gained in importance. This is where large banks and foreign universal banks have taken the lead in servicing SME market which was one presumed niche territory for small. Basically, through multi-service and economies of scale, large banks can diversify credit risk, generate enough data to build more sophisticated CRM tools, an added advantage over small banks.

The evidence unravelled from the two (2) complimentary surveys is novel and unique for Zimbabwe as it is generated from two (2) competing but corroborating sources and captures both the demand- and supply-side dimensions. Anecdotal evidence shows that while banks-SMEs relationships are increasingly becoming stronger, but SMEs seem yet unable to obtain sustainable access to loans secured by collateral or long-term fixed-interest rate loans in domestic currency. This has led to SMEs seeking for such financing from non-banking institutions. This induces cut-throat competition in the lending business thereby threatening pursuance for profitability by banks. There is need for banking to develop underwriting technologies that do not depend on the availability of collateral especially start-ups.

For further analyses of the bank survey are contained in Appendix A while for SME survey is in Appendix B.

CHAPTER 6: CRM MODELLING

6.1 INTRODUCTION

This chapter captures the modelling of the CRM. It covers the key features of modelling procedures followed up to the validation stage of model built. It starts by exploring the credit data provided by a bank in SME financing, in preparation for a preliminary CRM modelling, since the procedure requires clean and less noisy data.

For CRM modelling, variable selection is a crucial preparatory procedure therefore is therein explained using WoE and IV, in addition to operational and business considerations. WoE is a quantitative method for combining evidence in support of a statistical modelling process hypothesis (Good, 1985). It is widely used in CRM modelling to separate between good accounts and bad accounts. It makes comparative analyses of the proportion of good accounts to bad accounts at each attribute level and quantifies the strength of the attributes of an independent variable in separating good and bad accounts. On the other hand, IV is a metric divergence measure that can be calculated at attribute level.

The chapter explains how the preliminary model was built from the determined risk profile, a consequence of variable selection process through pre-processing and transformation of independent variables. It furthermore discusses the key concepts of reject Bound and Collapse Bayesian reject inference technique employed which embodies the estimation of the missing data mechanism as well as the choice of priors for its implementation. This resulted into the formation of an AGB development sample where the “weak” and “strong” final CRM models were built. Some validation procedures are therein explained up to back-testing of the final developed model for its robustness.

6.2 DATA

Data for CRM modelling was collected at ZimSME bank, which provided detailed data on the SME owners’ characteristics, enterprises’ characteristics, business performance and enterprises’ financial position together with corresponding credit quality or payment behaviour of each client in their loan register. The portfolio contained 354 observations for all SMEs that were operational under the stewardship of the current ownership during the 2010-2012 performance window. This database formed the original KGD development sample for the preliminary CRM modelling. Currently to reach a selection decision to grant loan to an SME clientele, the bank assessed loan qualifying clients on 26 risk factors, which are scored using judgemental/expert scorecard or the discretion of the credit analysts. To develop an objective CRM model for the bank, the data was supposed to be divided into a training sample (80%) and hold-out (20%) samples but because of small size, the whole SME loan portfolio (100%) was used as development sample for the preliminary model and an arbitrary 70% hold-out sample was extracted for validation of the eventual model (Siddiqi, 2006).

The envisaged CRM model must be accurate on average across the range of the borrowers or facilities to which the bank is exposed and there must be no known material biases (BCBS, 2003). For fear of confidentiality and competition, the majority of banks were reluctant to volunteer their bank information for CRM modelling purposes regardless of ethical assurance by the researcher. But ZimSME bank volunteered credit registers of its SME portfolio provided the researchers observed the ethical code of conduct signed prior to data collection. In sum, the prominent feature of this portfolio was that the bad rate was 20.62%, that is, 73 SMEs were delinquent on at least one business obligation in the performance window of 2 years (2010-2012).

The data collected was initially noisy therefore had to be processed and transformed to be applicable for potential CRM modelling, so was first explored and cleaned. CRM data is very sensitive to noisy data therefore needed thorough cleaning to remove none desirable features and outliers.

6.3 DATA EXPLORATION FOR CLEANING PURPOSES.

Data quality was paramount to the successful development of the envisaged CRM model, for the process is highly susceptible to irrelevant or noisy and unreliable input data. Therefore, for substantial knowledge discovery during CRM model development, data cleaning became a crucial preliminary process. Data cleaning was done at attribute-level, where some data were not included for model development process because they were removed on basis that they did not add any value and efficacy to the envisaged model. Such data included the SME customer code or application number. Again, SME clients with 50% or more of missing values of characteristic variables were removed from the portfolio. For case of ZimSME data there were no clients that had 50% or more missing values therefore the initial development sample size remained at 354 after thorough data cleaning.

6.4 DATA CLASSING/BINNING OR DATA DISCRETISATION

Owing to fact that some ranges of continuous intervals were so wide or due to the presence of outliers, the data was discretised. Effective discretization reduced the demand for modelling memory and enhancing efficiency of data mining by making the extraction of predictor variables from discretized data easy to understand. In this case data classing was used as a way of discretization process. This was done using R programming for validation.

Automatic data classing involved dividing the interval of values of a numeric attribute into a number of smaller intervals, whereby each of the resultant smaller intervals was treated as one value of a categorical attribute. Such refinement helped better understanding of the relationship between the attributes concerned with the dependent variable, the credit quality. The discretisation process was interactive that it was guided by recommendations from ZimSME bank credit risk manager and the staff of the department. As a result of this consultation, binning was performed in two (2) distinct iterative steps which were fine classing and coarse classing. For fine classing, the raw data was examined for its reliability and suitability and then categorized into smaller groups.

The second step involved coarse classing. Here all the 26 characteristics were aggregated into stable and statistically significant classes. In fact, each data was converted into standardized good/bad odds ratio (WoE) which was calculated using R programming. Coarse classing was performed on each attribute with the aim to minimize the drop in the characteristic's IV without breaching coarse classing standards. For this study we used ZimSME bank own coarse classing standards of having a minimum of 5% 'bad' for each interval. The process of fine and coarse classing was recursive on each attribute until coarse classing standards were satisfied.

Consequentially a new set of clean data was attained in which the values of the attributes were represented by their corresponding WoE. This was done in preparation to the crucial stage of variable selection or variable reduction to achieve parsimony in model development process. For preliminary modelling, at least weak to too "good" characteristic variables were selected for stepwise logistic regression model building.

6.5 INITIAL VARIABLE SELECTION ANALYSIS OF KGB SAMPLE

In modelling of the CRM for ZimSME bank, there was initially a temptation to include all the 26 predictor variables the bank was currently using to select its SME clientele for loan granting, to explain the response variable, credit quality, adequately. But the principle of parsimony is contrary to that temptation, therefore required that the number of predictor variables be reasonably small, not to over-fit the model. Harrell (2001) identified that over-fitting of the data results in model unreliability, quite common in datasets with relatively large number of variables as candidates for categorical predictors yet the dataset was relatively small. For CRM modelling, a parsimonious model would require between 8 and 15 decision variables (Siddiqi, 2006) to make far-reaching credit granting decisions.

Careful selection of risk predictor variables was a crucial prerequisite to exclude unimportant or redundant decision variables from the initial set of 26 characteristic variables. The statistical CRM model was supposed to reflect the inherent interrelations between different borrower's characteristics. Therefore, the quality of the resultant model was dependent both on the forecasting capacity of the borrower's characteristics in use and the link of these characteristics that would take part in the process of development of the model. This gave great weight to the variable selection procedure as a premise of an efficient and purposeful CRM model.

The case studied ZimSME bank currently assessed on 26 characteristic variables to select its SME clientele through expert/judgmental scorecard and the discretion of credit analysts for loan granting. The number of input variables was large, therefore called for rigorous and effective variable selection process; to avoid overestimation of the main effects in the analysis. To extract significant predictors for the model, variable selection was meticulously done, using different approaches, to help the bank make credit granting decisions with great ease but with far reaching effects. The inclusion of unnecessary attributes would affect the efficiency of the resultant model, which would be drastically reduced.

Variable reduction process was done, in parallel, using other multivariate variable reduction methods in the form of WoE and IV approaches. These are the mostly used as pertinent tools for variable selection in CRM model development. The output of the predictor variables IV rank-ordered are shown in Table 6.1.

Table 6.1: Characteristics rank-ordered by information value (IV)

	CHARACTERISTIC	IV	COMMENT
1	Stock T/O	2.285801316	Suspicious or too good to be true
2	Net Profit Margin	1.858083826	Suspicious or too good to be true
3	Liquidity Ratio	1.376789969	Suspicious or too good to be true
4	Age	0.723543789	Suspicious or too good to be true
5	Annual Turnover	0.386202057	Strong predictor
6	Interest Rate	0.350522094	Strong predictor
7	Debtor days	0.277656269	Medium predictor
8	Gearing	0.266338835	Medium predictor
9	Income	0.194543471	Medium predictor
10	Loan Amount	0.192823414	Medium predictor
11	Length of Relationship	0.171665483	Medium predictor
12	Experience (Yrs.)	0.146297777	Medium predictor
13	Creditor days	0.141425936	Medium predictor
14	No. of Employees	0.123270680	Medium predictor
15	Asset Size	0.105491386	Medium predictor
16	Local Trade	0.092097831	Weak for prediction
17	Qualifications	0.070571085	Weak for prediction
18	Sector	0.069287810	Weak for prediction
19	Number of Directors	0.047231991	Weak for prediction
20	Collateral	0.037774106	Weak for prediction
21	Technology -Level	0.009776749	Useless for prediction
22	Export	0.009564626	Useless for prediction

	CHARACTERISTIC	IV	COMMENT
23	Other Accounts	0.006746730	Useless for prediction
24	Tenor of Loan	0.005768000	Useless for prediction
25	New firm	0.001602608	Useless for prediction
26	Loan Purpose	0.001413792	Useless for prediction

The WoE and IV statistics helped identify predictor variables or characteristic variables from application forms and other sources that yielded good estimates of the probabilities of default (PDs) of the SME clients in the ZimSME bank loan portfolio. The two (2) (2) statistics (WoE and IV) have the following advantages over other commonly used variable reduction techniques. They are used to assess the predictive power of categorical, ordinal, discrete and continuous data simultaneously. The two (2) (2) statistics were calculated for missing values of each characteristic, leading to the development of the preliminary CRM model for ZimSME Bank.

From the rank-ordered variables by IV, we considered variables which were at least weak predictors to the too good to be true predictors since the number of the variables was small. We used backward logistic regression model building technique to further screen out insignificant variables. Therefore, we used univariate screening of the 26 original characteristic variables which ZimSME bank was currently using to extract the initial credit risk profile (Table 6.1). The criteria used to gauge the predictive power of each characteristic variable included the following among others; WoE, range and logical trend of WoE across group, IV and operational or business considerations. The process of constructing this credit risk profile was interactive with the bank's credit risk department.

Variable Selection: Backward Logistic regression results:

Coefficients:

Variable	Estimate	Std. Error	z-value	Pr (>|z|)
(Intercept)	-15.109916	6.559383	-2.304	0.0212 *
`Annual T/O`	-0.178854	0.349305	-0.512	0.6086
`INTEREST RATE`	7.380678	3.376932	2.186	0.0288 *
`Debtor days`	0.009553	0.013492	0.708	0.4789
Gearing	1.331374	0.592452	2.247	0.0246 *
`Loan Amount`	0.048566	0.188751	0.257	0.7969
`Length of relationship	0.023770	0.021464	1.107	0.2681
`Experience (Yrs)`	-0.017539	0.021234	-0.826	0.4088
`Creditor days`	0.021578	0.010209	2.114	0.0346 *
`Number of Directors`	0.290517	0.145966	1.990	0.0466 *
`Asset Size`	-0.257311	0.195318	1.317	0.1877
Income	1.820393	0.467363	3.895	9.82e-05 ***
Age	0.092995	0.022115	4.205	2.61e-05 ***
`Net Profit Margin`	-8.959870	1.965926	-4.558	5.17e-06 ***
`Liquidity Ratio`	-0.261197	0.132063	-1.978	0.0479 *
`Stock T/O`	-1.658440	0.318545	-5.206	1.93e-07 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 360.30 on 353 degrees of freedom

Residual deviance: 183.21 on 338 degrees of freedom

AIC: 215.21

Number of Fisher Scoring iterations: 7

Table 6.2: Risk profile

No.	Characteristic variable
1	Stock T/O
2	Net Profit Margin
3	Liquidity Ratio
4	Age
5	Creditors days
6	Interest rate
7	Debtor days
8	Gearing
9	Income
10	Loan Amount
11	Number of directors

6.6 PRELIMINARY MODEL DEVELOPMENT BASED ON KGB SAMPLE

Owing to the binary character of the response variable, credit quality (default/no-default); logistic regression technique was adopted for the development of a preliminary CRM model for the ZimSME bank. In fact, the preliminary model was developed by the application of the logistic regression on KGB sample using the risk profile in Table 6.2. It suited well since the predicted variable, credit quality, was categorical and dichotomous. This generalized linear model was used in its multivariate form since the binary outcome; good/bad, depends on a number of predictor variables in the risk profile, as deduced from initial characteristic analysis of the KGB sample. The logit transformation, the logarithm of the odds was used to linearize posterior probability and to limit the outcome of estimated probabilities in the linear model to be between 0 and 1.

The parameters were jointly estimated by using the maximum likelihood estimation (MLE) method in R programming software and appropriate interpretation of these estimated parameters. In the process of selecting a good fit model, various regression diagnostic techniques were employed in the form of t-values, p-values and AIC to assess the suitability of each and every suspected predictor variable. The preliminary model is as shown in Equation 98:

Equation 98: Preliminary CRM model

$$\boldsymbol{P(y_i = 1|x_i)} = \frac{e^{-20.418+6.725x_1+1.181x_2+0.019x_3+0.302x_4+1.946x_5+0.086x_6-9.188x_7-0.272x_8-1.613x_9}}{1+e^{-20.418+6.725x_1+1.181x_2+0.019x_3+0.302x_4+1.946x_5+0.086x_6-9.188x_7-0.272x_8-1.613x_9}}$$

where

$x_1 - interest\ rate$, $x_2 - gearing\ ratio$, $x_3 - credotors\ days$, $x_4 - number\ of\ directors$, $x_5 - income$, $x_6 - age$, $x_7 - net\ profit\ margin$, $x_8 - liquidity\ ratio$ and $x_9 - stock\ turnover$.

6.7 APPLICATION OF BOUND AND COLLAPSE TO CREATE AGB SAMPLE

The preliminary CRM model developed was applied to the original KGB sample of 354 SME clients to simulate a credit granting policy, by setting a 500 - CRM cut-off score policy to obtain a “accepted subsample” and a “rejected subsample”. From the application of the preliminary model, the “accepted subsample” contained 298 while the “rejected subsample” contained 56 clients respectively. This translated to a bad rate of 15.82%, a drop from the original judgemental model, which had a bad rate of 20.62%. This implied there was a 23.28% improvement from the judgmental CRM model the bank was originally using to the preliminary model developed on the same KGB sample.

To develop a model capable to measure credit risk of the TTD population, there needed first to apply the BC methodology to create a statistically random AGB sample representative of the population of SME loan applicants at ZimSME bank. To simulate a robust model befitting banking industry, two (2) granting policies were set up by applying two (2) different thresholds or cut-off CRM scores such that the extent of missingness was made different between the two (2) resulting samples. This also had bearing on the potential sample selectivity bias such that it was also different between the two (2) resultant selection strategies. We obtained two (2) selections which we referred to as “weak” and “strong” based on high and low cut-off selection criteria respectively. This is in line with banking industry because a bank either raises or reduces the credit cut-off in response to any adversity in the bad rate. In emulating a real situation at ZimSME bank, a credit rejection policy was simulated in such that the acceptance rate was equal to $(1 - bad\ rate)$ of the original portfolio. After applying the preliminary model on the original KGB sample our bad rate was found to be 17.8% $\left(\frac{63}{354}\right)$, which implied that the acceptance rate was 82.2% and therefore the cut-off score was 396 CRM score. This selection strategy was referred to as "weak selection" which meant that all SME clients with CRM score less than 396 constituted the "selected subsample" and those with scores greater than 396, made up the "rejected subsample". For the "strong selection" strategy we doubled the bad rate of the weak selection strategy to 35.6% and the acceptance rate was 64.4% and therefore the respective cut-off CRM score was 153. Similarly, the "selected subsample" was made up of clients with credit scores less than 153 and those whose scores were greater than 153 constituted the "rejected subsample". To create a missing data problem, it was assumed that the credit quality of applicants was only observable in the respective simulated "selected subsamples" while missing in the simulated "rejected subsamples".

The institution of the "weak" and "strong" selections created two (2) subsamples respectively: "accepted sample" and "rejected sample" under respective rejection policies. It is assumed that the respective "rejected subsamples" were made up of SME loan applicants whose credit quality were unknown, engendering missing data problem. The simulated missing data needed to be imputed. This needed, first, the estimation of the missingness function and choice of the non-informative prior embedded in the Dirichlet function of BC. In CRM modelling, to estimate the missing data mechanism (MDM), the probability of being bad was used as a logical proxy, because it is justifiably true that the probability of being bad is equivalent to the probability that the credit application is rejected. This implied that the original CRM score was considered as the probability of being missing in the "selected subsample".

6.7.1 Estimation of the missing data mechanism (MDM)

The Bound and Collapse Bayesian reject inference methodology is an imputation model-based technique capable of dealing with incomplete samples based on Bayesian inference analysis. For this case it was the technique employed to create an AGB development sample on which the final CRM model was to be estimated. To accomplish that task, we needed to, first of all, estimate probabilistically the missing data mechanism (missingness function). To achieve that crucial procedure, possible sources of information for the envisaged missingness function were initially identified. In credit risk modelling, it is suggested that the original score contains all "external" information vital for approximating this missingness function and also that the developed preliminary CRM score contains "internal" information useful for the inference of the anticipated missingness function. To source explicit information from the "external" source, it was assumed that the original credit risk scoring model had sufficient classification power, to the extent that the probability of being bad in the selected sample was trustworthy for future decision-making process. In this thesis we used the external source only due to lack of enough data.

Once we obtained the simulated "accepted" and "rejected" subsamples from the institution of "weak" and "strong" selection strategies, an assumption that the credit quality or credit risk was only observable in the "accepted" sample and not observable in the "rejected" sample was made. Then to estimate the missing credit quality in the "rejected" sample, the bad rates from the "accepted" region were used. The estimation was done by means of a simple linear extrapolation model between the mid-credit score range and the bad rate, where the bad rate was the dependent variable and mid-score range was the independent variable. The resultant model was extrapolated to infer the missingness probabilities of the predicted bad rate values in the "rejected" region. In fact, the predicted bad rates in the "rejected" regions were interpreted as the probabilities of missingness:

For the weak selection strategy, the simple regression equation used to estimate missingness, $\varphi_{j|i}$ was given by:

$$Bad\,rate = 0.00066Score + 0.0056,\ R^2 = 0.9581$$

While for the strong selection strategy, the simple regression equation used to estimate missingness, $\varphi_{j|i}$ was given by:

$$Bad\,rate = 0.0012Score - 0.00445,\ R^2 = 0.7037.$$

6.7.2 Selection of priors

For complete development of an AGB sample by the application of BC there was need to select appropriate precision parameter of the Dirichlet distribution, a generalised Beta distribution embedded in the BC methodology. In fact, the BC uses the multivariate generalisation of the Beta distribution, the Dirichlet, as the conjugate prior distribution (see equation 85). For simplicity, we chose a non-informative prior distribution $\alpha_{ij} = 0$ for all i and j. This choice was also necessitated by the fact that this prior distribution meant that the expected probability of being bad would be determined by the MDM when there were no observations in some CRM score ranges. At the same time, in score bands where there were observations then equation (88) was implemented to calculate respective estimated probabilities of being bad. Therefore, an AGB sample was achieved where both good and bad were elements making it more representative of the TTD population of SME loan applicants. It was on this resultant sample where the final CRM model was eventually built. Estimated probabilities of being bad conditional on each score range, $\hat{p}_{j|i}$, on the rejected samples for both strong and weak selections are given in Table 6.3 and Table 6.4.

Table 6.3: Observed and predicted bad rate for strong selection

Score range	Mid class	α_{ij}	α_{l+}	n_{ij}	n_{i+}	m_i	$\varphi_{j\|i}$	$\hat{p}_{j\|i}$
154-194	174	0	0	0	0	1	0.1643	0.1643
195-235	215	0	0	0	0	5	0.2135	0.2135
236-276	256	0	0	0	0	1	0.2627	0.2627
277-317	297	0	0	0	0	4	0.3119	0.3119
318-358	338	0	0	0	0	4	0.3611	0.3611
359-399	379	0	0	0	0	3	0.4103	0.4103
400-440	420	0	0	0	0	1	0.4595	0.4595
441-481	461	0	0	0	0	1	0.5087	0.5087
482-522	502	0	0	0	0	2	0.5579	0.5579
523-563	543	0	0	0	0	2	0.6071	0.6071
564-604	584	0	0	0	0	3	0.6563	0.6563
605-645	625	0	0	0	0	1	0.7055	0.7055
646-686	666	0	0	0	0	2	0.7547	0.7547
687-727	707	0	0	0	0	3	0.8039	0.8039
728-768	748	0	0	0	0	4	0.8531	0.8531
769-809	789	0	0	0	0	3	0.9023	0.9023
810-850	830	0	0	0	0	5	0.9515	0.9515

Score range	Mid class	α_{ij}	α_{l+}	n_{ij}	n_{i+}	m_i	$\varphi_{j\|i}$	$\hat{p}_{j\|i}$
851-891	871	0	0	0	0	6	1.0007	1.0007
892-932	912	0	0	0	0	5	1.0499	1.0499
933-973	953	0	0	0	0	4	1.0991	1.0991
974-1000	987	0	0	0	0	7	1.1399	1.1399

Table 6.4: Observed and predicted bad rate for weak selection

Score range	Mid class	α_{ij}	α_{l+}	n_{ij}	n_{i+}	m_i	$\varphi_{j\|i}$	$\hat{p}_{j\|i}$
397-437	417	0	0	0	0	1	0.2558	0.2558
438-478	458	0	0	0	0	1	0.2804	0.2804
479-519	499	0	0	0	0	2	0.305	0.305
520-560	540	0	0	0	0	2	0.3296	0.3296
561-601	581	0	0	0	0	3	0.3542	0.3542
602-642	622	0	0	0	0	1	0.3788	0.3788
643-683	663	0	0	0	0	2	0.4034	0.4034
684-724	704	0	0	0	0	3	0.428	0.428
725-765	745	0	0	0	0	4	0.4526	0.4526
766-806	786	0	0	0	0	3	0.4772	0.4772
807-847	827	0	0	0	0	4	0.5018	0.5018
848-888	868	0	0	0	0	6	0.5264	0.5264
889-929	909	0	0	0	0	7	0.551	0.551
930-970	950	0	0	0	0	4	0.5756	0.5756
971-1000	991	0	0	0	0	7	0.6002	0.6002

6.8 VERIFICATION OF THE BOUND AND COLLAPSE REJECT INFERENCE

After having inferred two (2) complete random development samples; weak and strong, some verification of the reject inference technique and parameter testing procedure were carried out before the initial characterization of the AGB samples. We split the respective inferred selected censored samples (weak and strong) into arbitrary "accepts" and "fake rejects" in the ratio 70% to 30%. We developed respective logistic models on 70% of the "accepted" of weak and strong selection strategies, using backward regression model building technique. The resultant respective models were then applied on the 30% "fake rejects" whose credit quality was known. Therefore, any observable misclassification due to the application of respective developed logistic models was used to gauge the performance of the respective Bound and Collapse Bayesian reject inference technique to resolve the selectivity bias. For the weak inferred sample, a misclassification of 9% was achieved whilst for the strong inferred sample, a 21.23% misclassification was got, implying that the inferred strong selection model had a higher misclassification than the weak one. These samples are referred to as censored samples, samples consisting of only the known-goods-bads applicants.

The misclassification was not bad for both cases (weak and strong selections) Therefore the sub-samples of the approved and the inferred rejects respectively were combined to create respective AGB samples, which were not truncated but were random samples on which respective weak and strong final CRM models were eventually built. These were the samples on which selectivity bias issue was resolved, restoring the expected randomness character of a representative sample, a basis for any statistical inference. The resultant weak and strong AGB samples were assumed better representation of the TTD population of the SME loan applicants at ZimSME bank. The respective final CRM model developed from respective AGB samples was applicable to the whole of the TTD potential SME loan applicants at ZimSME bank. Using the same procedure as for the preliminary model development, some initial characterization of the respective AGB samples was done prior to final model development (weak and strong).

6.9 INITIAL CHARACTERISATION OF AGB SAMPLE

After the once missing credit quality for the rejected SME loan applicants were imputed, initial characteristic analysis and statistical modelling procedures were carried out to generate final risk profile for the respective final CRM models. There was no limit to characteristics selected in the preliminary risk profile, as some characteristics became weaker and some stronger after incorporation of imputed (inferred) missing credit quality.

This implied that variable selection was repeated, exploring the respective AGB samples. Unlike the preliminary model development, which was constructed on the KGB development sample, the final models were derived after performing initial characteristic analysis and running the backward logistic regression onto the respective AGB samples. The resultant logistic regression models, parameter estimates, and model performance statistics were the major consequential outcomes from the post-inferred samples of the respective AGB samples. Respective resulting risk profiles are given in Table 6.5.

Table 6.5: Risk profiles for the AGB development samples (weak and strong selections)

Variable (Weak Selection)	Variable (Strong Selection)
Liquidity **	Liquidity***
Creditors days*	Net profit margin***
Net profit Margin ***	Creditors days**
Asset size*	Asset size**
Age***	Length of relationship
Income***	Income***
	Age ***

6.10 FINAL MODEL DEVELOPMENT BASED ON THE AGB SAMPLE

Having obtained respective risk profiles for the weak and strong AGB samples, logistic regression was run on both to obtain respective final CRM models. The models were achieved by using backward elimination logistic regression model building technique. Respective models are given in Equation 99 and Equation 100. To choose the final model between the two (2) candidate CRM models, we evaluated both models based on misclassification and model strength as well as using the Akaike's Information Criterion (AIC). The AIC penalizes for the addition of parameters, and

Therefore selects a model that fits well but has a minimum number of parameters, that is, a model which is simple and parsimonious. The model with the lowest AIC is a "better" model between the two (2) models specified for the data at hand. The AIC focuses on the strength of evidence and gives a measure of uncertainty for each model. Measures from the confusion matrix and F1-scores respectively were also used in line with the ZimSME bank policies with SME financing involvement.

Equation 99: Final CRM Model for the Weak Selection

$$\boldsymbol{P(y_i = 1|x_i)} = \frac{e^{-39.766-24.878x_1-2.500x_2+3.068x_3+0.168x_4+0.029x_5+0.706x_6}}{1+ e^{-39.766-24.878x_1-2.500x_2+3.068x_3+0.168x_4+0.029x_5+0.706x_6}}$$

where

$x_1 - net\ profit\ margin$, $x_2 - liquidity\ ratio$, $x_3 - income$, $x_4 - age$, $x_5 - creditors\ days$ and $x_6 - asset\ size.$

Equation 100: Final CRM Model for the Strong Selection

$$\boldsymbol{P(y_i = 1|x_i)} = \frac{e^{-38597-28.633x_1-2,169x_2+0.160x_3+3.076x_4-0.060x_5+0.036x_6-0.685x_7}}{1+ e^{-38597-28.633x_1-2,169x_2+0.160x_3+3.076x_4-0.060x_5+0.036x_6-0.685x_7}}$$

where

$x_1 - net\ profit\ margin$, $x_2 - liquidity\ ratio$, $x_3 - age$, $x_4 - income$,

$x_5 - length\ of\ relationship$ and $x_6 - creditors\ days$ and, $x_7 - asset\ size.$

6.10.1 Misclassification

The misclassification statistics were employed to assess the predictive prowess of the two (2) final models (weak and strong). For operational use of this statistic, a maximum level of acceptable approval rate of 50% was chosen as a "cut-off", equivalent to 500 cut-off score. Loan applicants whose CRM score was above the set "cut-off" point were declined loan and tagged as potential defaulters. The AIC focuses on the strength of evidence and gives a measure of uncertainty for each model. These measures for misclassification were derived from respective Confusion Matrices in Tables 6.6 and 6.7. The measures included accuracy, error rate, sensitivity and specificity.

Table 6.6: Confusion matrix for "Weak" model

	PREDICTED		
ACTUAL	**GOOD**	**BAD**	**TOTAL**
GOOD	289	8	297
BAD	11	46	57
TOTAL	**300**	**54**	**354**

Table 6.7: Confusion matrix for "Strong" model

	PREDICTED		
ACTUAL	**GOOD**	**BAD**	**TOTAL**
GOOD	290	7	297
BAD	9	48	57
TOTAL	**299**	**55**	**354**

A good final CRM model was the one where the "true" cases are maximised, and "false" cases are minimised. The measures were calculated as follows in Table 6.8.

Table 6.8: Misclassification measures for the models

Measures	Weak Model	Strong Model
$Accuracy$	0.9463	0.9548
$Error\ rate$	0.0537	0.0452
$Sensitivity$	0.9633	0.9700
$Specificity$	0.8519	0.8725

Based on these measures, if the bank were to decide to strategically maximize the rejection of bads, with the aim to reduce losses (risk averse) therefore would choose a CRM model that maximizes specificity. If the bank, like ZimSME bank, is a risk-taker to get a higher market share by even approving some bads for credit services, that is, the bank chooses to minimize the rejection of goods by choosing a final model that maximizes sensitivity. The specificity and sensitivity statistic measures are used as decision criteria to achieve business goals of the bank through the choice of a corresponding model that achieves those goals. These are important misclassification measures. Based on misclassification and business strategy of ZimSME bank, the better CRM model was the weak selection model, which is also confirmed by the AIC measure. The weak selection model had AIC of 96.93 while the strong selection model had AIC of 126.64.

Table 6.9: AIC values across models developed

Model	Preliminary Model	"Strong" Model	"Weak" Model
AIC	207.78	126.64	96.93

Table 6.10: F1-Scores for the Weak and strong Models

Measures	**Weak Model**	**Strong Model**
$Precision$	0.9733	0.9700
$Recall$	0.9730	0.9764
$F1 - Score$	0.9681	0.9764

For model the strength, the F1-score statistic was used to make a choice between the weak and strong CRM models (Table 6.10). Based on the business and operational goals of ZimSME bank and as well as the level of misclassification and model strength measures, the weak CRM model was found to fit well in all respects for the final model for the case-studied bank. Therefore, the final CRM model for ZimSME bank was weak selection model (Equation 99).

6.1 VALIDATION OF THE FINAL CRM MODEL

When the final model was built and chosen, the next procedure was validation. The validation process was carried out to confirm whether the developed CRM model was serving the purpose it was built for. Was the final model applicable to the TTD population of SME loan applicants? The other purpose of validation was to check whether the resultant CRM model was not over-fitted. In our case the CRM model was developed on the 100% of the portfolio which constituted the development sample and no pre-selected holdout sample was set. This was done because the loan portfolio for ZimSME bank was small (354).

The holdout sample used for validation process was an arbitrary 80% of the working sample. This was a process whereby the distributions of measured goods and bads across the development and holdout samples were comparable. To carry out the comparisons, some additional measure of classification power was applied. The Receiver Operation Characteristic (ROC) was applied on the final model. This is a plot of the true positive rate (TPR) against the false positive rate (FPR) for different cut-off points of a diagnostic test. In the case of CRM model, the area under the curve (AUC) of the ROC is a good measure of the classification power of the model. Figure 12 shows the ROC for the final CRM model with AUC equal to 0.9782 indicating strong classification power of the adopted model.

Area under the curve: 0.9782

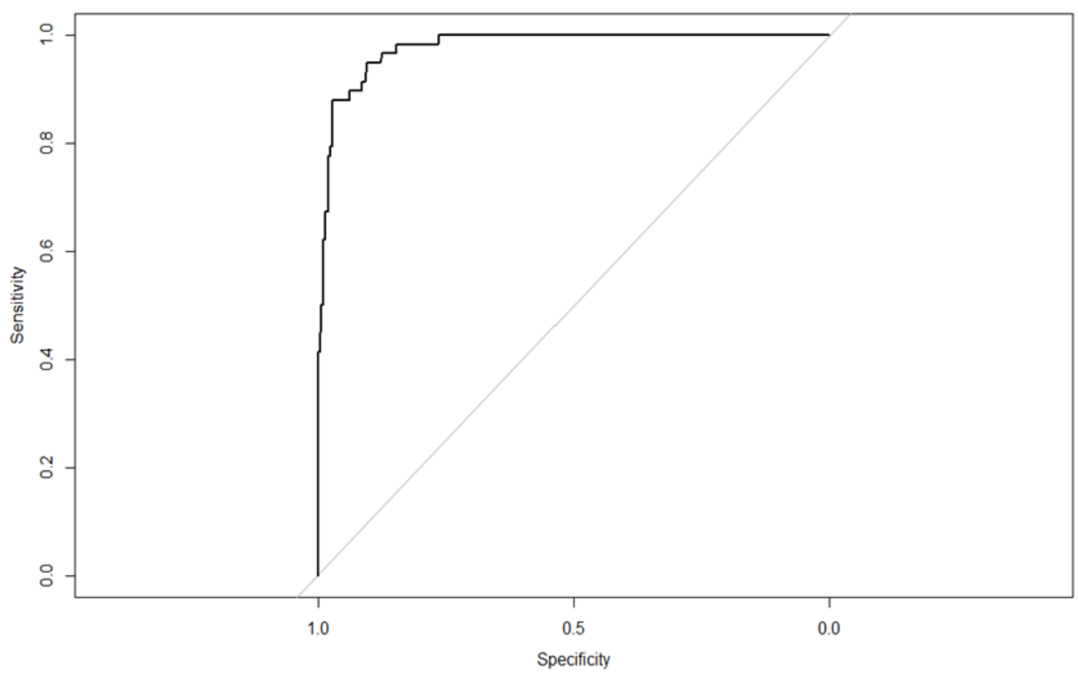


Figure 6.1: ROC for the weak selection model

6.2 STRESS TESTING

Typically, CRM models are developed on static data, often collected at the time of loan application and during a defined performance window. As result for a CRM model to be versatile in out-of-sample, there is need to credit risk stress test the final adopted model. In fact, stress testing of a CRM model is an essential element of the Basel II framework, and is formulated by making explicit reference to economic cycle. Stress testing is a process of determining effect of a change to a portfolio due to extreme realistic events (sensitivity analysis). Generally, considerable scepticism with respect with the merits of a sophisticated CRM models arise as a result of large number of write-downs in the banking sector and financial distress banks experience during financial crisis. This could be caused by risk factors adopted in the model and the existing parameter calibration, which are dependent on historical data. Consequentially such limitations eventuate in severe underestimation of risk due to rapidly changing market conditions. According to the Basel guidelines, CRM models must be complemented by independent risk sensitivity analysis and stress testing for robustness.

Stress testing as a process of determining the effect of a change to a portfolio due to realistic events, there was need to assess the eventual CRM model adopted for ZimSME bank. The types of stress tests include sensitivity analyses and scenario analyses among other tests. For CRM models, the predominant methods are sensitivity analyses and scenario analysis. Sensitivity analysis involves assessment of the impact of a large movement on a single factor or parameter of the model. This method basically is used to evaluate a CRM model's effectiveness of potential hedging strategy. On the other hand, scenario analysis looks at a full representation of future situations to which the portfolio may come across or exposed to. This means looking at simultaneous movements of a set of factors to assess the resultant scenario thereby getting better understanding of current or future situations.

Our final SME CRM model at ZimSME bank was stress tested for robustness. So far, the model has passed the statistical and classification power tests, and ZimSME bank could not adopt the model until, according to the Basel Capital Accord, the stress testing was done. In the model there is no any economic risk factor statistically picked. The model was supposed to be stress tested at the extreme level of economic factor to find out how the final model would have behaved in response to such extreme events. We would have tested on whether the classification power of the model would have been heavily affected by the economic risk factor level being at the extreme lowest in the country and would have done the same when the economic factor level was at the extreme highest.

In the development of the AGB we used the Bound and Collapse Bayesian reject inference to impute the missing credit quality for the rejected applicants. The BC embeds a Dirichlet distribution which uses two (2) priors as conjugate priors. In the development of the final model we used the prior distribution $\alpha_{ij} = 0$ for simplicity reasons, therefore there was need to find whether there was any effect on the final model if the $\alpha_{ij} = 1$ prior distribution was used. The results of the sensitivity analysis are given in Table 6.11 showing that the choice of the prior distribution function does not affect the performance of the model developed.

Table 6.11: Choice of prior distribution

	$\alpha_{ij=0}$	$\alpha_{ij} = 1$
Accepted	303	295
Rejected	51	59

These stress tests performed implied that the model is adoptable although further tests could be done to assess the robustness of the model in response to population drift and to rapidly transforming market conditions.

Using real data and through rigorous statistical procedures a non-random KGB sample was transformed, to obey the basic sampling, into an AGB samples on which the final CRM models were built. The transformation principles were possible through a theoretically based reject inference technology, the Bayesian theoretically model-based BC methodical approach was employed. Missing data is no longer considered a nuisance but play a significant part in the final inference especially if the pattern of missingness is MNAR, which is non-ignorable. Credit quality of rejected loan applicants is normally missing not at random (MNAR) in CRM domain. Using the generalised beta distribution, Dirichlet distribution, the BC can estimate the credit quality of the rejected thereby transforming the KGB sample into a random AGB sample. Prior to the application of reject inference after, variable selection and reduction is performed through coarse and fine classing of predictive variables. Using metrics of WoE and IV, initial independent variables are assessed of their explanatory power on the credit risk outcome of each loan applicant. This variable reduction help develop parsimonious preliminary and final models which were validated for goodness of fitness and applicability.

To validate the final models, different cut-off points based on bank rejection policy varying levels were proposed, and their respective effect on model's predictive accuracy was studied in a confusion matrix regarding Type I and Type II errors. ROC curve plots and AUROC analysis were used for assessing model performance and validating models' predictive accuracy respectively. The capability of the weak and strong selection models was examined to forecast PD through a 2-year performance window (horizon) based on different cut-offs groups.

Overall, the final CRM models demonstrated better performance than the preliminary model since the AGB sample included both good and bad applicants or respective cut-off point was selected so that an acceptance rate was relatively low, otherwise model's predictive accuracy would decline. Through the AIC metric the weak selection model presented better predictive accuracy with higher acceptance rates and performed better when it was used to predict default in the same performance window. When considering a loan to a company, a bank wants to know the likelihood default for duration of loan.

CHAPTER 7 RESEARCH RESULTS

7.1 INTRODUCTION

The primary objective of this thesis has been to develop an empirical CRM model specifically for SMEs, a tool which can adopted by banks that venture in the non-traditional lending market. To achieve this goal, the research was set to seek solutions to the two (2) research questions enunciated in Chapter 1. For robust achievement to this goal, which has attracted increased attention of not only academic researchers and financial practitioners but also regulators, the research has been split into mutually exclusive interconnected two (2) phases. The respective results of the two (2) phases have been summarised in chapter 5 and chapter 6 respectively. The research findings of the thesis are the responses to the research questions and are reported herein. In fact, this chapter of the thesis is an in-depth discussion of the overall findings from both phases, thereby providing an eventual holistic and comprehensive development of the SME CRM model from the perspective of the bank in accordance with the dictates of Basel II Capital Accord. In that respect, the findings are entwined to the objectives to answer the research questions therein.

The contributions of the thesis are explored in terms of theoretical and practical connotations for all the stakeholders in SME financing web. Varum and Rocha, (2013) report that the SME sector is gathering unprecedented momentum in today's global economy because of its recognised role as an economic stabilizer during difficult periods, making Zimbabwe scholarly suitable for the study. As a result of fierce competition in retail and corporate lending segments coupled with economic downturn, banks in Zimbabwe are fast moving into the untapped high potential for the growth of the SMEs lending market. However, banks are still sceptical about how to tame the credit risk that characterise the sector, leaving many questions unanswered, such as how banks should quantify SME credit risk and what CRM models they should adapt in order to derive profitability from the SMEs portfolios. The majority of the current CRM tools apply only to retail and corporate assets of the bank, whilst a few models have been designed for SMEs application (Altman & Sabato, 2007). In Zimbabwe, CRM for SMEs is a grey area, a gap this thesis attempts to fill.

7.2 KEY RESEARCH FINDINGS AND CONCLUSION

The problem statement and research questions have been re-visited to highlight the key research findings. In literature, the problem of SME financing has been studied either from the perspective of SMEs or the banks, in most cases recognition is given to the banks (OECD, 2007; Ramlee & Berma, 2013). Again, the focus is mostly on constraints faced by SMEs without querying the ability and capacity of the same SMEs to borrow from banks. In order to determine the extent at which the banks are involvement with SME financing in Zimbabwe, a double perspective approach was used in this thesis for the preliminary phase I. Sample survey approach was used to gather primary data from the two (2) indispensable economic players (banks and SMEs). This dovetailed into Phase II which was about empirical development of a CRM model for SMEs from real credit data from a bank in Zimbabwe. The phases are intrinsically bound to achieve a holistic understanding of the issue of the SME CRM gap in Zimbabwe and other economies at large.

In Phase I, data analysis was wholly descriptive statistics of bank and SME surveys' data. The underlying objective was to identify factors that influence bank-SMEs economic relation, with the aim to measure the extent of the financial interaction between them. The Phase II consists of the empirical development of a CRM tool specifically for SMEs on credit data collected from a bank, as a case-study in Zimbabwe. In this phase, the thesis takes the perspective of the bank only because it is the banks that currently are holding the knowledge regarding the CRM processes; therefore, it has been seen necessary to take that stance.

The data analysis in Phase I (Chapter 5) confirmed that the extent of banks involvement with SME financing was determined by factors originating from both banks and SMEs. From SMEs viewpoint, collateral and financial information provision were picked as factors inhibiting bank involvement. Limited banks' involvement with SMEs financial constrains the latter's growth and development. Yet globally, the SME sector is recognised as a poverty alleviation tool through employment and income generation (Holmes, Hutchinson, Forsaith, Gobson & McMahon, 2003; Rao, 2003; Shinozaki 2014). The same is what the government of Zimbabwe considers them to be as the sector is a major constituent of the private sector, a conduit to socio-economic growth of the country (Nyamwanza, 2014; Wangmo, 2015). Berger and Udell (2006) report that SMEs are financially constrained, thereby hampering their growth and development and the hardest hit are SMEs in emerging economies like Zimbabwe. To achieve the ultimate goal of this thesis, the following key research questions were answered:

- What are the factors that cause SME financing constraints from the perspectives of SMEs and the banks in Zimbabwe?
- How do banks measure credit risk of SMEs in Zimbabwe?
  - ✓ How do we minimise selectivity bias present in the KGB sample?
  - ✓ What are the criteria used by banks when measuring credit risk of SMEs in Zimbabwe?

7.3 PART I :BANK INVOLVEMENT WITH SMES IN ZIMBABWE

The guiding integration process to substantiate and validate the final findings was triangulation, which incorporates evidence from varied sources or methods in the economic web of SME financing (Creswell & Plano-Clark, 2007; Ivankova, Creswell & Stick, 2006; Teddlie & Tashakkori, 2003). The qualitative data analysis in Chapter 5 revealed the extent to which banks are involved with SME financing against the backdrop of conventional wisdom that banks in Zimbabwe do not do business with SMEs. In fact, the analysis revealed the challenges and difficulties encountered from the perspectives of both supply and demand sides of the financing equation. The results of this phase formed the basis for the empirical development of a CRM model for SMEs in Phase II and the final analytic results of the thesis.

7.3.1 Bank Involvement with SME financing in Zimbabwe Analysis Results

The qualitative analysis of interviews with bank managers, credit managers and loan officers, on one hand and with SME owners and administrators on the other revealed factors that determine the extent of banks involvement with SME financing. These analyses provided complementary and corroborative findings about the extent to which the two (2) economic players interact with each other for the development of the national economy and entrepreneurship. This is contrary to a similar survey done by World bank in Zimbabwe in 2012 on Micro Small and Medium Enterprises (MSMEs) in Zimbabwe. The World bank survey focused on obstacles faced by MSMEs only without querying the ability and capability of MSMEs to borrow from banks. This is consistent with literature which confirms that the issue of SMEs financing has always been looked at either from the perspective of banks or SMEs (OECD, 2007; Ramlee & Berma, 2013). Wangmo (2015) notes that literature on SME financing issue from the perspective of both economic players is lacking thereby limiting the integration of the two (2) indispensable components of the delicate issue. Both surveys were carried out with the aim to answer the research question; What are the factors that cause SME financing constraints from the perspectives of SMEs and the banks in Zimbabwe?

7.3.2 Research question 1

The first research question explored the banks involvement with SME financing in Zimbabwe. It was answered based on data extracted from bank and SMEs surveys. The qualitative data analysis established that the extent of banks involvement with SME financing is determined not only by factors originating from the supply side but also by factors originating from the demand side. The constraint involvement of banks with SME financing was discovered to be caused by information asymmetry and credit risk on the demand side (SMEs) and the imposition stringent loan conditions in the form of exorbitant collateral and demand for audited financial reports by banks. In accordance with analysis of bank survey data, lack of SMEs sector knowledge (information opacity), insufficient business planning, and small credit size were pointed as serious challenges for the banks (Appendix A). These factors incapacitate banks to measure the creditworthiness of their SME portfolios in protection of depositors' money.

The study revealed anecdotal evidence that there is great information opacity in Zimbabwean SMEs as manifested through poor business planning and unaudited financial statements, as a result of the sector's poor financial record keeping and incompetent financial management. The rampant information asymmetry and inadequate knowledge of the sector by banks implied low loan provision to the sector by banks, thereby increasing credit risk. Consequently, information asymmetry and credit risk lead to adverse selection and severe credit rationing from the supply side. This, in turn, exposes the banks to high credit risk when doing business with the SMEs sector, since it is insurmountable to measure the SME credit quality (St-Pierre & Bahri, 2011). In order to protect depositors' money and their intermediary role, banks resort to SME unfriendly and strict lending mechanisms to minimise the adverse effects of non-performing loans. Some of the measures include among others; requirement of exorbitant collateral, shorter term loans and prohibitively high interest rates. In response to strict lending mechanisms adopted by banks to safeguard depositors' money, SMEs sometimes indulge into risky business ventures to commensurate the high cost of finance due to high interest rate and collateral (Brent & Addo, 2012). In turn, this moral hazard affects banks after the initiation of the loan transaction as this will culminate in loan non-performance by the SMEs.

Conclusively, banks are risk averse when it comes to dealing with SMEs due to information asymmetry, therefore adopt stringent measures to protect their ultimate commercial goal of profit generation. On the other hand, such measures counter adequate bank involvement with SMEs financing issue. For SMEs borrowing from banks comes with prohibitively additional costs in the form of interest rate and collateral. As a result, some SMEs do not borrow from banks but resort to internal funding for their investments. Although not adequate, internal financing is less costly as compared to bank loan. This affected banks involvement with SME finance from the SMEs viewpoint. Therefore, the need to device the means to lessen the adverse effects of information asymmetry between banks and SMEs and the imposition of stringent lending mechanisms.

7.3.3 Fulfilment of Contributions

7.3.3.1 Theoretical Contributions

Most of the studies on bank involvement with SME financing are limited to middle-income countries such as in Latin America and Europe and with scarce applicability to emerging economies like Zimbabwe which are characterised by low economic and financial market developments. The Zimbabwean economy is small, whose economic activities are handled majorly by SMEs and supported by a simply structured small financial services sector. From literature, there are scarce academic bank-SME surveys available in reference to emerging economies like Zimbabwe, thereby creating a gap of knowledge. To the researcher's knowledge, this double perspective bank-SME study on banks involvement with SME financing is the first academic study carried out to calibrate the extent of interaction between the two (2) indispensable economic players. This study was exhaustive and corroborative in that the SMEs financing issue was looked at from the perspective of both SMEs (demand side) and banks (supply side) and was implicitly built on economic theory of information asymmetry between the two (2) sides in one economy. This was different from the 2012 World Bank survey on MSMEs in Zimbabwe which was done from the perspective of the SMEs only and focused on obstacles faced by the borrowers. Therefore, the study contributes theoretically on existing literature on banks' involvement with SMEs in middle-income economies (de la Torre *et al.*, 2010) and applies it to the context Zimbabwe, an emerging economy with substantially social fabric.

Prior studies on bank-SME financing issues have been either from the perspective of SMEs' borrowing difficulties or from the perspective of the banks. The integration of these two (2) surveys is lacking in most SME financing literature. Therefore, the bank-SME survey approach was a holistic approach to calibrate the extent at which banks are involvement with SME financing by examining both SMEs and banks' perspectives through corroborating surveys. By doing so, the corroborative research findings were with substantial weightage by overcoming any bias and deficiencies that may have arisen if only one of the key components of SME financing gap was queried. The triangulation process generates anecdotal evidence and exhaustive findings as the perspectives of SMEs and the banks were simultaneously considered.

The academic contribution was that the primary data was collected through a bank level survey on one side and SMEs survey on the other side. The data collected from these surveys corroboratively revealed a real picture of the extent at which banks are involved with SME financing in Zimbabwe at face value rather than inferring from the archival financial data, a method used in most prior studies and in phase II. The nature of the primary data enabled validation, comparison and corroboration of data from either survey, while prior studies data was either from bank survey or SME survey.

7.3.3.2 Practical Implications

The ultimate goal of this bank-SME study was to nurture the growth and development of SMEs in Zimbabwe, a goal with policy implications. Literature confirms that banks involvement with SME financing issue is a conspicuous daunting challenge confronted by emerging economies like Zimbabwe. Therefore, the study of querying the extent of banks involvement with SME financing and consequentially the issue of SMEs' inaccessibility to bank loans produces a huge socio-economic significance. The study provides anecdotal evidence and better picture of SME financing issue from all the stakeholders, SMEs, banks, and government and non-government agencies involved in SME development particularly at the policy level (Wangmo, 2015). If such practical implications of the study are reviewed and put into action appropriately, growth and development of SMEs is assured to the benefit of the economy at large. In emerging economies, characterised with small and growing financial services sector, as seen in Zimbabwe, the leading role of the government is fundamental for the conducive development and operation of SMEs. Therefore, findings of this study have major contribution for policymakers involved in the growth and development of SMEs to help create a conducive and supporting financing environment for both SMEs and banks, be it at government or non-governmental levels. The conclusions drawn from Zimbabwean's case study may be applicable to sub-Saharan Africa and any other emerging countries with small and weak economies.

7.4 PHASE II: SUMMARRY OF RESEARCH FINDINGS

After the global financial crisis and the advent of Basel II Accord, CRM has increasingly become indispensable in banks' risk management systems. In response, banks all over the world have devoted much of their meagre resources to developing IRB CRM models to better measure their financial risks, customer profitability analysis, risk-based pricing, active portfolio management and capital structure decision. Bank regulators, playing their supervisory role, encourage such prudence for the soundness of the delicate financial industry. Guided by the Basel II, such efforts have culminated in banks and FIs moving swiftly into the field of CRM modelling especially for non-traditional assets like the SMEs asset.

On that background, this thesis has explored CRM modelling for SMEs in Zimbabwe, the case of an emerging economy. Literature acknowledges that SMEs CRM models belong to a class of models that lie between between the corporate CRM and retail CRM models. But, research on this domain, SME CRM modelling is scarce irrespective of the socio-economic significance of the sector in any national economic space, according to Berger and Frame (2005). Despite of the fact that, this is the sector which has the greatest inaccessibility to credit due to problems emanating from adverse information asymmetry, opacity and poor financial handling, among other challenges as found in Phase I of this thesis. The ultimate goal of this phase and the thesis has been to develop an empirical development of a CRM model tailored for SMEs. This is a topic which has aroused the interest of not only academic researchers and financial practitioners but also regulators in the CRM domain. This is because in all economies, developed or developing, the majority of their economic activities are done by SMEs, therefore, consistent and apt supply of credit to this sector will definitely translate into global economic boom.

7.4.1 Background of the Case-study Bank

A bank with pseudonym ZimSME is a bank in Zimbabwe which is attempting to enter into SME lending market as a result of the 2007 global financial crisis and heightened competition in the retail and corporate banking. In pursuance of profits, the bank is making inroads into the non-traditional lending market, SMEs market. The move calls for innovative banking. Like many banks, ZimSME lacks strategic capacity and appropriate business models to adequately deal with this emergent market, (Al Baz, 2017; Wangmo, 2015). In pursuit of that goal, ZimSME bank has decided to adopt the IRB approach to tame the envisaged high credit risk which characterises the SMEs lending terrain. Therefore, offered its SME loan portfolio for this thesis as long ethical issues with data collection and analysis were strictly observed. The researcher signed an ethical code of conduct to uphold with the entailed ethical requirements of bank data. Framework of Basel II in relation to CRM.

The Basel II Capital Accord is the most recognised and significant transformation in CRM domain in the legal and economic framework of bank lending business. This has given way to proliferation of measures and technologies to address pertinent issues to do with credit risk components, risk exposures, IRB approach, the potential impact on banking systems and practical implementation issues. Its major impact is linking the credit risk associated with borrower to the amount of regulatory capital that is required. For the first time, banks and FIs are allowed to select their CRM approach from either the standardised approach or the IRB approach as explained in Chapter 2. Standardised approach is meant for FIs which lack capacity to internally develop own CRM tool, and therefore seek for ratings provided by ECAIs. In such a scenario, it is the prerogative of the responsible regulator to link the agency ratings to capital requirements by setting an average PD. On the other hand, the IRB approach gives banks and FIs room to develop their own CRM models to quantify credit risk through internal calculation of PDs. In sum, the Basel II Accord recognises three (3) procedures for linking credit risk grade to a PD:

- Historical average PDs based on internal default experience;
- Mapping internal grades to external rating agencies;
- Statistical default models.

Contemporary banks either apply the standardised or the IRB approaches, whereby the PD for a single firm is a function of its CRM, either externally or internally calculated.

Limited literature (Al Baz, 2017, Chen & Astebro, 2012; Horstedt & Linjamaa, 2015) has discussed on the issue of banks implementation of the Basel II Accord on SMEs and how to facilitate SMEs access to finance. The Basel II induced revolutionaries are calling for a paradigm shift in bank lending process, as banks currently offer more loans to matured segment of the economy at the disadvantage of SMEs, unrated enterprises in Basel I Accord. Literature (Lin, 2007; Altman & Sabato, 2007; Wangmo, 2015) suggests that under Basel II, SMEs credit risk capital requirements are likely to fall as banks increasingly use of IRB as a basis for loan pricing decisions, thereby inducing drastic cut in the cost of finance.

7.4.2 Research Question 2

In addressing the second research question, the banks are seen as the ones holding the knowledge regarding CRM processes, therefore the analysis is limited from the perspective of the banking sector only. The Basel II Accord's major improvement was that banks were given the autonomy to develop internally their own CRM tools to evaluate the creditworthiness of their own clientele (Antão & Lacerda, 2011). Since the 1990s, banks have been hit hard by heavy losses as a result of upsurge of non-performing loans, prompting the need for objective and accurate CRM (Zhou, Lai & Yu, 2010). It has also been noted that banks suffer the most even when a small miscalculation of credit risk is made during CRM process, a process when banks measure the credit quality of the firm (Horstedt & Linjamaa, 2015). In that regard, to maintain such a perspective, the credit data to develop the eventual CRM model was gathered through interviews with the management of a bank offering credits to SMEs in Zimbabwe and from an SME loan portfolio of the same bank.

From the analysis of interviews with credit management of ZimSME bank it was discovered that SMEs' credit accessibility in Zimbabwe was a function of the enterprise, owner and financial characteristics. This discovery is consistent with the outcomes of Phase I and with literature (Al Baz, 2017; Hostedt & Linjamaa, 2015). Another noteworthy finding is that transformation (fine and coarse-classifications) of predictor variables leading to improved variable selection for eventual parsimonious model building. This provided the first step to analyse the data for further model construction. Through in-depth scrutiny of the 26 characteristic variables, the bank is currently using to assess its clientele for loan granting, the list was cut to only 6 factors which is consistent with literature (Siddiqi, 2006; Kennedy, 2013). The eventual factors were found to be asset size, and credit days (enterprise); income and age (owner) and net profit margin and liquidity ratio (financial). The effect of these characteristic variables on SMEs' eligibility to bank loans was investigated through the application of BC reject inference technology to develop an AGB sample on which the final logistic regression model was built. The key finding of Phase II was that an

SME’s PD was the critical metric in determining its accessibility to bank loan. This finding has a significant financial implication to banks to resolve the SME credit risk problem in emerging economies such as Zimbabwe.

The primary objective of Phase II has been to develop an empirical tool for SMEs CRM. In that respect, the study provided empirical evidence and support for a binding and strong relationship between an SME’s financial capacity and bank credit accessibility. Collateral and owner’s equity were not among the credit predictive characteristic variables in the final logistic regression model (Equation 99), a result inconsistent with Wangmo (2015). This is not a disparity but is consistent with the Basel II Accord which rules out generic attribute of IRB CRM models. In view of SMEs’ high information asymmetry and credit risk, traditionally banks impose heavy lending mechanisms (high collateral, high interest rate and sizeable owner’s equity) to mitigate exposure to non-performing loans. To the contrary, both collateral and owner’s equity requirement have not been found as credit risk drivers to mitigate exposure to non-performing loans. Using CRM models, banks may extend loans to SMEs without relying on the collateral and the size of the owner’s investment in the business. Phase II provided anecdotal evidence that SMEs in Zimbabwe and emerging economies at large can access bank loans without the imposition of the current inhibiting lending mechanisms.

When putting the results in Chapter 6 into relation to the research gap established, the lack of research regarding CRM models used by banks when measuring credit risk of SMEs in Zimbabwe was evident. It is therefore argued, in this thesis, that with this new knowledge regarding the CRM model developed for ZimSME provided in equation 99, the research gap has been minored. In emerging economies, this thesis provides the basis for further exploration of complex and advanced CRM technologies to still minor the research gap since SMEs are a developmental asset class characterised by intense growth dynamism. They have potential to grow from a very small to a large sized enterprise, if consistently well-financed. There is a great potential for FIs to derive higher returns on assets by applying accurate and well-performing CRM models tailored for these SMEs. The CRM model developmental procedure in Chapter 6 remains the same and only the criteria that are used in the measurement of credit risk are bank specific so change among banks. The thesis develops a statistical CRM model to assess the credit risk of Zimbabwean SMEs with ambitions to grow in their businesses into national or international corporations. Such a CRM technology can be used to build large lending portfolio for this important SME sector to the benefit of the national economy. The CRM model developed can also play a vital role in developing large databases for accurate modifications of future CRM models as well as the long-term development of these SMEs.

The first sub-question of the second question was formulated, “How do we minimise selectivity bias present in the KGB sample?” In terms of answering this question it was established that theoretically supported reject inference techniques do better to resolve the inherent problem of selectivity bias common in CRM domain. Using bad rate, as metric for assessing the classification performance of a CRM model, the decrease of bad rate from the preliminary model (equation 98) to final model after applying BC (equation 99) demonstrate the superiority of AGB sample over KGB sample which is truncated and non-random. The adverse effects of selectivity bias are far-reaching in that, it compromises the ground on which statistical principles are built, thereby breeding biased estimates hence wrong selection decision for loan granting. Horstedt and Linjamaa (2015) note that FIs suffer the most even when a small miscalculation of credit risk is made during loan process, a process when FIs decide on who to grant loan or not at application stage. If selectivity bias is not taken care of, CRM models will not serve the role they are developed for as they are applied on TTD population not on a “cherry-picked” population.

In this thesis, the BC, was the answer to minimisation of selectivity bias. From literature, conclusions regarding other reject inference techniques (re-weighting methods, methods of extrapolation, Heckman’s bivariate two-stage model) are discouraging because they are hinged on tenuous assumption (Crook & Banasik, 2004; Nguyen, 2015; Chen & Astebro, 2012). The BC, unlike other reject inference techniques, uses Bayesian statistics theory which builds on a model of missing data developed by Rubin (1976). Its strength is in its ability to incorporate the impact of the data source through imputation of missing data of the dependent variable, in this case, credit quality of the rejected loan applicant, based on the estimated missingness probability function. Again, the BC allows exogeneous supplementary information about the rejected loan applicants to be flexibly incorporated into model (Sebastian & Ramoni, 2000; Chen & Astebro, 2012; Nguyen, 2015). In sum, the BC is a model-based technology, theoretically hinged and supported, unlike other reject inference techniques built on tenuous assumptions. It is where it extracts its advantage and utility value over other reject inference methods. It is also implementable with any data as demonstrated in this thesis.

The sub-question of the second question was formulated, “What are the criteria used by banks when measuring credit risk of SMEs in Zimbabwe?” To answer this question, a list of the criteria that were used in the empirical development of the CRM model was established to be used by ZimSME bank in SMEs CRM. The list is as presented in Table 6.5. (Weak Selection). As earlier mentioned in literature, there is little empirical work available concerning the means and methods employed by banks when measuring credit risk of SMEs. Therefore, the researcher also wanted to contribute in terms of clarifying the CRM process such that SME owners understand what factors that they are measured upon when seeking credit and Therefore become better prepared when seeking a bank loan from ZimSME bank. Therefore, reducing information asymmetry which is rife between banks and SMEs. By demonstrating empirical evidence of how the CRM process takes place up to the actual CRM model, the researcher is bound to believe that this aimed sub-research question has been successfully answered.

7.5 CONTRIBUTIONS OF PHASE II

The primary goal of this thesis has been to develop effective approach for SMEs CRM. This is a goal which draws attention of not only academic researchers and bankers but also regulators in the area of CRM. Despite that this thesis has a focus on a bank in Zimbabwe, the developed CRM model in Chapter 6 (Equation 99) can also be applied to other banks in many other economies, given that the SMEs lending challenge is global. The aim of the thesis is not to explain the current CRM processes employed by ZimSME bank, but rather to provide an initial foundation of fact as to build further research upon through exploration. In fact, the SME CRM model, a product of this thesis is applicable to other banks in Zimbabwe and other countries because the procedure followed to achieve it is pursued. This thesis makes several contributions to both academic and business spheres.

Academically the thesis designed a CRM model in accordance with Altman and Sabato’s (2007) model to quantify PD for Zimbabwe SMEs which carry different characteristics from SMEs in the developed economies. To the knowledge of the researcher, this is for the first time that an IRB CRM model has been developed for Zimbabwe SMEs. From a business viewpoint, the empirical contribution of the thesis is that the developed CRM model can be adopted for SMEs in Sub-Saharan Africa region or any other emerging countries with similar setup to Zimbabwe. For successful development of a CRM model for SMEs, credit data is the crucial component, therefore, FIs with sizeable SMEs loan portfolios can develop improved CRM models by just following the procedures outlined in Chapter 6. The application of theoretically hinged reject inference techniques, like BC, to build the statistically random AGB sample prior to eventual logistic regression modelling, leads to more accurate CRM models which dovetails to significant improvement of the risk management of the financial system as a whole. In that regard, empirical CRM tool must help FIs' capability to approve credit to eligible SMEs thereby translating into good returns and minimal losses. The tool must be robust for SMEs such that banks and FIs comply with the Basel II requirements for capital adequacy (Altman, 2018).

7.5.1 Fulfilment of Contributions

7.5.1.1 Theoretical Contributions

The thesis also aimed to make several contributions in terms of new valuable insight regarding the scarcely studied research topic. It is confirmed from literature that there is relatively little empirical research available regarding development of SME CRM models for banks, especially in emerging economies, where the successful development of SMEs is needed most. Hence, the need to make certain theoretical contributions which in turn form the basis for further research. Through answering the research question 2 enunciated in chapter 1 and the application of BC to construct an AGB sample prior to the CRM modelling, produces valuable theoretical contribution to the CRM domain. This provides valuable foundation on which further empirical research would be based giving increased incentive for continuous studies in the CRM domain.

7.5.2 Practical Contributions

By answering the stated research question 2, and consequently fulfilling the second purpose of the thesis at hand, the researcher also aimed to make certain practical contributions towards the Zimbabwean banking sector. Given the empirical findings made in Phase II, the researcher was able to initiate a description of the internal CRM processes and practices employed by ZimSME bank and present a coherent procedure used within these CRM processes. Therefore, the findings of the Phase II of the thesis, contribute with a sound basis of information that may be used as a benchmark against which both the banks and FIs offering SME loans can compare their own practices with.

7.6 SUMMARY

In this chapter, the primary research objective and questions were re-visited and then entwined with research findings. The aim of the thesis was to empirically develop a CRM model for SMEs guided by the Basel II Capital Accord. Prior to the ultimate development of the model, bank-SME surveys were carried out to calibrate the extent of banks involvement with SME financing in Zimbabwe, justifying why a CRM model specifically for SMEs is undeniably necessary. The prior discussion provided anecdotal evidence on the SME financing gap from the perspectives of both the supply side (banks) and the demanders (SMEs). The conventional wisdom that banks in Zimbabwe do not do business with SMEs and that SME lending market was for small banks which use relationship lending technologies was proved null and void. Regardless of the fact that their bank-specific and SME-specific challenges, all banks regardless of size and ownership are into SME financing dedicating resources into developing arms-length technologies and units to handle doing business with the unsaturated lending market. That gave impetus to the development of a CRM tool to help banks and other FIs to enter a lending market they virtually have little knowledge of on the background of the devastating global financial crisis of 2007.

In sum, the results of the preliminary study revealed that banks of all sizes and ownership are involved with SMEs financing regardless of difficulties from either side. It was discovered that SMEs' accessing of bank loans was influenced by the SME specific attributes, its owners' attributes and its financial information provision. The major inhibitors of SMEs' accessibility to bank credit have been found to be collateral and hugely information asymmetry between SMEs and the banks. Information asymmetry breeds all other financing constraints. Start-up SMEs, with little or no historical reputation in the market and an absence of standard accounting practices, are deemed risky investment by banks. In response to the absence authentic financial information, the banks adopted strict lending mechanisms to overcome the high risk involved in SME lending, such as increasing loan interest rates, collateral size and owner's equity in the business, thereby inhibiting SMEs from seeking external funding from the custodial suppliers.

From the SME perspective, owners must work hard, through their association, to provide their financial information in standard format as per bank requirement such that their creditworthiness is measurable at loan application stage. This is possible if they have full knowledge about what financial aspects that they are evaluated on (Al Baz, 2017). Meanwhile, the banks as the custodian suppliers of credit need to realign their mandate and address the specific demands and needs of the SME sector by re-engineering their CRM tools without resorting to exorbitant collateral demands. The facilitating role of the government is paramount in creating conducive business and legal environment by developing bank-SME friendly policy measures that reduces information asymmetry between the two (2) economic players. The preliminary study's final findings concluded that banks are highly involved with SME financing, contrary to conventional wisdom that banks are reluctant to do business with SMEs. To fulfil the self-presenting mandate, banks are investing into arms-length CRM technologies specifically for the SME asset and are setting SME-specific units responsible for the sector financing issue. These findings gave impetus to phase II of the thesis, whose major goal was to develop an empirical CRM model specifically for SMEs, guided by the prudence of the Basel II Capital Accord.

In the modelling phase of this work, the researcher demonstrated the impact of selectivity bias on SME CRM performance and eventually profitability. Selectivity bias has been suggested to pose a sizeable peril to bank profitability owing its inherent implications on either population drainage or biased estimates. The larger the amount of data screened out, the more important to use some of reject inference techniques, as models built on non-random, pre-screened samples become less and less able to represent the complete TTD population of loan applicants. In fact, reject inference technique, BC, adopted in this work is theoretical hinged on Bayesian Statistics theory and considered a case of missing data (Chen & Astebro, 2012; Sebastian & Ramon, 2000). In fact, the BC, a Bayesian imputation procedure, has been used to incorporate auxiliary credit-policy into the estimation of the probability of missingness, unlike other reject inference procedures which ignore the impact of missingness. For successful application of BC, reject inference is considered as a case of missing MNAR, a non-ignorable case. Using real data, BC has demonstrated superior to simply deleting missing data and to the banking industry that suggested reduction in selectivity bias are of economic value. Comparing the preliminary model (equation 98) trained on KGB sample to the final model (equation 99) trained on AGB sample revealed underperformance of the former model as portrayed by high AIC value. In fact, BC provided reduction of the bad rate on the portfolio of accepted borrowers as shown in Table 7.1.

Table 7.1: Bad rates across models

Models	Expert scorecard	Preliminary Model	Final Model (Weak)
Bad rate	20.62%	15.82%	14.69%

The CRM modelling phase gave answers to sub-research questions of research question 2. It has been demonstrated that the bank must not continue to use all the 26 characteristic variables to classify their SME clientele into good and bad categories prior to loan granting but use 6 variables which best describe and explain credit quality of its clientele and reduce the bad rate as demonstrated from the preliminary model to the final model (Table 6.5). Therefore, SME owners have full knowledge about what aspects that they are evaluated on.

The Phase II empirical modelling results have demonstrated that it is possible to quantify credit risk of SMEs and that there is great improvement in the classification power of the logistic model when developed on a random and representative AGB sample of the TTD pool of SME loan applications. Most reject inference procedures fail to perform satisfactorily when credit quality is MNAR, but the modelling phase of this thesis has demonstrated that the BC works well under such data condition, as it is capable of compensating for data MNAR condition unlike other reject inference techniques. The incorporation of the missingness function into BC model gives room for the incorporation of any auxiliary credit related policies to model the probability function of the missingness.

CHAPTER 8 CONCLUSIONS AND FUTURE WORK

8.1 INTRODUCTION

SMEs CRM modelling has seen many successes in recent years, especially in developed economies which have embraced the Basel II Capital Accord as well as in few emerging economies. The modelling is being done in a variety of ways: artificial neural network (ANN), discriminant analysis (DA), genetic algorithm (GA), machine learning (ML), support vector machine (SVM)and data enveloping analysis (DEA) just to mention a few. Despite these successes, the topic modelling is still relatively new and there remains much to be explored especially in emerging economies where such technology is much needed. One of the most noticeable absences from most of the previous work on CRM modelling in general is the statistical development of the model, from where an objective and far-reaching CRM model is developed to help banks in the objective classification of good and bad loan applicants from the TTD population of loan applicants. The use of theoretically rooted reject inference techniques in constructing a random AGB sample for developing a CRM model has been missing in literature. Studies on bank involvement with SME financing have been done piecemeal in various emerging economies but this study queried the Bank-SME relations by interacting with both stakeholders of the SME financing equation.

This thesis has two (2) phases: Bank involvement with SME financing and the SME CRM modelling. The two (2) phases are entwined in that the former phase exposed the fundamental relationship that exists between banks and SME which the latter phase builds on by empirically developing an instrument that can facilitate the smoothening of the constraining co-existence of the banks and SMEs for the eventual prosperity of the national economy. In fact, phase one unravelled the intricacies of the existing relationship between the banks on one hand and the SMEs on the other. Furthermore, traditionally FIs stereotype SMEs as high risk for lending since SMEs do not have sufficient financial and business records as well as riddled with informality and opacity. The relations between banks and SMEs have always been very similar to those between an old couple who constantly blame each other yet must live together. This is true for emerging economies where banks have traditionally dominated financial systems, leaving little leeway SMEs seeking alternative financing to bank loans (Derreumaux, 2009).

In chapter 2, we looked at the theoretical background of the study in so far as the historical perspective of CRM in general, the rise of the need of CRM, the synchronisation of the definitions of a Small and Medium Enterprise (SME) and the support given to SMEs growth by the Basel Capital Accords (Basel I and II). Literature confirms that default/credit risk is one of the oldest financial risks entailed by lending but interest in it by researchers, practitioners and academics was only visible in the 1990's. Before that, bankruptcy prediction studies focussed on single financial ratios until in 1966, when Beaver gave birth to multivariate credit risk forecasting. This was ensued by a proliferation quantitative CRM methodology in the form of the Z-score by Altman (1968), probit, logit, neural networks, genetic algorithm, machine learning, artificial neural networks, support vector machine approaches. The surge of CRM was due to number secular forces which included; global structural increase in the number of bankruptcies, more competitive margin on loans and decline value of collateral in several financial markets. To forecast credit risk banks, use subjective CRM tools in the form of 4C's or 5C's which are expert underwriting tools or qualitative credit measurement tools, which are majorly dependent on "soft" information.

We also looked at the definition of an SME, a definition which is heterogeneous. This current definition which is based on turnover, number of employees and assets, has far reaching adverse consequences in the financing of the sector by banks, in the allocation of funding for private sector development and strength of the sector as a force to reckon within an economy. This call for adoption of an SME definition that answers to the following questions: where do large firms come from?; How does a country diversify its economy?; and what is an SME?. Critiquing the common non-universal definition of the SME, this study viewed the following definition by Gibson and van der Vaart, 2008, as the best objective descriptive definition for an SME:

An SME is a formal enterprise with annual turnover, U.S. dollar terms, of between 10 and 1000 times the mean per capita gross national, at purchasing power parity, (GNI/PPP) of the country in which it operates"

The merit of this definition is that it captures the functionality and behavioural aspects of the SME sector as dynamic and development al asset class. It distinctively demarcates an SME from a micro enterprise and a corporate firm extreme, which calls for appropriate treatment of the sector for bank financing far from the current tradition of either treating an SME as a retail asset or corporate asset. The current crisis in SME-Bank financing could be attributed to the heterogeneity of the

SME definition. The definition fits well the pronouncement of the Basel II Accord which aims to enhance the impact on SMEs sectors and how to improve SMEs access to finance (Lin, 2007).

We also looked at Basel Capital Accords visa viz their contribution to CRM for SMEs. These Accords came into being as cushion against credit risk, emphasising on the change of the lending activity, adoption of risk-adjusted return approach to loan and favourable and deliberate treatment of SME exposure. Basel I was a one-size-fit-all accord which was indiscriminate of the SME exposure. It was the advent of the Basel II which ushered a new era for SMEs, giving a favourable background for objective IRB CRM for robust creditworthiness SME assessment. It allowed banks, through the IRB framework, to employ own internal estimates of borrowers' creditworthiness to measure credit risk in their portfolio.

We lastly looked at binary models, best candidates for CRM modelling. Owing to the dichotomous nature of credit risk, logit and probit binary choice models were seen to be more appealing than linear probability models.

Chapter 3 presented the exploratory literature on CRM models which have been implemented in credit risk from the time when bankruptcy prediction studies were instrumental for banks to forecast firms which could not honour their loan obligations due to bankruptcy. These studies were based on single financial ratios (univariate) therefore failed in majority of cases to predict firms' bankruptcy thereby engendering great loss for FIs. Owing to large numbers of bank clients applying for loans in some assets, there was great need to automate the underwriting process which led to credit scoring. Owing to market data deficiency, statistical credit scoring took centre stage in client classification among other methodologies. The credit scoring modelling lifecycle was explained in detail.

The major problem in CRM modelling was found to be selectivity bias, where CRM models agree developed from non-random truncated, KGB samples which result in that most CRM models are biased when applied on the TTD pool of loan applicants. This led to the introduction of reject inference techniques which came as panacea to selectivity bias which appeared with certainty in CRM domain. The truncated samples, on which most CRM models are built, were explained from the loan process. Literature has demonstrated that regardless of the introduction a host of reject inference techniques in CRM domain, the majority of the methodologies have not yielded substantive results, due to their failure to recognise the missing data mechanism and that do not have firm theoretical base. Most of these techniques are based on tenuous assumptions rendering them ineffective to tackle selectivity bias problem.

The last part of the chapter 3 has been dedicated to the literature and theory on missing data and Bayesian inference on incomplete samples. The theory premised the introduction of the Bound and Collapse (BC), a Bayesian model-based reject inference technique. The BC technique is theoretically supported unlike other reject inference techniques (augmentation, interpolation, re-weighting etc.) which are built upon tenuous assumptions and are not supported by authentic theory.

Chapter 4 explained the methodologies used for the respective phases entailed in this thesis. The first phase consisted of an exploratory study of SME financing situation in Zimbabwe as a case study of emerging market. The research design adopted for this phase was explained and justified. The survey-based research designs were used to extract data from the bank and SME populations. For the bank survey, the selection units or units of analysis were the banks whilst the reporting units were bank managers, credit risk managers as well as loan officers. For the SME survey, the units of analysis were the SMEs where the SME owners and administrators were the reporting units. The methodology of data analysis for both surveys was based on descriptive statistical analysis to either explore or confirm relationships and inter-connections between or among variables of interest. In both surveys, the instrument of measurement was the questionnaire. In bank survey, the questionnaire was an epitome of instruments used by the World Bank across emerging countries and by other researchers, thereby facilitating the validity and reliability of the instrument used.

The research design used for the second phase of the research was a case-study of a bank in Zimbabwe, a design grounded on the Basel II IRB principles, where a bank must measure credit risk of its clientele. Real credit data was used to demonstrate the empirical application of the BC Bayesian reject inference technique in dealing with sample selectivity bias which appears with certainty in CRM domain. The BC is a model-based reject inference methodology which resolves the problem of selectivity bias inherent in CRM theory by imputing the missing data, which is credit quality of the rejected loan application. The procedure yielded substantial results regardless of limited dataset. This success could be attributed to its firm theoretical base as well as its malleability to accommodate the missingness probability function of the data, a character which other reject inference techniques fail to recognise.

8.2 SUGGESTIONS FOR FURTHER STUDIES

8.2.1 Phase I: Banks involvement with SME financing.

From a research point of view, research limitations of the study give way for future research opportunities. Owing to the unavailability of well-defined sampling frame for SMEs in Zimbabwe, the sample design and sample size used for this thesis were limited. Therefore, future research needs to incorporate probability sample design and determine plausible sample size for survey of SME owners and administrators This helps generate substantial data required and for better generalisation of bank involvement with SME to the whole SME population.

SMEAZ is a voluntary association of urban SMEs, therefore using its list of members as a proxy of the sampling frame of SMEs in Zimbabwe is a limiting factor. Therefore, for future studies, it is suggested that both rural and urban-based SMEs be included in a random sample, to counter the regional imbalance of this current study. This would facilitate a comparative study between SMEs in urban and rural areas in Zimbabwe would be another interesting area to be explored. A representative sample of scientifically determined sample size is expected to achieve more in-depth information about the banks' involvement with SME financing from an SME viewpoint. This would give conclusions applicable to the whole SME population of Zimbabwe.

Gender of owner of an SME is also an interesting factor, therefore it is suggested that gender-based study can be another area of further studies. Such a study would provide valuable evidence and inroads into gender role in as far as banks involvement with SME financing. The interviews with SME owners of different genders can generate valuable information with respect to their challenges in accessing financing, from gender point of view.

8.2.2 Phase II: CRM modelling

The topic of this thesis is quite unexplored especially in emerging economies like Zimbabwe, therefore there are various fronts and openings that could be further studied and explored for more insights and knowledge. This study was limited by data availability as well limited to developing a CRM model for SMEs thereby determine criteria the banks uses for loan underwriting. The study never went to evaluate each and every of the 6 criteria in the final CRM model (equation 99). This is an initial step into an unexplored area, therefore further study is suggested to scrutinise the criteria found to evaluate statistically whether some characteristic variables are more significant than others and if the same variables are applicable to a different bank other than the case-studied bank.

Unlike phase I of thesis, where both SMEs and banks' perspectives were considered, phase II only the perspective of the bank was taken and how it measures SMEs credit risk. This limitation calls for a replica of thesis, taking the SMEs perspective only to evaluate if the criteria the bank uses commensurate with what SMEs perceive that they are assessed on. This could be very interesting because SMEs would easily get loans from banks if only if they know exactly what the banks require to assess them for credit granting decisions and also to find out whether the bank share the same view of the criteria they use to assess them.

In terms of the first sub-question, this thesis limited itself to only use single imputation technique to resolve the pertinent issue of selectivity bias rife in CRM domain. In fact, missing data are a pervasive problem in CRM domain. Missing data can be a serious impediment for the development of an objective CRM. Owing to credit quality of the rejected loan applicants has given way to selectivity bias which in turn compromises statistical modelling even other methodological approaches like ML and SVM just to mention a few. In this study the Bound and Collapse Bayesian reject inference was used to resolve sample selection bias to achieve an AGB sample on which the final CRM model was built. The BC was instrumental in building an objective SME CRM model at ZimSME bank. To compare or improve the classification power of the eventual SME CRM model, further study is proposed to use of multiple imputation methodology as an alternative to BC methodology.

8.3 SUMMARY

This thesis provided an empirical analysis for applying theoretically model-based imputation approach to modelling CRM of SMEs and proposed some directions for further research in the improvement of reject inference technologies. Two (2) major rejection policies: weak selection (for risk-taker bank) and strong selection (for risk-averse bank) models were compared to choose the final CRM model for the case-studied bank. Predictive variables were statistically selected for suitability and inclusivity in the final model based on predictivity and explanatory power to determine the PD, a measure of credit risk of a loan applicant. The main limitation of this research was perceived to be due to the truncation of the development sample and scarcity of publicly available information on SMEs. Irrespective of these limitations, the thesis provided a much-needed pre-analysis of development sample prior to the modelling of CRM of the most important economic player, SMEs. Most research has been on other bank assets (retail and corporate) with limited work on the SME CRM modelling, therefore this work fills-in the gap in literature. Possible extensions to the current research may develop along the following directions:

- Inclusion of audited financial predictors of the enterprise into the models;
- Inclusion of economic factors and industry specific information developing;
- Use of multiple imputation technology for reject inference
- Use of other approaches to model validation;
- Performing cost-benefit analysis for the optimal cut-off selection;
- Modelling credit risk of small private firms.

REFERENCE LIST

Abdulrahman, A. 2018. Optimal regulation of banking system's advanced credit risk management by unified computational representation of business processes across the entire banking system. *Cogent Economics & Finance*, 6(1): 1-22.

Abdulsaleh, A.M.A. 2016. Bank Financing for Small and Medium Enterprises in Libya. PhD thesis. Griffith University, Queensland, Australia.

Abiodun, E.A. & Entebang, H. 2015. Small and medium business management-financial sources and difficulties. *International Letters of Social and Humanistic Sciences,* 58: 49-57.

Abor, J. & Quartey, P. 2010. Issues in SME Development in Ghana and South Africa. *International Research Journal of Finance and Economics*, 39: 218-228.

Ackah, J. & Vuvor, S. 2011. The challenges faced by Small and Medium Enterprises (SMEs) in Obtaining Credit in Ghana. MBA Thesis. Blekinge Tekniska Hogskola School of Management.

Agrawal, K. & Maheswari, Y. 2019. Efficacy of Industry factors for corporate default prediction. *IIMB Management Review*, 31: 71-77.

Ahn, J-H. 2016. The Impact of the Banking Competition on funding and lending Technology. *Cairn Info*, 67(6): 1117-1139.

Al Baz, N.A. 2017. Modelling Credit Risk of Small and Medium Enterprises in Saudi Arabia. DBA thesis, Cranfield University. Cranfield.

Aliu, M. & Sahiti, A. 2016. The Effect of Credit Management on Banks Profitability in Kosovo. *European Journal of Economic Studies*, 18(4): 492-515.

Altman, E.I. 2018. Applications of Distress Prediction models: What have learned After 50 years from the Z-score models. *International Journal of Financial Studies*, 6(3): 1-15.

Altman, E.I. & Saunders, A. 1998. Credit risk measurement: Development over the last 20 years. *Journal of Banking & Finance*, 21: 1721-1742.

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

Altman, E.I. & Sabato G. 2005. Effects of the New Basel Capital Accord on Bank Capital Requirements for SMEs, *Journal of Financial Services Research,* 28(1): 15-42.

Altman, E.I. & Sabato, G. 2006. Effects of the New Basel Capital Accord on bank capital requirements for SMEs. *Journal of Financial Services Research*, 28(1/2): 15-42.

Altman, E.I. & Sabato, G. 2007. Modelling credit risk for SMEs: evidence from the U.S. market. *Abacus*, 43(3): 332-357.

Altman, E.I., Sabato, G. & Wilson, N. 2009. *The value of Non-financial Information in SME Risk Management*. Credit Scoring and Credit Control, XI Conference, Edinburgh.

Akelola, S. 2012. Fraud in the Banking Industry: A Case study of Kenya. PhD thesis. Nottingham Trent University, Nottingham.

Akhavein, J., Frame, W.S. & Lawrence, J.W. 2005. The diffusion of financial innovation: An examination of the adoption of small business credit scoring by large banking organisations. Journal of Business, 78: 577-596.

Akinyooye, R.F. 2006. The Challenges of Implementing the Basel II Accord in Nigerian Banks. PhD thesis. St. Clements University.

Amidu, M. & Wolfe, S. 2013. The impact of market power and funding strategy on bank-interest margin. *The European Journal of Finance*, 19(9): 888-908.

Anagnostopoulos, T., Skouloudis, A., Khan, N. & Evangelinos, K. 2018. Incorporating Sustainability Considerations into Lending Decisions and the Management of Bad Loans: Evidence from Greece, *Sustainability*, 10: 4728-4743.

Anderson, R. 2007. *The Credit Scoring Toolkit: Theory and Practice for Retail Credit Risk Management and Decision Automation*. London: Oxford University Press.

Antão, P. & Lacerda, A. 2011. Capital requirements under the credit risk-based framework. *Journal of Banking and Finance*, 35 (6): 1380-1390.

Arnoud, B. & Matej, M. 2008. The Evolving Landscape of Banking. *Industrial and Corporate Change*, 17: 1173-1202.

Arturo, E. 2000. Credit Ratings and complementary sources of credit quality Information. Basel Committee on Banking Supervision Working Paper 2000.

Asad, U.K., Asad, B., Abdullah, H. & Khalid, I. 2018. Revisiting Barriers to External Finance for SMEs. *Journal of Advanced Research in Business and Management Studies*,11(1): 24-32.

Atristain-Suarez, C. 2012. Employment generation and economic development through increased operational efficiency of SME in Mexico: Some research perspective. *International Journal of Business Competition and Growth 2042-3845*, 2(2): 181-199.

Asrat, H. 2018. Assessment of Credit Management Practices at United Bank S. C. MBA thesis. College of Business and Economics, Addis Ababa University, Ethiopia.

Astrom, Z.H. O. 2015. Credit Related Practices of Islamic Banking in Comparison with theory: Case Study of Turkey. *International Journal of Islamic Economics and Finance Studies*, 1(2): 39-60.

Avery, R.B. 1977. Credit Scoring Models with Discriminant Analysis and Truncated Samples. Unpublished paper.

Aysa I.E. 2019. Determinants of perceived bank financing accessibility for SMEs: evidence from an emerging market. *Economic Research-Ekonomska Istraživanja,* 32(1): 690-716.

Aubier, M. 2007. Examining the impact of Basel II on supply of credit to SMEs, Tresor-Economics, No. 13 2007/2 (178-179):
168.

Babak, B.L. & Xu, M. 2017. How Small and Medium Enterprises in fluence in Bilateral Economic and Commercial Relation of Iran and China. *Asian Journal of Business and Management Sciences.* 4(12): 1-15.

Baily, M.N., Litan, R.E. & Johnson, M.S. 2008, Origins of the Financial Crisis. Initiative on Business and Public Policy st Brookings. Fixing Finance Series - Paper 3. November 2008.

Balkehol, B. & Evans-Klock, C. 2002. Private Equity and Capitalisation of SMMEs in South Africa: Quo vadis? Working paper #34.

Ball, R. 2016. IFRS - 10 years later. *Accounting and Business Research*, 46(5): 545-578.

Bakhtiari, S., Breung, R., Magnani, L. & Zhanga, J. 2020. *Financial Constraint and SMEs: Review*. IZA Institute of Labour Economics. (Discussion Paper No. 12936).

Bamper, P., Fernandez-Stark, K., Gereffi, G. & Guinn, A. 2014. *Connecting Local Producers in Developing Countries to Regional and Global Value Chains*: Update. Paris: OECD Publishing (OECD Trade Policy Papers, No. 160.

Banasik, J. & Crook, J. 2006. Reject Inference, Augmentation and Sample Selection. Working Paper Series No. 05/04. The Credit Research Centre, The School of Management, The University of Edinburgh.

Banasik, J. & Crook, J. 2004. Does Reject Inference Really Improve the Performance of Application Scoring Models?. *Journal of Banking and Finance*, 28: 857-874.

Banasik, J.L., Crook, J.N. & Thomas, L.C. 2003. Sample selection bias in credit scoring models. *Journal of the Operational Research Society* 54: 822-832.

Banasik, J., Crook, J.N. & Thomas, L.C. 1999. Not if but when will borrowers default. Journal of the Operational Research Society, 50: 1185-1190.

Bank for International Settlements (BIS), 2001. *The New Basel Capital Accord: An Explanatory Note*. Basel: Basel Committee on Banking Supervision (April 1999).

Banwo, A.O., Du, J. & and Onokala, U. 2017. The determinants of location specific choice: small and medium-sized enterprises in developing countries. *Journal of Global Entrepreneurship Research*, 7 (16): 1-17.

Barakova, I., Glennon, D. & Palvia, A. 2013. Sample Selection Bias in Acquisition Credit Scoring Models: An Evaluation of the Supplemental-Data Approach. *Journal of Credit Risk,* 9: 1-43.

Bartels, J. C. 2005. Basel II and the survival of SME. Are lenders and borrowers ready to comply with Basel II? *Business Credit*, 104(10): 48-49.

Ba Zhang, C.H. 2009. Review of the Literature on credit risk modelling: Development of the recent 10 years. One-year-Master thesis in Applied Statistics, Hogskolan Dalarna.

Bbenkele, E.K. 2007. An investigation of SMEs' perceptions Towards Services offered by Commercial Banks in South Africa. *African journal of accounting, economics, finance and banking research,* 1(1): 13-25.

BCBS. 1998. *Basel I: International convergence of capital measurement and capital standards*. Basel, Switzerland: Bank for International Settlements.

BCBS. 1999. *Credit risk modelling: current practices and applications*. Basel, Switzerland: Bank for International Settlements (Report 49).

BCBS. 2003. *The New Basel Capital Accord -Third consultative paper. (CP3)*. Basel, Switzerland: Basel Committee on Banking Supervision (April 2003).

BCBS. 2005. *International convergence of capital measurement and capital standards: A revised framework.* Basel, Switzerland: Bank for International Settlement.

BCBS. 2006. *Basel II: International convergence of capital measurement and capital standards: A revised framework - Comprehensive version.* Basel, Switzerland: Bank for International Settlements.

Beaver, W.H. 1967. Financial ratios as predictors of failure. Empirical research in accounting: Selected studies. *Supplement to Journal of Accounting Research,* 4: 71-111.

Beck, T., Demirguc-Kunt, A., Laeven, L. & Levine, R. 2008. Finance, firm size, and growth. *Journal of Money, Credit and Banking*, 40(7): 1379–1405.

Belyaeva, E. 2014. On a new logistic regression model for bankruptcy prediction in the IT branch. U. U. D. M. Project Report 2014: 35. Uppsala Universitet.

Bensic, M., Sarlija, N. &, Zekic-Susac, M. 2005. Modeling Small-Business Credit Scoring by Using Logistic Regression, Neural Networks and Decision Trees. *Intellectual Systems Accounting and Financial Management,* 13(3): 133-150.

Benzin A., Truck S. & Rachev S.T. 2003. Approaches to credit risk in the New Basel Capital Accord, Contributions to Economics: 1-33.

Berg, G. & Fuchs, M. J. 2013. Bank Financing of SMEs in Five Sub-Saharan African Countries: The Role of Competition, Innovation, and the Government (August 1, 2013). World Bank Policy Research Working Paper No. 6563.

Berger, A.N. 2004. Potential Competitive Effects of Basel II on Banks in SME Credit Markets in the United States. Wharton Financial Institutions Center, Philadelphia, USA.

Berger, A.N. 2006. Potential competitive effects of Basel II on banks in SME credit markets in the United States. *Journal of Financial Services Research*, 29(1): 5-36.

Berger, A.N. & Frame, S.W. 2005. Small Business Credit Scoring and Credit Availability. Federal Reserve Bank of Atlanta, Working Paper Series Nr. 10.

Berger, A.N. & Udell, G.F. 1995. Relationship lending at lines of Credit in Small Firm Finance. Journal of Business, 68(3): 351-382.

Berger, A.N. & Udell, G.F. 2006. A more complete conceptual framework for SME finance. *Journal of Banking and Finance* 30: 2945-2966.

Berisha, G. & Pula, J.S. 2015. Defining Small and Medium Enterprises: a critical review. *Academic Journal of Business, Administration, Law and Social Sciences*, 1(1): 17-28.

Blundell-Wignall, A. & Atkinson, P. 2010. Thinking beyond Basel III: Necessary Solutions for Capital and Liquidity. *OECD Journal: Financial Markets Trend*, 2010(1): 9-33.

Bogale, E. 2018. Effect of Credit Risk on Profitability and Lending Decision of Commercial Banks in Ethiopia. *Research Journal of Finance and Accounting*, 9(21): 23-26.

Bolton, C. 2009. Logistic Regression and its Application in Credit Scoring. PhD thesis. University of Pretoria, Pretoria.

Bolton, P., Cecchetti, S., Danthine, J. P. & Vives, X. 2019. *Sound at Last? Assessing a Decade of Financial Regulation. The Future of Banking 1*. University of Navarra: CEPR Press.

Brent, W. H. & Addo, C. K. 2012. Minimizing information asymmetry: does firm's characteristics matter? *Academy of Banking Studies Journal*, 11(1): 43-54.

Bruni, M.E., Beraldi, P. & Iazzolino, G. 2014. Lending decisions under uncertainty: A DEA approach. *Journal of Production Research*, 52(3): 766-775

Bushe, B. 2019. The causes and Impact of Business Failures among small to micro and medium enterprises (SMMEs) in South Africa. *Africa's Public Service Delivery and Performance Review,* 7(1): 1-26.

Butt, S. 2017. Taming Financial Capital: The Role and Limitations of Basel Capital Regulation in Pakistan. PhD thesis. Royal Dock School of Business and Law. University of East London.

Cardone-Reportella, C. & Trujillo-Ponce, A. 2007. Efectos del avalde las SGRs en la financiacion de las PYME y los requerimientos de la capital Basilea II. *Revista Espanola de Financiacion y Contabilidad* 36: 753-785.

Cardone-Riportella, C., Trujillo-Ponce, A. & Briozzo, A. 2011. *What do Basel Capital Accords mean for SMEs*?, Working Paper No. 10. Departamento de Economía de la Empresa, Business Economic Series 04, Universidad Carlos III de Madrid, Madrid, April.

Chandler G.G. & Coffman, J.Y. 1977. Using credit scoring to improve the quality of consumer receivables: legal and statistical implications. Paper presented at the Financial Management Association meetings, Seattle, Washington.

Chatzigakis N. 2016. How the replacement of Basel II by Basel III has an Effect on Economic Growth, Regional *Science Inquiry,* VIII (3): 147-157.

Chen, G.G. & Astebro, T. 2012. Bound and collapse Bayesian reject inference for credit scoring. *Journal of the Operational Research Society*, 63: 1374-1387.

Chen, S. & Haziza, D. 2018. Recent Developments in Dealing with Item Non-response in a Survey: Critical Review. *International Statistical Review*, 87(51): 192-218.

Chin, Y-W. & Lim, E-S. 2018. SME Policies and Performance in Malaysia. Yusof Ishak Institute, Economics Working Paper No. 2018-3. Unpublished.

Chionsini G., Marcucci, J. & Quagliariello, M. 2010. The Treatment of Small and Medium Enterprises in Basel 2: So Right, So Wrong?, working paper presented at Procyclicality and Financial Regulation Conference, University of Tilburg, 11- 12 March 2010.

Claessens, S. & Kodres, L. 2014. Regulatory Responses to the Global Financial Crisis: Some uncomfortable Questions. IMF Working Paper WP/14/46.

Coetzee, F. & Buys, P.W. 2017. The Impact of the independent review on Small and Medium Enterprises access to banks finance: the case of South Africa. *Banks and Bank Systems*, 12(1): 135-142.

Collins, D., Morduch, J., Rutherford, S. & Ruthven O. 2009. *Portfolios of the Poor*. Princeton University Press.

Copas, J.B. & Li, H.G. 1977. Inference for non-random samples (with discussion). *Journal of the Royal Statistical Society*, B(59): 55-95.

Creal, D., Schwaab, B., Koopman, S. J. & Lucas, A. 2014. Observation-Driven Mixed Measurement dynamic factor models with an application to credit risk. *Review of Economics & Statistics,* 96 (5), 898-915.

Creswell, J. W. & Plano-Clark, V. L. 2007. *Designing and conducting mixed methods research*, Sage Publications Ltd., London, United Kingdom.

Crook, J.N. 1996. Credit scoring: An overview. Working paper series No. 96/13, British Association, Festival of Science. University of Birmingham, The University of Edinburgh.

Crook, J. & Banasik, J. 2004. Does reject inference really improve the performance of application scoring models? *Journal of Banking & Finance*, 28(87): 857-874.

Cruickshank, D. 2000. Competition in UK Banking: A report to the Chancellor of the Exchequer. Norwich. HM Treasury.

Curcio, D., Gianfrancesco, I. & Malinconico, A. 2011. Investigating implied asset correlation and capital requirements: empirical evidence from the Italian banking system. *Banks and Bank Systems*, 6(2): 116 -125.

Dada, E.G., Bassi, J.S., Chiroma, H., Abdulhamid, S.M., Adetunmbi, A.O. & Ajibuwa, O.E. 2019. Machine Learning for email span filtering: review, approaches and open research problems. *Heliyon*, 5(2019): 1-23.

Dambaza, M & Kruger, J.W. 2018. *Bank Involvement with Small and Medium Enterprises (SMEs)*. Proceedings of the 131St IASTEM International Conference, Bali, Indonesia, 30-31st July 2018.

Dainelli, F., Giunta, F. & Cipollini, F. 2013. Determinants of SME credit worthiness under Basel rules: the value of credit history information. *PSL Quarterly Review,* 66 (264): 21-47.

de la Torre, A., Soledad Martínez Pería, M. & Schmukler, S.L. 2010. Bank Involvement with SMEs: Beyond Relationship Lending. *Journal of Banking and Finance,* 34(9): 2280-2293.

de Noni, I., Lorenzon, A. & Orsi, L. 2007. Measuring and Managing Credit Risk in SMEs: A Quantitative and Qualitative rating Model. Working Paper No. 2007-36. University of Milan, Italy. Unpublished.

Demirguc-Kunt, A., Maksimovic, V., Beck, T. & Laeven, L. 2006. The determinant of financing obstacles. *International Journal Money Financing*, 25(6): 932-952.

Dempster, A.P., Laird, N.M. & Rubin, D.B. 1977. Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Association*, B(39): 1-38.

Derreumaux, P. 2009. *The difficulties banks face in financing SMEs in Sub-Saharan Africa: Who is to blame? Issue May 2009 SME financing in Sub-Saharan Africa*, Private Sector Development, PROPARCO Magazine.

Dietsch, M. & Petey, J. 2004. Should SME exposures be treated as retail or corporate exposures? A comparative analysis of default probabilities and asset correlations in French and German SMEs. *Journal of Banking & Finance*, 28: 773-778.

Dimitras, A.I., Papadakis, S. & Garefalakis, A. 2017. Evaluationof Empirical attributes for credit risk forecasting from numerical data. *Investment Management and Financial Innovations*, 14(1): 9-18.

Dinh, T.H.T. & Kleimeier, S. 2007. A credit Scoring Model for Vietnam's Retail Banking Market. *International Review of Financial Analysis*, 16(5): 471-495.

Ditrich, J. 2015. *Selection Bias Reduction in Credit Scoring Models*. The 9th International Days of Statistics and Economics, University of Economics, Prague W. Churchill Sq. 4, 130 67 Prague 3, Czech Republic. September 10-12, 2015.

Dong, Y. & Peng, C.Y. J. 2013. Principled Missing Data Methods for Researchers. *SpringerPlus*, 2(222): 1-19.

Dynan, K. & Sheiner, L. 2018. GDP as a measure of Economic well-being, Hutchins Centre, Brookings: Hutchin Center on Fiscal and Monetary Policy (Working Paper # 43).

Du Jardin, P. 2009. Bankruptcy Prediction Models: How to choose the most relevant variables? *Bankers, Markets & Investors,* 98: 39-46.

Edmister, R. 1972. An empirical test of financial ratio analysis for Small Business failure prediction. *Journal of Financial and Quantitative Analysis,* 7(2): 1477-1493.

Eisenbeis R.A. 1977. Pitfalls in the application of discriminant analysis in business, finance and economics. *Journal of Finance,* 32: 875-900.

Ekpu, V. U. 2015. *Determinants of Bank Involvement with SMEs: A Survey of Demand-Side and Supply-Side Factors.* Springer International Publishing.

Fatoki, O.O. 2014. Factors Influencing Financing of Business Start-ups by Commercial Banks in South Africa. *Mediterranean Journal of Social Sciences*. 5(20): 94-100.

Fatoki, O. and Van Aardt, A. 2011. Constraints to Credit Access by New SMEs in South Africa: A supply-side analysis. African Journal of Business Management, 5(4): 1413-1425.

Farrar, J.H. 2010. The global financial crisis and the governance of financial institutions. *Australian journal of corporate law,* 24 (3): 227-243.

Feelders, A. 2000. Credit scoring and reject inference with mixture models. *International Journal of Intelligent Systems in Accounting, Finance & Management*, 9*:*1-8.

Ferreiro, J. O. 2016. Modelling Default Risk Charge (DRC): Internal Model Approach. Masters Dissertation, University of Valencia, Madrid. Spain.

Financial Service Agency. 2003. Progress of Action Program for Strengthening Relationship-base Banking. Japan: (as of the first half of FY2003).

Finlay, S. 2011. Multiple classifier architectures and their application to credit risk assessment. *European Journal of Operational Research*, 5(210): 368-378.

Fink, A. 1995. *How to Analyze Survey Data*. SAGE Publications, Inc, Thousand Oaks, CA; London.

Fin Mark Trust. 2006. Fin scope small business survey report. [Online] Available: http://www.finmarktrust.org.za [accessed 15 July 2010].

Flannery, M. J. & Bliss, R. R. 2018. Market Disclosure in Regulation: Pre- and Post-Crisis.
Oxford: Forthcoming Oxford Handbook of Bank (3rd edition, 2019).

Fogarty, D. 2000. Intelligent e-Imputation: A New Methodology for Data Quality Management of Commercial Data Warehousing Applications Using Artificial Neural Networks to Support the Critical Role of Business Intelligence in e-Business, Unpublished Ph.D. dissertation, Leeds Metropolitan University.

Fogarty, D. & Blake J. 2002. Utilizin g Recent Advancements in Techniques for the Analysis of Incomplete Multivariate Data to Improve the Data Quality Management of Current Academic Research, *Journal of Quantity and Quality,* 36(3): 277-289.

Fotache, M., Fotache, G., Ocneanu, L. & Bucşă, R. C. 2011. SMEs in the Current Economic Environment. *Economy Transdisciplinary Cognition*, 15(2): 95-106.

Forster, J. J. & Smith, P. W. F. 1998. Model-based inference for categorical survey data subject to non-ignorable non-response (disc: P89-102). *Journal of the Royal Statistical Society, Series B, Methodological*, 60: 57–70.

Foxcroft, M., Wood, E., Kew J., Herrington, M. & Segal, N. 2002. Global Entrepreneurship Monitor. South African Executive Report to University of Cape Town Graduate School of Business. Unpublished.

Fraser, J. & Simkins. B.J. 2010. *Enterprise Risk Management. The Robert W. Kolb Series in Finance.* John Wiley & Sons, Inc.

Fullenkamp, C. & Rochon, C. 2014. Reconsidering Bank Capital Regulation: A new combination of Rules, Regulations and Market Discipline. International Monetary Fund Working Paper No. WP/14/169.

Gakure, R.W., Ngugi, J.K., Ndwiga, P.M. & Waithaka, S.M. 2012. Effect of Credit Risk Management Techniques on the Performance of unsecured Bank Loans Employed by Commercial Banks in Kenya. *International Journal of Business and Social Research (IJBSR)*, 2(4): 221-236.

Gambacorta, L. & Karmakar, S. 2016. Leverage and Risk Weighted Capital Requirements. Monetary and Economics Department. BIS Working Paper No. 586.

Gambacorta, L., Huang, Y., Qiu, H. & Wang, J. 2019. How do Machine Learning and No-traditional data affect Credit Scoring? New Evidence from a Chinese Fintech firm. BIS Working Papers No 834. Monetary and Economic Department. December 2019.

Gelman, A., Jakulin, A., Pittau, M.G. & Su, Y-S. 2008. A weakly Informative Default prior distribution for logistic and other regression models. *Annals of Applied Statistics*, 2(4): 1360-1383.

Gelman, A., John, B. C., Hal, S. S. & Rubin, D. R. 1995. *Data Analysis*. Chapman & Hall.

Gherghina, S.C., Botezatu, M.A., Hosszu, A. & Simionescu, L.N. 2020. Small and Medium-Sized Enterprises (SMEs): The Engine of Economic Growth through Investments and Innovation. *Sustainability*, 12: 347-369.

Gibson, T. & van der Vaart, H. J. 2008. Defining SMEs: A less imperfect way of defining Small and Medium Enterprises in Developing Countries. Brookings Global Economy and Development, 11. September.

Gilley O .W. & Leone R. P. 1991. A Two-Stage Imputation Procedure for Item Nonresponse in Surveys. *Journal of Business Research*, 22: 281-291.

Gillham, B. 2000. *The Research Interview.* New York: Continuum.

Giovannini, A., Laeopata, M. & Minneti, R. 2013. Financial Markets, Banks, and Growth: Disentanglingthe Links. *Cairn Info*, 2013/5, No. 131:105-147.

Giuseppe, S. & Ughetto, E. 2010. The Basel II Reform and Provision of Finance for R&D Activities in SMEs: An Analysis for a sample of Italian Companies. *International Small Business Journal*, 28: 65-89.

Goh, R. Y. & Lee, L. S. 2019. Credit Scoring: A review on Support Vector Machines and Metaheuristic Approaches. *Advances in Operations Research*, 2: 1-30.

Gombola, M., Haskins, M., Ketz, J. & Williams, D. 1987. Cash flow in bankruptcy prediction. *Financial Management* 16(4): 55-65.

Gonzalez-Watty, A. 2016. The Quest of Accountability in Transnational Regulatory Networks: The case of Basel Committee on Banking Supervision (BCBS). PhD thesis. Wolfson College, University of Oxford, London.

Good, I. J. 1985. Weight of Evidence: A Brief Survey. *Bayesian Statistics*, 2: 249-270.

Gordy, M. 2002. A Risk-Factor Model Foundation for Ratings-Based Bank Capital Rules. Board of Governors of the Federal Reserve System.

Gottschalk, R. 2014. *Institutional Challenges for Effective Banking Regulation and Supervision in Sub-Saharan Africa*: Shaping Policy for Development Workshop, Accra, 10-11th September 2013.

Griffith-Jones, S. 2003. Implications of the Basel II for Stability and Growth in Developing Countries; Proposal for further Research and Action. Paper prepared for Ibase Rio Meeting on Financial Liberalisation and Global Governance: The Role of International Entities, 19-20 March 2007.

Gwangwava, E., Faitira M., Gutu K., Chinoda T. & Frank R. 2014. An Assessment of Risk Management Practices in SMEs in Zimbabwe: A Review and Synthesis. *Journal of Humanities and Social Science*, 19(8): 6-14.

Habiby, A.S. & Coyle, D.M. 2010. The High-Intensity Entrepreneur. *Harvard Business review*, 88(9): 75-78.

Haga, E. 2017. The Role of Small Businesses (Small Scale Economic Projects) in Alleviating the Acuity of Unemployment. *International Business Research*, 10 :120-132.

Hagos, M. 2010. Credit Management. Case study of Wegagen Bank Share Company in Tigray Region. MSc in Finance and Investment thesis. Mekelle University, Ethiopia.

Haldar, A. & Stiglitz, E. 2016. Group Lending, Jointly Liability and Social Capital: Insights from the Indian Microfinance Crisis. *Politics and Society*, 44(4): 459-497.

Hand, D.J. 1998. *Reject Inference in Credit Operations*. In Credit Risk Modeling Design and Application, ed. E. Mays. Chicago: Glenlake Publishing: 181-190.

Handley, G., Higgin, K. & Sharma, B. 2009. Poverty and Poverty Reduction in Sub-Saharan Africa: An Overview of Key Issues. Overseas Development Institute. (Working Paper 299).

Hanson, S.G., Pesaran, H.M. & Schuermann, T. 2008. Firm heterogeneity and credit risk diversification. *Journal of Empirical Finance*, 15: 583-612.

Hao, C., Alam, M.M. & Carhy, K. 2010. Review of the Literature on Credit Risk Modelling: developments of the past 10 years. *Banks and Bank Systems*, 5(3): 43-60.

Harrell, F. E., Jr. 2001. *Regression modeling strategies, with application to linear models, logistic*

Hasumi, R. & Hirata, H. 2010. Small Business Credit Scoring: Evidence from Japan. RIETI Discussion Paper Series 10-E-029.

Hawkins, J. & Turner, P. 1999. *Bank restructuring in practice: An overview*. BIS Policy Paper No. 6. August: 6-105.

Heinze, G., Wallisch, C. & Dunkler, D. 2017. Variable Selection: A review and recommendation for practising statistician. *Biometrical Journal*: 431-449.

Henley, W.E. 1995. The Statistical Aspects of Credit Scoring, Unpublished PhD thesis. The Open University. UK.

Henneke, J. & Truck, S. 2006. Asset Correlations and Capital Requirements for SME in the Revised Basel II Framework. *Banks and Bank Systems*, 1(1): 75-92.

Hoadley, B. 2001. Statistical Modeling: The Two Cultures: Comment. Statistical Science, 16, 220-224.

Holmes, S., Hutchinson, P., Forsaith, D., Gobson, B. & McMahon, R. 2003. *Small enterprise finance,* John Wiley & Sons, Milton Qld, Australia.

Horstedt, M. & Linjamaa, J. 2015. *Credit Evaluation of Swedish Small and Medium Enterprises A Banking Sector Perspective*. Degree Project. Umea University. Umea, Sweden.

Hoskins, S.M. & Labonte, M. 2015. An Analysing of the Regulatory Burden on Small Banks. Congressional Research Service. CRS Report 7-5700.

Hosmer D.W. & Lemeshow, S. 2000. *Applied Logistic Regression*. 2nd Ed., John Wiley and Sons, New York.

Huyghebaert, N. & van de Gucht, L.M. 2007. The Determinants of Financial Structure: New Insights from Business Start-Ups. *European Financial Management,* 13(1): 101-133.

Ibe, S.O., Moemena, I.C., Alozie, S.T. & Mbaeri C.C. 2015. Financing Options for Small and Medium Enterprises (SMEs): Exploring Non-Bank Financial Institutions as an Alternative Means of Financing. *Journal of Educational Policy and Entrepreneurial Research (JEPER),* 2(9): 28-37.

IMF. 2019. Financial Inclusion of Small and Medium Enterprises in the Middle East and Central Asia.

Ishtiaq, M. 2015. Risk Management in Banks: Determination of Practices and Relationships with Performance. PhD thesis. University of Bedfordshire. Bedfordshire, UK.

Ivankova, N.V., Creswell, J.W. & Stick, S.L. 2006. Using mixed-methods sequential explanatory design: From theory to practice. *Field Methods,* 18(1): 3-20.

Jacobson, T. & Roszbach, K. 1999. Bank Lending Policy, Credit Scoring and Value at Risk, Sveriges Riksbank Working Paper Series, No. 68, Sveriges Riksbank, Stockholm.

Japkowicz, N. & Shah, M. 2011. *Evaluating Learning Algorithms: A Classification Perspective*. Cambridge University Press.

Jimenez, G. & Saurina, J. 2004. Collateral, Type of Lender and Relationship Banking as Determinants of Credit Risk. *Journal of Banking and Finance*, 28: 2191-2212.

Jones, E. 2014. *Global banking standards and low-income countries: Helping or hindering effective regulation?* University of Oxford, Global Economic Governance Programme (GEG), Oxford. (GEG Working Paper, No. 2014/91).

Jones, E. & Zeitz, A. 2018. Regulatory Convergence in Financial Periphery: How interdependence shapes Regulator's Decisions. *International Studies Quarterly*, 63: 908-922.

Kanapickiene, R. & Spicas, R. 2019. Credit Risk Assessment Model for Small and Micro-Enterprises: The Case of Lithuania. *Risks*, 7: 67-90.

Karels, G. & Prakash, A. 1987. Multivariate normality and forecasting of business bankruptcy. *Journal of Business Finance & Accounting,* 14(4): 573-593.

Karadag, H. 2016. The Role of SMEs and Entrepreneurship on Economic Growth in Emerging Economies within the Post Crisis Era: An Analysis from Turkey. *Journal of Small Business and Entrepreneurship Development*, 4: 22-31.

Katuka, B. & Dzingirai, C. 2015. Micro and Macro Drivers of Credit risk: The case of Zimbabwean Banking Industry 2009-2013. *MEFMI Research and Policy Seminar Journal*, 2: 51-76.

Kazi, R.H. 2016. Development of a Credit Scoring Model for Retail Loan Granting Financial Institutions from Frontier Markets. *International Journal of Business and Economics Research,*5(5): 135-142.

Kennedy, K. 2013. Credit Scoring Using Machine Learning. PhD thesis. Technological University Dublin, Dublin.

Kerr, S.P., Kerr, W.R. & Xu, T. 2017. Personality Traits of Entrepreneurs: Review of Recent Literature. Harvard Business School (Working Paper 18-047).

Khaled, A. & Wahab, N.S.A. 2019. Credit risk management and business intelligence approach of the banking sector in Jordan, *Cogent Business & Management,* 6(1): 1-9.

Kipsang, Y.W. 2014. The Relationship between Credit Risk Management and the Financial Performance of Microfinance Institution in Kenya. Master's thesis. School of Business. University of Nairobi.

Koh, H., 1987. *Prediction of going concern status: A probit model for the auditors*. Ph.D. thesis, Virginia Polytechnic Institute and State University.

Kolari, J. W. & Shin, H. G. 2004. *Assessing the Profitability and Riskiness of Small Business Lenders in the U.S. Banking Industry*, Available at: www.sba.gov. [Accessed 23 September 2017].

Konovalova, N., Kristovska, I. & Kuduriska, M. 2016. Credit Risk Management in Commercial Banks. *Polish Journal of Management Studies*, 3(2): 90-100.

Kraus, A. 2014. *Recent Methods from Statistics and Machine Learning for Credit Scoring.* Masters Dissertation. Universitat München.

Kritikos, A.S. 2014. Entrepreneurs and their impact on jobs and economic growth. *IZA World of Labour*, 8: 1-10.

Kritzinger, N. & van Vuuren, G.W. 2017. An Optimised credit scorecard to enhance cut-off score determination. *South African Journal for Management Sciences*, 21(1): 1571-1585.

Kuritzkes, A. 2002. Operational risk capital: A problem of definition. *Journal of Risk Finance*, 4(1): 47-56.

Kushnir, K., Mirmulstein, M.L. & Ramalho, R. 2010. Micro, small and medium enterprises around the world: How many are there, and what affects the count? MSME Country Indicators, World Bank/IFC. Available Online: www.ifc.org [Accessed 15 May 2018].

Lagat, K., Mugo, R. & Otuya, R. 2013. Effect of Credit Risk Management Practices on Lending Portfolio among Savings and Credit Cooperatives in Kenya. *European Journal of Business Management,* 5(19): 92-109.

Lai, K.E. & Kuo, C.J. 2010. How to Gauge the Credit Risk of Bank Loans: Evidence from Taiwan. *International Research Journal of Finance & Economics,* 39: 7-14.

Lall, R. 2009. *Why Basel II failed and why Basel III is doomed*. GEG Working Paper No. 2009/52. University of Oxford, Global Economic Governance (GEG) programme, Oxford.

Lawrence, J.R., Pongsatat, S. & Lawrence, H. 2015. The Use of Ohlson's O-Score for Bankruptcy Prediction in Thailand. *The Journal of Applied Business Research,* 31(6): 2069-3078.

Lesle, V. L. & Avramova, S. 2012. Revisting Risk-Weigheted Assets. IMF Working paper WP/12/90.

Lewis, E. 1992. An introduction to credit scoring. Athena Press: San Rafael, CA. 71.

Leung, K. H. K. 2008. An Investigation of Artificial Immune Systems and Variable Selection Techniques for Credit Scoring. Ph.D. thesis, Royal Melbourne Institute of Technology University, Australia.

Li, F. & Zou, Y. 2014. The Impact of Credit Risk Management on Profitability of Commercial Banks: A Study of Europe. Degre Project. Umeå School of Business and Economics.

Lin, S-M. 2007. *Small and Medium Enterprises Credit Risk Modelling from Internal Rating Based (IRB) Approach in Banking Implementation of Basel II Requirement*. PhD thesis. University of Edinburgh, Edinburgh.

Little, R. J. A. 1993. Pattern mixture models for multivariate incomplete data. *Journal of the American Statistical Association,* 88: 125-134.

Little, R. J. A. & Rubin, D. B. 1987. *Statistical Analysis with Missing Data,* John Wiley & Sons.

Löffler, G., Posch, P. N. & Schöne, C. 2005. *Bayesian methods for improving credit scoring models*. Technical report, Department of Finance, University of Ulm, Germany.

Loughrey, J. & Cunningham, P. 2005. Overfitting in wrapper-based feature subset selection: The harder you try the worse it gets. In Research and Development in Intelligent Systems XXI: Proceedings of AI-2004, the 24th Sgai International Conference on Innovative Techniques and Applications of Artificial Intelligence, 33-43, Springer-Verlag New York Incorporated. 76.

Ma, T. 2016. Basel III and the Future of Project Finance Funding. *Michigan Business & Entrepreneurial Law Review*, 6(1): 109-126.

Madeira, C 2018. Explaining the cyclical volatility of consumer debt risk using a heterogeneous agents model: The case of Chile. *Journal of Financial Stability*, 39: 209-220.

Magali, J.J. 2014. Effectiveness of Loan Portfolio Manangement in Rural SACCOS: Evidence from Tanzania. *Business and Economical Research*, 4 (1): 299-318.

Maiti, M. 2019. Is idiosyncratic risk ignored in asset pricing: Sri Lankan evidence? *Future Business* Journal, 5(5): 1-12.

Mamo, A. Q. 2011. Applicability of Altman (1968) Model in Predicting Financial Distress of Commercial Banks in Kenya. MBA thesis. Business University of Nairobi, Nairobi.

Martens, D., van Gestel, T., De Bdcker, M., Haesen, R., Vanthienen, J. & Baesens, B. 2010. Credit rating prediction using ant colony optimization. *Journal of the Operational Research Society*, 61(72): 561-573.

Maseko, N. & Manyani, O. 2011. Accounting Practices of SMEs in Zimbabwe: An Investigative study of Records Keeping for Performance Measurement (A case Study of Bindura). *Journal of Accounting and Taxation*, 3(8): 171-181

Mays, E. 2004. *Credit scoring for risk managers: The handbook for lenders*. Thomson/South-Western, OH, USA.

McNab, H. & Wynn, A. 2000. *Principles and practice of consumer credit risk management*. Chartered Institute of Bankers Publishing, Canterbury.

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

Merton, R.C. 1974. On the pricing of corporate debt: the risk structure of interest rates. *Journal of Finance* 29:449-70.

Merwin, C. 2017. Financing small corporations in five manufacturing industries, 1926-1936. New York: *Open Journal of Business and Management*, 6(1): 1-6.

Mester, L. 1997. What's the point of credit scoring? *Federal Reserve Bank of Philadelphia Business Review:* 3-16.

Mileris, R. 2012. The Effects of Macroeconomic conditions on loan portfolio credit risk and banking system interest income. *Ekonomika*, 91(3): 85-100.

Mills, K. G. & McCarthy, B. 2016. *The State of Small Business lending: Innovation & Technology and the Implications of Regulation*. Harvard Business School. Working Paper 17-042.

Mlachila, M., Dykes., Zajc, S., Aithnard, P.H., Beck, T., Ncube, N. & Nelvin, O. 2013. Banking in sub-Saharan Africa: Challenges and opportunities. *Regional Studies and Roundtables, European Investment Bank (EIB),* Luxembourg.

Mok, J.M. 2009. Reject Inference in Credit Scoring, BMI paper. Findings from the National Survey of Small Business Finances. *Small Business Economics, 27*(2): 157-168.

Moradi, S. & Rafiei, F.M. 2019. A Dynamic Credit Risk assessment model with data mining techniques: Evidence from Iranian Banks. *Financial Innovation,* 5(15): 1-27.

Mullineux, A.W. & Murinde, V. 2014. Financial Sector Policies for Enterprise Development in Africa. *Review of Development Finance,* 4(2): 66-77.

Muriithi, S.M. 2017. African Small and Medium Enterprises (SMEs) contributions, challenges and solutions. *European Journal of Research and Reflections in Management Science,*5(1): 36-48.

Muritala, T. & Taiwo, A.S. 2013. Credit Management Spur Higher Profitability: Evidence from Nigerian Banking sector. *Journal of Applied Economics and Business,* 1(2): 46-53.

Mustafa, K. & Perssons, V. 2017. *Credit Risk Model for loans to SMEs in Sweden*. Master's thesis. Umea University. Umea, Sweden.

Mutezo, A.T. 2015. *Small and Medium Enterprise Financing and Credit Rationing: The Role of Banks in South Africa*. PhD Thesis. University of South Africa. Pretoria.

Muthinja, M.M. 2016. *Financial Innovations and Bank Performance in Kenya: Evidence from Branchless Banking Models*. PhD in Finance thesis. University of Witwatersrand, Johannesburg.

Nehale, F.M. 2016. *Financing Small and Medium Enterprises in Developing Countries*. The 2016 WEI International Academic Conference Proceedings, Boston, USA: 169-177.

Nemoto, N., Yoshino, N., Okubo, Y., Inaba, D. & Yanagisawa, K. 2018. *Credit Risk Reduction Effect on Small and Medium-Sized Enterprise Finance through the Use of Bank Account Information.* ADBI Working Paper 857. Tokyo: Asian Development Bank Institute.

Newman, D. 2014. Missing Data. *Organisational Research Methods,* 17(4): 372-411.

Ngwa, E. 2010. *Credit Risk Management in Banks as Participants in Financial Markets*. Masters Dissertation. Umea School of Business, Umea Universiteit.

Nguyen, H-T. 2016. *Reject Inference in Application scorecards evidence from France.* Document de Travail Working Paper UMR 7235. Universite de Paris Ouest Nanterre La Defense.

Nie, N .H., Hull, C. H., Jenkins, J. G., Steinbrenner, K. & Brent, D. H. 1975. *SPSS,* 2nd edition. McGraw-Hill, New York.

Nkonge, B.K. 2013. Challenges faced by SME suppliers when bidding for tenders: A Case of Thika District. *International Journal of Academic Research in Business and Social Sciences*, 3(12): 194-220.

Novales, A. & Chamizo, A. 2019. Splitting Credit Risk into Systemic, Sectorial and Idiosyncratic Components. *Journal of Risk and Financial Management,* 12(129): 1-33.

Novianti, P. W., Jong, V. L., Roes, K. C. & Eijkemans, M. J. C. 2015. Factors affecting the accuracy of a class of prediction models in gene expression data. *BMC Bioinformatics*, 16(2015): 199-210.

Nyamwanza, T. 2014. *Strategy Implementation for Survival and Growth among Small to Medium-sized Enterprises (SMES) in Zimbabwe.* PhD Thesis. Midlands State University. Gweru, Zimbabwe.

Nyankomo, M. 2014. Micro, Small and Medium Enterprises' External Financing Challenges: The Role of Formal Financial Institutions and Development Finance Intervention in Tanzania. *International Journal of Trade, Economics and Finance*, 5 (3): 230-234.

OECD. 2007. *SME financing gap: theory and practices*. Paris: Organisation for Economic Co-operation and Development.

OECD. 2015. *New Approaches to SME and Entrepreneurship financing: Broadening the range of instrument*. Paris: Organisation for Economic Co-operation and Development.

OECD. 2017. *Enhancing the Contributions of SMEs in a Global and Digitalised Economy*. Meeting of the OECD Council at Ministerial Level. Paris: 7-8th June 2017.

OECD. 2018. *Financial Markets, Insurance and Private Pensions: Digitalisation and Finance*. Paris: Organisation for Economic Co-operation and Development.

OECD. 2018. *Enhancing SME access to diversified financing Instruments.* Plenary Session 2. SME Ministerial Conference. Mexico City. Organisation for Economic Co-operation and Development. 22-23 February 2018.

Ogechukwu, O.L., Akunio, A. & Goldman, G.A. 2015. Financial Schenesto boost SMEs- Sources of Finance by the Nigerian Government: a commentary. *Banks and Bank Systems,* 10(3): 49-60.

Ohene, J. 2015. *An Examination of the Credit Management Practices of Rural Banks: Case study of Asokore Rural Bank Limited*. MBA thesis, Nkwame Nkrumah University of Science and Technology, Accra.

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

Ojo, A.T. 2003. *Partnership and Strategic Alliance Effective SME Development.* Small and Medium Enterprises Development and SMIEIS: Effective Implementation Strategies, CIBN Press Ltd, Lagos.

Oppong, M., Owiredu, A. & Churchill, R.Q. 2014. Micro and Small-scale Enterprises Development in Ghana. *European Journal of Accounting, Auditing and Finance Research*, 2(6): 84-97.

Osano, H. M. & Languitone, H. 2016. Factors influencing access to finance by SMEs in Mozambique: Case of SMEs in Maputo central business district. *Journal of Innovation and Entrepreneurship*, 5(13): 1-16.

Ozili, P.K. & Outa, E. 2017. Bank Loan loss provisions research: A review. *Borsa Istanbul Review*, 17(3): 144-163.

Padachi, M., Howorth, C. & Narasimhan, M.S. 2012. Working Capital Financing Preferences: The case of Mauritian Manufacturing Small and Medium Enterprises (SMEs). *Asian Academy of management Journal of Accounting and Finance.* 8(1): 125-157.

Petersen, M. A. & Rajan, R. G. 1994. The benefits of lending relationships: evidence from small business data. *Journal of Finance*, 49(1): 3-37.

Pratt, A.C. & Virani, T.E. 2015. *The Creative SME: a cautionary tale.* Creative works London Working Paper Series Number 14. Arts and Humanities Research Council.

Pretorius, M. 2008. Critical Variables of Business failure: A review and classification framework. *South African Journal of Economic and Management Sciences (SAJEMS)*, 11(4): 408-430.

Radmehr, E. & Bazmara, M. 2017. A Survey on Business Intelligence Solutions in Banking Industry and Big Data Applications. *International Journal of Mechtronics, Electrical and Computer Technology*, 7(23): 3280-3298.

Rahman A., Rahman M.T. & Ključnikov A. 2016. Collateral and SME financing in Bangladesh: an analysis across bank size and bank ownership types, *Journal of International Studies*, 9(2): 112-126.

Ramlee, S. & Berma, M. 2013. Financing gap in Malaysian small-medium enterprises: A supply-side perspective. *South African Journal of Economic and Management Sciences*, 16(5): 115-126.

Sebastiani, P. & Ramoni, M. 2000. Bayesian inference with missing data using bound and collapse. *Journal of Computational and Graphical Statistics*, 9(4): 779-800.

Ranganathan, P., Pramesh, C.S. & Aggarwal, R. 2017. Common Pitfalls in Statistical Analysis: Logistic Regression. *Perspective in Clinical Research*, 8(3): 148-151.

Rankhumise, E.M. & Letsoalo, M.E. 2019. Owners' Perspective of factors associated with performances of Small and Medium Enterprises. *International Journal of Entrepreneurship*, 23(3): 1-17.

Rao, P.K. 2003. *Development finance*, Springer, Heidelberg.

Reeg, C. 2013. *Micro, Small and Medium Enterprise upgrading in low-and-middle-income countries: A literature review.* German Development Institute, Discussion Paper 15/2013.

Rehman, Z.U., Muhammad, N., Sarwar, B. & Raz, M.A. 2019. Impact of risk management strategies on the credit risk faced by commercial banks of Balochistan. *Financial Innovation*, 5(44): 1-13.

Reserve Bank of Zimbabwe (RBZ). 2011. *Technical Guidance on the Implementation of the revised Capital Adequacy Framework in Zimbabwe*. Harare, Zimbabwe (Guideline No. 1-2011/BSD).

Rezende, F.F., da Silva Montezano, R., de Oliviera, F.N. & de Jesus Lameira, V. 2017. Predicting financial distress in publicly traded Companies. *Revista Contabilidade & Financias*, 28(75): 390-406.

Rhoads, C.H. 2012. Problems with Tests of the Missingness Mechanisms in Quantitative Policy Studies. *Statistics, Politics and Policy*, 3(6):1-23.

Rocha, R., Farazi, S., Khouri, R., & Pearce, D. 2011. *The Status of Bank Lending to SMEs in the Middle East and North Africa Region: Results of a Joint Survey of the Union of Arab Bank and the World Bank.* World Bank Policy Research Working Paper 5607: The World Bank.

Roelofs, R. 2019. *Measuring generalisation and overfitting in Machine Learning*. PhD thesis. University of California. Berkeley, USA.

Roggi, O. 2015. *Risk, Value and Default.* Volume of the World Scientific Series in Finance. World Scientific, 2015.

Roggi, O. & Altman, E.I. 2013. *Managing and Measuring of Risk: Emerging Global Standards and Regulations After the Financial Crisis*. Volume of the World Scientific Series in Finance. World Scientific, 2013.

Rommer, A.D. 2005. *Accounting based credit-scoring Models: Econometric Investigations*. PhD thesis. University of Copenhagen, Copenhagen.

Ross, A.S. 2015. *The Impact of Home Loan, Keep facts Sheets on Borrowers' Comparisons of Loan Costs.* PhD Thesis. Queensland University of Technology. Queensland.

Ross, G., Ligang, S. & Cai, F. 2018. *China's 40 years of reform and development: 1978-2018*, China Update Book Series: ANU Press, Acton.

Roth, P. 1994. Missing Data: Conceptual Review for Applied Psychologists, *Personnel Psychology*, 47: 537-560.

Rubin, D. B. 1976. Inference and Missing Data, *Biometrika*, 63: 581-592.

Saeed, M.S. & Zahid, N. 2016. The Impact of Credit Risk on Profitability of the Commercial Banks. *Journal of Business & Financial Affairs,* 5(2): 192-199.

Sasaki, Y. 2007. The Truth of the F-measure. School of Computer Science, University of Manchester.

Saunders, A. & Allen, L. 2002. *Credit Risk Measurement*. New York: Wiley & Sons.

Saurina, J. & Trucharte, C. 2004. The impact of Basel II on lending to small and medium-sized firms: A regulatory policy assessment based on Spanish credit register data. *Journal of Financial Services Research*, 26: 121-144.

Sebastiani, P. & Ramoni, M. 2000. Bayesian inference with missing data using bound and collapse. *Journal of Computational and Graphical Statistic*s, 9(4): 779-800.

Seliane, T.N. & Sello, M.N. 2015. *The architecture of the Basel Accords: Perspectives on Evolution and Adaptation in the context of Lesotho.* Working Paper No. WP01/15. Supervision Department, Central Bank of Lesotho, Maseru.

Schmuklers, S., de la Torre, A & Martinez, P.M. 2008. Bank Involvement with SMEs: Beyond Relationship Lending. *Journal of Banking and Finance,* 34: 2280-2293.

Schoen, E.J. 2017. The 2007-2009 Financial Crisis: An Erosion of Ethics: A Case Study. *Journal of Business Ethics,* 146: 805-830.

Schouten, R.M., Lugtig, P. & Vink, G. 2018. Generating Missing values for Simulation purposes: a multivariate computation procedure. *Journal of Statistical computation & Simulation*, 88(15): 2909-2930.

Schwaiger, W.S.A. 2002. *Basel II: Quantitative Impact Study on Austrian Small and Medium-sized Enterprises*. Technical University of Vienna.

Serov, V. 2017. *Revisting Credit Risk Assessment of Small and Medium Enterprises.* Masters thesis. The British University. Dubai.

Sharma, P. & Gounder, N. 2012. Obstacles to Bank Financing of Micro and Small Enterprises: Empirical Evidence from The Pacific with some Policy Implications. *Asia Pacific Journal,* 19(2): 49-75.

Shen, S-W., Nguyen, T-D. & Ojiake, U. 2013. Modelling the predictive performance of credit scoring. *Acta Commercii*, 13(1): 1-12.

Shinozaki, S. 2014. *Capital market financing for SMEs: a growing need in emerging Asia*, Asian Development Bank, Manila, Philippines.

Sibanda, K, Hove, S. P. & Shava, H. 2018. The Impact of Small and Medium Enterprises access to Finance and Performance on Exporting Behaviour at firm level. A case of furniture Manufacturing SMEs in Zimbabwe. *Actaa Commercii,* 18(1): 1-13.

Siddiqi, N. 2006. *Credit risk scorecards: Developing and implementing intelligent credit scoring*. Hoboken, NJ, USA: John Wiley & Sons.

Simpasaand, A. & Pla, L. 2016. *Sectorial Credit Concentration and Bank Performance in Zambia.* African Development Bank Group Working Paper No. 245. December 2016.

Sitharam, S. & Hoque, M. 2016. Factors affecting the performance of small and medium enterprises in KwaZulu-Natal, South Africa. *Problems and Perspectives in Management,* 14(2-2), 277-288.

Smorfitt, R. 2009. *SMEs in South Africa: why is finance difficult to access?* [Online] Available: http://innoveur.blogspot.com [accessed 23 November 2018].

Smith, A. & Elkan, C. 2004. *A Bayesian network framework for reject inference.* In: Proceedings of the tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining: 286-295.

Smith, R. and A. Winakor. 1935. Changes in Financial Structure of Unsuccessful Industrial Corporations. Bureau of Business Research, Bulletin No. 51. Urbana: University of Illinois Press.

Song, Y. 2016. *Performance Management in Chinese Commercial Banks.* PhD thesis. Kent Business School. University of Kent, Kent.

Sousa, M. R., Gama, J. & Brandao, E. 2016. A dynamic modelling Framework for Credit risk assessment. *Expert Systems with Applications,* 45: 341-351.

Spiegelhalter D. J. & Lauritzen S. L. 1990. Sequential Updating of Conditional Probabilities on Directed Graphical Structures. *Networks*, 20: 579-605.

Spuchľáková, E., Valašková, K. & Adamko, 2015. The credit risk and its measurement: Hedging and Monitoring. *Procedia Economic and Finance*, 24: 675-681.

Statistics South Africa. 2006. *Labour Force Survey, September 2005*. Available at: www.statssa.gov.za [Accessed 25 June 2017].

Stein, P., Goland, T. & Schiff, R. 2010. *Two Trillion and Counting: Assessing the Credit Gap for Micro, Small, and Medium Enterprises in the Developing World*. Washington DC Publisher: International Finance Corporation/McKinsey & Company.

Stepanyan, E. 2018. A Survey on Loanwards and Borrowings and their roles in the Reflection of Cultural Values and Democracy Developments: The American Paradigm. *European Journal of Marketing and Economics*, 1(2): 77-86.

St-Pierre, J. & Bahri, M. 2011. The determinants of risk premium: the case of bank lines of credit granted to SMEs. *Journal of Developmental Entrepreneurship*, 16(4): 459-476.

Sudhakar, M. & Reddy, C. V. K. 2016. Two steps credit risk assessment model for retail bank loan applications using Decision Tree Data Mining Technique. *International Journal of Advanced Research in Computer Engineering and Technology,* 5(3): 705-718.

Taiwo, J.N. & Falohun, T.O. 2016. SMES Financing and its effects on Nigerian Economic Growth. *European Journal of Business, Economics and Accountancy*, 4(4): 37-54.

Teddlie, C. & Tashakkori, A. 2003. *Handbook of mixed methods in social & behavioral research*, Sage Publications, Inc., Thousand Oaks, CA.

Thomas, L. C. 2009. Operations research in consumer finance: Challenges for operational research. *Journal of Operational Research Society*, 61: 41-52.

Thomas, L. C. 2000. A Survey of Credit and Behavioral Scoring: Forecasting Financial Risk of Lending to Consumers. *International Journal of Forecasting*, 16(2): 149-172.

Tsai, S. B., Li, G., Wu, C. H., Zheng, Y. & Wang, J. 2016. An Empirical Research on evaluating banks credit assessment of corporate customers. *SpringerPlus*, 5(2088): 1-13.

Tserng, H.P., Chen, C-Y., Huang, C-W., Tran, Q.H., Lei, M.C. & Zhang, L. 2014. Prediction of default probability for construction firms. *Journal of Civil Engineering and Management*, 20(2): 247-255.

Tumkella, K. 2003. *The Challenge of Globalisation and SME Sector in Nigeria: Repositioning through Technology and Innovation.* Paper presented at National Summit on SMIEIS organised by the Bankers' Committee and Lagos Chambers of Commerce and Industry (LCCI), Lagos.

Tursory, T. 2018. *Risk Management processing in Banking Industry.* Munich Personal RePec Archive (MPRA) paper No. 86427 posted 02 May 2018.

Udell, G.F. 2004. *SME lending: defining the issues in a global perspective.* Working Paper.

Vabalas, A., Gowen, E., Poliakoff, E. & Casson, A. 2019. Machine Learning algorithm of validation with limited sample size. *PLOS ONE*, 14(11): 1-20.

Vandenberg, P., Chantapacdepong, P. & Yoshino, N. 2016. *SMEs in Developing Asia: New Approaches to Overcoming Market Failures*. Asian Development Bank Institute.

Van der Meijs, A. 2018. *Missing Data Imputation: Predicting Missing Values*. Master's Thesis, Tilburg University.

Varum, C. & Rocha, V. 2013. Employment and SMEs during crises. *Small Business Economics,* 40(1): 9-25.

Veiga, M.G. & McCahery, J.A. 2019. Financing of Small and Medium Enterprises (SMEs): An Analysis of the financing Gap in Brazil. *European Business Organisation Law Review*, 20: 633-664.

Verstraeten, G. & van den Poel, D. 2005. The impact of sample bias on consumer credit scoring performance and profitability. *Journal of the Operational Research Society,* 56(8): 981-992.

Uzoigwe, D.C. 2007. *Economic Development in Nigeria Through Agriculture, Manufacturing and Mining Sectors: An Econometric Approach*. PhD. University of Pretoria, Pretoria.

Wagenvoort, R. 2003. SME finance in Europe: Introduction and overview. *EIB Paper*, 8: 11-20.

Wagner, N. 2008. *Credit Risk Models, Derivatives and Management*. CRC: Press Taylor & Francis Group.

Wang, Y. 2011. *Corporate Default Prediction: Model Drivers and Measurement.* PhD thesis. University of Exeter.

Wang, Y. 2013. *Credit Risk Management in Rural Banks in China.* PhD thesis. Edinburgh Napier University.

Wangmo, C. 2015. Small Medium Enterprise (SME) bank financing constraints in developing countries: a case study of Bhutan. *International Journal of Arts and Sciences*, 8(5): 569-590.

Wattanapruttipaisan, T. 2003. Four Proposals for Improved Financing of SME Development in Asian. *Asian Development Review*, 20(2): 1-45.

Wehinger, G. 2013. SMEs and Credit crunch: Current Financing Difficulties, Policy measures and A review of Literature. *OECD Journal. Financial Market Trends,* 2013(2): 115-148.

Wojakowski, R., Ebrahim, M.S., Jaafar, A. & Salleh, M.O. 2019. Can Loan Valuation Adjustment (LVA) Approach Immunize collateralised debt from defaults. *Financial Markets, Institutions & Instruments*, 28: 141-158.

Wolfe, S. & Amidu, M. 2012. The impact of market power and funding strategy on bank-interest margins. *European Journal of Finance*, 19(9): 888-908.

Wood, A.P. 2012. The Performance of Insolvency Prediction and credit risk models in the UK: A Comparative study, development and wider application. PhD thesis. University of Exeter, Exter.

WorldBank. 2005. *Credit risk measurement under Basel II: An overview and implementation issues for developing countries.* World Bank Policy Research (Paper No. 3556).

Wu, Y., Gaunt, C. & Gray, S. 2010. A Comparison of Alternative Bankruptcy Prediction Models. *Journal of Contemporary Accounting and Economics*, 6(1): 34-45.

Xu, Y. & Goodacre, R. 2018. On splitting Training and Validation set: A comparative study of cross-validation Bootstrap and Systematic Sampling for estimating the generalisation performance of supervised learning. *Journal of Analysis and Testing*, 2: 249-262.

Yeung, G. 2009. How Banks in China Make Lending decisions. *Journal of Contemporary China*. 18(59): 285-302.

Yi, J.M. 2017. *Identification of Relevant Predictors of Loan Defaults using the Elastic Net Model.* PhD thesis. University of Adelaide, Adelaide.

Yi, J. 2019. Corporate Distress Prediction in China: A Machine Learning Approach. *Accounting and Finance,* 58 (4): 1063-1109.

Yoshino, N. 2013. *The Background of Hometown Investment Trust Funds. In Hometown Investment TrustFunds: A Stable Way to Supply Risk Capital*, N. Yoshino and S. Kaji, eds. Tokyo: Springer.

Yoshino, N. & Taghizaden-Hesary, F. 2016. *Major Challenges facing SMEs in Asia and Solutions for Mitigating them.* ADBI Working Paper Series No. 564 (April 2016).

Yoshino, N. & Taghizaden-Hesary, F. 2017. *Solutions for Small and Medium Enterprises (SMEs) Difficulties in Accessing Finance*. Asian Experiences. Asian Development Bank Institute (ADBI) Working paper Series No. 768 (August 2017).

Yoshino, N. & Taghizadeh-Hesary, F., 2014. Analytical Framework on Credit Risks for Financing Small and Medium-sized Enterprises in Asia. *Asia-Pacific Development Journal*. 21(2): 1-22.

Yoshino, N., Taghizadeh-Hesary, F. Charoensivakorn, P. & Niraula, B. 2016. Small and Medium-sized Enterprise Credit Risk Analysis Using Bank Lending Data: An Analysis of Thai SMEs. *Journal of Comparative Asian Development*, 15(3): 383-406.

Yoshino, N. & Taghizadeh-Hesary, F. 2016. Japan's Lost Decade. In: eds. *Lessons for Asian Economy:* Tokyo: Japan. Springer.

Yoshino, N. & Yamagami, H. 2017. *Monetary Economics: Practice and Theory (in Japanese)*. Keio University Press. Tokyo. Japan.

Yurdakul, F. 2014. Macroeconomic Modelling of Credit Risk for Banks. *Procedia - Social and Behavioural Sciences,* 109: 784-793.

Zhou, L., Lai, K.K., & Yu, L. 2010. Least squares support vector machines ensemble models for credit scoring. *Expert Systems with Applications*, 37(1): 127-133.

Zmijewski, M.E. 1984. Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research*, 22: 59-86.

APPENDIX A: BANK SURVEY RESULTS

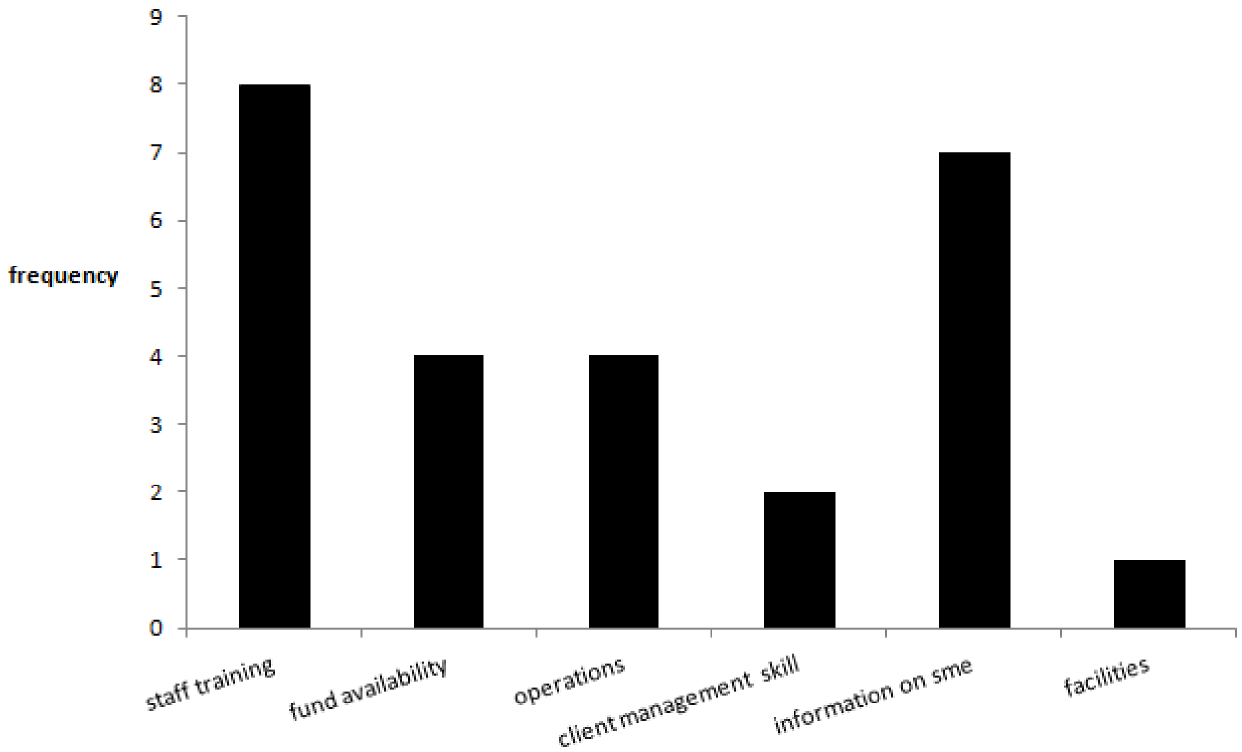
Variable	Category	Frequency	Proportion (%)
What type of financial service do you provide to SMEs?	short-term	3	37.5
	long-term	1	12.5
	Overdraft	4	50.0

Involvement of govt assessment

Variable	Category	Frequency	Proportion (%)
Would you say it has	improved a lot	3	37.5
	improved a little	2	25.0
	stay about the same	2	25.0
	got a bit worse	1	12.5

1a) Problems/Challenges encountered in dealing with SME Sector.

Variable	Frequency	Percentage (%)
Staff training and capacity building	8	50.0
Fund Availability	4	25.0
Operations	4	25.0
Client management skill	2	12.5
Information on MSME	7	43.8
Facilities	1	6.3



1b) Areas considered as a challenge to the Bank in dealing with SMEs

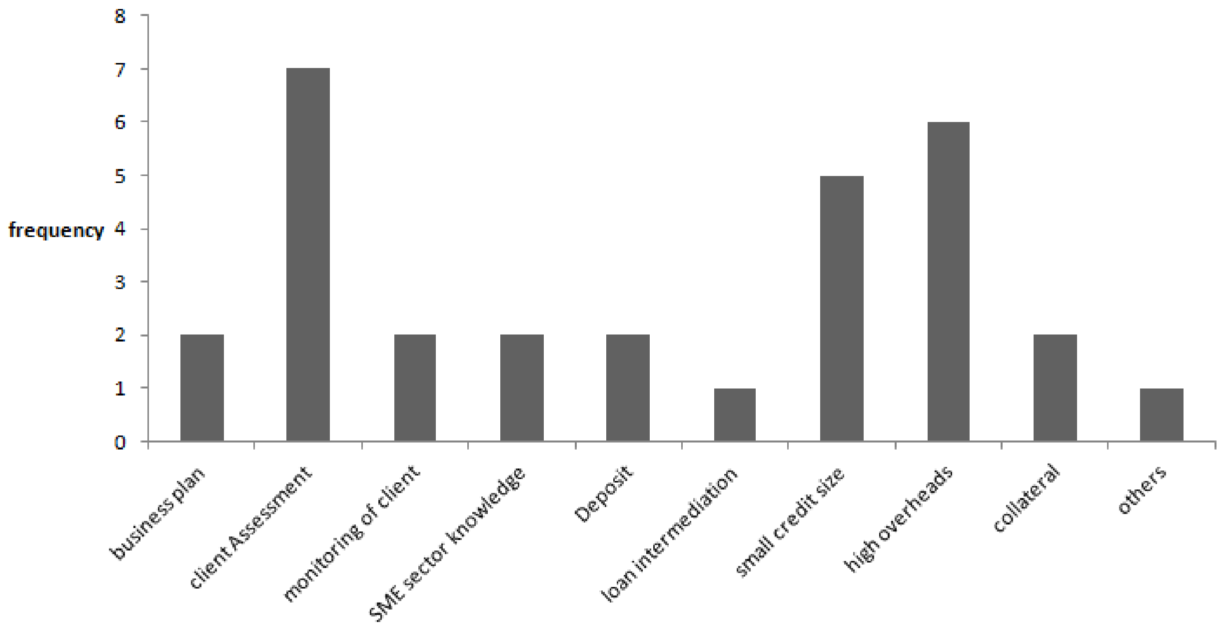
Variable	Frequency	Percentage (%)
Staff training and capacity building	4	25.0
Fund Availability	5	31.3
Operations	5	31.3
Client management skill	3	18.8
Information on MSME	6	37.5
Facilities	4	25.0

2a) Operational Constraints experienced in effective delivery of services to SMEs

Operational Constraints	Frequency	Percentage (%)
Business plan/ feasibility study	6	37.5
Client assessment	2	12.5
Monitoring of client	6	37.5
MSME sector knowledge	7	43.8
Fund or deposit	2	12.5
Loan intermediation between commercial and community banks	4	25.0
Small credit size	4	25.0
High overheads costs	4	25.0
Collateral	4	25.0
Others	4	25.0

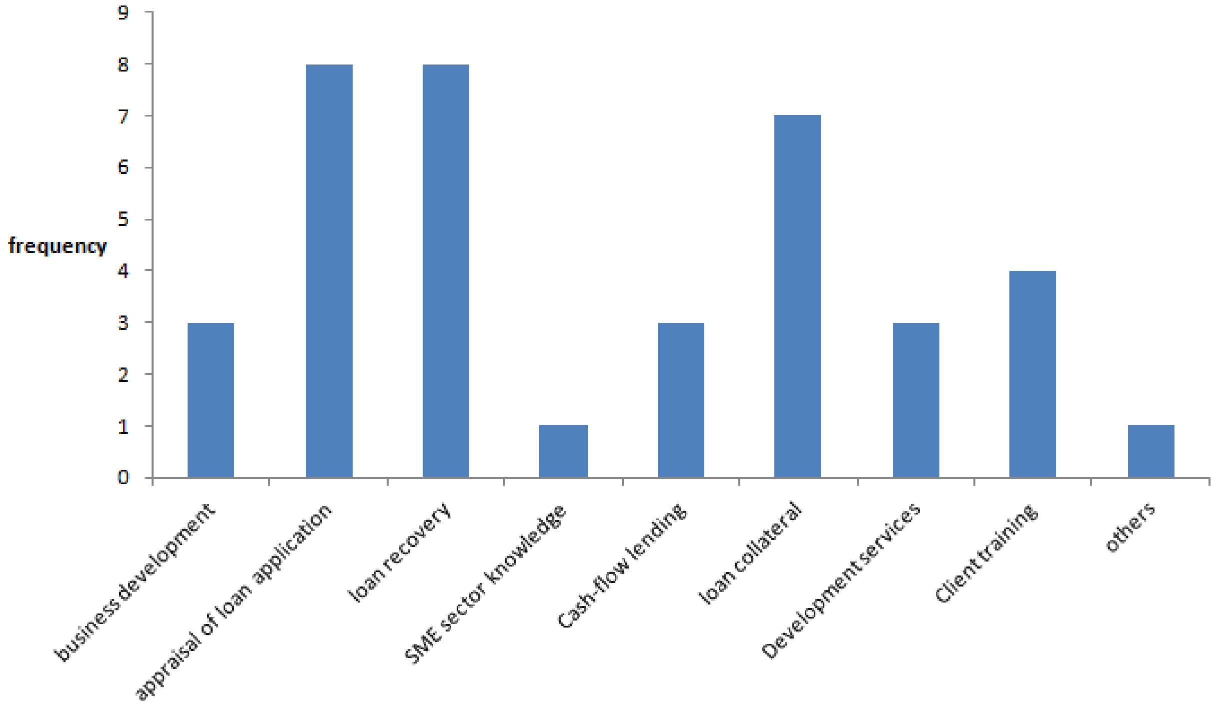
b) The Major operational constraints experienced in effective delivery of SME services

Operational Constraints	Frequency	Percentage (%)
Business plan/ feasibility study	2	12.5
Client assessment	7	43.8
Monitoring of client	2	12.5
MSME sector knowledge	2	12.5
Fund or deposit	2	12.5
Loan intermediation between commercial and community banks	1	6.3
Small credit size	5	31.3
High overheads costs	6	37.5
Collateral	2	12.5
Others	1	6.3



3a) Specific Training Areas in dealing with SMEs

Variable	Frequency	Percentage
Business development	3	18.8
Appraisal of loan application	8	50.0
Loan recovery	8	50.0
SME sector knowledge	1	6.3
Cash flow lending	3	18.8
Loan collateral	7	43.8
Business development service	3	18.8
Capacity building	4	25.0
Others	1	6.3

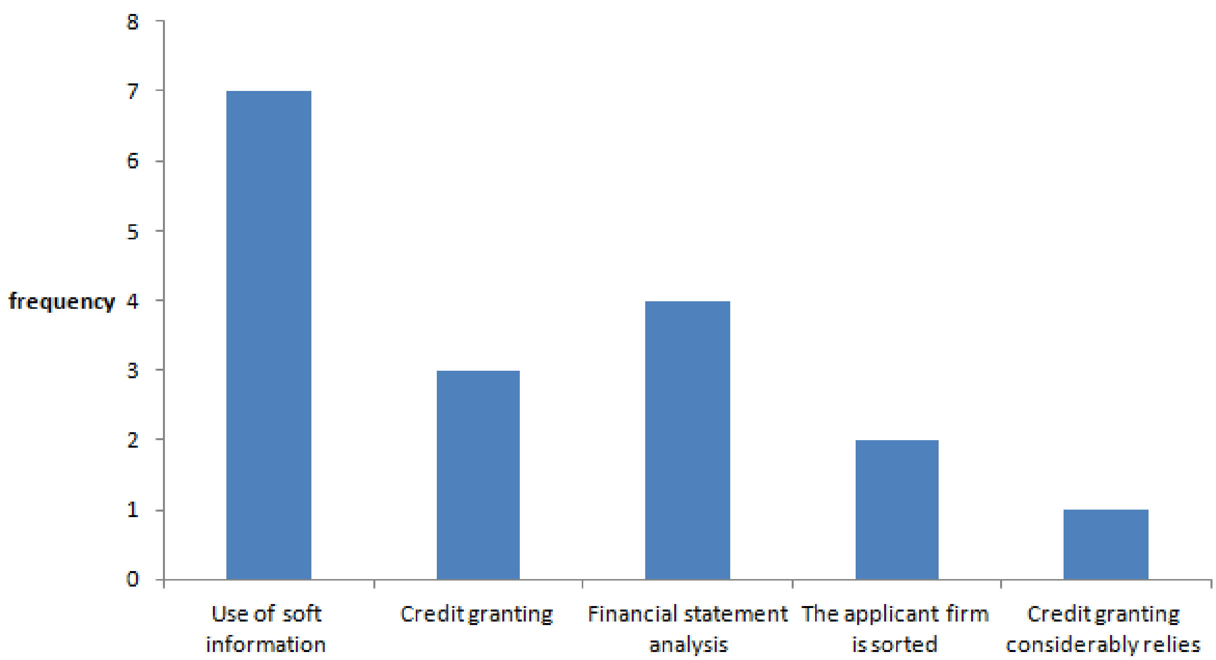


3b) Important Training Areas to strengthen capacity of Banks in serving SMES

Variable	Frequency	Percentage
Business development	2	12.5
Appraisal of loan application	5	31.3
Loan recovery	4	25.0
MSME sector knowledge	3	18.8
Cash flow lending	2	12.5
Loan collateral	6	37.5
Business development service	2	12.5
Capacity building	2	12.5
Others	2	12.5

C5 Techniques used Bank to underwrite SMEs

Variable	Category	Frequency	Proportion (%)
What types of techniques are generally used by your institution?	Use of soft information directly collected by loan officials from past	7	41.2
	Credit granting considerably relies upon external ratings of the applicant firm	3	17.6
	Financial statement analysis with strong focus at the applicant firm level	4	23.5
	The applicant firm is sorted into credit merit cluster by means of quantitative methods based on hard information	2	11.8
	Credit granting considerably relies upon external ratings of the applicant firm	1	5.9

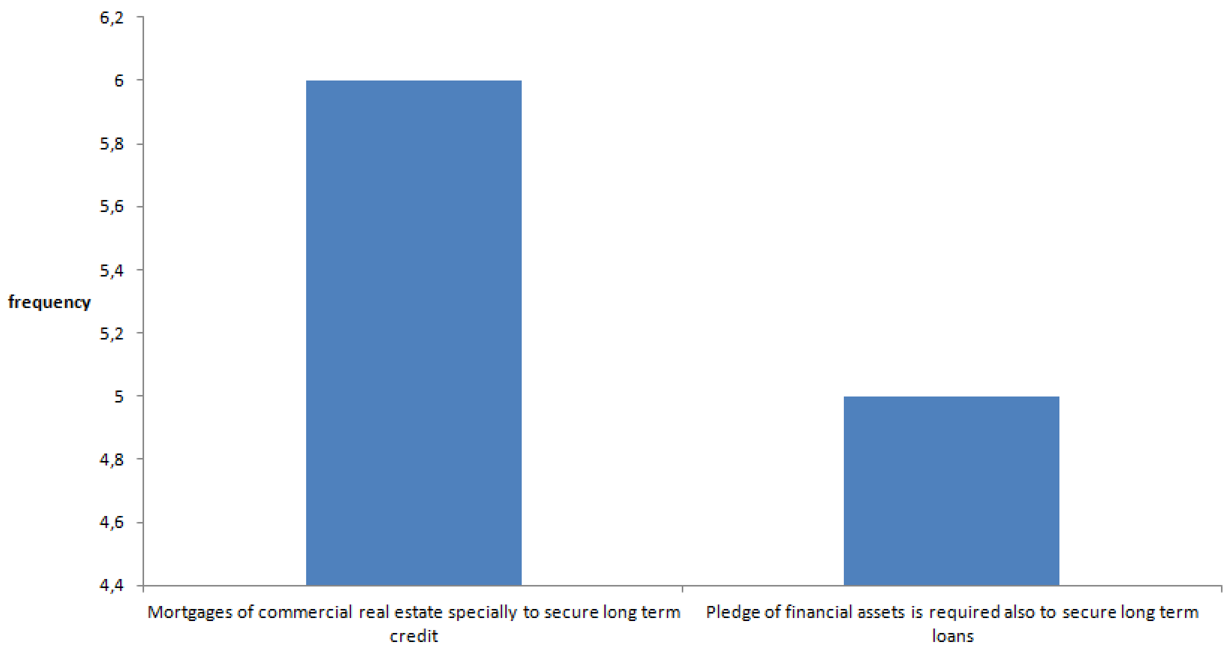


C6 Source of Database for underwriting

Variable	Category	Frequency	Proportion (%)
Creditworthiness by processing hard information in the form of financial statements	Your own bank database	5	55.6
	Public credit	2	22.2
	Private external credit bureaus	2	22.2

C7 Risk mitigation techniques used by the Bank

Variable	Category	Frequency	Proportion (%)
The risk mitigation techniques generally used by your bank to secure exposures to SMEs	Mortgages of commercial real estate specially to secure long-term credit	6	75.0
	Pledge of financial assets is required also to secure long term loans	2	25.0



FA11 Business model: centralised aspects

Variable	Category	Frequency	Proportion (%)
Sales of non-lending products	Only done at headquarters	1	12.5
	Done primarily at branches	3	37.5
	Bank allows for both	4	50.0
Loan pre-screening	Only done at headquarters	1	12.5
	Done primarily at headquarters	2	25.0
	Done primarily at branches	1	12.5
	Bank allows for both	4	50.0
Loan approval	Only done at headquarters	2	25.0
	Done primarily at headquarters	4	50.0
	Bank allows for both	2	25.0
Risk management	Only done at headquarters	1	12.5
	Done primarily at headquarters	2	25.0
	Bank allows for both	5	62.5
Non-performing loan recovery	Only done at headquarters	3	37.5
	Done primarily at headquarters	2	25.0
	Bank allows for both	3	37.5

FA16 Business model: factors used to identify potential SME clients

Variable	Category	Frequency	Proportion (%)
Rely on existing deposit clients	most important	5	62.5
	3rd choice	2	25.0
Use information from existing firm databases	2nd choice	2	25.0
	3rd choice	3	37.5
Attracting clients with bank credit	most important	2	25.0
	2nd choice	1	12.5
	3rd choice	3	37.5
Focus on attracting SMEs that are clients of existing clients	most important	1	12.5
	2nd choice	5	62.5
	3rd choice	1	12.5

Business model: criteria to determine SME clients

Variable	Category	Frequency	Proportion (%)
Company size	Most important	3	75.0
	Important	1	25.0
Geographic area	Not important	3	100.0
Industry Sector	Important	4	100.0
Products needs	Important	1	50.0
	Not important	1	50.0
Expected profitability	Most important	1	50.0
	Not important	1	50.0
Exposure size	Most important	2	100.0
Credit quality	Most important	2	100.0

Business model: products offered

Variable	Category	Frequency	Proportion (%)
Banks indicate whether they offer SME clients standardised or tailored products	Standardised	5	62.5
	Standardised/tailored	2	25.0
	Tailored	1	12.5

Risk management systems

Variable	Category	Frequency	Proportion (%)
Automated	Yes	2	25.0
	No	6	75.0
Done by Analyst	yes	6	75.0
	no	2	25.0
Separated from sales	yes	7	87.5

Variable	Category	Frequency	Proportion (%)
	no	1	12.5
Done at head office	yes	3	37.5
	no	5	62.5

Risk management: collateral (1)

Variable	Category	Frequency	Proportion (%)
If collateral requirements are higher for SME than large corporate, indicate reasons which below apply	SMEs are more unstable	1	12.5
	SMEs are more informal	2	25.0
	SMEs have worse management	2	25.0
	SMEs are harder to evaluate	2	25.0
	SMEs collateral more difficult to seize in case of default	1	12.5

Monitoring methods

Variable	Category	Frequency	Proportion (%)
Which of the following characterize some of the ways the bank monitors credit risk outlook over time for SME?	there preventive triggers that monitored to detect a possible deterioration in the credit outlook of an SME	4	50.0
	some triggers are automatically generated for monitoring purposes	3	37.5
	monitoring of client health relies on the diligence of the relationship manager analyst	1	12.5

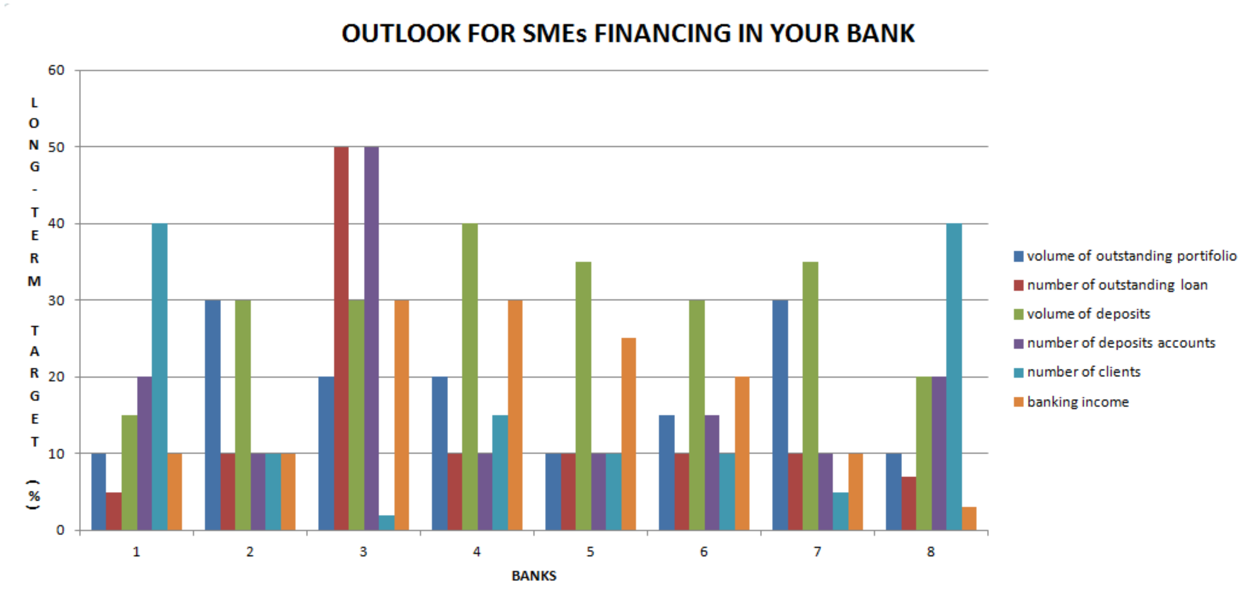
Items monitored

Variable	Category	Frequency	Proportion (%)
Which are of the following items are monitored by the bank?	Evergreens in overdraft	3	37.5
	Switching practices	2	25.0
	Deterioration in the cash flow receivables recovery of the SME	1	12.5
	Total debt outstanding	1	12.5
	Repayment frequency is a means to check on ongoing solvency of the MSME	1	12.5

Drivers of banks' involvement with SMEs

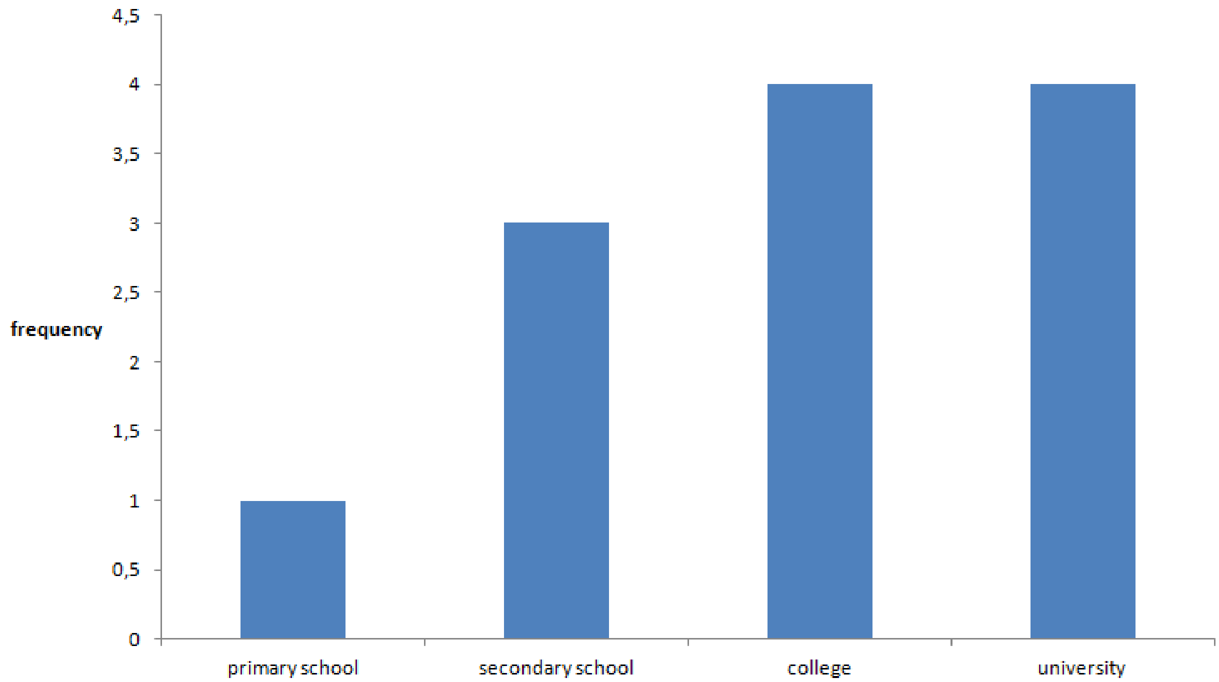
Variable	Category	Frequency	Proportion (%)
Bank need to be involved in a collaborative provision of financial services	Somewhat agree	113	37.5
	Neither agree nor disagree	113	37.5
	Strongly disagree	50	25.0
Bank requires training on technicalities of the provision of financial services	Strongly agree	300	100.0
Bank requires information on types and nature of financial services	Strongly agree	4	50.0
	Somewhat agree	2	25.0
	Neither agree nor disagree	2	25.0
Bank requires information on business	Strongly agree	8	100.0
There is growing demand for SME financial services from banks	Strongly agree	7	87.5
	Somewhat agree	1	12.5
	Strongly agree	6	75.0
	Somewhat agree	1	12.5

Variable	Category	Frequency	Proportion (%)
Banks should consider cash-flow lending option for SME loan request	Neither agree nor disagree	1	12.5
Banks should work through intermediaries	Somewhat agree	3	37.5
	Neither agree nor disagree	3	37.5
	Strongly disagree	2	25.0

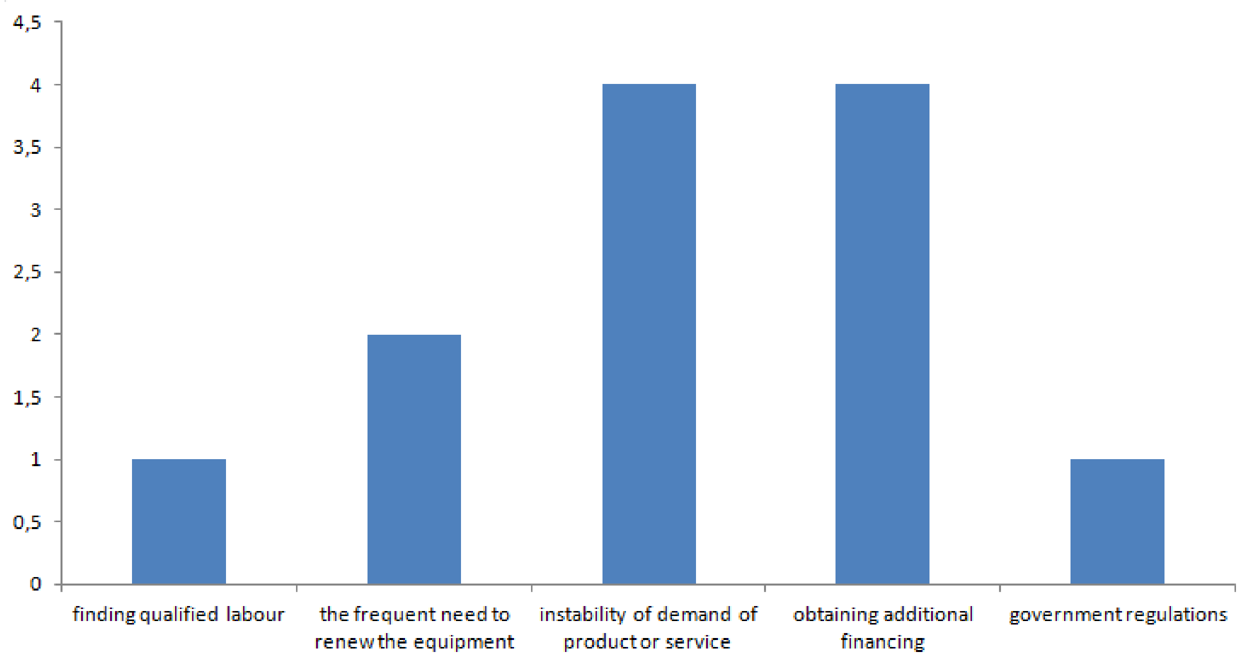


APPENDIX B: SME SURVEY ANALYSIS

Variable	Category	Frequency	Proportion (%)
Highest level of education you completed	Primary school	25	8.3
	Secondary school	75	25.0
	College	100	33.3
	University	100	33.3

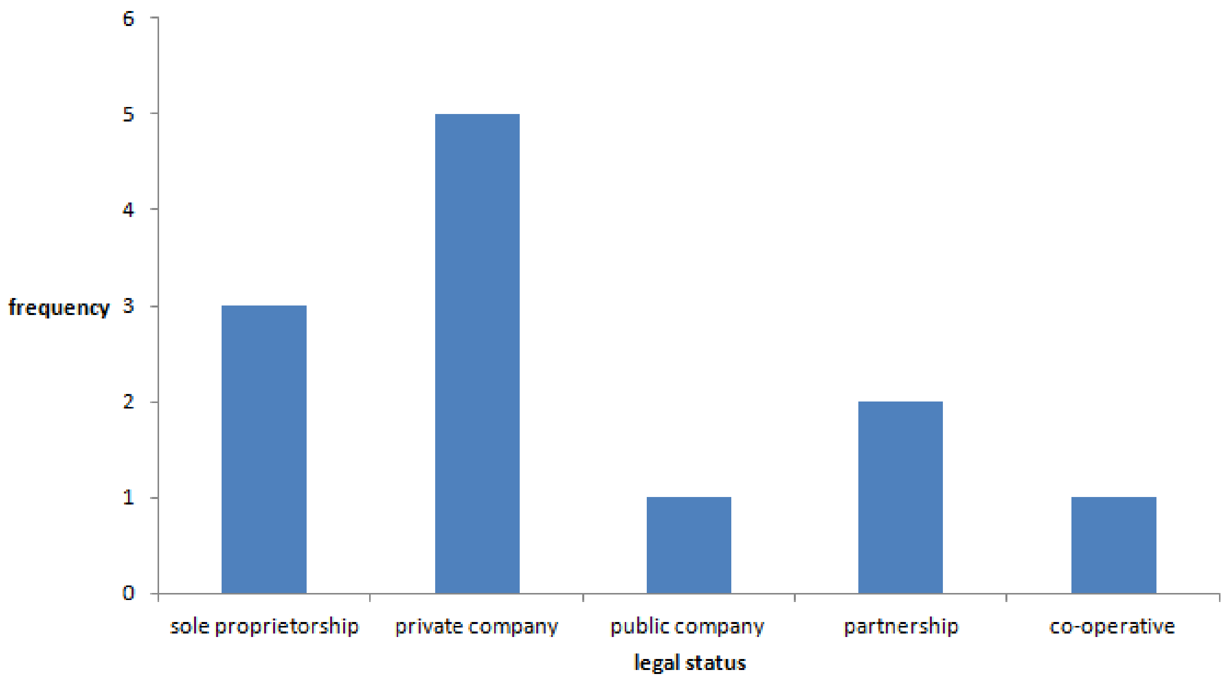


Variable	Category	Frequency	Proportion (%)
The biggest obstacles to the growth of your enterprise	finding qualified labour	25	8.3
	the frequent need to renew the equipment	50	16.7
	instability of demand of product or service	100	33.3
	obtaining additional financing	100	33.3
	government regulations	25	8.3

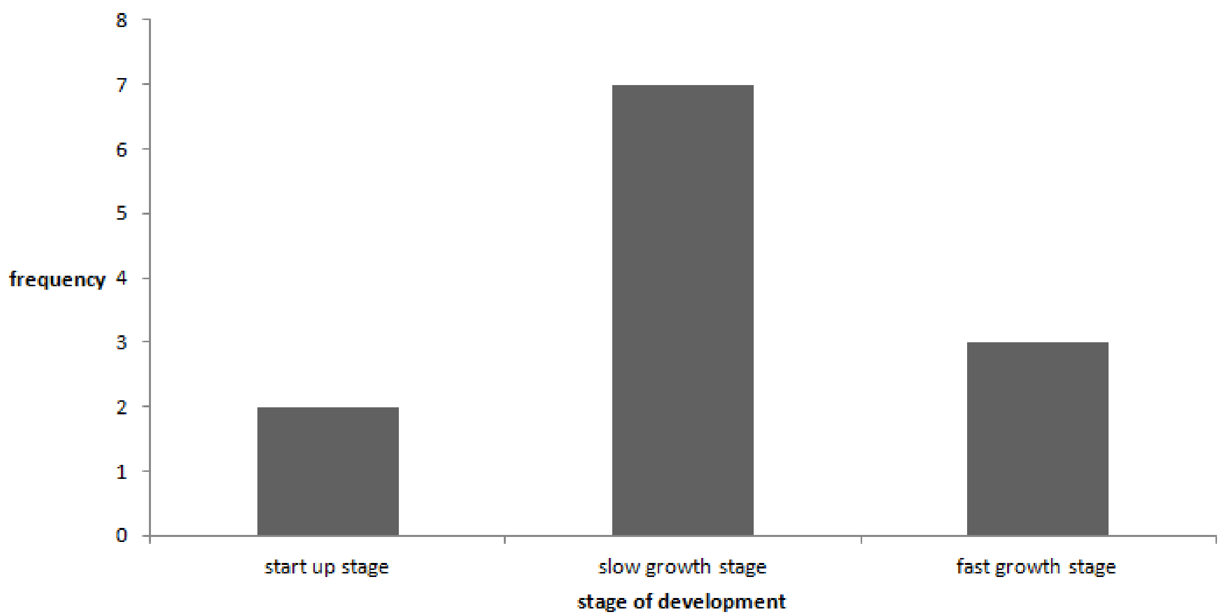


Legal Status

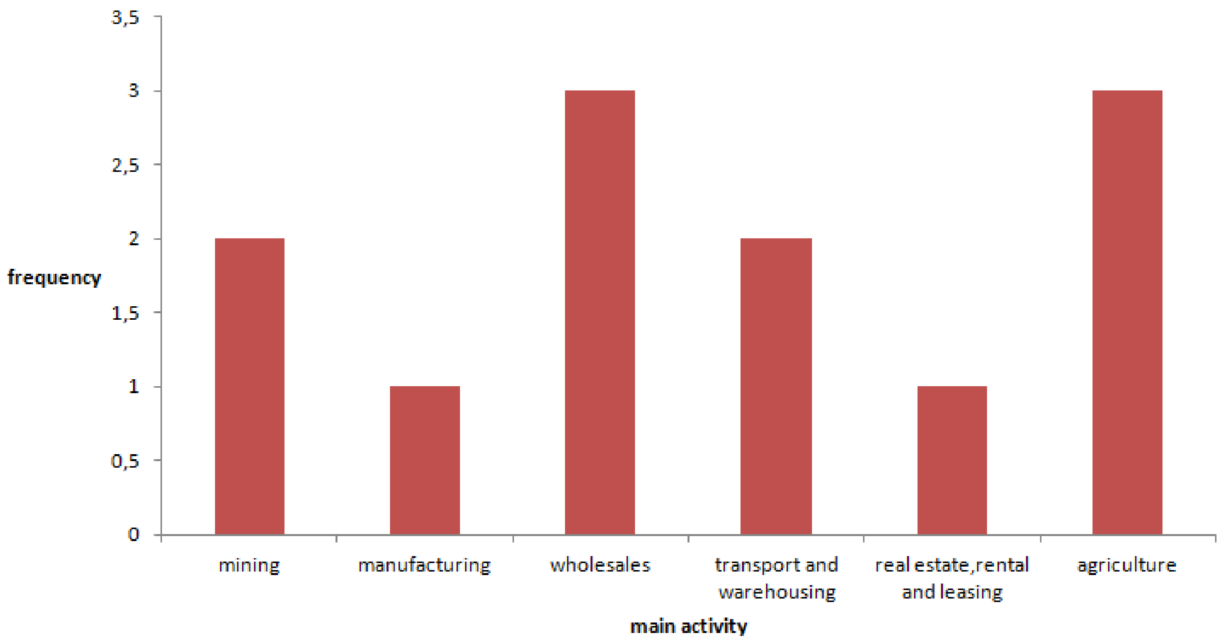
Variable	Category	Frequency	Proportion (%)
What is the business legal status?	sole proprietorship	75	25.0
	private company	125	41.7
	public company	25	8.3
	partnership	50	16.7
	co-operative	25	8.3



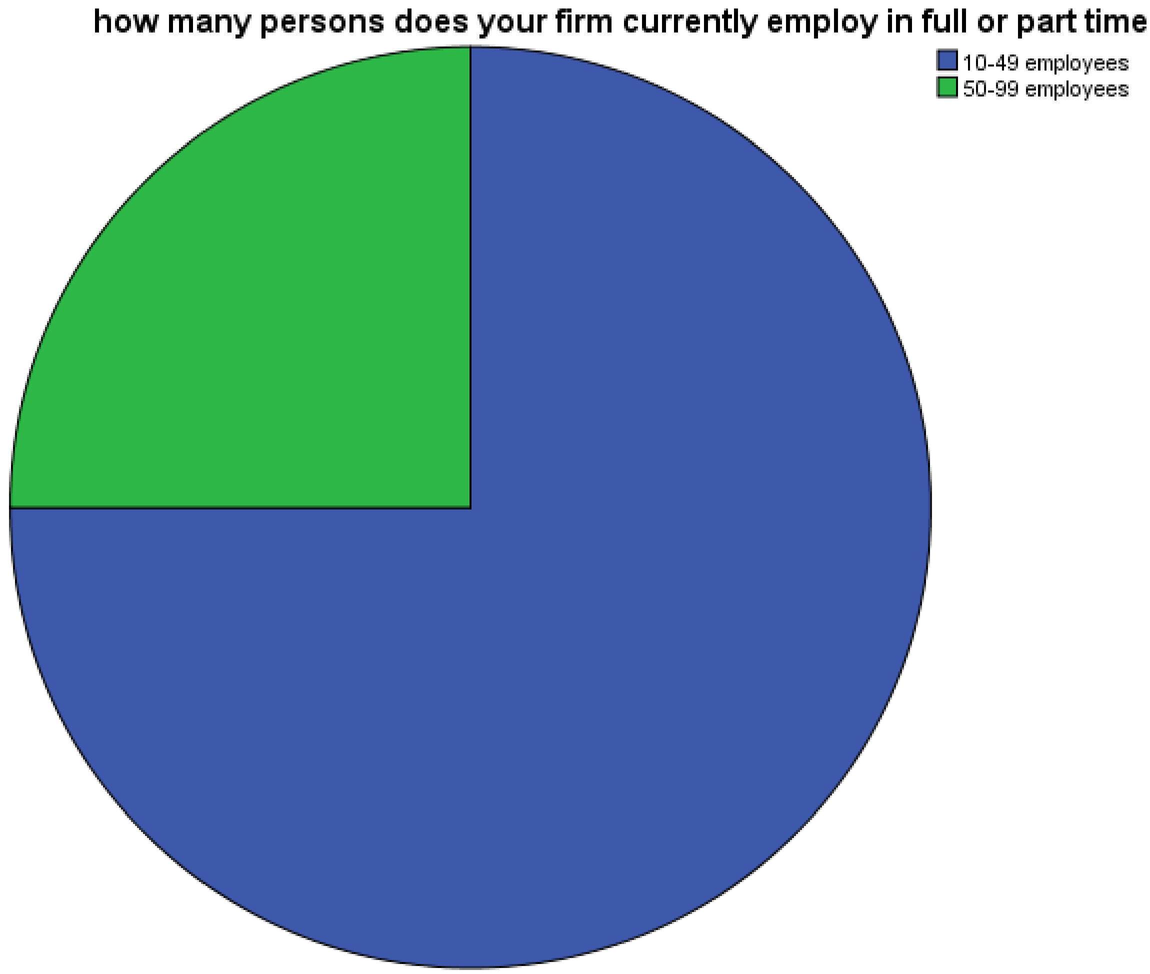
Variable	Category	Frequency	Proportion (%)
What stage of development would you say business is in at the present time?	Start-up stage	50	16.7
	Slow growth stage	175	58.3
	Fast-growth stage	75	25.0



Variable	Category	Frequency	Proportion (%)
What is the main activity of your company?	Mining	50	16.7
	Manufacturing	25	8.3
	Wholesale or retail	75	25.0
	Transport, warehousing and couriers	50	16.7
	Real estate and rental and leasing	25	8.3
	Agriculture	75	25.0



Variable	Category	Frequency	Proportion (%)
How many persons does your firm currently employ in full or part time?	10-49 employees	225	75.0
	50-99 employees	75	25.0



Variable	Category	Frequency	Proportion (%)
What is the business legal status?	public company	1	8.3
	co-operative	1	8.3
	partnership	2	16.7
	sole proprietorship	3	25.0
	private company	5	41.7
What stage of development would you say business is in at the present time?	start-up stage	2	16.7
	fast-growth stage	3	25.0
	slow growth stage	7	58.3
	manufacturing	1	8.3

Variable	Category	Frequency	Proportion (%)
What is the main activity of your company?	real estate and rental and leasing	1	8.3
	mining	2	16.7
	transport, warehousing and couriers	2	16.7
	wholesale or retail	3	25.0
	agriculture	3	25.0
Who are the owners of your firm	other firms	1	8.3
	venture capital firm	1	8.3
	one owner who is a female	2	16.7
	one owner who is a male	3	25.0
	family	5	41.7

Financing of the firm

Variable	Category	Frequency	Proportion (%)
Internal funds	Used	250	83.3
	Instrument is not applicable to my firm	25	8.3
Grants or subsidised bank loan	Used	100	33.3
	Did not use	150	50.0
	Instrument is not applicable to my firm	25	8.3
Bank overdraft, credit line or credit cards overdraft	Used	175	58.3
	Did not use	25	8.3
	Instrument is not applicable to my firm	75	25.0

Variable	Category	Frequency	Proportion (%)
Bank loan	Used	100	33.3
	Did not use	125	41.7
	Instrument is not applicable to my firm	50	16.7
Trade credit	Used	100	33.3
	Did not use	75	25.0
	Instrument is not applicable to my firm	75	25.0
Other loan	Used	125	41.7
	Did not use	25	8.3
	Instrument is not applicable to my firm	125	41.7
Leasing or hire purchase of factoring	Did not use	125	41.7
	Instrument is not applicable to my firm	150	50.0
debt securities issued	Did not use	75	25.0
	Instrument is not applicable to my firm	175	58.3
	Not applicable	25	8.3
Subordinated loans, participation loans or similar financing instruments	Used	25	8.3
	Did not use	25	8.3
	Instrument is not applicable to my firm	200	66.7
	Not applicable	25	8.3
Equity issuance or external equity investors	Did not use	25	8.3
	Instrument is not applicable to my firm	225	75.0
	Not applicable	25	8.3

Variable	Category	Frequency	Proportion (%)
Other	Did not use	50	16.7
	Instrument is not applicable to my firm	200	66.7

Types of financing

Variable	Category	Frequency	Proportion (%)
Bank loans	Increased	150	50.0
	Decreased	75	25.0
	Instrument not applicable to my firm	50	16.7
Trade credit	Increased	100	33.3
	Remained unchanged	25	8.3
	Decreased	50	16.7
	Instrument not applicable to my firm	100	33.3
Equity investments in your firm	Increased	50	16.7
	Remained unchanged	50	16.7
	Decreased	25	8.3
	Instrument not applicable to my firm	125	41.7
	Not applicable	25	8.3
Debt securities issued	Remained unchanged	25	8.3
	decreased	50	16.7
	Instrument not applicable to my firm	175	58.3
	Not applicable	25	8.3
Other	Increased	25	8.3

Variable	Category	Frequency	Proportion (%)
	Remained unchanged	50	16.7
	Instrument not applicable to my firm	175	58.3
	Not applicable	25	8.3

Firm's needs for external financing over past 6 months

Variable	Category	Frequency	Proportion (%)
Fixed investment	Increased needs for external financing	250	83.3
	No impact on needs for external financing	25	8.3
Inventories and working capital	Increased needs for external financing	225	75.0
	Decreased needs for external financing	25	8.3
	Not relevant, did not occur	25	8.3
Internal funds	Increased needs for external financing	50	16.7
	No impact on needs for external financing	75	25.0
	Decreased needs for external financing	125	41.7
	Not relevant, did not occur	25	8.3
Mergers and acquisitions and corporate restructuring	Increased needs for external financing	25	8.3
	Decreased needs for external financing	25	8.3
	Not relevant, did not occur	225	75.0

Applied ways of financing over the past 6 months

Variable	Category	Frequency	Proportion (%)
Bank loan	Applied	150	50.0
	Did not apply because of possible rejection	75	25.0
	Did not apply because of sufficient internal funds	25	8.3
	Did not apply for other reasons	25	8.3
Trade credit	Applied	50	16.7
	Did not apply because of possible rejection	75	25.0
	Did not apply because of sufficient internal funds	50	16.7
	Did not apply for other reasons	50	16.7
	Not applicable	50	16.7
Other external financing	Applied	100	33.3
	Did not apply because of possible rejection	75	25.0
	Did not apply because of sufficient internal funds	50	16.7
	Did not apply for other reasons	25	8.3
	Not applicable	25	8.3
Bank loan	Applied and got everything	25	8.3
	Applied but got part of it	150	50.0

Variable	Category	Frequency	Proportion (%)
	Applied but refused because cost too high	50	16.7
	Applied but was rejected	50	16.7
Trade credit	Applied and got everything	25	8.3
	Applied but got part of it	50	16.7
	Applied but refused because cost too high	125	41.7
	Not applicable	75	25.0
Other external financing	Applied and got everything	75	25.0
	Applied but got part of it	125	41.7
	Applied but refused because cost too high	25	8.3
	Applied but was rejected	25	8.3
	Not applicable	25	8.3

Availability of ways of financing

Variable	Category	Frequency	Proportion (%)
Bank loans	Improved	50	16.7
	Remained unchanged	125	41.7
	Deteriorated	100	33.3
Trade credit	Improved	50	16.7
	Remained unchanged	100	33.3
	Deteriorated	25	8.3

	Not applicable	75	25.0
	Not applicable	25	8.3
Equity investments in your firm	Improved	50	16.7
	Remained unchanged	50	16.7
	Deteriorated	75	25.0
	Not applicable	100	33.3
Other	Improved	50	16.7
	Remained unchanged	25	8.3
	Deteriorated	50	16.7
	Not applicable	75	25.0
	Not applicable	50	16.7

Terms and conditions of the bank financing available

Variable	Category	Frequency	Proportion (%)
Level of interest rates	Was increased by the bank	225	75.0
	Was decreased by the bank	25	8.3
	Not applicable	25	8.3
Level of cost of financing other than interest rates	Was increased by the bank	175	58.3
	Remained unchanged	25	8.3
	Was decreased by the bank	25	8.3
	Not applicable	25	8.3
Available size of loan credit line	Was increased by the bank	100	33.3
	Was decreased by the bank	175	58.3

Variable	Category	Frequency	Proportion (%)
Available maturity of the loan	Was increased by the bank	125	41.7
	Remained unchanged	50	16.7
	Was decreased by the bank	100	33.3
Collateral requirements	Was increased by the bank	125	41.7
	Remained unchanged	75	25.0
	Was decreased by the bank	75	25.0
Other	Was increased by the bank	50	16.7
	Remained unchanged	25	8.3
	Was decreased by the bank	175	58.3
	Not applicable	25	8.3

Financing with

Variable	Category	Frequency	Proportion (%)
Banks	Yes	125	41.7
	No	175	58.3
Equity investors	Yes	200	66.7
	No	100	33.3

External financing would prefer most

Variable	Category	Frequency	Proportion (%)
External financing	Bank loan	175	58.3
	Loan from other sources	75	25.0

Variable	Category	Frequency	Proportion (%)
	Subordinated loans and participation	25	8.3
What amount of financing would you aim to obtain	Less than $25,000	125	41.7
	$25,000 - $100,000	75	25.0
	$100,000 - $1,000,000	100	33.3
What do u see as the most important limiting factor to get this financing	There are no obstacles	75	25.0
	Insufficient collateral	50	16.7
	Interest rates or price to high	150	50.0
	Reduced control over the firm	25	8.3
What do u see as the most important limiting factor to get this financing	Interest rates or price to high	100	33.3
	Reduced control over the firm	50	16.7
	Financing not available at all	25	8.3
	Other	25	8.3
	Not applicable	25	8.3
Internal funds	Will improve	150	50.0
	Will remain unchanged	100	33.3
	Will deteriorate	50	16.7
Bank loans	Will improve	125	41.7
	Will remain unchanged	50	16.7
	Will deteriorate	125	41.7
Equity investments in your firm	Will improve	125	41.7
	Will remain unchanged	50	16.7

Variable	Category	Frequency	Proportion (%)
	Will deteriorate	50	16.7
	not applicable	75	25.0
The biggest obstacles to the growth of your enterprise	Finding qualified labour	25	8.3
	The frequent need to renew the equipment	50	16.7
	Instability of demand of product or service	100	33.3
	Obtaining additional financing	100	33.3
	Government regulations	25	8.3
How many persons does your firm currently employ in full or part time	10-49 employees	225	75.0
	50-99 employees	75	25.0
What is the business legal status	Sole proprietorship	75	25.0
	Private company	125	41.7
	Public company	25	8.3
	Partnership	50	16.7
	Co-operative	25	8.3
What stage of development would You say business is in at the present time	Start-up stage	50	16.7
	Slow growth stage	175	58.3
	Fast-growth stage	75	25.0
Does firm have a board of directors that meets more than twice a year	Yes	125	41.7
	No	175	58.3
Besides the owner, is there anyone in the firm in charge of finance	Yes	275	91.7
	No	25	8.3

Variable	Category	Frequency	Proportion (%)
What is the main activity of your company	Mining	50	16.7
	Manufacturing	25	8.3
	Wholesale or retail	75	25.0
	Transport, warehousing and couriers	50	16.7
	Real estate and rental and leasing	25	8.3
	Agriculture	75	25.0
Who are the owners of your firm	Family	125	41.7
	Other firms	25	8.3
	Venture capital firm	25	8.3
	One owner who is a male	75	25.0
	One owner who is a female	50	16.7
What is currently the most pressing problem your company is facing	Competition	25	8.3
	Access to finance	275	91.7
Turnover	Increased	125	41.7
	Remained unchanged	150	50.0
Mark up	Increased	25	8.3
	Remained unchanged	175	58.3
	Decreased	100	33.3
Profit	Increased	50	16.7
	Remained unchanged	200	66.7
	Decreased	50	16.7
	Increased	150	50.0

Variable	Category	Frequency	Proportion (%)
Would u say the amount of dept compared to the Assets of your company	Remained unchanged	50	16.7
	Decreased	75	25.0
	No debt	25	8.3
Net interest	Increased	125	41.7
	Remained unchanged	100	33.3
	Decreased	75	25.0

APPENDIX C: CREDIT RISK MODELLING OUTPUT

Preliminary CRM Model

Variable	Coefficient	Std Error	z-value	PR (>|z|)	Signif code
Interest rate	6.724869	3.058489	2.199	0.0279	0.05
Gearing ratio	1.180722	0.575546	2.051	0.0402	0.05
Creditors days	0,019070	0.009285	2.054	0.0400	0.05
No. Of Directors	0.302409	0.133585	2.264	0.0236	0.05
Income	1.945908	0.455356	4.273	1.93e-05	0.001
Age	0.086533	0.030238	4.276	1.90e-05	0.001
Net profit margin	-9.188423	1.979402	-4.642	3.45e-06	0.001
Liquidity ratio	-0.271903	0.136638	-1.990	0.0466	0.05
Stock T/O	-1.612757	0.293155	-5.501	3.77e-08	0.001
Intercept	-20.417667	4.316626	-4.730	2.25e-06	0.001
				AIC	207.78

Final CRM Model for the Weak Selection

Variable	Coefficient	Std Error	z-value	PR (>|z|)	Significance code
Net profit margin	-24.87811	6.10027	-4.078	5.54e-05	0.001
Liquidity ratio	-2.50004	0.79952	-3.127	0.0177	0.01
Income	3.06757	0.68494	4.479	7.51e-06	0.001
Age	0.16778	0.03867	4.338	1.44e-05	0.001
Creditors days	0.02939	0.01431	2.053	0.04004	0.05
Asset size	0.70584	0.31879	2.214	0.02682	0.05
Intercept	-39.76557	8.62166	-4.612	3.98e-06	0.001
				AIC	96.926

Final CRM Model for the Strong Selection.

Variable	Coefficient	Std Error	z-value	PR (>|z|)	Significance code
Net profit margin	-28.63312	5.02270	-5.701	1.19e-08	0.001
Liquidity ratio	-2.16914	0.49887	-4.348	1.37e-05	0.001
Age	0.16023	0.03139	5.104	3.32e-07	0.001
Income	3.07624	0.58444	5.264	1.41e-07	0.001
Length of Relationship	-0.06008	0.03549	-1.693	0.09051	0.1
Creditors days	0.03567	0.01339	2.664	0.00773	0.01
Asset size	0.68512	0.26513	2.584	0.00976	0.01
Intercept	-38.59686	7.12788	-5.415	6.13e-08	0.001
				AIC	126.64

Current Model vs Preliminary Model

Models	Accepted	Rejected	Bad rate	% Improvement of Bad rate
original	281	73	20.62%	
				23.28%
Preliminary model	298	56	15.82%	

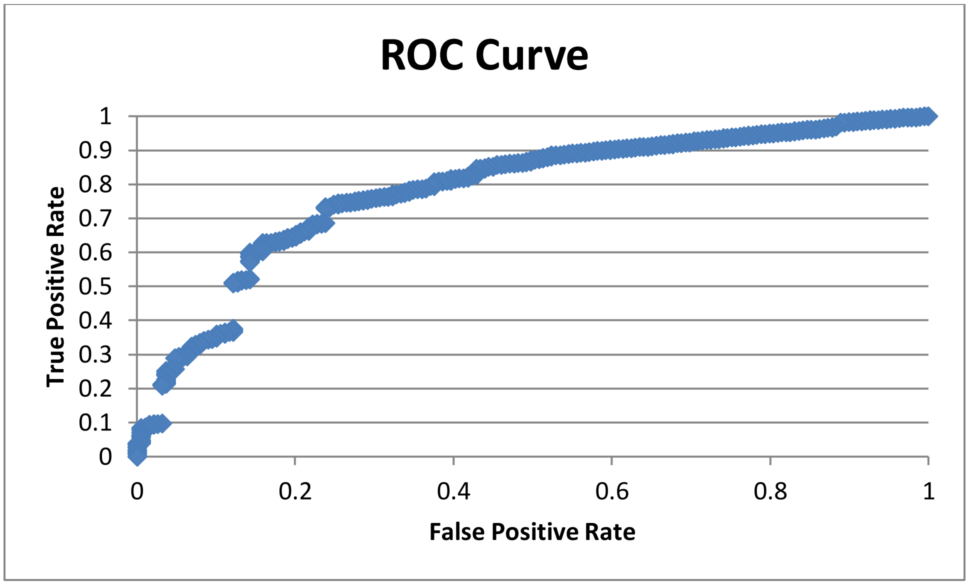
Notes:

Simulation of "accepted sample" and "rejected sample" - credit granting policies

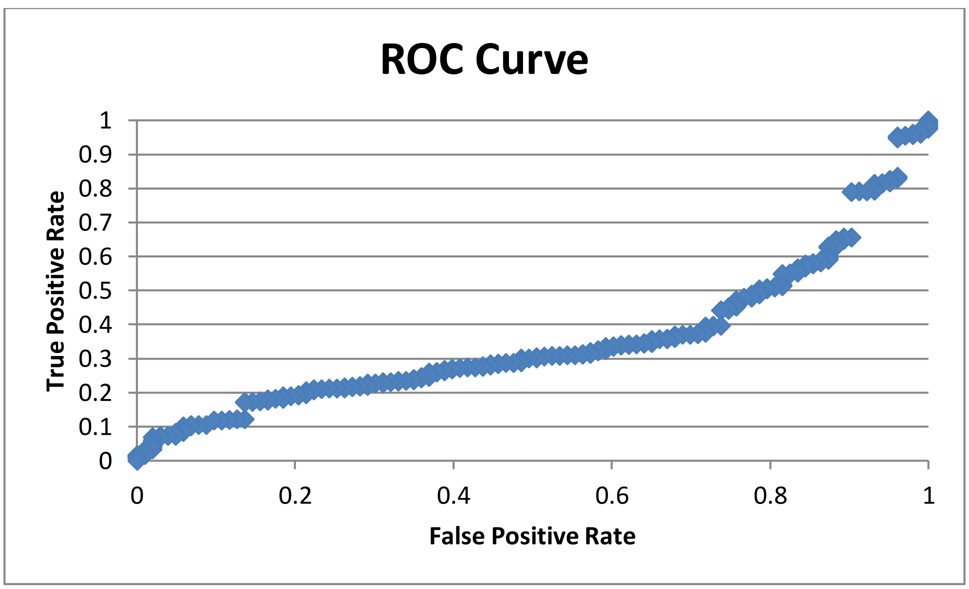
1) **Weak selection**: We simulated credit rejection criterion by instituting a 500-credit score cut-off: the accepted sample had 295 and rejected sample had 29
2) **Strong selection**: We simulated credit rejection criterion by instituting a 500-credit score cut-off: the accepted sample had 230 and rejected sample had 62
3) Application of the BC with prior distribution $\alpha_{ij} = 0$
4) The regression equation used to estimate missingness ($\varphi_{(i,1)}$) is given by

   $y = 0.0010836x - 0.0167039.$ with an R squared value of 74.71%, where y is the bad rate for the accepted and x is the midpoint of the score ranges.
5) The regression equation used to estimate missingness ($\varphi_{(i,1)}$) is given by

   $y = 0.0013982x - 0.0827607.$ with an R squared value of 53.46%, where y is the bad rate for the accepted and x is the midpoint of the score ranges
6) Weak Selection Model developed by backward modelling on all-good-bad (AGB) development sample



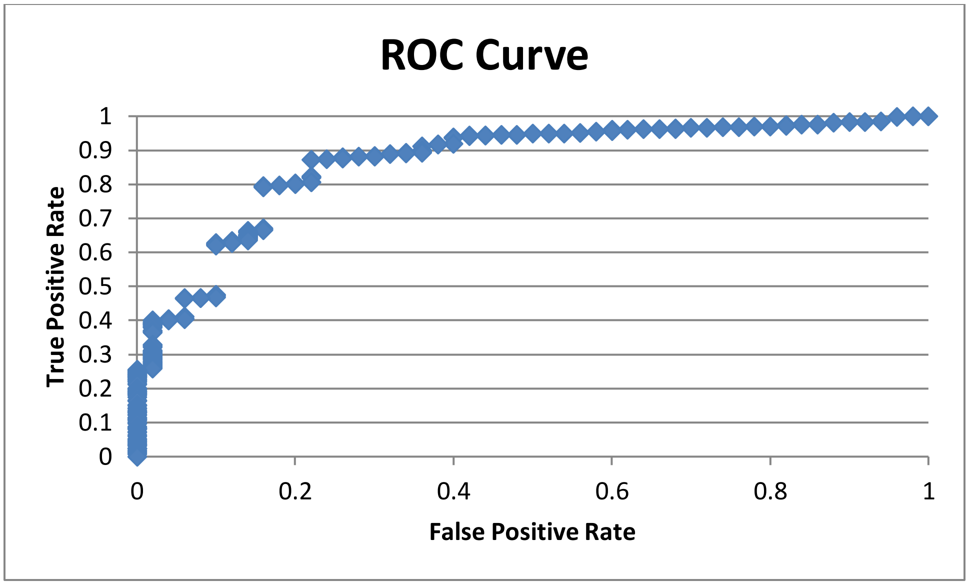
Weak selection ROC



Strong Selection ROC

Validation (80% hold-out) ROC

0.56101814


ROC Curve
True Positive Rate
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.2
0.4
0.6
0.8
1
False Positive Rate

APPENDIX D: TURNITIN REPORT

Originality Report Turnitin

Dissert by Marx Dambaza

From Dissert (Psyco)

- Processed on 22-Mar-2018 10:45 AM SAST
- ID: 934326400
- Word Count: 101513

Similarity Index 7%

Similarity by Source

Internet Sources: 8%

Publications: 3%

Student Papers: 1%

sources:

1 3% match (Internet from 23-Sep-2012)

http://www.era.lib.ed.ac.uk/bitstream/1842/4914/1/Lin%202007.pdf

2 2% match ()

http://pims.math.ca/birs/workshops/2003/03w5023/Asterbro_Bound.pdf

3 1% match (Internet from 06-Jul-2012)

http://interstat.statjournals.net/YEAR/2006/articles/0609001.pdf

4 1% match (Internet from 06-Nov-2012)

http://www.scribd.com/doc/80592557/Emilie-Rose-Breathless-Passion

5 1% match (Internet from 23-Sep-2005)

http://www.statistik.uni-karlsruhe.de/htm/mitarbeiter/trueck/sme.pdf