`

Adapting a Developed Market Credit Risk Model for the Understanding and Estimation of Consumer Credit Losses in South Africa

by

Geoffrey John Kimetto

submitted in accordance with the requirements

for the degree of

DOCTOR OF BUSINESS LEADERSHIP

at the

UNIVERSITY OF SOUTH AFRICA

SUPERVISOR: Dr A D Joseph

May 2023

DECLARATION

Name: Geoffrey John Kimetto

Student Number: 64101304

Degree: Doctor of Business Leadership

Adapting a Developed Market Credit Risk Model for the Understanding and Estimation of Consumer Credit Losses in South Africa

I declare that the above thesis 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.

28.05.2023

SIGNATURE DATE

ACKNOWLEDGEMENTS

I thank God, the maker of heaven and earth, for the grace to carry out this research, and my wife who supported and encouraged me when things were not going well. My children were patient and encouraging and this was invaluable.

The role of the supervisor, Dr Angelo Joseph, in guiding me to this stage was distinctive; his energy, enthusiasm, sharp mind, and positive demeanour were more than encouraging to say the least. He went the extra mile to avail himself for discussions and reviews on occasions without number.

The contribution of Unisa, especially regarding the value added of the preparatory workshop of April 2018 was invaluable.

I appreciate Ms Tumelo Seopa, the DBL Administrator, and Mrs Megan Scheepers, the Librarian, for helping me with administrative issues and access to Library resources respectively.

My student colleague, Mr. Garikai Gondo was of immense help with resolving IT issues, and Mr. Tungamirai Tambandini, a fellow student who is supervised by my supervisor, was of help, especially with sharing SBL information. Many other DBL student colleagues contributed positively in one way or another on this journey. I thank them all. I also thank Khabo Phage of the South African Reserve Bank, who made time to discuss the availability of data from the Reserve Bank and the use of proxy data for credit losses.

I appreciate the financial support that I received from the Unisa Masters and Doctoral Bursary administration office. It went a long way in making the research possible.

ABSTRACT

The aim of this research was to build a developed market credit risk model and adapt it for the understanding and estimation of consumer credit losses in the emerging market of South Africa. The developed market of the United States of America was chosen since it has good quality data (the research is quantitative in nature). Credit, a contractual agreement in which a person or institution that is a party to the agreement, borrows something of value today with an undertaking to repay (often with interest) the other party at a later date(s), has existed and evolved for thousands of years. With almost every financial transaction comes credit risk. Credit risk is measured using credit risk models and reviewed literature indicates that models that have been developed have poor estimation accuracy levels. Model errors are important as they directly affect profitability, solvency, shareholder value, macro-economy, and society. From publicly available data at big databases such as the FRED (Reserve Bank Economic Data (US)), the SARB (South African Reserve Bank) and the World Bank, independent variables (with data of monthly and quarterly frequencies and spreading from 2008-2021 and 1987-2021 for SA and the US respectively) were selected. Multivariable regression analysis, amongst other analyses, was done to establish relationships between individual or sets of the selected independent variables and credit losses. Explanatory variables that capture Sentiment (defined in this research as reactions or behaviour of credit market participants in far-from-equilibrium situations - for example recessions, financial booms and busts) were coupled with economic variables and obligor characteristics. The establishment of a model building analyses blocks framework and a relatively accurate consumer credit risk model for the emerging market of South Africa (with an R-squared value of 85% and back test estimation accuracy of 88% on average) were amongst the key results of the analyses. The universal applicability, subject to the availability of quality data, of the credit risk modelling methods used in this research could motivate policymakers in South Africa and other emerging markets to adopt the data collection, storage, and organisation formats used in the FRED database.

KEY TERMS:

Emerging Markets, Consumer Credit Risk Model, Charge Off Rate, Loss Given Default, Sentiment, Credit Risk Measurement, Credit Losses, Impairments, Probability of Default, and Components of Credit Risk.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

DEFINITIONS

ABS: Asset Backed Securities.

CDO: Collateralised Debt Obligation are securitisations of assets into tranches with different levels of risks.

CDO Squared: CDO backed by a portfolio of other CDOs.

CDS: Credit Default Swaps are standard market insurance contracts on loans or bonds issued by a company.

Countercyclical: A counter cyclic (or countercyclical) variable moves in the opposite direction of conventional business cycle indicators.

Credit Rating Agencies: Rating agencies assign ratings on creditworthiness to corporate and sovereign issuers of debt. The three main agencies are Moody's, Fitch, and Standard & Poor's (S&P).

Delinquency Rate: The percentage of loans of consumers whose payments are delinquent.

EAD: Exposure At Default is the amount outstanding on the loan in the event of a payment default.

Impairment: A permanent reduction in the value of a company asset or the amount by which the carrying amount of an asset (or a cash-generating unit or group of assets) exceeds its recoverable amount.

LGD: Loss Given Default is the expected percentage of loss incurred by the lender in the event of default on a loan.

Loan Charge Off: The net amount of loans a bank charges off minus any recoveries of previously charged-off loans. Therefore, LCOs are realised loan losses resulting from the application of impairment rules to bank loans, wherein such loan losses represent a bank's realised credit risk.

Loan Charge Off Rate: The ratio of loan loss charge-offs to loans issued.

MBS: Mortgage-Backed Securities.

Obligor: The counterparty to a loan obliged to repay the debt.

PD: Probability Of Default.

R-squared: A relative statistical measure of the percentage dependent variable variance explained by the regression model.

Adjusted R-squared: The R-squared adjusted for the number of independent variables.

SPV: Special Purpose Vehicle is a bankruptcy remote company that is set up to facilitate a trade.

ACRONYMS

BCBS	Basel Committee on Banking Supervision
BIS	Bank for International Settlements
FASB	Financial Accounting Standards Board
FED	Federal Reserve Bank of US
FRED	Federal Reserve Economic Data
IFRS	International Financial Reporting Standards
IMF	International Monetary Fund
IASB	International Accounting Standards Board
OECD	Organization for Economic Co-operation and Development
SA	South Africa
SARB	South African Reserve Bank
US	United States of America
USBLS	U.S. Bureau of Labour Statistics
USEIA	U.S. Energy Information Administration
WB	World Bank

CHAPTER 1: ORIENTATION

1.1 INTRODUCTION

Credit, a contractual agreement in which a person or institution that is a party to the agreement, borrows something of value today with an undertaking to repay (often with interest) the other party at a later date(s), has existed and evolved for thousands of years (Sanders, 1992; Oliver and Hand, 2005; Swartz, 2010). With credit exposure comes credit risk. (Langley, 2008). This is the risk that a borrower – also referred to as an obligor – defaults on a debt obligation. Furthermore, defaults have ramifications on people, businesses, and the economy (Eubank, 2012; Baesens, 2015). It is therefore important for credit providers to be able to measure credit risk. The literature reviewed indicates that building credit risk models remains a work in progress as existing models do not accurately estimate credit risk losses (Bae and Kim, 2015; Mester, 2015; Baesens, Rosch and Scheule, 2016). Baesens, Rosch and Scheule (2016) stated that the low R-squared values of the Loss Given Default (LGD) models (see Section 3.3.3 and Table 3.1) suggest that there is a considerable amount of variation that credit risk models do not explain and searching for more precise models would take many years. Antoniades (2018) argues that there is no greater intellectual challenge than that of how to harness the power of credit while mitigating the risk of default and its ramifications. The aim of this study was to develop a consumer credit risk model that would improve the estimation of credit losses, by using data from the developed market of the United States of America, and use the insights gained to develop a similar model for the emerging market of South Africa. The layout of this chapter is as follows: Section 1.2 is a synopsis of credit exposure levels for adult – 15 years and older – populations globally, the credit situation in the emerging market of South Africa and the significance and impact of the research. The research problem, the aim of the research, the research questions, the research objectives, the scope of the research and delimitations are set out in Sections 1.3, 1.4, 1.5, 1.6, 1.7, and 1.8 respectively. A summary of the research methodology is given in Section 1.9 and the research theoretical and conceptual frameworks are outlined in Sections 1.10 and 1.11 respectively. Mentions of the research output and chapter summary are made in Sections 1.12 and 1.13 respectively.

1.2 CONSUMER CREDIT OVERVIEW

Credit is a substantial and growing aspect of individuals’ lives in all countries (Demirgüç-Kunt *et al.*, 2015, 2018; Antoniades, 2018; NCR, 2018). According to Demirgüç-Kunt et al., (2015;2018) and as shown in Figure 1.1 and Figure 1.2, in 2017, about 50% of the world’s population of 15 years and older had obtained credit in the preceding 12 months. This included both formal and informal borrowing of money. In developed economies, 90% of the borrowers obtained their credit from financial institutions, while in developing countries about half of the credit consumers borrowed the money from family and friends (Demirgüç-Kunt *et al.*, 2018). The reasons for borrowing varied from buying land or a home, health, or medical purposes to starting, operating, or expanding a business. The graphs on consumer credit exposure as at the end of Quarter Two (2) 2018 and consumer credit standing as at the end of Quarter Four (4) 2017 (see Figure 1.3 and Figure 1.4) for the emerging market of South Africa contextualise the credit landscape (NCR, 2018). This picture – of elevated levels of borrowings and delinquency – reflects similar situations prevalent worldwide (Hyman, 2012; Angel and Heitzmann, 2015; Apanga, Appiah and Arthur, 2016). Credit risk is therefore a problem that is not unique to emerging economies.

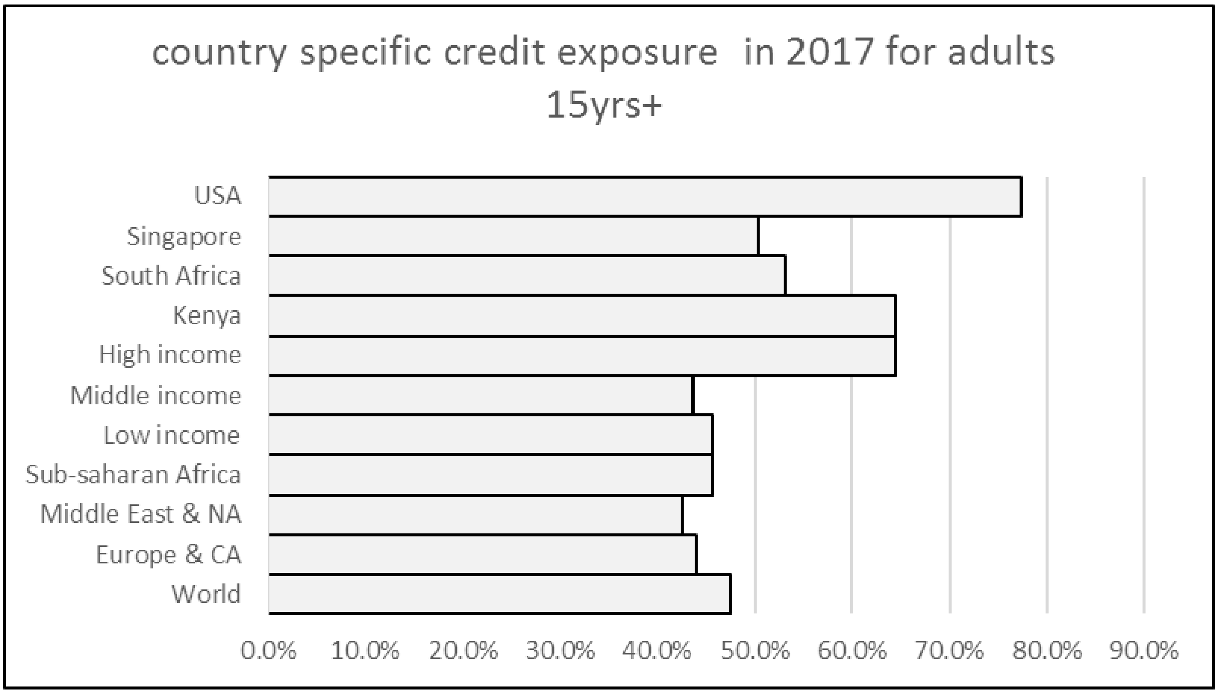


Figure 1.1: Cross Section of Global Consumer Credit Exposure

(Adapted from: Demirgüç-Kunt et al., (2018, p.76)).

As shown in Figure 1.1, the proportion of the population 15 years and older that borrowed money in the preceding 12 months is relatively high globally (>40% level). This statistic appears to occur irrespective of the country, region, or income level.

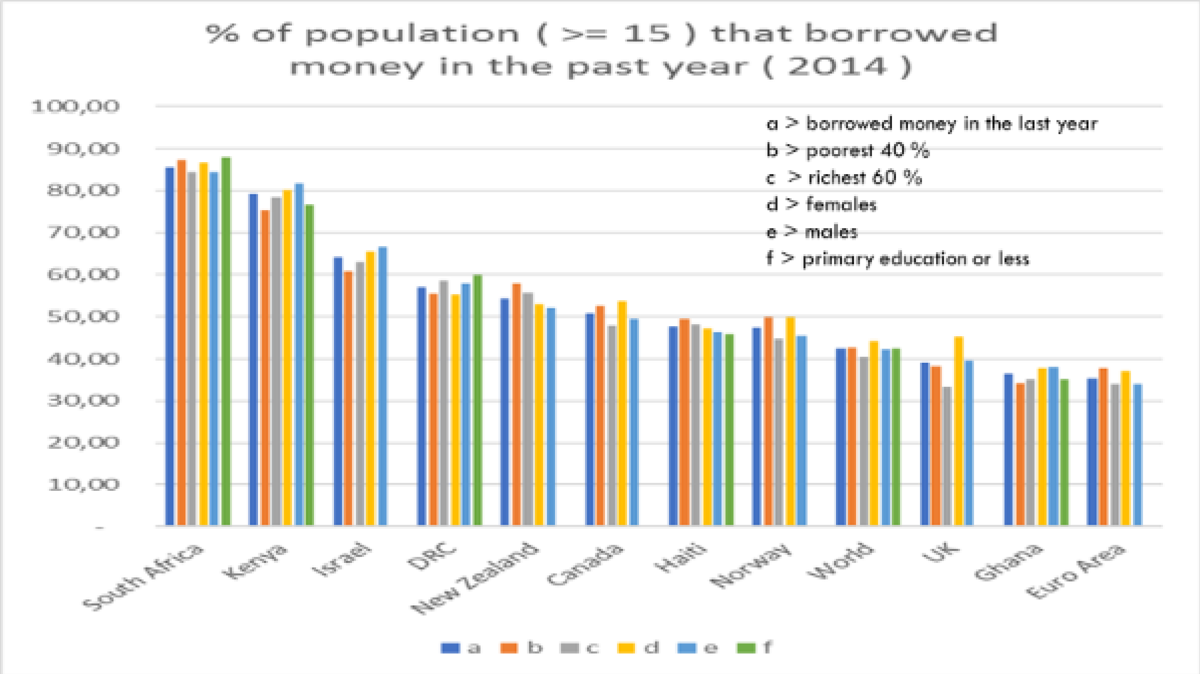


Figure 1.2: Profile of Consumer Credit across the World

(Adapted from: Demirgüç-Kunt et al. (2015, p.48)).

The general level of borrowing[1] within a country, as shown in Figure 1.2, appears not to be dependent on factors such as education levels, gender, or income levels.

1.2.1 Consumer Credit in the Emerging Market of South Africa

The emerging market of interest in this study is the South African market. According to the National Credit Regulator (NCR) 2017/2018 annual report, as shown in Figure 1.3, as at Quarter Two (2) 2018, South Africans owed a total of R1.8 trillion to creditors (NCR, 2018). Of this amount R922 billion (51.2%) was in mortgages, R416 billion (23.1%) was in secured credit (pension-backed, insurance-backed, furniture and motor vehicle accounts), R231 billion (12.8%) was in credit facilities (credit cards, store cards, overdrafts), R179 billion (10%) was

---

[1] The lower level of borrowing in Ghana (40%) compared to South Africa (86%) or Kenya (75%) may be due to similarly significant lower level of account ownership (41%) amongst her adult population. Account ownership stood at 70% and 75% for South Africa and Kenya respectively (Demirgüç-Kunt *et al.*, 2015). Account ownership is an indicator of financial inclusion.

in unsecured credit, R3 billion (0.17%) in short-term credit and R47 billion (2.6%) in developmental credit such as school fee loans. The amount of debt and the proportion of the indebted population is technically not an issue per se. Debt becomes a problem when a consumer becomes overindebted or in other words, when the consumer is no longer able to meet their financial obligations (Swartz, 2010; Angel and Heitzmann, 2015). This may pose a problem to both the borrower and the credit provider as they now face the possibility of loss arising from non-payment of the amount outstanding.

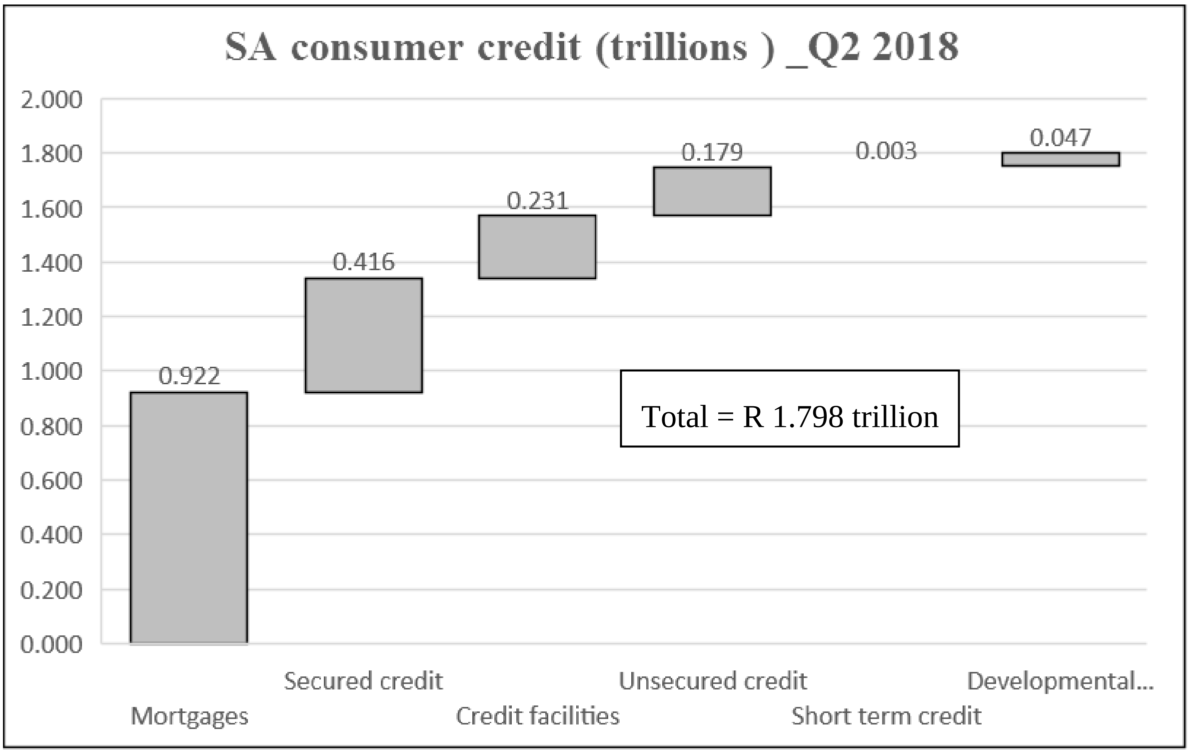


Figure 1.3: Credit Distribution for the Emerging Market of South Africa

(Data from: National Credit Regulator (2018, p.44)).

The data in Figure 1.3 indicates that as at Quarter Two (2) 2018, South Africans owed a total of R1.8 trillion to creditors (NCR, 2018), the largest of which was mortgages at R 0,922 trillion.

The NCR’s 2017/2018 annual report (NCR, 2018) had the number of credit-active consumers as at the end of 2017 at 25.3 million (Figure 1.4) of which 51.7 % or 15.6 million were in good standing (current or maximum two (2) months in arrears) while the rest, totalling 9,7 million in number, had impaired records. Making up the 38.3% borrowers with impaired records, were 21.7 % (5.5 million) with three (3) or more months in arrears, 11.1 % (2.8 million) in adverse listings, and 5.5 % (1.4 million) in judgements and administration.

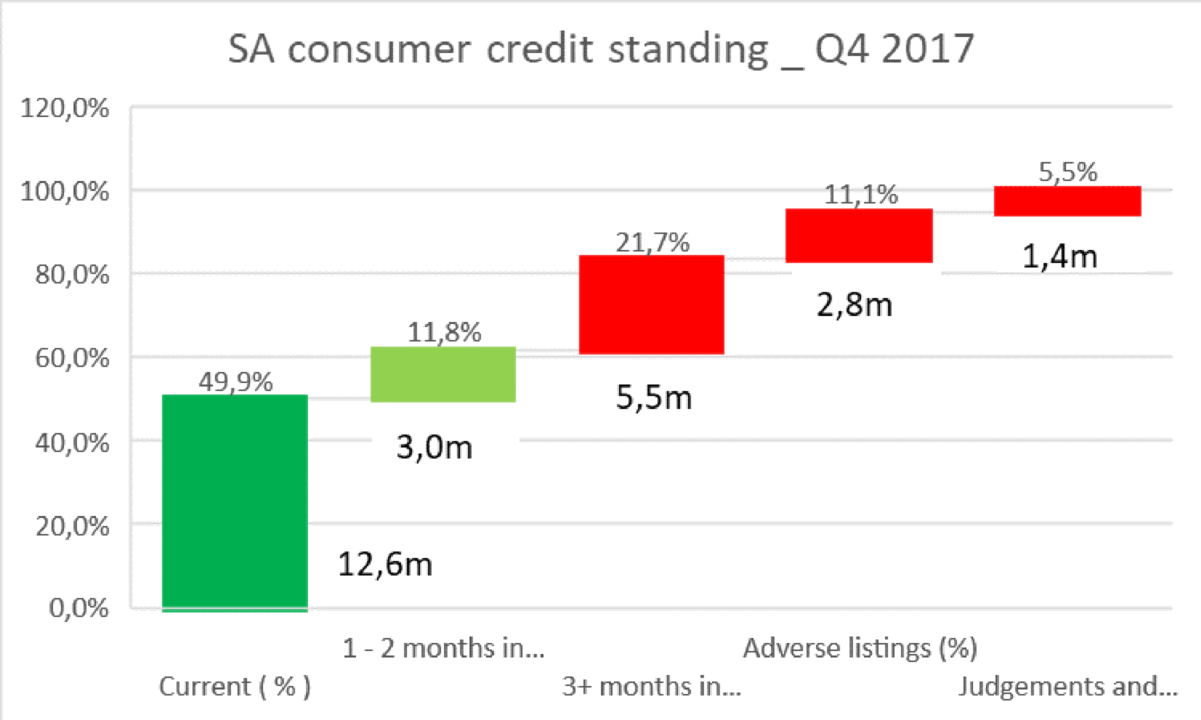


Figure 1.4: Credit Maturity Profile for the Emerging Market of South Africa

(Data from: National Credit Regulator (2018, p.46)).

1.2.2 Research Significance and Impact

The number of South Africans with impaired credit as at the end of 2017 was high; it stood at 9.7 million or 38.3 % of the 25.3 million credit-active consumers (NCR, 2018). The picture in other countries such as the United States of America is in principle similar (Hyman, 2012; Angel and Heitzmann, 2015; Apanga, Appiah and Arthur, 2016). This spells potential financial loss to the lender and/or the credit consumer. The need to estimate and mitigate losses is therefore important to banks, and other credit providers (Barth *et al.*, 2018; Jin *et al.*, 2021). The literature reviewed revealed that existing credit risk models have low estimation accuracy levels (Bolton, 2009; Mester, 2015; Baesens, Rosch and Scheule, 2016). A model that can be used to improve the estimation of consumer credit losses and is universally applicable with adaptation, would therefore be a significant contribution to the stability of the consumer credit market.

1.3 RESEARCH PROBLEM

Through the literature reviewed, the use of credit was traced as far back as 3000 years ago (Oliver and Hand, 2005). Credit is a facility, agreement or offer to access money, goods or services with the understanding that it will be paid later with or without interest (Simpson and

Weiner, 2022). With every financial transaction comes credit risk, a fundamental problem especially in the banking industry (Chopra and Bhilare, 2018). Measuring and managing credit risk is essential for the stability of financial institutions and the economy (Barth *et al.*, 2018; Jin *et al.*, 2021). The need for accurate credit risk models increased following the 2007-2008 credit crisis, as one of the key factors that contributed to the crisis was inaccurate credit risk models (Ishikawa, 2009; Lowenstein, 2011; Jarrow, 2012; Athanassiou and Theodosopoulou, 2015; Honohan, 2016). Existing literature indicates that credit risk models that have been developed have poor accuracy levels (Bae and Kim, 2015; Mester, 2015; Baesens, Rosch and Scheule, 2016; Nehrebecka, 2021). Model errors, which pose model risk, directly affect profitability, solvency, shareholder value, macro-economy, and society (Baesens, 2015). For example, the use of an inaccurate credit risk model can mean that financial institutions overestimate provisions to cover expected losses and hold higher regulatory capital to insure against unexpected losses. For the home loan consumer-borrower, whose home is foreclosed because of default (credit approval inferred from an inaccurate credit risk model), this may mean wealth destruction from which the borrower may take a long time to recover.

1.4 AIM OF THE RESEARCH

The aim of this research was to develop a consumer credit risk model, using data from the developed market of the United States of America, which would improve the estimation of consumer credit losses, and use the insights gained to develop a similar model for the emerging market of South Africa. The minimum requirements of such models are that they should be simple, easy to use, effective, affordable, and legally compliant in the jurisdictions in which they are used (Baesens, 2015).

1.5 RESEARCH QUESTIONS

To find the optimal way to generate the knowledge needed for partially or fully solving the research problem (Wisse and Roeland, 2022) and attaining the aim of the research, the research questions below were formulated.

1.5.1 Main Research Question

What new credit risk modelling approach and key independent (explanatory) variables will improve the estimation of a consumer credit risk model developed using data from the developed market of the United States of America and how can such a model be adapted for use in the emerging market of South Africa?

1.5.2 Sub-Research Questions

In addition to the main research question, the following sub-questions were addressed.

(1) What are the gaps in data quality that currently exist within the emerging market of South Africa and how can they be closed?

(2) Can coupling of proxy variables for sentiment (defined in this research as the behaviour or reaction of borrowers to events in far-from-equilibrium (disequilibrium) situations) with macro-economic variables, and obligor characteristics improve the estimation of a consumer credit risk model?

(3) What are the key commonalities between the developed market of the United States of America consumer credit risk model and that of the emerging market of South Africa?

(4) Can the insights gained in developing a consumer credit risk model for the emerging market of South Africa be useful to financial institutions, regulators and policymakers in South Africa?

(5) Can the insights gained in developing a consumer credit risk model for the emerging market of South Africa be used to develop general guidelines and standards for measuring and managing consumer credit risks in emerging markets?

1.6 RESEARCH OBJECTIVES

To answer the research questions, the following objectives were set.

1.6.1 Primary Research Objective

The primary objective of this research was to develop a consumer credit risk model that would improve the estimation of consumer credit losses, using data from the developed market of the United States of America, and adapt the model for use in the emerging market of South Africa.

1.6.2 Secondary Research Objectives

(1) Identify or develop proxy data that can be used to bridge data gaps that may exist in the emerging market of South Africa.

(2) Identify explanatory variables that capture sentiment (defined in this research as the behaviour or reaction of borrowers to events in far-from-equilibrium (disequilibrium) situations), suitable macroeconomic variables, and obligor characteristics that improve the estimation of a consumer credit risk model.

(3) Establish the key commonalities between the developed market of the United States of America’s consumer credit risk model and that of the emerging market of South Africa.

(4) Use insights gained in developing a consumer credit risk model for the emerging market of South Africa to make pertinent recommendations that can be used by financial institutions, regulators and policymakers in South Africa.

(5) Use insights gained in developing a consumer credit risk model for the emerging market of South Africa to develop general guidelines and standards for measuring and managing consumer credit risks in emerging markets.

1.7 SCOPE OF THE RESEARCH

This research was confined to building a consumer credit risk model with improved estimation accuracy using regression analysis and other techniques on data from the developed market of the United States of America and adapting it for the emerging market of South Africa. The literature reviewed revealed that there are primarily two (2) strategies for improving the estimation accuracy of credit risk models. One is to use sophisticated models such as neural networks, ensemble models, and support vector machines (Gouvêa, 2007; Allen and Powella, 2011; Klieštik and Cúg, 2015; Chopra and Bhilare, 2018; Nehrebecka, 2021; Takawira and Mwamba, 2022). While these models may produce marginally better outcomes, it is non-trivial to relate the outcomes to the inputs. This loss of interpretability renders the models illegal in many jurisdictions (Baesens, 2015; Nehrebecka, 2021). The other is to assure the quality of data and use simple models whose estimations (EAD, LGD, PD) are related to the inputs. This study aimed to contribute to the ability of financial institutions to estimate consumer credit losses with more accuracy by researching a better consumer credit risk model using the latter strategy.

1.8 DELIMITATIONS OF THE RESEARCH

Credit risk is a wide subject. This study focused on the adaptation of a developed market credit risk model for the understanding and estimation of consumer credit losses in the emerging market of South Africa. In the study no specific pre-existing model was targeted for adaptation. Instead, a credit risk model was built – with an effort made to achieve a relative improvement - using data from the developed market of the US and adapted for the emerging market of South Africa. The study, therefore, constituted the following delimitations.

In this research, the focus was neither on counterparty nor wholesale credit risk. Counterparty and wholesale credit risks (as explained in Section 2.6.1) are about modelling and measurement of institutional credit risk profiles.

In this research, the focus was mainly on quantitative consumer credit risk research. However, given that consumer credit risk (especially in South Africa) is an emerging yet important research area in Finance, publicly available data to calibrate credit risk models are sparse (Apanga, Appiah and Arthur, 2016). Moreover, institutions that have good credit data either invested large amounts of money and resources to improve the data quality or generate data of such superior quality that they keep it highly confidential. To circumvent this challenge, the research in one case made use of econometric data under statistical modelling assumptions together with qualitative input from a financial supervisor with the South African Reserve Bank.

In this research, the focus was not only on banks. Other businesses whose growth may be directly or indirectly affected by the probability of the consumer defaulting on payment can also be considered in the light of credit risks. This includes telecommunications and retail outlets that offer debt facilities to their clients.

1.9 RESEARCH METHODOLOGY

This research used data that is found (the data is publicly and freely available for educational research and other non-commercial use) at the Federal Reserve Bank of St. Louis, US, the South African Reserve Bank (SARB), and the National Credit Regulator of South Africa (NCR) websites (SARB, 2020; FED, 2022; National Credit Regulator, 2022) to model consumer credit risk losses. The modelling entailed multi-step multivariable regression analysis of selected explanatory variables on the credit losses which is the target or independent variable. The explanatory variables were selected using correlation, bivariable regression, and ANOVA analyses. Quantitative research, which is elaborated further in the research methodology chapter, is the main method that was used in this research. The methodology chapter also includes the proof-of-concept modelling techniques and concepts, ethical considerations, research validity and reliability, and foreseen limitations of the research, and proposed remedies.

1.10 THEORETICAL FRAMEWORK OF THE RESEARCH

The theoretical framework outlines the theories, hypotheses, and concepts upon which this research is constructed. It situates and contextualises the theories used in the study (Adom, Hussein and Agyem, 2018).

1.10.1 The Efficient Market Hypothesis and the Theory of Reflexivity

Figure 1.5 is a diagrammatic representation of the two key theories – the Efficient Market Theory and the Theory of Reflexivity – that are part of the framework.

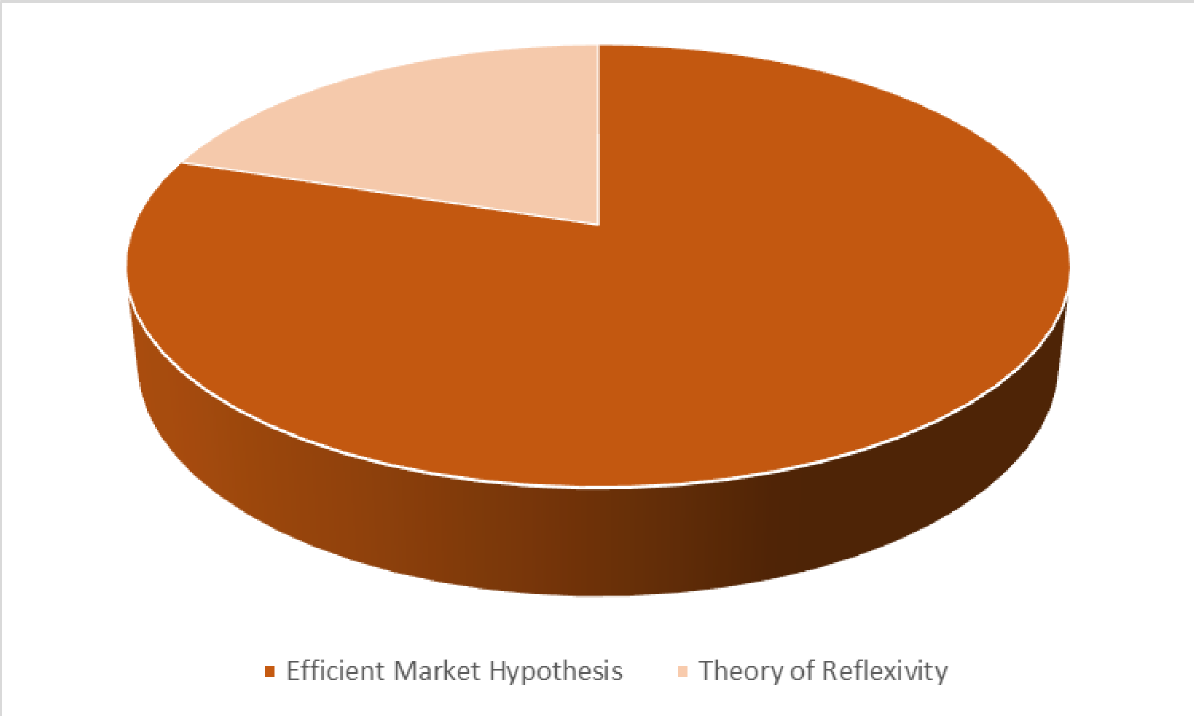


Figure 1.5: Diagrammatic Representation of the Application of the Efficient Market Hypothesis and the Theory of Reflexivity.

(Source: Own preparation based on Soros (2012), Liew (2014) and Delcey (2019)).

Figure 1.5 is a diagrammatic representation showing that the Efficient Market Hypothesis applies in most of the markets situations (equilibrium situations) but not in the less frequent far-from-equilibrium situations (such as recessions, and financial booms and busts) while the Theory of Reflexivity only holds in such situations.

As aforementioned, the Efficient Market Hypothesis (EMH) or Theory and the Theory of Reflexivity are the two (2) key theories underpinning this research. Efficient Market Hypothesis posits that the price of a security in the financial market reflects all publicly available information about its fundamental value (Russel and Torbey, 2002). Such securities include those that derive their value from underlying credit assets like mortgages, credit card loans, personal loans and other consumer loans (Ishikawa, 2009). The financial crisis of 2007-2008, for example, was caused by issues of subprime mortgages that were then securitised and sold to investors in the financial markets (Jarrow, 2012). There is therefore a link between the market prices of certain securities and the credit risks inherent in their underlying credit assets (Isaacs, 2014). This relationship is discussed further in the next chapter. Since it was first postulated by Fama and Samuelson in 1965 (Delcey, 2019), the EMH has come under trial by academics, especially in the 1970s and 1980s (Russel and Torbey, 2002; Malkiel, 2003). Some of them now hold the view that the hypothesis does not hold at all times and in all situations (Liew, 2014; Rossi and Gunardi, 2018). One of those is far-from-equilibrium (disequilibrium) situations like the financial crisis environment of 2007-2008. On the other hand, the Theory of Reflexivity espouses two principles – the principle of fallibility and the principle of reflexivity (Soros, 1992, 2012). It postulates that in events that have thinking participants (as is the case in financial markets), the participants' understanding of those events is always incomplete and distorted. This is the principle of fallibility. This misunderstanding leads to inappropriate action that can influence the situation to which it relates. This is the principle of reflexivity (Soros, 2014).

Thinking serves two functions; one is to understand the situation, and this is referred to as the cognitive function while the other is to manipulate the situation in the participants' favour (manipulative function). The two functions connect thinking in one direction and reality in the opposite direction. When the two functions operate simultaneously, there is interference between them, and the deprivation of each other's explanatory variable needed to determine the value of the explained variable. As a result, their understanding of reality and the actual course of events becomes uncertain and the two phenomena become reflexive, taking the form of feedback loops (Soros, 2014). Participants' understanding of reality drives the course of events and the course of events drives the participants' understanding of reality. The influence is continuous and circular, turning it into a feedback loop. These feedback loops can reinforce one another in one direction or the other leading to a boom or a bust[2] as is the case in financial

[2] Credit cycles shown in Figure 2.3 are an example of a boom-and-bust phenomenon.

markets (Soros, 2012; Marks, 2022). The theory of reflexivity only applies in events with thinking participants and far-from-equilibrium situations (Soros, 2014). As stated earlier, the EMH on the other hand does not apply in far-from-equilibrium situations but holds well in equilibrium situations. This study uses this coincidence to select explanatory variables that mimic the two situations: economic factors and obligor characteristics that apply in equilibrium situations and proxy variables for sentiment (or participants' views of the market in far-from-equilibrium situations) (Marks, 2022). This is elaborated further in Section 3.3.1 of the methodology chapter.

1.10.2 The Basel II/III Capital Accords

The credit risk modelling done in this research conforms to the model-building guidelines of the Basel Committee on Banking Supervision (BCBS). The Committee was founded by the group of the 10 most developed nations (G10) in 1974 and meets at the headquarters of the Bank for International Settlements based in Basel, Switzerland. The guidelines are part of the Basel capital accords set by the Basel Committee on Banking Supervision. The Committee's main aim is to help banks determine and set aside regulatory capital (BCBS, 2017). Amongst other recommendations, the Committee proposed that the burden is on the bank to satisfy its supervisor that the model has good predictive power and accuracy and the variables used must form a reasonable set of predictors. The Committee also expects the bank to have a data vetting process, and such data used to build the models must be representative of the borrowers. A process must be in place to combine human judgement and model results and the bank must have procedures for human review of model-based rating assignments. Credible ongoing efforts to improve the model's performance must be in place and the bank must have a regular cycle of model validation, monitoring, review, and testing of model outputs against the outcome. The Basel capital accords are covered in more detail in Section 2.6.6 of the chapter on extensive literature review.

1.10.3 Epistemology and Ontology

The Oxford English Dictionary defines epistemology simply as the part of philosophy that deals with knowledge and ontology as the branch of philosophy that deals with the nature of existence (Simpson and Weiner, 2022). Epistemology is the philosophical study of the nature, origin and limits of knowledge (Abduholiqovna, 2021). The main research philosophies related to business research are pragmatism, positivism, realism, and interpretivism (Dudovskiy, 2018) (Dudovskiy, 2018). Pragmatism is the view that both observable phenomena and subjective meanings can lead to acceptable knowledge, while in positivism only observable phenomena are deemed to provide credible data or facts. In realism, observable phenomena are accepted as leading to knowledge, but the focus is on explaining within context or contexts. In interpretivism, the focus is on the details of the situation and the reality behind these details. The philosophical stance of this research is positivism and the hypothetico-deductive reasoning or scientific method was used (Walliman, 2011). The research is quantitative in nature (McNeil, Frey and Embrechts, 2005).

1.11 CONCEPTUAL FRAMEWORK OF THE RESEARCH

The modelling framework adapted for this research is shown in Figure 1.6. It is a process that has seven (7) steps of which the first is selection of explanatory variables to predict the credit losses (dependent variable). Data available on the variables – firstly from the developed market of the United States of America (US) –were then prepared for ease of analysis in the third step. The preparation included, however, was not limited to ensuring that data for the respective variables covered the same period and that they were of the same frequency. To gain insight into the relationships between credit losses and each of the explanatory variables and the collinearity of the variables, the data were then analysed using various analysis techniques: correlation, bivariable regression, ANOVA, and multivariable regression. In the fourth step, statistics from the multivariable regression analysis were extracted and incorporated into the proof-of-concept credit risk model for the developed market of the US. This model was back tested by comparing the credit losses that it estimated over a period and the actual losses incurred in the period. The process was repeated when new variables were added. Once the best proof-of-concept model had been created from available data, this process was repeated for the creation and back testing of the emerging market of South Africa’s consumer credit risk model.

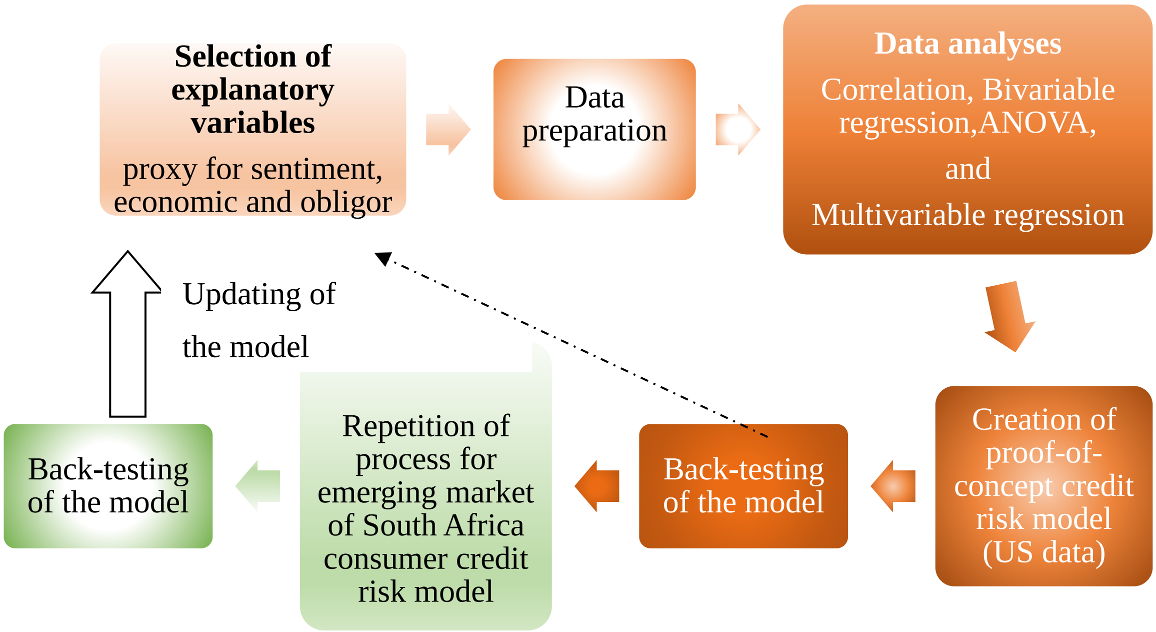


Figure 1.6: Conceptual Framework for Developing and Adapting a US Credit Risk Model for The Emerging Market of South Africa.

(Source: Own preparation based on Baesens (2015, p.6)).

Starting with the selection of explanatory variables, Figure 1.6 shows the framework that captures the essential steps that are followed in creating the proof-of-concept credit risk model, using data from the developed market of the US and adapting it for the emerging market of South Africa.

1.12 RESEARCH OUTPUT

A paper inspired by this research was accepted for presentation at the 7th International Conference on Business Management (ICoBM) held in Pakistan in the year 2020 (Joseph and Kimetto, 2020). A copy of the attendance certificate is shown in Annexure A. The researcher was afforded the opportunity to edit the paper that was subsequently published by the Journal of Risk and Financial Management. Additionally, the researcher assisted the supervisor to review risk journal articles – as requested by the journal editors - written and submitted for publication by other authors. This has put the researcher in good stead to write own/joint papers for publication.

1.13 CHAPTER SUMMARY

In this chapter of the thesis, the state of consumer credit generally around the world and specifically in South Africa was addressed. The significance and impact of the intended solution to the problem that was identified as being poor estimation accuracy of existing consumer credit risk models were examined. Next, the aim of the research, the research questions, the primary and secondary objectives of the research, its scope, delimitations, and an introduction to the methodology that was used to carry out the research, were outlined. Also covered were the theoretical and conceptual frameworks, and research output. In the next chapter, an extensive literature review is presented.

CHAPTER 2: LITERATURE REVIEW

2.1 INTRODUCTION

The previous chapter was an introduction to this study. In the chapter, the researcher made the case for the study – that credit and credit risk are increasingly a big part of a sizeable portion of the population worldwide (Demirgüç-Kunt *et al.*, 2018) and that credit risk is difficult to measure since existing credit risk models are inaccurate (Baesens, Rosch and Scheule, 2016). The aim of the study, the questions and objectives set to solve the problem, the scope, and the delimitations were addressed. The methodology, theoretical, and conceptual frameworks used in the study were introduced. Also mentioned was the output of the research. This chapter is a report of the extensive review of the literature found to be relevant to this research on consumer credit risk models. The content of the literature is divided into six (6) broad sub-categories. The theories on which the research is founded are discussed in Section 2.2. In Section 2.3 credit in general is presented, whereas consumer credit in particular is reviewed in Section 2.4. Section 2.5 is a presentation of the growth and development of consumer credit, and Section 2.6 is about consumer credit risk and its measurement. In Section 2.7 the gaps identified in the literature reviewed and the research contribution are addressed. Section 2.8 is the summary of the chapter.

2.2 THEORIES ON WHICH THE STUDY IS FOUNDED

In this section, a more in-depth review is presented of the theories that were introduced in chapter one (1) and which informed the study. They are the Efficient Market Hypothesis (EMH) or Theory, and the Theory of Reflexivity. According to Delcey (2019), the EMH was first proposed by Fama and Samuelson in papers that they separately wrote in 1965. Fama was then a professor at the Chicago Graduate School of Business while Samuelson was a professor at the MIT School of Industrial Management – now the MIT Sloan Business School of Management (Delcey, 2019). At the time, there was a focus on the study of economic finance and its impact on investment practice and policy making and the two schools were renowned in the field. Fama focused on the application of EMH in investment decisions while Samuelson concentrated on its implications for policymaking (Delcey, 2019). As a result of its application, the EMH has had far-reaching effects on both the investment field and policies that have been made or proposed to be made at the government level (Ang, Goetzmann and Schaefer, 2010).

The EMH[3] postulates that security prices in financial markets reflect all publicly available information about their fundamental value and therefore, trying to come up with investment ideas – like active trading using charting (following market trends) or value investing (the idea of identifying and investing in undervalued security assets using fundamental analysis) to beat the market is a futile exercise. It also informed the thinking behind small government – the idea that since the markets are best at determining asset prices, there is no need for governments to intervene. This ideology resulted in the growth of passive investing (index investing) and the reduction in the role of government in financial markets through deregulation and the sale of previously government-owned businesses and assets (Russel and Torbey, 2002; Ang, Goetzmann and Schaefer, 2010; Liew, 2014; Rossi and Gunardi, 2018). However, in the 1970s and 1980s, EMH came under opposition (Russel and Torbey, 2002; Liew, 2014).

Academics and practitioners raised questions on its validity or limitations in certain market situations. In a 1984 speech – that raised doubts about the Efficient Markets Hypothesis – to commemorate the 50th anniversary of the book entitled ''Security Analysis'' written by Benjamin Graham, Warren Buffet listed seven (7) investment professionals (and two (2) investment funds), all of which practised value investing (the notion of buying undervalued assets in the financial markets that are mispriced due to market inefficiency), and whose investments had outperformed the markets for extended periods (Buffet, 1984). He also stated that if the markets were efficient, he would have been a beggar instead of the rich man that he became through investing in mispriced or undervalued financial securities over an extended period (Rattner, 2013). Academics also questioned the EMH's inability to explain the market crash of 1987 and the financial crisis of 2007-2008 as well as other market anomalies (Malkiel, 2003; Ang, Goetzmann and Schaefer, 2010; Rossi and Gunardi, 2018).

While the debate on the EMH may continue (Liew, 2014), the proponents, and the critics of EMH generally agree that it is a useful theory and applies in equilibrium and near-equilibrium

---

[3] This is the semi-strong form of EMH which has three forms: the strong form, the semi-strong form and the weak form. The strong form postulates that prices of securities in the financial markets reflect all information including private information. However, insider traders who hold information not yet available in the market are known to profit from them. The weak form suggests that prices or returns of securities reflect past prices or returns and that future prices or returns can be predicted using accounting or macro-economic variables. Research evidence do not support the latter part of this suggestion. This research is therefore based on the semi-strong form as is the case with most empirical research (Russel and Torbey, 2002).

situations. It does not apply in far-from-equilibrium[4] situations (economic recessions, financial booms and busts) (Buffet, 1984; Russel and Torbey, 2002; Rossi and Gunardi, 2018) while the Theory of Reflexivity only holds in such situations. The Theory of Reflexivity postulates that the prices of security assets in the financial markets in far-from-equilibrium situations reflect the perceptions of the participants in the markets (Soros, 2014; Lawson, 2015). Since participants' views – according to the theory and as elaborated on in Section 1.10.1 – are always incomplete and distorted, the prices can diverge far from reality, resulting in a boom or a bust. Soros (2012), the key proponent of the Theory of Reflexivity, was influenced by the ideas of Popper[5], whose writings he read while studying at the London School of Economics, and where he graduated with Bachelor and Master of Science degrees in philosophy in 1951 and 1954 respectively. Popper also proposed the hypothetico-deductive reasoning or scientific method of research (Walliman, 2011; Soros, 2014), and argued that the attainment of perfect knowledge is impossible since human beings are inherently fallible (Soros, 1992). In addition, he proposed the idea of the unity of method, the notion that phenomena and events occurring in natural and social science can be treated similarly.

This researcher agrees with Soro's argument that there is a fundamental difference between the two: Social phenomena have thinking participants while natural phenomena do not. Soros (1992) also claimed that there is a special affinity between credit and reflexivity in that credit depends on expectations and expectations involve bias. Credit is therefore a phenomenon that allows bias to have a causal effect in the course of events leading to its association with booms and busts (Soros, 1992; Russel and Torbey, 2002; Marks, 2022).

---

[4] Exactly when far-from-equilibrium situation is entered in a boom and bust process can only be established in retrospect (Soros, 2003).

[5] Karl Raimund Popper, born on 28 July 1902 in Vienna, is generally regarded as one of the greatest philosophers of science of the twentieth century. From 1937 to 1945, he taught philosophy at the University of Canterbury, New Zealand. In 1946, he moved to England to teach at the London School of Economics and became professor of logic and scientific method at the University of London in 1949. Popper was knighted in 1965, and retired from the University of London in 1969 (Thornton, 2022).

2.2.1 Relationship Between Risk and Return

As stated in Section 2.2, this research is underpinned by the Efficient Market Theory – also referred to as Efficient Market Hypothesis (EMH). EMH holds that there is a direct relationship between risk and return; that the higher the risk associated with an investment, the higher the return and vice versa. This is derived from the hypothesis that the price of a financial instrument reflects all available information about its fundamental value – meaning that above-average returns can only be obtained with increased risk. In the real world, this simplified relationship does not exist. Market participants have imperfect information (Soros, 1992) and are forced to deal with perceived risk and expected return. As the level of the perceived risk increases, the range of possible outcomes also increases. This makes it difficult to predict outcomes with certainty.

Actual returns and the expected could differ widely. Moreover, having perfect information and analysis is no guarantee that taking on greater risk results in greater future returns (Isaacs, 2014). This is in line with the argument that the traditional line graph (represented by the straight line in Figure 2.1) is incomplete (Marks, 2006). It shows a positive correlation between risk and returns; however, it does not communicate the uncertainty involved. Normal curves of increasing spreads or standard deviations – the higher the risk – are added to the line to capture the fact that as risk increases, so does the uncertainty of return and the possibility of loss (Marks, 2006, 2022). Moreover, as EMH does not hold in disequilibrium situations, high returns can be obtained with decreased risks (when the market under-prices financial assets) or with disproportionately greater risk (when the market prices are much higher than the fundamental value) (Buffet, 1984).

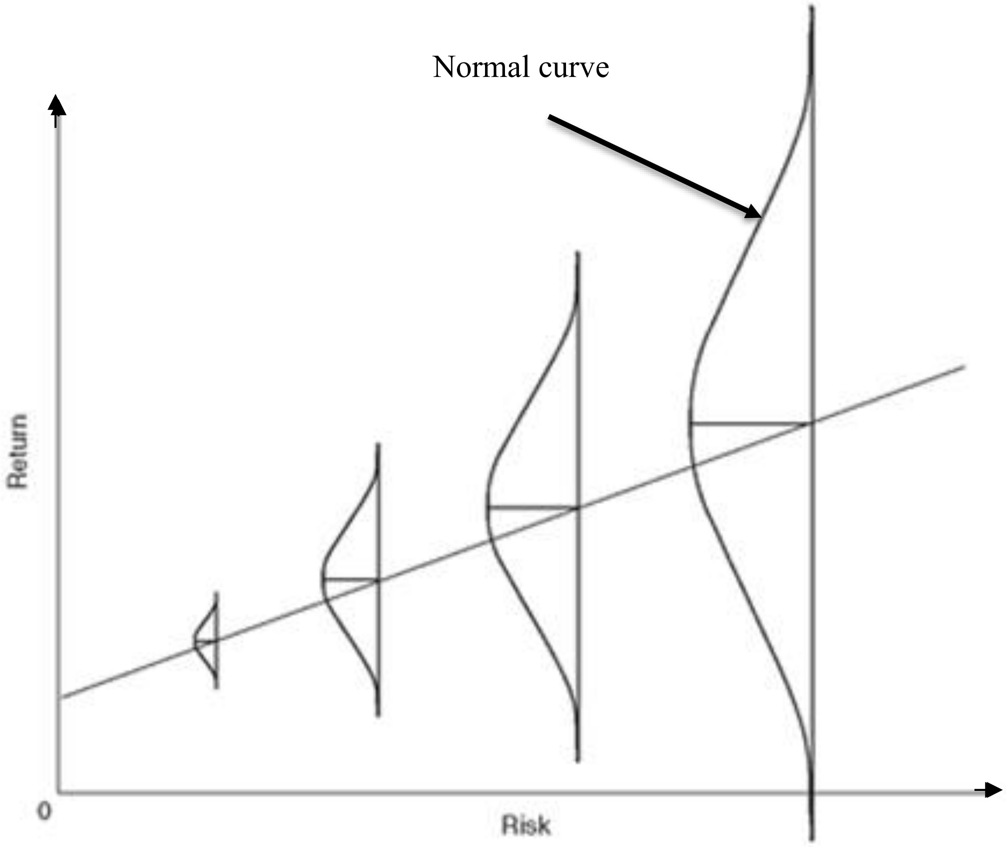


Figure 2.1: Relationship Between Risk and Return.

(Source**:** Marks (2006, p.2)).

The solid line graph and the normal curves of Figure 2.1 indicate that while return may increase as risk increases, the range of potential outcomes (and therefore the uncertainty of those outcomes) rise – from high positive through zero to high negative returns.

2.3 CREDIT IN GENERAL

Credit is the most important factor in the growth and development of a country (Dalio, 2013; BCBS, 2017; Gabeshi, 2022). Credit is based on trust (Borio, 2019) and the level of trust has been found to affect the cost of credit (Meng and Yin, 2019). Credit is also known to affect the standard of living for individuals depending on how it is used. Debt magnifies the rewards for those who use it to leverage their positions to increase capital gains in the property and financial markets (Sgambati, 2022). These are mainly the middle and upper classes of the population who borrow relatively more than the other classes. It turns out that class struggles concerning debt are not between rich creditors and the poor masses to whom they lent their surplus funds

(Swartz, 2010) but between debtors who borrow (mainly for consumption) and debtors who borrow even more for investments in productive assets (Sgambati, 2022). This difference in the use of credit may, at least partially, explain the wealth gap between the rich and the poor (DW Documentary, 2017).

2.3.1 Types of Credit

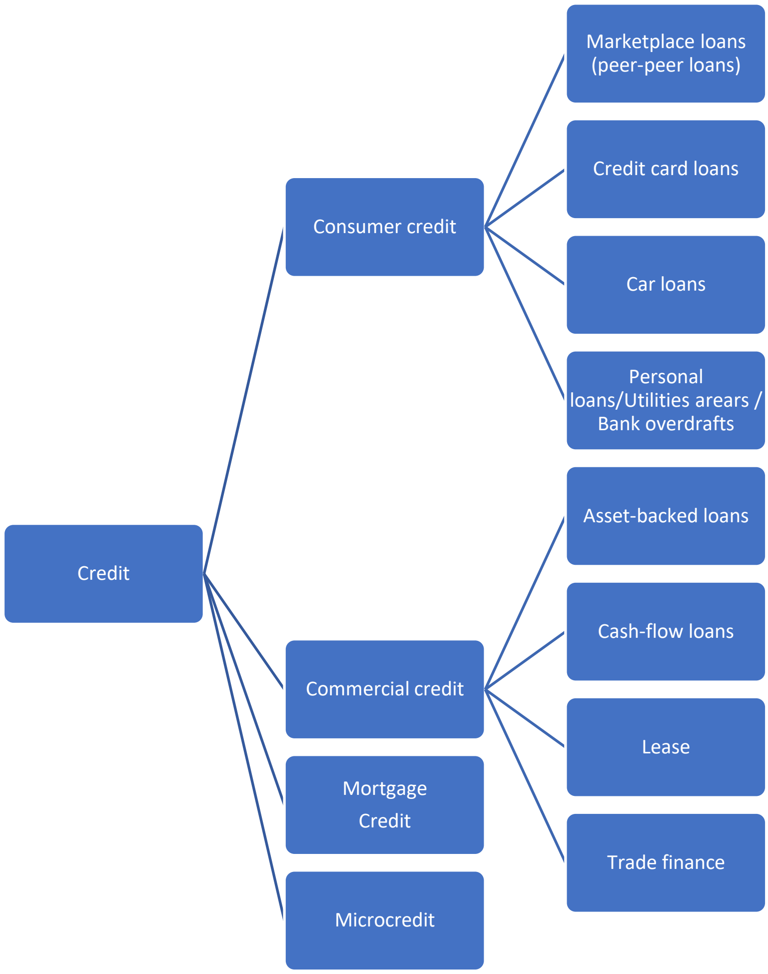


Figure 2.2: Types of Credit

(Source: Own preparation based on Ivashina, Laeven and Moral-Benito, (2020) and Dueñas-Peña, Barbosa-Guerrero and Rozo, (2022)).

Figure 2.2 shows the classifications that most of the credit fall into (Ivashina, Laeven and Moral-Benito, 2020; Dueñas-peña, Barbosa-Guerrero and Rozo, 2022). However, it may not include all types of credit.

Credit is the ability of a customer to obtain money, goods, or services before payment with the promise that payment will be made in future (Simpson and Weiner, 2022). Most of the credit issued by banks falls into three categories: consumer, commercial, and mortgage (Dueñas-peña, Barbosa-Guerrero and Rozo, 2022). Consumer credits are loans borrowed by individuals to meet their cost of consumption of goods and services. They include credit card, personal, car, and store card loans, utilities in arrears, and bank overdrafts. Commercial credits are loans extended to institutions for working capital, investment, and operating expenses. As shown in Figure 2.2 commercial credits consist mainly of asset-backed loans, cash-flow loans, trade finance and leasing (Ivashina, Laeven and Moral-Benito, 2020). The main difference between them is the type of collateral that backs them up.

The core characteristics of collateral are liquidation value, pledgeability, and durability. Asset-backed loans are backed by physical assets like property or aeroplanes and cash-flow loans are backed by a first (or senior) charge on the sales proceeds of unencumbered assets of a company in the event of default. In the assessment for the issue of such a loan, the emphasis is on the borrower's ability to pay rather than the value of the collateral. Therefore, there is usually no specific list of collateral. Trade finances are business-to-business loans that are backed by the goods in the transaction. Leases are arrangements in which the lender purchases a property and lets the borrower use it for an agreed periodical fee. However, the borrower does not have ownership of the property under the lease, as the title to the property remains with the lender. The advantage of this arrangement is that in case of default, the property can easily be repossessed (Ivashina, Laeven and Moral-Benito, 2020).

A mortgage loan can be either consumer or commercial in type and is a loan granted for the purchase of a property. The property purchased or built becomes the collateral for the loan (Dueñas-peña, Barbosa-Guerrero and Rozo, 2022). It is treated as a separate type of loan since it generally accounts for the largest portion of a bank's loan portfolio (Oliver and Hand, 2005). Micro credit and marketplace (peer-to-peer) loans are a type of innovative consumer loans. Micro credits are loans extended to borrowers with weak or no collateral. They are normally small in size as their name suggests. Marketplace loans are a recent development and are loans that are normally available through platforms whereby borrowers access loans entirely online from private lenders. The way it works is that a borrower applies for the loan on an online platform. The platform managers then assess the creditworthiness of the borrower and prices the loan (indicate the interest rate). Investors on the platform then decide whether to lend money to the borrower (Dore and Mach, 2019).

2.3.2 Importance of Credit

The importance of credit, more than anything else, is reflected in its magnitude: As at the end of 2021, global debt is estimated to have stood at US$ 303 trillion equivalent to 360% of global GDP (Sgambati, 2022). In 1980 it was three times less as a percentage of GDP (at 120%). Credit is therefore arguably the most important factor in an economy (Dalio, 2013). The economy consists of a few parts amongst which are transactions that are repeated many times over. The economy is the sum of all the transactions that make it up. Each transaction is made by a buyer exchanging money or credit for goods, services, or financial assets with a seller. These transactions are driven by people and create the forces that drive the economy: productivity growth, short-term and long-term debt cycles (Dalio, 2013). The total of all money and credit used to make purchases form the total expenditure. Expenditure drives economic growth. Economic growth can therefore be increased by increasing the amount of credit in the economy.

There is a limit to how much debt can be issued in the economy as it is dependent on income for its repayment. When debt rises too high above the income, the prices of goods, services and financial assets rise. To lower inflation, the Central Bank then raises interest rates to make credit more expensive (Borio, 2019). When credit becomes expensive, fewer people can afford it and less of it is taken up. As people, businesses and government spend less, economic growth slows down. This leads to deleveraging in the economy. This process of leveraging and deleveraging of credit in the economy repeats itself continually. In the short term, the net outcome of the rises and falls (short-term debt cycle) in credit is a net increase. Over the long term, therefore, debt still becomes unsustainable with respect to income. It then is forced to deleverage resulting in a long-term debt cycle. Long-term debt cycles similarly recur. These cycles are shown in Figure 2.3 (Dalio, 2013).

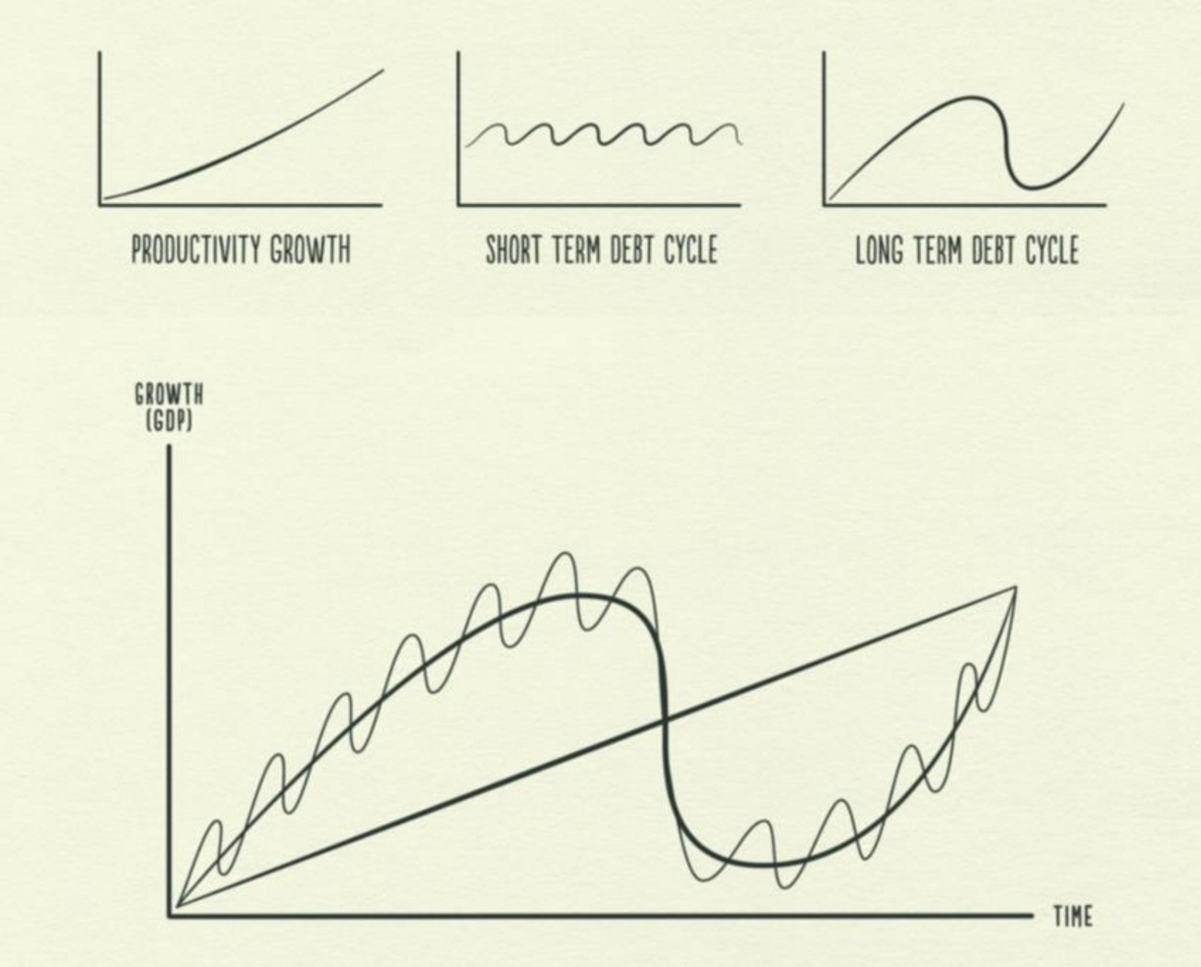


Figure 2.3: Short and Long-Term Credit Cycles

(Source: Dalio (2013)).

In general, short term debt cycles occur every five (5) to eight (8) years while long term debt cycles occur every 75 to 100 years (Dalio, 2013). These cycles are associated with the booms and busts in the financial markets and the real economy. Examples include the great depression of the 1930s and the financial crisis of 2007-2008.

2.3.3 The Emergence of Credit Risk Modelling

With the increase in the number of applications for loans and credit cards, it became difficult for banks to hold onto the traditional method of bank supervisors assessing the creditworthiness of applicants on a qualitative, one-on-one and face-to-face basis. Banks, therefore, came up with application credit-scoring systems that used information supplied by applicants on loan applications to assign a creditworthiness score to an applicant (Oliver and Hand, 2005). Loans were then issued to applicants whose credit scores were equal to or above a predetermined minimum score. With time the banks also developed similar scoring systems based on the performance of loan repayments. These are referred to as behavioural scorecards. Today banks have many other models on credit for different purposes (Koulouridi, 2020). For example, there are models for predicting the profitability of a borrower, the probability of default of a portfolio of loans, and the credit losses in a period, etc.. These are covered in more detail in Section 2.6.3. This study sought to develop, using data from the developed market of the US, a consumer credit risk model with improved capability to estimate consumer credit losses and to adapt the model to the emerging market of South Africa.

2.4 CONSUMER CREDIT

Antoniades (2018) in his study on the social power of money and debt, argues that ever larger parts of life and nature are turned into future cash flows; being indebted has become normal while the future of students has been monetised through student loans. This, according to him, creates convoluted debt structures that are unstable and unsustainable. Furthermore, according to Eubank (2012), consumer credit is often a source of unhappiness in an individual's and family's life. He opens his case study on the effects of consumer credit on the family by asking why consumer credit exists for the small wage-earner. In his opinion, such business should be prohibited altogether.

Angel and Heitzmann (2015) note that the route credit consumers take into and out of this situation is multi-dimensional and complex. And that the factors that influence the likelihood of a household becoming over-indebted can be traced to individual(s), the household and/or the total population of a country. With reference to the South African context, the National Credit Act (NCA) No. 34 of 2005 (Republic of South Africa, 2006; Renke, Roestoff and Haupt, 2007) was established to address the high levels of consumer over-indebtedness. Evidence of the NCA having had the intended effect is lacking (Botha, Booyens and de Wet, 2015). However, such macroeconomic variables as Gross Domestic Product, Interest Rate, Debt-To-Disposable Income, Consumer Consumption Expenditure, and Unemployment Rate are major explanatory factors of credit risk. Swartz (2010) argues that people whose economic model is based on interest end up having two (2) classes: the lending class who are excessively rich and the poor who cannot afford even the basic needs of life. He proposes the adoption of the Islamic economic model in which the bank receives a portion of the profits (if any) of the business instead of interest.

2.5 GROWTH AND DEVELOPMENT OF CONSUMER CREDIT

Over the years some religious societies were averse to credit that attracted interest. Canon law, for example, did not allow it altogether, and even today sections of Islamic society do not allow it (Swartz 2010). Moreover, Judaism forbade Jews from charging interest on loans to fellow Jews (Exodus 22:25; Leviticus 25:35–37; Deuteronomy 23:19-20) (Blblica Inc., 2011). Nevertheless, consumer credit has grown and developed apace. This has been helped by its profitability, technology and financialisation. Hyman (2012, p.40) argues that consumer credit has grown in both absolute terms as well as relative to investment credit since it became more profitable than other investment options. He points out that, for a bank or other financial institution, extending credit is an investment decision as the allocation of money to credit is made in the light of other uses for the money. More broadly, however, consumer credit has evolved in form and grown in depth and breadth in tandem with economic revolutions (Garrett, 2000). The economic revolutions have been both the reason for and the enabler of the expansion of consumer credit. In the hunter-gatherer age, credit may have taken the form of items borrowed with the understanding that they would be returned in the same form and/or quantity at the agreed-upon time. With the agricultural revolution, farmers structured credit deals enabling them to buy forward seed and other farm inputs which were repaid upon harvesting. Later, during the industrial revolution, credit in the form of capital was required to set up factories and the required machinery (Hyman, 2012; Ventura and Voth, 2016). In the technological revolution, credit has been instrumental in its development while at the same time technology has enabled more efficient and effective management of credit (Sanchez, 2017). Of the connection between the information technology revolution and the internationalisation of finance (and more so the speed at which this happens) Garrett (2000, p.956) argued that shrinkage of time and space is more apparent in international finance than in other fields as the modern means of communication, like the internet, have significantly reduced the costs of transmitting information. Besides technology, other factors that have accelerated the growth and development of consumer credit are credit scoring and reporting, and financialisation (Garrett, 2000; Langley, 2008; Burton, 2012; Mester, 2015). Langley (2008) argues that consumer credit has evolved from the notion of risk intermediation to the notion of risk decomposition, and risk transfer, and that the resulting transformed credit networks (formally defined as the mathematical representation of the real credit relationships within an economy) have an increasing build-up of credit risks. An example of such credit networks, composed of mortgage borrowers, banks (lenders), protection sellers, and investors, is next illustrated by how asset-backed securities were created.

A resurgence of interest in the study of credit risk occurred in the aftermath of the 2007-2008 financial crisis. The crisis was precipitated by, amongst many other factors, the bundling of US mortgage loans by banks into residential mortgage-backed securities (RMBS) that were sold to investors. Bankers ostensibly transferred credit risk to other parties through credit default swaps (CDS), a typical credit derivative. They took on more risk by issuing subprime mortgages – loans issued to borrowers with poor creditworthiness. These financial instruments were similarly securitised, therefore creating a spiral of an increasing credit risk bubble (Langley, 2008; Ishikawa, 2009; Lowenstein, 2011; Hull, 2012; Jarrow, 2012). Figure 2.4 is a diagrammatic representation of a simpler example of one of the many credit derivatives created and sold to investors in the run-up to the 2007-2008 financial crisis. The diagram shows the five (5) steps through which these derivatives were created.

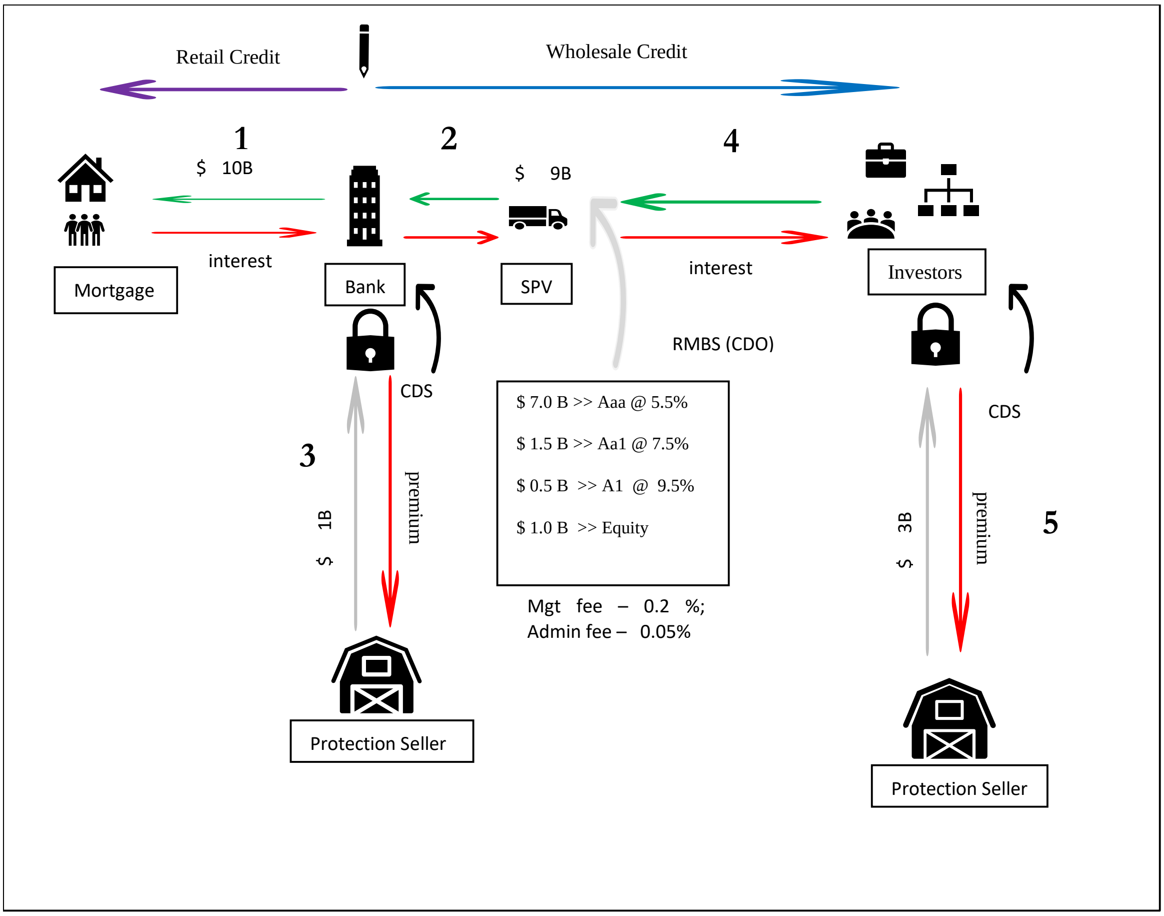


Figure 2.4: Residential Mortgage-Backed Securities (RMBS)
(Source: Own preparation based on Ishikawa (2009, p.66, p.87)).

The financial assets created included collateralised debt obligations, credit default swaps, and other credit derivatives. In the first step, a bank issued loans worth a total of US $10 Billion to, for example, 40,000 consumers of various credit worthiness (rating), to buy homes *(see point no 1, in* Figure 2.4*)*. The loans were secured by the homes purchased. In return, the bank received interest and a portion of the principal amount on an agreed schedule. The interest may have been fixed or variable (in most cases it was variable). The bank formed a special purpose vehicle (SPV) which pooled and sliced the loans into tranches according to their credit risk ratings *(see point no 2, the rating box in* Figure 2.4 *and Annexures B and C)*. These were sold to investors who in return for the cash injection into the SPV would receive interest payments from the streams of mortgage payments being made by the consumers. It is to be noted that what gave the tranches the ratings is the slicing and tranche categorisation in terms of seniority with respect to the order of access to the cash flows.

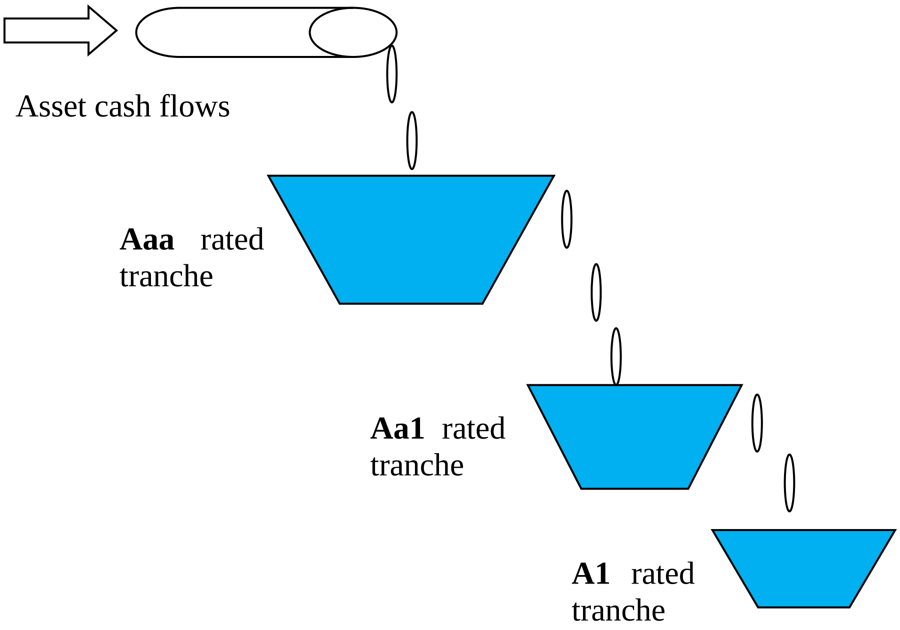


Figure 2.5: Typical Waterfall of an Asset Backed Security

(Adapted from: Hull (2012)).

As shown in Figure 2.5, cash flows (of interest and principal) flow to the tranches in order of their credit ratings (Moody's rating scale shown in Annexure B was used). The cash flows are not guaranteed, and the A1-rated tranche is more likely to lose a portion or the whole of it compared to the Aaa-rated tranche (credit risk increases with ratings' movement from Aaa to A1) . The Aaa[a]-rated securities were the first to receive the promised principal and interest in the waterfall arrangement of Figure 2.5 (Hull, 2012). Once these securities had received the full allocation of the promised principal or interest cash flows (there were separate waterfalls for the principal and interest), then any additional amounts available would be paid to the next rated tranche (Aa1 in this example) until they had received the full amount promised. This arrangement similarly applied to the A1-rated securities. Viewed from the perspective of losses, if there were any, this meant that the lowest-rated securities would be the first to absorb the losses in proportion to their promised principal or interest amounts, while the highest-rated securities would be the last.

As depicted in Figure 2.4, the bank was able or chose to sell US$9 billion of the loans and keep the balance of US$1 billion in its books. It subsequently purchased credit default swaps (CDS) *(see point no 3, in Figure 2.4)*, insurance, to cover this balance. In so doing, it effectively transferred all the credit risk associated with the home loans to credit derivative investors. The CDS derivative seller in return received a premium for the provided cover. In case of no default, the CDS seller paid nothing and the premiums it received went into its profit and loss account (P&L). If, however, the mortgage holders defaulted, the seller paid compensation to the buyer for the agreed notional cover amount less any recoverable money.

The SPV *(see point no 4, depicted in Figure 2.4)* sold the tranches of securities to investors in various combinations depending on their risk appetites. In return, the investors received an agreed periodical interest payment as a reward for their investment and the initial capital at the end of the investment period. The highest-rated tranches were the first to absorb the cash flows and the last to take a knock on the losses. Since they were perceived to have lower default probabilities, they received a lower rate of interest compared to the riskier tranches.

Some of the investors may, out of choice or developing cold feet, have bought CDS to cover part of their investment *(see point no 5, depicted in Figure 2.4)*. In that case, they would have had to part with a portion of the interest that they were receiving from their investment to pay for the premiums. Others, which may include banks, would yet still use their own SPVs to create and sell new configurations of tranches of Collateralised Debt Obligations Squared (CDO^2) backed by the cash flow streams from CDOs. The next iteration in the creation of these products would have been cubed (CDO^3). These products were in so much demand at the time even though most investors claimed they did not understand them. As can be seen from these steps, the parties to the chain of contracts assumed that the credit ratings based on which the securities (tranches) were created, accurately reflected underlying credit risks. This assumption was erroneous, and it cost investors billions of dollars in losses (Jarrow, 2012). With more and more iterations of CODs, CDS and so on created, the correlations between them increased so that when the mortgage holders started defaulting on their payments, the entire system of interconnected debt obligations fell like a row of dominoes.

2.6 CONSUMER CREDIT RISK AND ITS MEASUREMENT

The Oxford Dictionary (Simpson and Weiner, 2022) defines the term risk as a chance or possibility of danger, loss, injury, or other adverse consequences. When considered per se, risk is random as its occurrence is uncertain (McNeil, Frey and Embrechts, 2005). Risk can therefore be represented by equations from the probability field of mathematics. Figure 2.6

illustrates the various categories and classes of risk. Risk can be broadly divided into financial and non-financial risks.

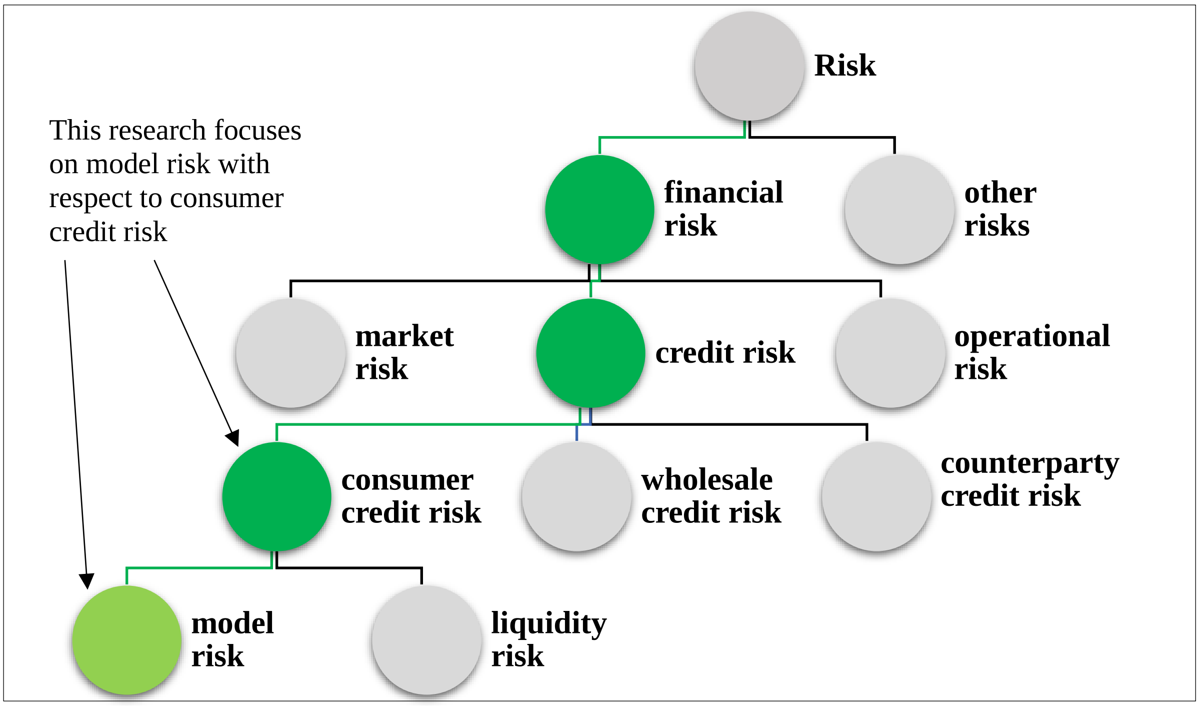


Figure 2.6 Breakdown of Risk Types
(Source: Own preparation based on McNeil, Frey and Embrechts (2005, pp.327-381)).

As depicted in Figure 2.6 (light green shading) model risk, the risk that the model credit loss estimations can be wrong is the focus of this research.

2.6.1 Consumer Credit Risk

Financial risk is defined as the measurable likelihood of loss or less-than-expected-returns (McNeil, Frey and Embrechts, 2005). The main risk types that are encountered in the financial industry are market, operational and credit risks (Figure 2.6). Market risk is the risk that the value of a financial position will change due to the changes in the values of the components on which the position depends, while operational risk is the risk that losses will occur due to partial or complete failure of the control processes and systems that the company has put in place or events outside of it. Credit risk is the risk that an obligor will fail to make the agreed-upon repayments on a loan (Smit, Swart and Niekerk, 2003). These three (3) risks do overlap and the notion of liquidity risk and model risk surface in all of them (McNeil, Frey and Embrechts, 2005). It is worth noting that they do not represent an exhaustive list of risks facing a financial institution. Credit risk can be further categorised into counterparty, wholesale, and consumer credit risks. Counterparty credit risk is the risk that a counterparty to a transaction with a bilateral risk of loss could default before the final settlement of the transaction (BCBS, 2020). Wholesale credit risk is the risk that an institution or a high-net-worth individual to which or whom – principally – a bank has extended a large credit default on the loan (Joseph, 2018). Consumer credit risk is the risk that a borrower will default on a consumer credit product like a mortgage loan, credit card debt, and such other consumer loans.

2.6.2 Components of Credit Risk

For a single financial transaction, components of credit risk are default risk, loss or recovery risk and exposure risk at a certain maturity (Gestel and Baesens, 2009). Default risk is the risk that an obligor fails to honour a commitment to repay the loan, generally within 90 days or longer; loss risk is the risk of loss when default occurs or the risk that the credit provider fails to recover the full amount of the loan outstanding at the time of default (Engelmann, 2021). In mathematical terms, Loss Given Default, LGD = 1 – recovery rate, where the recovery rate is the portion of the loan outstanding that is recovered in the event of a default (Gestel and Baesens, 2009). On the other hand, risk exposure is the notional amount of loan outstanding (exposure) at the time of default (Engelmann, 2021). Maturity refers to the duration of a financial instrument and by or at the end of which the notional principal and/or interest are fully repaid (Sauder and Penas, 2006).

2.6.3 Credit Scoring, Credit Rating and Credit Risk Modelling

Each component of credit risk can be measured on an ordinal scale using credit scoring. Credit scores are ordered (200 to 450 for example application scorecard of Table 2.1) credit risk assessments of obligors with high scores generally implying a low risk of default and low scores denoting a high risk of defaulting on a loan obligation (Gestel and Baesens, 2009). In the example application scorecard of Table 2.1, a customer with age 45, income of 50 000 and the residential status tenant was assigned a score of 150 + 140 + 100 = 390 points. If the loan application approval cut-off point was 300, then the loan application of this customer would have been approved. If, however, the cut-off point was 400, the loan application would have been rejected but may be subject to review by a human expert as the customer's score was close to the cut-off point.

Table 2.1: Example Application Scorecard

Characteristic	Range	Scorecard Points
Age	Up to 30	80
	30 – 40	120
	>40	150
Income	Up to 10 000	50
	10 000 – 100 000	140
	> 100 000	170
Residential Status	Owner	130
	Tenant	100
	With Parents	70

(Source: Gestel and Baesens (2009, p.95)).

The credit scoring system was initially designed by banks to discriminate good from bad payers at the point of application. These were referred to as application scorecards. The system has since been developed to measure other aspects of obligor credit risk as well as for other business purposes. Nowdays there are, for example, performance scores that measure the performance of the existing portfolio of loans, behavioural scores that measure, using payment behaviour, the risk of a borrower defaulting on the loan and profitability scores that measure the profitability of an obligor for the bank. These scores are used by banks to make credit risk management decisions on an obligor, subsequent new loan applications by an obligor, and targeted marketing amongst other uses (Gestel and Baesens, 2009).

When pools of homogeneous borrowers (with similar characteristics) having the same credit score or scores within a certain range are aggregated, the credit risk assessment assigned to the pool is referred to as a credit “class” or “rating”. Credit ratings were first used to distinguish investment grade (above BBB) from non-investment grade bonds (below BBB – see S&P credit ratings in Annexure B). The credit ratings were provided by external credit rating agencies like Moody’s, Standard and Poor’s, and Fitch. Nowdays there are credit ratings of the instruments and issuers. The credit rating of an instrument gives the risk of the issuer defaulting on the specific instrument and the risk of loss on default. The credit rating of an issuer gives an assessment of the overall risk of an issuer defaulting on its obligations, usually referenced to its unsecured debt (see Table 2.2). The credit rating is determined by a team of experts after doing a thorough analysis of the issue or issuer’s financial and other business information, both quantitative and qualitative. In the aftermath of the Basel II capital accords (see Section 2.6.6), banks have had the incentive to develop internal credit ratings across all their credit portfolios. Basel II gave banks, depending on size and the level of sophistication and subject to regulatory approval, the option to use a standardised approach, the internal-ratings-based approach (IRBA) or the advanced internal-ratings-based approach (AIRBA) in calculating their regulatory capital requirements (Oliver and Hand, 2005). However, banks, especially the internationally active ones, still use the ratings of external credit rating agencies to assess their credit risk exposure to counterparts and to benchmark their internal credit ratings. Long-term issuer default ratings by Moody's, Standard & Poor's, and Fitch are shown in Table 2.2.

Table 2.2: Long-Term Issuer Default Ratings by Moody's, Standard & Poor's and Fitch.

Moody's	**S&P**	**Fitch**	**Credit quality**	
Aaa	AAA		Extremely strong	
Aa1	AA+	AA+		
Aa2	AA	AA	Very strong	
Aa3	AA−	AA−		
A1	A+	A+		Investment
A2	A	A	Strong	Grade
IA3	A−	A−		
Baa1	BBB+	BBB+		
Baa2	BBB	BBB	Adequate	
Baa3	BBB−	BBB−		
Ba1	BB+	BB+		
Ba2	BB	BB S	Speculative	
Ba3	BB−	BB−		
B1	B+	B+		
B2	B	B	Highly speculate	
B3	B−	B−		Non-investment
Caa1	CCC+	CCC+		Grade
Caa2	CCC	CCC	Vulnerable	
Caa3	CCC−	CCC−		
Ca	CC	CC	Highly vulnerable	
C	C	C	Extremely vulnerable	
RD	SD	RD	Selective, restrictive default	
D	D	D	Default	

(Source: Gestel and Baesens (2009, p.116)).

While credit scores and ratings are an ordinal measure of credit risk, credit risk models are used to quantify actual credit components expressed by the credit scores and ratings. Probability of Default (PD) model for example is used to quantify the probability of an obligor defaulting on its obligation to pay a loan. The simplest way of determining such quantities is to look at historical figures of actual default rates to determine the probability of default. However, there are many ways, as described in Section 2.6.4, in which credit risk can be estimated using credit risk models. Credit scores and ratings are mainly used for risk assessments of a single financial transaction while credit risk models are used to estimate loss and other components of credit risk on a portfolio of credits.

2.6.4 Credit Risk Models

Credit risk is normally measured through a qualitative assessment or using models that are a simplified and hence imperfect representation of the real financial transactions and how the participants in the transactions carry them out (see Figure 2.7 and Figure 2.8). The latter (quantitative credit risk models) can be categorised into dynamic and static credit risk models. Dynamic models determine loss distributions in continuous time, for example, in the analysis of derivatives whose payoff depends on the default occurring at a specific time. On the other hand, static models focus on the loss distribution within a fixed period, say, one year[6]. Both these two (2) categories of models can be classified further into structural/company value models and reduced-form models as shown in Figure 2.7.

[6] Time horizon typical of Basel regulatory requirements.

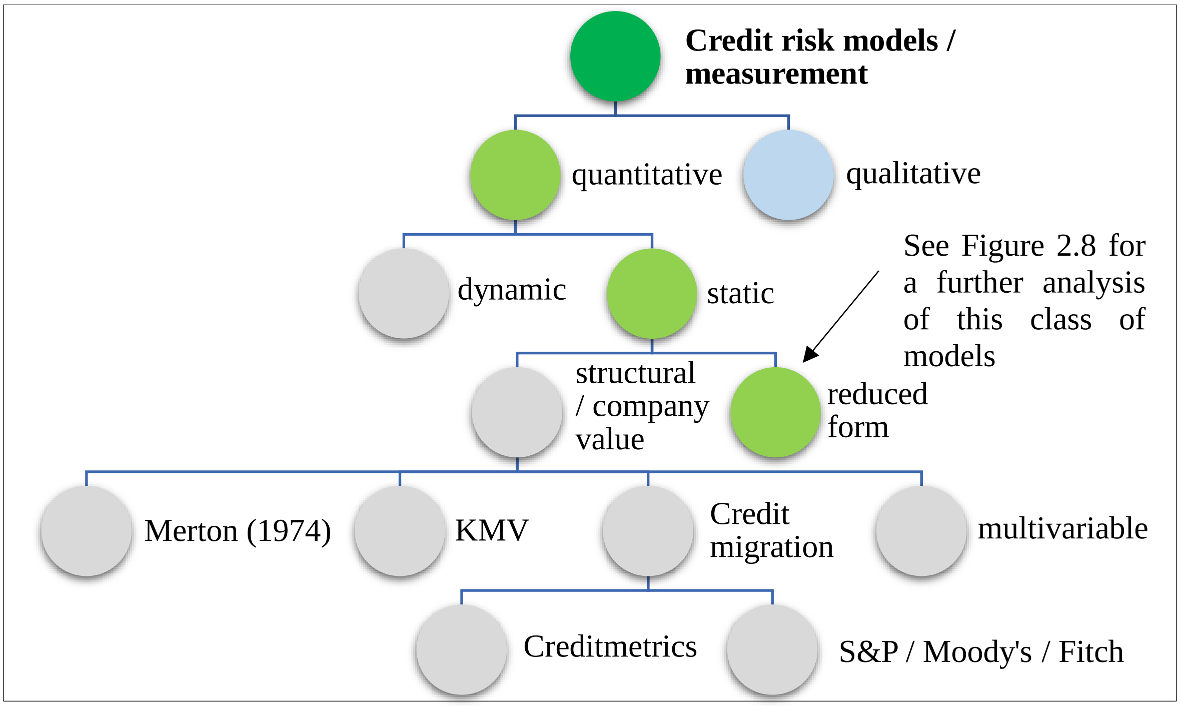


Figure 2.7 Credit Risk Models

(Source: Own preparation based on McNeil, Frey and Embrechts (2005, pp.327-381); Allen and Powell (2011); Hao, Alam and Carling (2010); Klieštik and Cúg (2015)).

In Figure 2.7, green circles depict model types (or similar) of focus in this research.

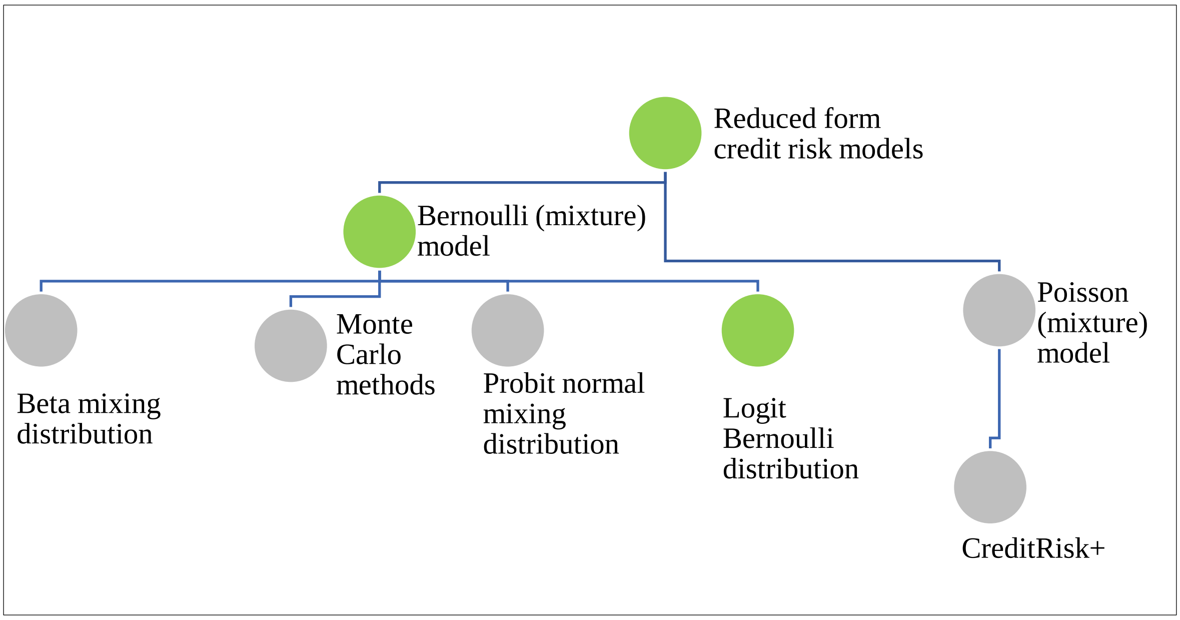


Figure 2.8 Reduced-Form Credit Risk Models

(Source: Own preparation based on McNeil, Frey and Embrechts (2005, pp.327-381); Allen and Powell (2011); Hao, Alam and Carling (2010); Klieštik and Cúg (2015)).

As shown in Figure 2.8, models similar to Logit models (Peng, Lee and Ingersoll, 2002) are the focus of this research. Structural/company value models assume a price mechanism leading to default; for example, Merton (1974) assumes the existence of a relationship between the default of a company and the values of its assets and liabilities at the end of a given period (Hao, Alam and Carling, 2010). Merton (1974), KMV, Credit migration (S&P/Moody's/Fitch) and Multivariable models are all examples of structural/company value models (Black and Scholes, 1973; Merton, 1974; Allen and Powella, 2011; Klieštik and Cúg, 2015). In reduced-form models, there is no assumption of the cause of default. In Table 2.3, samples of credit risk models are described with an outline of their strengths and weaknesses.

Table 2.3: Samples of Credit Risk Models

S/N	Model Type	Description / Strengths / Weaknesses
1	Merton	The Merton model is a credit risk assessment model proposed by Robert C Merton in 1974. It models the company's equity as a call option on its assets. The model is the progenitor and prototype of all structural/company-value models. Its strength is in its ability to capture market fluctuations which vary with industry risk. While it is elegant mathematically, it may over/under-state depending on market volatility.
2	KMV	KMV is a structural approach-based credit risk model developed by a company named after its creators Kealhofer, McQuown and Vasicek. It calculates the expected default frequency (EDF) of a company using information contained in a company's stock price and balance sheet that is translated into implied risk of default. Its advantage over its progenitor, the Merton model is its capability to capture company value variations as well as macroeconomic changes. Its weaknesses are that it is sensitive to equity price movements and is applicable only to publicly traded companies.
3	CreditMetrics	CreditMetrics is a credit risk model developed by J P Morgan (now J P Morgan Chase) and the CreditMetrics Group in the late 1990s to measure credit risk. It measures value at risk (VaR) due to credit upgrades, downgrades, or defaults in a portfolio. It is based on external ratings which include detailed financial analysis. It does not capture market fluctuations, uses complex techniques and its accuracy is high at the beginning and lower with time.
4	Multivariable	These models are analogous to the Merton model as applied at the portfolio level rather than at a single company level. The logic, advantages and disadvantages are similar to those of Merton.
5	CreditRisk+	CreditRisk+ is a credit risk model developed by Credit Suisse First Boston in the late 1990s. It is an industry example of the reduced-form family of credit risk models and focuses on default (and not on credit rating migration), and the default and loss distribution. It uses quantitative methods used in the insurance industry to model default as a sudden event. Its advantage is that it

S/N	Model Type	Description / Strengths / Weaknesses
		can be used to model all types of credit risks – wholesale, retail, bonds, third-party et cetera.
6	S&P/Moody's/ Fitch	See Annexures B and C, showing a table of the three Credit Rating Agencies' credit rating scales and an example of the S & P credit risk migration table respectively. These models are based on detailed financial analysis, and industry factors are incorporated at the time of ratings. The ratings are readily and publicly available. They do not capture market fluctuations. They have high accuracy at the time of rating that becomes lower with time. Model errors in these models are blamed for the 2008 financial crisis (International Monetary Fund, 2010; Honohan, 2016).

(Source: Various sources. (J.P. Morgan, 1997; Smit, Swart and Niekerk, 2003; McNeil, Frey and Embrechts, 2005; Hao, Alam and Carling, 2010; Allen and Powella, 2011; Klieštik and Cúg, 2015; Lawson, 2015)).

These models have evolved over time beginning with the bond ratings by Moody's and Standard & Poor's in 1909 and 1922 respectively. In 1924, Fitch first published the AAA to D ratings system that was adapted by Moody's and Standard & Poor's. The first mathematical credit risk models were developed by Beaver in 1967. This was followed by the development of the first generation of structural form models with that of Merton in 1974 being their progenitor. The second generation of the structural form models emerged in 1993 with Kim, Ramaswamy and Sundaresan presenting a model with the capability of predicting credit losses for a default that might happen any time. The first reduced-form credit risk model was published in in 1997, by Jarrow, Lando and Turnbull. Subsequently some institutions (see Table 2.3) developed Value-at-risk credit risk models, the most known of which are Credit Metrics, KMV's Credit Portfolio Manager, Credit Risk+ and Credit Portfolio View (Adamko, Klieštik and Birtus, 2014).

2.6.5 IFRS 9 and Credit Risk Modelling

Following the 2007-2008 financial crisis, in order to improve the stability of the international financial system (Engelmann, 2021), the Financial Accounting Standards Board (FASB) and International Accounting Standards Board (IASB) that set accounting standards applicable to the US companies and companies in the rest of the world respectively, collaboratively developed a new standard for the modelling of expected credit losses (ECL). The new standard developed was referred to by IASB as International Financial Reporting Standard 9 (IFRS 9) (IASB, 2019). It was first published in 2014 and took effect in 2018 (Gaffney and McCann, 2017). The FASB equivalent was named Current Expected Credit Loss (CECL) (FASB, 2016).

The IFRS 9 accounting standard aims to provide decision-useful information on the recognition and measurement of assets and liabilities of financial instruments to investors and other users outside the company (Beerbaum and Ahmad, 2015; IASB, 2019). IFRS 9 replaced International Accounting Standard 39 (IAS 39). Under IAS 39, preparers of financial statements were only required to disclose actual impairments (recognised after an overdue period of 90 days) in the values of financial assets. ECL is a concept introduced in IFRS 9 to overcome the shortcomings of this practice that was partially blamed for the 2007-2008 financial crisis as being "too little too late" (Beerbaum and Ahmad, 2015; IASB, 2019). Under IFRS 9 framework, the ECL is modelled on the expected credit losses of a financial instrument at the time it is issued (and later at periodical intervals) using forward-looking information (Prorokowski, 2018). The only instruments subject to impairment under IFRS 9, are financial assets whose cash flows are made up of solely principal and interest and the business model is to hold these financial assets to benefit from contractual cash flows (measured at amortized cost) or a dual intention to benefit from contractual cash flows and the sale of such asset (measured at fair value through other comprehensive income). Impairment (ECL) does not apply to equity instruments and any financial assets measured at fair value through profit or loss. The stipulation that the implementation of ECL should satisfy both regulatory capital and accounting requirements meant that credit risk models would have to be updated to accommodate the change.

2.6.6 Basel I/II/III Capital Accords and Credit Risk Modelling

As introduced in Section 1.10.2, the Basel capital accords are a set of rules on the minimum regulatory capital that banks must set aside to protect depositors, bondholders and the financial system against losses arising from severe markets conditions (Gestel and Baesens, 2009; Beerbaum and Ahmad, 2015). It is to be noted that while a well-capitalised bank is less likely to fail, regulatory capital is not meant to protect the bank against financial failure. Besides regulatory capital, therefore, banks have to assess and set aside economic capital to insure against economic insolvency. Initially, the Basel capital accords were agreed upon amongst bank supervisors from the group of the 10 (G10) most developed countries. The committee has since evolved to include members from 28 countries/jurisdictions composed of central banks and bank supervisory authorities. The committee also has eight (8) observers from central banks (of three countries), multilateral financial institutions and other bodies. The first set of rules, Basel I, first released in 1988 and effective from 1992, comprised simple standardised risk weights (shown in Table 2.4) applicable to various financial assets of a bank

(Gestel and Baesens, 2009). Regulatory capital was calculated by multiplying 8% by the product of the risk weight and the exposure at default (in mathematical terms minimum Capital = 8% x risk weight x exposure). The product of risk weight and exposure is referred to as risk-weighted asset (RWA). If, for example, a bank is exposed on a mortgage of US$ 100 000 to a customer, the minimum regulatory capital that the bank would have to set aside to insure against the risk of the customer defaulting on the mortgage would be 8% x 50% x US$ 100 000 = US$ 4 000. Basel I applied only to credit risk. It was later updated to take other risks into account, be more sensitive to risks inherent in different asset classes and allow banks to use internally developed risk ratings. The Basel II guidelines that became effective from 2007/2008, enabled banks, depending on their size and capacity, to use one of three approaches to regulatory capital determination: standardised approach, internal-ratings based approach, and advanced internal-ratings based approach. The internal ratings are obtained based on PDs, LGDs, and EADs quantified using relevant credit risk models (Bluhm, Overbeck and Wagner, 2003). In Basel III, the committee improved the quality of regulatory capital and gave bank supervisors more leeway in determining regulatory capital levels for banks (BCBS, 2017; Engelmann, 2021).

Table 2.4: Risk weights for Basel I

Risk weight	Asset type
0%	Cash held
0%	Claims on OECD central governments (foreign currency)
0%	Claims on central governments (national currency)
20%	Cash to be received
20%	Claims on OECD banks
20%	Claims on non-OECD banks (<1 year)
20%	Claims on multilateral development banks
20%	Claims on foreign OECD public-sector entities
50%	Residential mortgage loans
100%	Claims on the private sector (firm debt, equity, . . .)
100%	Claims on non-OECD banks (≥1 year)

100%	Real estate
100%	Plant and equipment

(Source: Gestel and Baesens (2009, p.346)).

2.7 GAPS IDENTIFIED IN LITERATURE AND RESEARCH CONTRIBUTION

In the following sub-sections, two (2) main gaps that were identified in the reviewed literature are outlined. One is the low estimation accuracy of the current credit risk models, and the other is the limitation of the Efficient Market Hypothesis.

2.7.1 Credit Risk Model Estimation Accuracy

Credit risk modelling is an approximate representation of the real credit markets being modelled (Klieštik and Cúg, 2015) and therefore by definition is necessarily inaccurate. Many assumptions are made at every stage of model development (Bluhm, Overbeck and Wagner, 2003; Oliver and Hand, 2005). For example, an assumption is made on the kind of distributions (see Figure 2.8) that the credit default losses will take, whether Poisson, Bernoulli, normal or other distributions. These are approximations of actual distributions that credit default losses will take. One point of convergence in the literature reviewed is the accepted view that these models fall short on estimation accuracy (Bolton, 2009; Mester, 2015; Baesens, Rosch and Scheule, 2016). Any incremental improvement in the estimation accuracy of credit risk models would be a welcomed contribution.

2.7.2 Efficient Market Hypothesis (EMH)

Credit risk models are anchored on the Efficient Market Theory (or Efficient Market Hypothesis (EMH)) which holds that the price of a financial asset reflects all available information about its fundamental value (Ang, Goetzmann and Schaefer, 2010). Researchers generally agree that the EMH holds in equilibrium (stable markets) situations, while some researchers now hold the view that while EMH is a very useful theory, it does not hold for all markets and at all times (Ang, Goetzmann and Schaefer, 2010; Lowenstein, 2011; Lawson, 2015), especially in disequilibrium ones. Moreover, credit risks are fundamentally different from equity price risks. Equity returns are symmetric and can therefore be well approximated by normal or gaussian distributions (Soros, 2014). On the other hand, credit returns are skewed and fat-tailed (J.P. Morgan, 1997; Soros, 2014). Therefore, in this study, the researcher incorporated proxy independent variables that capture sentiment (behaviour of credit consumers in disequilibrium situations – times of extreme fear, exuberance, and other such situations). The proxy independent variables used in this research are described in Sections 4.2.2 and 4.3.1.

2.7.3 Research contribution

The contribution of this research was therefore to build, using data from the developed market of the United States of America, a consumer credit risk model that would incrementally improve the estimation accuracy of credit risk losses through developmental choices anchored on EMH and the inclusion of proxy variables to capture consumer sentiment in disequilibrium situations. The model was adapted for use in the emerging market of South Africa.

2.8 CHAPTER SUMMARY

This chapter was a summary of the literature that was reviewed and found to be relevant to the research on consumer credit risk modelling. It included theories on which the study is founded, credit in general, consumer credit in particular, growth and development of consumer credit, consumer credit risk and its measurement, gaps identified in literature and research contribution. In the next chapter, the researcher outlined the research methods, the conceptual frameworks as well as the research design that were used to carry out this research. Also included are ethical considerations, validity, and reliability of the research as well as its limitations and proposed remedies.

CHAPTER 3: RESEARCH METHODOLOGY

3.1 INTRODUCTION

The previous chapter was about the extensive review of literature relevant to this study. This included literature on relevant theories, credit in general, consumer credit in particular, growth and development of consumer credit, consumer credit risk and its measurement, gaps identified in the literature review and research contribution. This chapter is a detailed outline of the research methods and the conceptual frameworks that were used to carry out this research as well as the design – how they were arranged to answer the research questions and hence attain the objectives of the research. In the chapter, the researcher also addressed ethical considerations, validity, and reliability of the research as well as its limitations and proposed remedies. Credit risk modelling is difficult (Oliver and Hand, 2005; Klieštik and Cúg, 2015) and especially so in emerging economies (Apanga, Appiah and Arthur, 2016). Generally, emerging market (EM) data is sparse due to the illiquidity in emerging financial markets. Available data often cover relatively short periods, which may not include sufficient economic expansion and contraction cycles (Apanga, Appiah and Arthur, 2016).

3.2 RESEARCH METHODS

Measuring and managing credit risk is important for the stability of financial institutions and markets (Chopra and Bhilare, 2018). The need for good credit risk models has increased following the 2007–2008 financial crisis as one of the key contributing factors was the use of inaccurate models (Ishikawa, 2009; Lowenstein, 2011; Jarrow, 2012). Literature that was studied in this research indicates that credit risk models that have been developed so far have poor estimation accuracy levels (Bae and Kim, 2015; Mester, 2015; Baesens, Rosch and Scheule, 2016; Chopra and Bhilare, 2018). The purpose of this research, therefore, was to develop a consumer credit risk model, using data from the developed market of the United States of America, which would improve the estimation of credit losses and adapt it for the emerging market of South Africa. This research lineage is given in Figure 3.1. From the figure, it can be seen that research can be divided into two (2) main types, basic research, and applied research. Basic research is concerned with solving a theoretical problem that need not be of immediate application, while applied research seeks to address a practical problem. This thesis is applied research as it aims to estimate consumer credit risk, a practical problem that financial institutions and in turn the economy, face (Baesens, Rosch and Scheule, 2016; Honohan, 2016; Chopra and Bhilare, 2018).

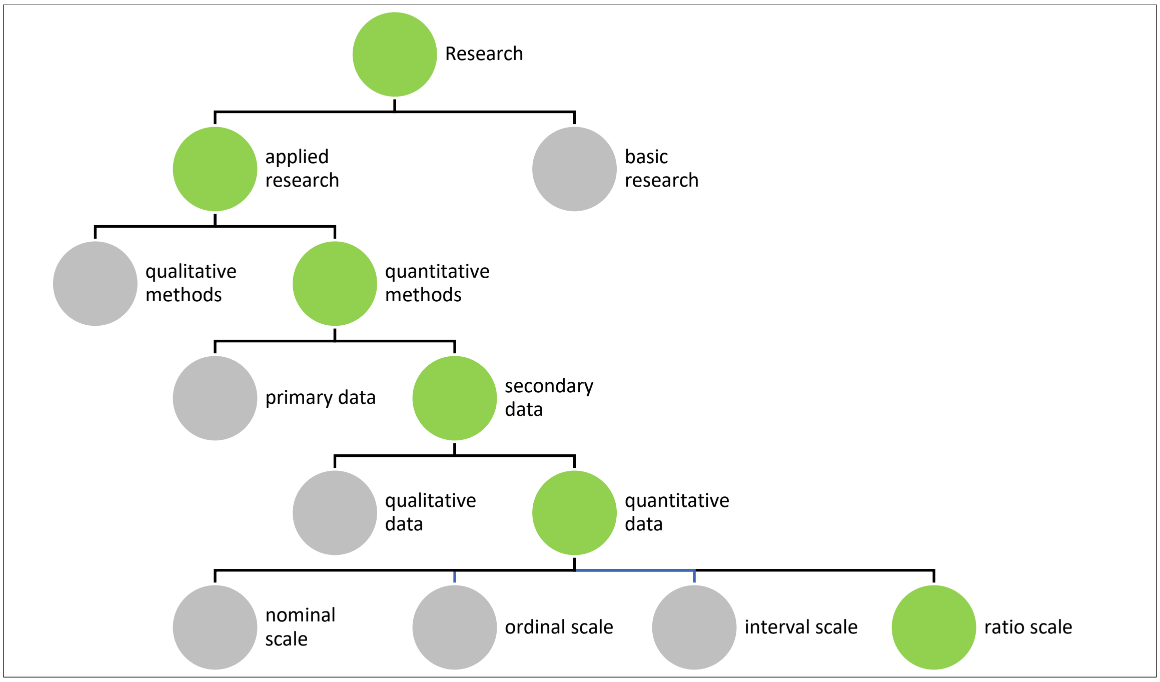


Figure 3.1 Research Methods

(Source: Own preparation based on Walliman (2011, pp. 5-28)).

The green circles in Figure 3.1 denote the methods followed in this research. Qualitative, quantitative, or a combination of both methods (called mixed methods) can be used to carry out research. This research, however, primarily used quantitative methods with secondary data of a quantitative nature and the ratio mode of measurement. The quantitative research approach is a statistical manner derived from science and mathematics that covers vast scientific activities (Leedy and Ormrod, 2015). There are various research designs, and the choice of which one to use is dependent on the research questions and objectives (Walliman, 2011). The credit risk modelling of this research was based on relationships between explanatory variables and realised credit losses (see Annexures D and E). Therefore, for this research, comparative, correlation, and simulation research designs were used. A comparative design helps the researcher to compare past and present or different parallel events to identify analogous situations (Allen and Powella, 2011) so that predictions can be made about similar ones recurring in the future. A correlation research design is used to examine two (2) concepts for association or causal relationships; in causal relationships, the cause is referred to as the independent variable while the affected is the dependent variable (Bolton, 2009). On the other hand, simulation design involves the creation of a representative small and simplified form (model) of a situation that can be manipulated to determine possible outcomes (what-if scenarios) (Bluhm, Overbeck and Wagner, 2003; Allen and Powella, 2011; Chopra and Bhilare, 2018).

3.3 CONCEPTUAL FRAMEWORK

In Section 1.11, the Conceptual Framework of this study was introduced together with its diagrammatic representation of Figure 1.6. The series of actions (that are captured in the framework) (Adom, Hussein and Agyem, 2018) which the researcher carried out in the study were: (1) Selection of explanatory variables (2) Data preparation (3) Data Analyses (4) Construction of proof-of-concept credit risk model (5) Back testing of the proof-of-concept credit risk model (6) Repetition of the process for the construction of the emerging market consumer credit risk model, and (7) Back testing of the emerging market consumer credit risk model. These actions are next elaborated on in more detail.

3.3.1 Selection of Explanatory Variables

The aim of this research was to build a developed market credit risk model and adapt it for the understanding and estimation of consumer credit losses in the emerging market of South Africa. The model was constructed by regressing realised credit losses on selected explanatory variables (simulation analysis method). According to Gestel and Baesens (2009), credit losses are affected by market perception and sentiment and the quality of the explanatory variables used as inputs in a model are the most important drivers of model performance. In this research explanatory variables that capture or represent sentiment (defined in this research as the behaviour of credit market participants in far-from-equilibrium situations such as economic recessions and financial booms and busts), the macroeconomy and obligor characteristics were carefully selected and coupled together in a multivariable regression analysis with consumer credit losses. The losses were in the form of charge off rates for the developed market of the United States of America (US) and impairments for the emerging market of South Africa. The novel idea of regressing credit losses on proxy explanatory variables for sentiment was postulated to improve estimation of credit losses. Existing models have variously used economic and obligor variables (Duffie, Saita and Wang, 2007; Baesens, Rosch and Scheule, 2016; Jin *et al.*, 2021). Additionally, the researcher used realised credit losses in the study models. To the best knowledge of the researcher, no specific attempt has been made in the literature to couple proxy variables for sentiment with economic and obligor characteristics, as depicted in Figure 3.2, in consumer credit risk modelling.

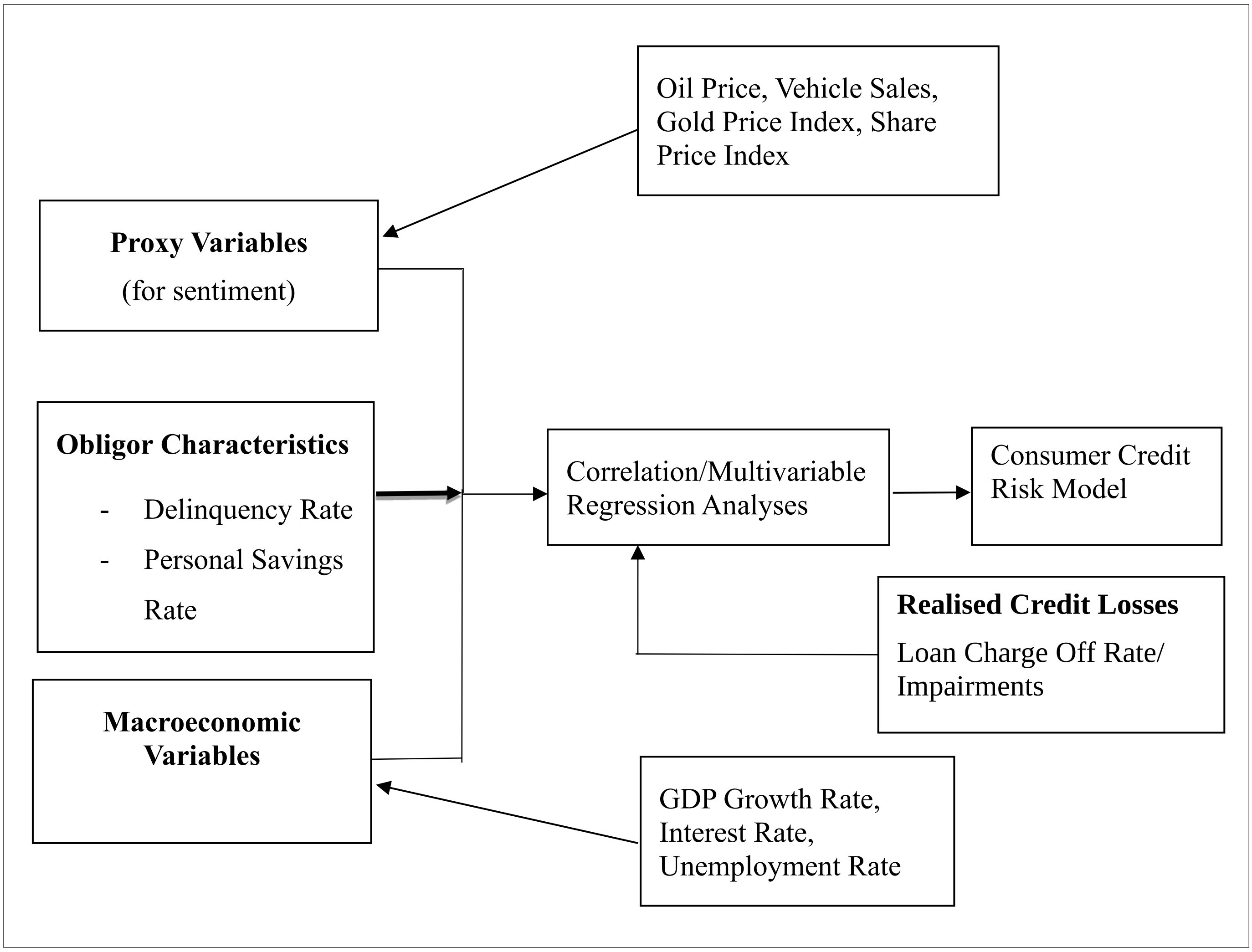


Figure 3.2: Diagrammatic Representation of the Coupling of Proxy Variables for Sentiment with Economic Variables and Obligor Characteristics
(Source: Own preparation based on Soros (2012) and Baesens (2015)).

Figure 3.2 is a diagrammatic representation of the coupling of proxy variables for sentiment with macroeconomic variables and obligor characteristics. Explanatory variables used in this study are described in Sections 4.2.2 and 4.3.1. Loan Charge Off Rate and Impairments served as proxies for consumer credit losses, the dependent variable in analyses of the developed market of the US and the emerging market of South Africa, respectively.

The three categories (sentiment, macroeconomic and obligor characteristics) of explanatory variables used in this research were selected based on relationships between them and the credit losses. The relationships were established through comparative and correlation analyses. Scatter diagrams, bivariable regression modelling (credit losses were regressed on each variable individually), and line graphs were used in the comparative analyses. The line graphs were obtained using an inbuilt analysis software available along with data on the Federal Reserve Bank of St. Louis, US, and the Reserve Bank of South Africa databases (SARB, 2020; FED, 2022). The correlation functions in the Microsoft Excel and Matrix Laboratory (MATLAB) software were used for the correlation analyses. From observation and studies of the resulting statistics from these analyses, explanatory variables that were representative or good proxies for sentiment, macroeconomic and obligor characteristics were then selected.

Some of the statistics under observation included R-squared values, p-values, and t-statistics. The selection of macro-economic variables was relatively straightforward – traditional indicators of economic performance like the Inflation Rate, Treasury Bill Yield, Consumer Price Index, Gross Domestic Product (GDP), and Unemployment Rate are known and their data readily available (DiGeorgia, 2001; Jareño and Negrut, 2016). However, the selection of variables to capture obligor characteristics and sentiment or the views of credit market participants, especially in far-from-equilibrium situations was not straightforward. Data for standard obligor characteristics like the occupation of the obligor, years at the present job, number of derogatory reports, and purpose of the loan (Baesens, Rosch and Scheule, 2016) were not readily available; what the researcher was able to access for this variable category was aggregate data for all credit consumers for the US and South African banks like Personal Saving Rate, Delinquency Rate and Personal Consumption Rate.

The identification and selection of variables that capture sentiment were based on the researcher's review of literature about variables that may reflect the reflexive phenomena (Soros, 1992; Fons Wijnhoven and Plant, 2019). The US auto sales, one of the variables selected as a proxy for sentiment, for example, are known to have fallen by nearly 40% during the financial crisis of 2007-2008. The spike in oil prices before the crisis and the fall in home prices could explain less than 40% of that variation. The rest could possibly be explained by a pessimistic outlook of the economic future by market participants (Dupor, 2019; Fons Wijnhoven and Plant, 2019; Dupor *et al.*, 2020). Similar analyses went into the choice of Oil Price and Gold Price as proxies for sentiment; the two variables are known to be volatile to different degrees during economic recessions (Kruger, Joseph and Aphane, 2012).

3.3.2 Data Preparation

A sound and acceptable consumer credit risk model should be universally applicable with adaptation within credit markets in all jurisdictions (Baesens, 2015; Baesens, Rosch and Scheule, 2016). Building a credit risk model based on emerging market data is challenging as that data is sparse (Apanga, Appiah and Arthur, 2016). Data, for example, from the National Credit Regulator of South Africa (National Credit Regulator, 2022) covers only a relatively short period, as the NCR was established in 2005 (Republic of South Africa, 2006). On the other hand, credit risk is a phenomenon that is extensively published on developed country data (Duffie, Saita and Wang, 2007; Jin *et al.*, 2021). This research, therefore, used two (2) sets of economic and other data series: one set, composed of 18 variables (Loan Charge Off Rate and 17 explanatory variables), from the developed market of the United States of America (US) and another set of 21 variables (Impairments and 20 explanatory variables), from the emerging market of South Africa. All the 18 developed market data series and 20 of the 21 emerging market data series are freely available from the Reserve Bank of St. Louis, US (FED, 2022) and are permitted for personal, educational and non-commercial purposes.

The Reserve Bank of St. Louis has over 816,000 data series (organised for ease of use in research) from over 100 sources of the US and other countries' economic and other data series in its data bank (Federal Reserve Economic Data or FRED). Over 1500 of the data series are for South Africa. The 20 South Africa economic and other data series in the custody of the FRED and used in this research were sourced mainly from the Organisation for Economic Cooperation and Development (OECD) (Organisation for Economic Cooperation and Development, 2022). The rest came from the World Bank (World Bank, 2022), and the International Monetary Fund (IMF) (International Monetary Fund, 2022). The one data series partially but not fully available from the Reserve Bank of St. Louis, US – of impairments on loans – was partly obtained from the South African Reserve Bank (SARB, 2020). The data in the custody of the Reserve Bank of St. Louis had been organised in daily, weekly, monthly, quarterly, and annual frequencies.

For the US part of the study (development of proof-of-concept consumer credit risk model), the researcher used the data with a quarterly frequency since the credit losses (Loan Charge Off Rates, the dependent variable) were only available on a quarterly frequency. However, all variable data series available covered long periods, the longest of which was 74 years (1948–2021), with the shortest being 34 years (1987-2021). Therefore, US proof-of-concept credit risk model was constructed from data from 137 quarters (Q1 1987–Q1 2021) since regression analysis can only be done with variable data covering equal periods. As already pointed out in this study, emerging market data is sparse (Apanga, Appiah and Arthur, 2016). The earliest period for which South African data for all variables used to build the emerging markets consumer credit risk model were available is 2008-2021, covering a total of 53 quarters (Q1 2008-Q1 2021) or 159 months (January 2008-March 2021). For this reason, the researcher used data with a monthly frequency. As the dependent variable (impairments as a proxy for credit losses) data was only available on a quarterly frequency, they were converted to monthly data by assuming that the variable values in a month were equivalent to those of the corresponding quarter.

The data for both markets were standardised to Z-score values for ease of comparative analysis. On each of the two data sets, one series was or representative of the realised credit losses, while the others were independent variables that explain variations in the losses. The independent variables were classified into the three (3) categories – sentiment, macro-economic and obligor. Sentiment variables capture the tone of the credit markets – how credit market consumers react to events in disequilibrium situations (recessions, financial booms and busts) that they observe (Kruger, Joseph and Aphane, 2012; Soros, 2014; Marks, 2022); macro-economic variables explain the impact of the macro-economy on credit losses (DiGeorgia, 2001; Jareño and Negrut, 2016), and obligor variables define the characteristics of the borrowers (Baesens, Rosch and Scheule, 2016).

It is worth noting that the FRED assures its data quality and integrity through a number of ways: first, it ensures that the process of collecting, analysing, storing and distributing the data is ethical and that this process is documented so that the work is reproducible. Secondly, it facilitates the continuous revision and updating of the data to ensure the elimination of errors and timeliness. Thirdly, all FRED data have citations describing its source so that the work can be replicated and checked for accuracy. And lastly, the FRED deploys enormous time and resources to the management of its data (Mendez-Carbajo, 2013b, 2013a).

3.3.3 Data Analyses

The classical statistical (Bae and Kim, 2015) way of measuring credit risk is through the determination of values of its components using component models (see Table 3.1). Components of credit risk are Probability of Loan Default (*PD*), Loss Given Default (*LGD*), and Exposure at Default (EAD) at a certain maturity. The relationship between these components and the losses (Expected Loss (*EL*) and Unexpected Loss (*UL*)) are given by the following formulae: $EL = PD \text{ x } LGD \text{ x } EAD$ and $UL = EAD \text{ x } [PD^2 \text{x } \sigma^2_{LGD} + LGD^2 \text{ x } \sigma^2_{PD}]^{1/2}$ (σ^2_{LGD} *and* σ^2_{PD} *are variances of loss given default and probability of default respectively)* (Harper, 2008; Joseph, 2018).

Table 3.1: Probability of Default and Loss Given Default Model Performance

No.	Context	Number of characteristics (independent variables)	Performance	Performance measure
1	PD application credit scoring	10 - 15	70 – 85 %	area under ROC curve
2	PD behavioural credit scoring	10 - 15	80 – 90 %	area under ROC curve
3	LGD	6 - 8	20 – 30 %	R-squared (Barth et al. 2018; Jin et al. 2021)

(Source: Baesens, Roesch & Scheule (2016)).

Data in Table 3.1 show that the amount of variation in the Loss Given Default (LGD) that is explained by the independent variables – as shown by R-squared of between 20% and 30% – is low. From the table, it is observed that the PD models can separate the positive from the negative class with an accuracy generally between 70% and 90% for both PD application and behavioural scoring, while variables used in the LGD models can explain only 20% to 30% of the variation in LGD (Baesens, Rosch and Scheule, 2016). Knowing that expected loss EL = PD * LGD * EAD, (Bluhm, Overbeck and Wagner, 2003) means that the credit loss estimation accuracy may be lower than 20%. This level of uncertainty and inaccuracy means that using these credit risk models in financial institutions poses model risk – the risk that the model estimations are wrong in themselves (Baesens, Rosch and Scheule, 2016; Honohan, 2016). However, this study used realised credit losses or their proxies (Loan Charge Off Rate for all consumer banks in the US market and impairments for all banks in the emerging market of South Africa) to construct a proof-of-concept consumer credit risk model for the US market and an emerging market of South Africa consumer credit risk model. The novelty of this approach is that realised credit losses incorporate PD, LGD and EAD (Jin *et al.*, 2021) eliminating the need to construct separate models for them and the associated assumption errors. It was expected that the model building approach would increase the accuracy of estimating credit losses and contribute an incremental improvement to credit risk measurement and management, especially in South Africa and other emerging markets.

3.3.3.1 Analysis Software and Techniques

The Reserve Bank of St. Louis, US, data bank has an inbuilt analysis software tool that was used to do a comparative analysis of the data in graphical representation (FED, 2022). The South Africa Reserve Bank has a similar albeit less comprehensive tool (SARB, 2020). Besides these tools, Excel software and IBM SPSS Statistics (Statistical Package for the Social Sciences) software were used to analyse the data. While, as the name suggests, the SPSS software was originally designed for the social sciences market, it has been adapted over time to be used in other scientific fields (Frey, 2017). Excel and SPSS were used primarily as the calculation engines – it is relatively straightforward to do comparative, correlation, and simulation (using regression) analyses with these tools. The research also used Matrix Laboratory (MATLAB) to make big data inference and factor lead and lag covariance analyses that gave more qualitative insights.

The research, therefore, has a strong bias towards quantitative econometric methods in credit risk. In the sample model analysis of Table 2.3, it was observed that the credit rating agencies' (S&P/Moody's/Fitch) credit rating and credit rating migration models have a high accuracy at the time of rating that becomes lower with the passage of time. These models are based on historical data that cover long periods, which include many of both economic expansions and contractions (International Monetary Fund, 2010). They also inherently incorporate all variables – individual obligor characteristics, micro-economic and macro-economic variables. Their disadvantage lies mainly in their static nature. This research borrowed from this insight to model relationships between realised credit losses and the universe (obligor, micro-economic, and macro-economic) of independent variables that specifically include variables that capture market fluctuations. This aligns with the research notion that the inclusion of proxy explanatory variables for sentiment with macroeconomic explanatory variables, and obligor characteristics, improves the estimation accuracy of a consumer credit risk model. Relationships between the credit losses and the explanatory variables were established using comparative, correlation, bivariable, and multivariable regression analyses.

Comparative analysis was done using the inbuilt data analysis tools of the FED and SARB. The tools generate graphical representation of the variables that were studied to gain an overview of the relationships between the credit losses and the explanatory variables. Excel software was also used to create scatter diagrams and conduct correlation and bivariable regression analysis of credit losses and each explanatory variable. This enabled the researcher to eliminate explanatory variables with the weakest relationships with credit losses. Those with strong links with credit losses were then put through an ANOVA and multivariable regression analysis. Credit risk is generally modelled in a regression setting (Duffie, Saita and Wang, 2007; Bolton, 2009; Barth *et al.*, 2018; Jin *et al.*, 2021). Regression modelling is a statistical method that is used to examine how one or more independent (explanatory) variables explain the dependent (predicted or target) variable (Bolton, 2009). Annexure D show scatter plots or diagrams of the correlations between Loan Charge Off rate and the selected 17 US explanatory variables. The technique of obtaining the mathematical equation (model) that best fits the data is called regression analysis (Mendenhall and Sincich, 2020). This is represented by the line of best fit on the scatter plots. Some data are on the line while others are outside (below or above) it. The absolute differences between the observed and the predicted values are referred to as deviations or errors of prediction. The one line for which the sum of the squares of these errors is least (or minimum) is called the least squares line, regression line or least squares prediction equation or model (Mendenhall and Sincich, 2020). The general form of this probabilistic model in regression is:

$$y = E(y) + \varepsilon,$$

Where y= dependent variable,

$E(y)$= mean (or expected value) of y,

ε = Unexplained or random error.

Multiple regression models are probabilistic models that include more than one independent variable, and they take the general form:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \varepsilon,$$

Where y = dependent variable,

$x_1, x_2, \ldots x_k$ are the independent variables,

$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k$ is the deterministic portion of the model

ε = Unexplained or random error.

Multiple regression models, like simple regression models, are fitted using the least square method. This is computationally more complex. This is therefore mainly done using regression analysis software. Testing whether the model is adequate for predicting y, the dependent variable is done using analysis of variance or ANOVA (Mendenhall and Sincich, 2020). Regression and ANOVA analysis results in several regression statistics, amongst which are the coefficient, the confidence interval of the coefficient, the standard error, the p-value and the R-squared (or R^2) value. The coefficient indicates whether there is a positive (+ve) or negative (-ve) correlation between the independent and the dependent variable and the rate at which the latter change with every unit change in the former. A+ve sign shows that as the independent variable increases, the dependent variable increases as well. Adjusted R-squared (or R_a^2) value is R^2 value adjusted for the sample size and the number of β parameters. The confidence level of the coefficient – 95% in this study – gives the range within which the real value of the coefficient being estimated falls in. The standard error is the absolute measure of the typical distance that the data points fall from the regression line. The t-statistic is the coefficient divided by its standard error, and the p-value tests the null hypothesis that the coefficient is equal to zero (no effect). A low p-value (< 0.05) indicates that the null hypothesis can be rejected. The R-squared is the relative measure of the percentage of the dependent variable variance that the model explains. Using correlations, R-square values, p-values and/or t-statistic values, the explanatory variables were reduced to the most significant ones (see Figure 3.3).

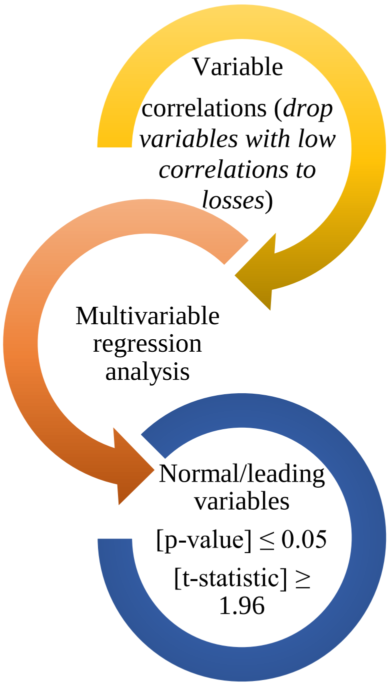


Figure 3.3: The Process Used to Determine the Significant Explanatory Variables.

(Source: Own preparation based on Bolton (2009)).

In this way, from a universe of 17 and 20 explanatory variables for the credit markets of the United States of America and South Africa respectively, the significant variables were determined.

3.3.4 Creation of Proof-of-Concept Credit Risk Model

Using variable data from the representative developed market of the US, a proof-of-concept consumer credit risk model was created. This was done by regression analysis of credit losses on the explanatory variables for as many times as was necessary to obtain a combination of significant explanatory variables. Significant explanatory variables at the chosen confidence level of 95% are those whose p-values were less than or equal to 0.05 (or with t-statistic of absolute values of greater or equal to 1.96). Once the best proof-of-concept consumer credit risk model had been created and back tested satisfactorily, the insights, concepts, techniques, and knowledge gained in the development were used to construct the emerging market of South Africa consumer credit risk model.

3.4 ETHICS – DATA COLLECTION AND MAINTENANCE

This research mainly used secondary data, and therefore ethical risks were low or negligible. Nevertheless, the Unisa guidelines on research ethics were strictly adhered to. This included, and was not limited to:

- Ensuring that all sources of information were acknowledged through citations and references.
- Obtaining Ethical Clearance from the Unisa Ethics Committee before the research analyses were undertaken (see Annexure F).
- Making sure that the data obtained was used only for the stated and duly approved purpose.
- Making sure that should the need arise to use data or methods that would change the ethical risk profile of the research, permission was sought from the Unisa Ethics Committee before such data or methods are used.
- Keeping Confidentiality.

3.5 RESEARCH VALIDITY AND RELIABILITY

Validity and reliability are two fundamental criteria of a measure. Validity refers to the accuracy of the measurement obtained while reliability refers to its repeatability (Mohajan, 2017). As this research used secondary data, the validity and reliability of the results were mainly dependent on the quality of data used in the creation of the model and any assumptions made about them. To ensure that the results are robust, and the model is applicable through the economic cycles, the quality aspects of the data were considered. The quality aspects referred to include (amongst many other measures) accuracy, objectivity, completeness, traceability as well as whether the data is representative, covering periods of economic expansion and contraction. The model data and assumptions were checked for the quality, validity, reliability and robustness aspects, and the model must be monitored and updated to account for any changes in the consumer credit environment. To enhance the validity and reliability of the research, data from reliable sources outlined in Section 3.3.2 were used. Finally, triangulation, back testing and benchmarking were also used to test and confirm from a different perspective the model validity and reliability.

3.6 RESEARCH LIMITATIONS AND PROPOSED REMEDIES

This research faced the limitation of the availability of quality data required to develop the consumer credit risk model. This was especially the case in South Africa, where data on aspects of credit risk of interest was not readily available and when available, do not cover a long enough period. In this research, secondary data that was not specifically collected for the purpose of credit risk measurement was also used (see Annexure G for recommended independent variables that can be used to build a consumer credit risk model assuming a mortgage). To overcome these challenges, the data used to develop, validate and back test the credit risk models, came from, amongst other prominent sources, two (2) secondary sources – the FRED, Federal Reserve Bank of St. Louis, US, and the South African Reserve Bank (SARB) (SARB, 2020; FED, 2022). Data from the FRED is of high quality, and easy to examine as it comes with an inbuilt analysis tool and covers relatively long periods.

3.7 CHAPTER SUMMARY

In this chapter, the researcher explained the methodology used to carry out this research. This included research methods and conceptual frameworks. Ethical considerations, validity and reliability of the research, research limitations and the proposed remedies were also discussed in the chapter. The next chapter outlines the analyses and results of the research.

CHAPTER 4: ANALYSIS AND RESULTS

4.1 INTRODUCTION

In the previous chapter, the researcher outlined the methods and the conceptual frameworks that were used in the research. Also included were discussions about ethical considerations, the validity and reliability of the research, and research limitations and the proposed remedies. This chapter is a record of the analyses of the data collected and the results of the research. The aim of this research is outlined in detail in Section 1.4. In summary, the purpose of the research was to build a developed market credit risk model that can improve the estimation of consumer credit losses and adapt it for the developing market of South Africa. The research sought a better credit risk modelling approach and key independent variables that would improve the accuracy associated with consumer credit risk measurement. This research's hypothesis was that including proxy variables for sentiment (the feeling or tone of a credit market) and macro-economic variables affecting consumer borrower characteristics would improve the ability of a credit risk model to predict consumer credit losses. The losses of interest are losses arising from consumers defaulting on their loan obligations (see examples of consumer credit in Figure 2.2) .

The chapter is divided into two (2) main sections: Section 4.2 is a record of the analyses of data from the developed market of the United States of America, and the credit risk models built using the data (named Model US1, US2, US3, US4 and US5 according to the sequence in which they were developed and for ease of reference); Section 4.3 outlines the adaptation of the US models to the emerging market of South Africa. Section 4.2 is sub-divided further into seven (7) Sub-sections: 4.2.1: Quality of Data and its Sources; 4.2.2: Description of US Data Variables; 4.2.3: Correlation Analyses of the Dependent and Independent Variables; 4.2.4: Standardisation of Independent Variables; 4.2.5: Bivariable and Multivariable Regression and ANOVA Analyses; 4.2.6: Back Testing Results for Model US1; 4.2.7: Back Testing Results for Model US2; and 4.2.8: Developed Market Consumer Credit Risk Models. In Section 4.3, the researcher focused on using the developed market of the US modelling approach in the analysis of the emerging market of South African data to develop Consumer Credit Risk Model SA2 discussed in Section 4.3.3. The chapter summary is described in Section 4.4.

4.2 DEVELOPED MARKET ANALYSES

In this section the US explanatory variables used in the research and the data collected on them are described. The analysis of the correlations between the dependent variable and the independent variables, standardisation of data, bivariable regression and ANOVA, and multivariable regression, are discussed next. Figure 4.1 shows the analyses blocks used to reduce the explanatory variables by elimination and build the most plausible credit risk model. The sizes of the blocks reflect the proportion of the number of variables used in each block.

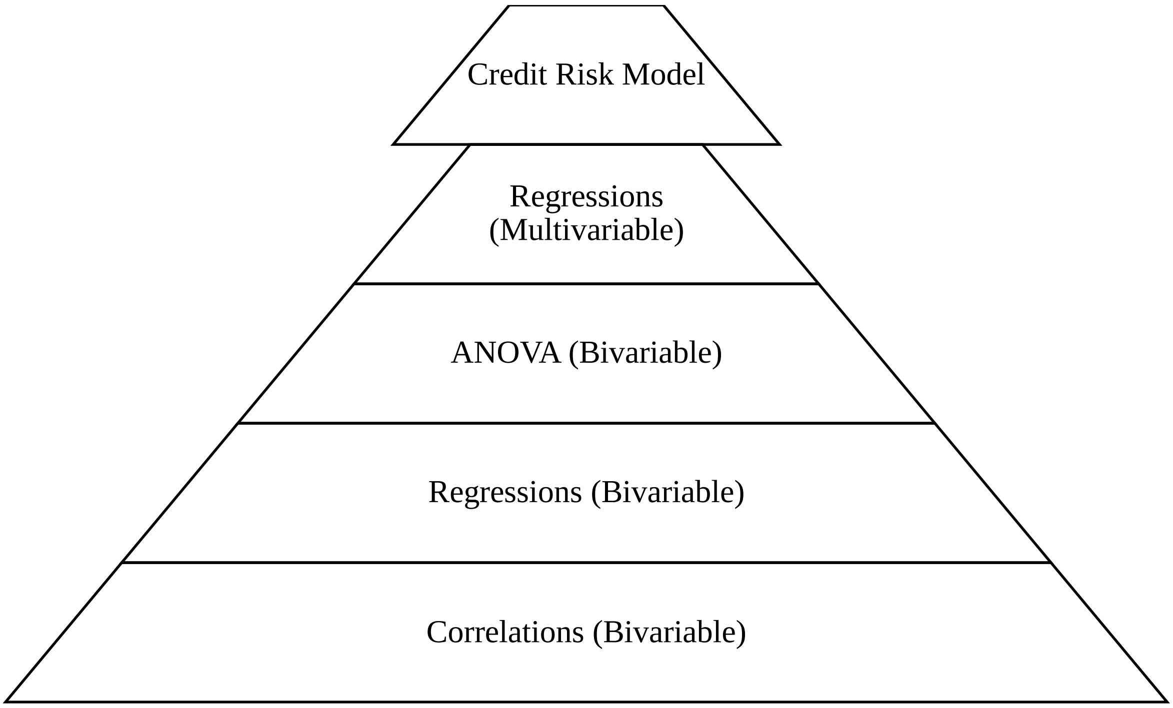


Figure 4.1: Model Building Analyses Blocks

The results of back testing the proof-of-concept consumer credit risk models US1 and US2 are also discussed, along with three additional models US3, US4 and US5.

4.2.1 Quality of Data and its Sources

The consumer credit risk research was done quantitively on big data meaning that the number of data elements considered was vast, surpassing 10,000 data points (see Section 3.3.2). The data used in the US market credit risk model were readily and freely available from the data bank of one of the 12 branches of the US federal bank system – the Reserve Bank of St. Louis. The data has been collected and organised into daily, monthly, quarterly, and annual frequencies. The quarterly data was used in the analysis as it covered a long period, 1987-2021.

The data bank also has an inbuilt analysis software. Data for the emerging market of South Africa was not easily available. Only one of the 21 data series ranging from 2008 to 2021 came directly from a South African source. Even then, one month's worth of data had to be extracted at a time from the Reserve Bank of South Africa data bank that is available on its website (SARB, 2020). The rest of the data came from the Reserve bank of St. Louis, US (FED, 2022), having been collected from OECD, the World Bank and IMF. It is possible that the data was originally collected by South African institutions. Proxy variables (such as Impairments) were used as loan loss ratio (consumer credit risk) in the emerging market of South African credit risk models to overcome the challenge of lack of ideal variables. Additionally, some of the data series were available only in quarterly frequencies, while the analyses were done on monthly data series. In those cases – Unemployment Rate, Real GDP, Residential Property Price Index, World Uncertainty Index (a proxy for volatility) and Constant Price GDP – the quarterly data were assumed to be constant throughout the months within each quarter.

To the best of the researcher's knowledge, no emerging market of South Africa consumer credit risk model has been developed using the methods and techniques used in this research (at the level of estimation accuracy that the models described in this research have demonstrated).

4.2.2 Description of US Data Variables

In this research, for the US data part of the study, the Charge Off Rates were used to represent the credit losses (the dependent variable). The Loan Charge-Off (LCO) is the amount of loan written off by a bank in one season (say a month, quarter, or year) less any recoveries of loans written off in the preceding season. The Loan Charge Off is therefore realised losses and represent the bank's realised credit risk (Jin *et al.*, 2021). For all practical purposes, the Loan Charge Off Rate was therefore adopted in this study as a measure of the credit risk for the US banks.

The 17 consumer credit risk independent (explanatory) variables listed and described in Table 4.1 were selected[7] from over 1500 South Africa economic and other data available at the Federal Reserve Economic Data (FRED) bank of the Reserve Bank of St. Louis, US (FED, 2022). They were selected on the basis of their fit into the three (3) categories that were outlined in Section 3.3.1. In principle, they are similar to variables used by Barth et al., (2018) to forecast Net-Charge Off Rates for two (2) small and two (2) large banks. The three (3) variable types are: sentiment, macro-economic and obligor explanatory variables (see Table 4.1). Sentiment is defined in this study as credit market participants' behaviour (or tone of the market) in disequilibrium situations (DiGeorgia, 2001; Soros, 2014; Marks, 2022). Proxy variables, that capture credit consumer sentiment, used in this study are the Auto and Light Truck Sales (VEH), Gold Price (WTI) (GOLD), Oil Price (OIL) and S&P Case-Shiller Home Price Index (S&P CSPI). During a recession, volatility is usually high as economic and financial markets are uncertain. Commodity prices (in particular the Oil Price) is a leading indicator of a recessionary environment as it generally spikes just before volatility increases (Venditti and Veronese, 2021). The Gold Price and the Oil Price are generally correlated (Kruger, Joseph and Aphane, 2012). To proxy the volatility or gauge the (recessionary or normal) tone of the credit market, the Gold and the Oil Prices were chosen.

Vehicle Sales are generally a barometer for the health of the economy (Dupor *et al.*, 2020). Home sales are like vehicle sales, a leading indicator of economic growth. If consumers have a pessimistic view of their economic future, they hold back purchases of vehicles and homes and if their view is optimistic, they do the opposite (DiGeorgia, 2001; Jareño and Negrut, 2016). For this reason, the Auto and Light Truck Sales, and S&P Case Shiller Home Price Index are chosen as viable proxies for the sentiment associated with economic growth. Macro-economic variables measure the state of the economy (DiGeorgia, 2001; Jareño and Negrut, 2016). Macro-economic variables used in this study are 10Y (expected) Inflation Rate (INFL), 10Y Treasury Constant Maturity Rate (T10Y), 10Y-2Y Treasury Constant Maturity Rate (10Y-

---

[7] The selection process was semi-automated; it entailed using the FRED inbuilt software to sort the data. First, the system was tasked with listing all the econometric time series specifically on South Africa. And then the list was screened to identify those that could be classified as economic, obligor or sentiment (as defined in this research) variables. Those selected in this way were next tested for accuracy, correlations significance, completeness and expected characteristics of each of the categories using the inbuilt software.

2Y), 2Y Treasury Constant Maturity Rate (T2Y), Consumer Price Index (CPI), Gross Domestic Product (GDP), and Unemployment Rate (UNRATE) (Barth *et al.*, 2018; Jin *et al.*, 2021). Obligor variables are variables developed and used by financial institutions' credit risk departments to estimate and monitor credit risks associated with their credit consumers (Baesens, Rosch and Scheule, 2016). Obligor variables or obligor proxy variables used in this study are Personal Consumption Expenditure (PCE), Personal Saving Rate (PSR), Household Debt Service Payment as a percentage of Disposable Personal Income (HD), Consumer Debt Service Payment as a percentage of Disposable Personal Income (CD), Delinquency Rate (DRATE) and Household and Non-Profit Companies' Consumer Credit Debt Liability Level (HNPO) (DiGeorgia, 2001; Baesens, Rosch and Scheule, 2016; Jareño and Negrut, 2016).

Table 4.1 Description of the Dependent and Independent Population Variables

Key	
Dependent variable (LCO)	(grey)
Sentiment explanatory variables	(yellow)
Macro-economic explanatory variables	(orange)
Obligor explanatory variables	(dark blue)

S/N	Dependent / Independent variable	Short name	Description
	Loan Charge Off Rate	LCO	Charge-Off Rate on Consumer Loans, All Commercial Banks (%)
1	Autos and L /Trucks	VEH	Autos and Light Trucks, Millions of Units
2	Gold Price	GOLD	GOLD /150, U.S. $ per Troy Ounce/150
3	Oil price (WTI)	OIL	Crude Oil Prices (WTI), Dollars per Barrel
4	S&P Case-Shiller home price index	S&P CSPI	S&P/Case-Shiller, Index Jan 2000=100
5	10Y Inflation rate	INFL	10-Year Breakeven Inflation Rate, (%)
6	10Y Treasury CM	T10Y	10-Year Treasury Constant Maturity Rate, (%)
7	10Y-2Y Treasury CM	10Y-2Y(Tilt)	10-Year Treasury Constant Maturity Minus 2-Year Treasury Constant Maturity, (%)
8	2Y Treasury CM	T2Y	2-Year Treasury Constant Maturity Rate, (%)
9	Consumer Price Index	CPI	Consumer Price Index: Total All Items for the United States, Growth Rate Previous Period
10	Gross Domestic Product	GDP	Gross Domestic Product, Billions of Dollars
11	UNRATE	UNRATE	Unemployment Rate, (%)
12	Personal cons. exp.	PCE	Personal Consumption Expenditures, Billions of Dollars
13	Personal saving rate	PSR	Personal Saving Rate, (%)
14	Household DSP / DPI	HD	Household Debt Service Payments as a Percent of Disposable Personal Income, (%)

15	Consumer DSP / DPI	CD	Consumer Debt Service Payments as a Percent of Disposable Personal Income, (%)
16	Delinquency rate	DRATE	Delinquency Rate on Consumer Loans, All Commercial Banks, (%)
17	HNPOCCLL	HNPO	Households and Non-profit Organizations; Consumer Credit; Liability, Level, Billions of Dollars

(Data from: FRED, Federal Reserve Bank of St. Louis, US).

Table 4.1 is a description of the dependent variable and the 17 independent (explanatory) variables that were initially selected for the building of the credit risk model – before the independent variable population was reduced to smaller samples of eight (8) and nine (9) for the consumer credit risk models US1 and US2 respectively. Loan Charge Off Rate represent the credit losses (dependent variable) that the independent variables explain.

4.2.3 Correlation Analyses of the Dependent and Independent Variables

Once the 17 explanatory variables were selected, a combination of a scatter diagram (comparative) and bivariable regression analyses of each variable with the Loan Charge Off Rate was done (see Annexure D). Four (4) explanatory variables were dropped on the basis of the results of the analyses. The four (4) variables were GOLD, S&P CSPI, INFL, and HNPO. GOLD was dropped since it had a low correlation with LCO as the R^2 value that measures how much it explains the losses was low at 3.78%; S&P CSPI was insignificant (p-value of 0,072) and explained only 2,36% of the variation in losses; INFL had a low R^2 value of only 0,06% and was marginally insignificant with a p-value of 0,508; HPNO had R^2 value that was marginally above 5%. Shown in Table 4.2 are the correlation matrix of the Loan Charge Off Rate (LCO) and the 13 independent variables that were not dropped.

As shown in the Charge Off Rate column (third column) of Table 4.2 and as expected, Delinquency Rate exhibits the strongest correlation at 54% with the Charge Off Rate. It is expected that borrowers who are delinquent – three (3) months or more behind on loan repayments – are the most likely to default leading to their loans being charged off (written off) by the bank. The next explanatory variable with a similarly high correlation with Charge Off Rate is Unemployment Rate at 50%. It is positively correlated with credit losses. That means as more people are laid off, loan losses increase as those who are dependent on income from their employment to meet their debt obligations may now find it difficult to do so. In terms of correlations among the explanatory variables, 10-year (T10Y) and 2-year (T2Y) Treasury

Constant Maturities are positively closely correlated with a correlation of 95%. This suggests that one of them could be used to explain the variation in loan losses (Charge Off Rate) instead of both. Similarly, GDP and Personal Consumption Expenditure (PCE) are 100% correlated.

Table 4.2 Correlation Matrix of the Loan Charge Off Rate and Independent Variables

	Serial number		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
Serial number	***Variable***	***COR***	***VEH***	***OIL***	***T10Y-2Y (Tilt)***	***T10Y***	***T2Y***	***CPI***	***GDP***	***UNRATE***	***PCE***	***PSR***	***HD***	***CD***	***DRATE***
	COR	100%													
1	VEH	-50%	100%												
2	OIL	39%	-7%	100%											
3	T10Y-2Y (Tilt)	49%	-39%	35%	100%										
4	T10Y	-31%	-19%	-66%	-28%	100%									
5	T2Y	-42%	-3%	-68%	-57%	95%	100%								
6	CPI	-16%	4%	-7%	-12%	27%	27%	100%							
7	GDP	21%	27%	69%	11%	-94%	-84%	-21%	100%						
8	UNRATE	50%	-75%	29%	67%	-18%	-37%	-7%	2%	100%					
9	PCE	23%	26%	70%	13%	-94%	-85%	-21%	100%	3%	100%				
10	PSR	-25%	-27%	-17%	1%	-10%	-8%	1%	11%	40%	11%	100%			
11	HD	30%	-9%	-12%	-7%	45%	40%	11%	-48%	-14%	-47%	-64%	100%		
12	CD	5%	48%	-21%	-29%	14%	22%	5%	-11%	-52%	-11%	-48%	68%	100%	
13	DRATE	54%	-61%	-27%	16%	55%	42%	7%	-64%	25%	-63%	-27%	65%	16%	100%

(Data from: FRED, Federal Reserve Bank of St. Louis, US).

The correlation analyses are shown in Table 4.2 between the dependent and each of the independent variables and amongst all the 13 explanatory variables – after four (4) variables were dropped on the basis of low explanatory power (R2 values of less than 5%) or insignificant relationships (p-values of less than 0,05) (see Annexure D). The correlations between the dependent and each of the independent (explanatory) variables and between one explanatory variable and another varied in strength and direction – whether positive or negative.

4.2.4 Standardisation of Independent Variables

Shown in Annexure H are the data for the nine (9) significant explanatory variables that were used to build the proof-of-concept credit risk model US2 for the US banks. Data for seven (7) of the variables used to build Model US1, which had eight (8) explanatory variables, are included in the set. The variables are of different sizes and measurement units. Some are less than one, while others are in billions. Most are measured in percentage terms, while others are in dollars. As they are, it is difficult to compare them on the same X-Y axis graph, for example. The data was therefore standardised by subtracting the mean of each data set from the value of the relevant variable and dividing the outcome by the standard deviation of the relevant data set. This resulted in the Z-score for each value of every variable. This is the number of standard deviations from the mean of a variable value. The outcome of standardising the values of the nine (9) significant explanatory variables of model US2 is given in Table 4.3. The Table was included in this chapter for ease of reference and quantitative overview comparative analysis. Each column of numbers has a positive and negative set whose average is zero, and most of the values are within two (2) standard deviations from their means. This indicates that the data is clustered around the mean with less extreme values.

Table 4.3 Dependent Variable and Standardized Independent (Explanatory) Variables

Date	LCO %	10Y-2Y %	10Y %	2Y %	Urate %	PCE (billions)	PSR %	HD %	CD %	Drate %
1987-01-01	1,47	0,29	1,08	1,02	0,43	1,50	0,71	0,68	1,17	0,28
1987-04-01	1,47	0,23	1,59	1,44	0,23	1,48	-0,04	0,81	1,21	0,18
1987-07-01	1,43	0,07	1,82	1,59	0,06	1,46	0,18	0,69	1,02	0,19
1987-10-01	1,49	0,05	1,94	1,64	-0,04	1,45	0,57	0,57	0,87	0,22
1988-01-01	1,45	0,04	1,62	1,40	-0,12	1,43	0,49	0,52	0,76	0,16
1988-04-01	1,47	0,07	1,84	1,60	-0,26	1,41	0,58	0,50	0,66	0,15
1988-07-01	1,52	0,53	1,93	1,82	-0,26	1,39	0,61	0,46	0,52	0,12
1988-10-01	1,46	0,98	1,86	1,92	-0,34	1,37	0,54	0,35	0,41	0,21
1989-01-01	1,57	1,51	1,97	2,18	-0,42	1,36	0,75	0,35	0,35	0,39
1989-04-01	1,55	1,49	1,77	2,00	-0,40	1,34	0,47	0,50	0,47	0,52
1989-07-01	1,54	1,25	1,49	1,68	-0,40	1,32	0,38	0,56	0,41	0,60
1989-10-01	1,64	1,21	1,40	1,59	-0,32	1,31	0,49	0,51	0,25	0,61
1990-01-01	1,64	1,20	1,63	1,79	-0,36	1,29	0,49	0,42	-0,00	0,64
1990-04-01	1,72	1,15	1,74	1,86	-0,34	1,27	0,64	0,38	-0,19	0,65
1990-07-01	1,83	0,57	1,75	1,69	-0,12	1,26	0,49	0,36	-0,31	0,94
1990-10-01	1,97	0,34	1,62	1,50	0,15	1,25	0,45	0,38	-0,43	1,23
1991-01-01	2,20	0,12	1,45	1,28	0,43	1,25	0,64	0,36	-0,56	1,39
1991-04-01	2,33	0,16	1,49	1,23	0,57	1,23	0,61	0,23	-0,79	1,52
1991-07-01	2,40	0,39	1,42	1,09	0,59	1,22	0,58	0,12	-1,00	1,46
1991-10-01	2,28	0,87	1,15	0,70	0,73	1,21	0,86	-0,06	-1,24	1,36
1992-01-01	2,39	1,07	1,13	0,62	0,89	1,18	0,97	-0,31	-1,53	1,16
1992-04-01	2,27	1,26	1,16	0,59	1,03	1,17	1,09	-0,47	-1,72	0,99
1992-07-01	2,10	1,60	0,83	0,19	1,05	1,15	0,86	-0,60	-1,86	0,80
1992-10-01	2,07	1,40	0,88	0,30	0,89	1,13	0,70	-0,67	-1,95	0,57
1993-01-01	1,84	1,20	0,67	0,18	0,75	1,11	0,67	-0,71	-1,96	0,48
1993-04-01	1,82	1,05	0,55	0,13	0,71	1,09	0,48	-0,77	-1,97	0,34
1993-07-01	1,72	0,64	0,39	0,12	0,55	1,07	0,17	-0,72	-1,86	0,06
1993-10-01	1,56	0,51	0,39	0,17	0,45	1,06	0,11	-0,75	-1,76	-0,26
1994-01-01	1,49	0,51	0,60	0,35	0,41	1,04	-0,04	-0,74	-1,60	-0,40
1994-04-01	1,42	0,20	1,04	0,83	0,19	1,02	0,02	-0,68	-1,42	-0,58
1994-07-01	1,41	0,00	1,14	0,98	0,06	1,00	-0,05	-0,55	-1,20	-0,65
1994-10-01	1,47	0,49	1,37	1,33	-0,16	0,98	0,07	-0,45	-1,03	-0,60
1995-01-01	1,44	0,86	1,21	1,31	-0,26	0,97	0,24	-0,28	-0,78	-0,42
1995-04-01	1,63	0,73	0,82	0,94	-0,14	0,95	-0,01	-0,10	-0,48	-0,23
1995-07-01	1,82	0,73	0,70	0,84	-0,14	0,93	-0,02	-0,00	-0,23	0,05
1995-10-01	1,99	0,81	0,51	0,70	-0,20	0,91	-0,11	0,03	-0,09	0,21
1996-01-01	2,14	0,53	0,52	0,61	-0,22	0,89	-0,09	0,05	0,03	0,26

Date	LCO %	10Y-2Y %	10Y %	2Y %	Urate %	PCE (billions)	PSR %	HD %	CD %	Drate %
1996-04-01	2,26	0,57	0,87	0,93	-0,24	0,87	-0,15	0,08	0,08	0,48
1996-07-01	2,33	0,57	0,90	0,96	-0,38	0,85	-0,09	0,12	0,15	0,67
1996-10-01	2,40	0,63	0,71	0,82	-0,34	0,83	-0,19	0,17	0,28	0,77
1997-01-01	2,56	0,66	0,81	0,91	-0,40	0,81	-0,23	0,13	0,18	0,86
1997-04-01	2,77	0,78	0,87	1,00	-0,54	0,80	-0,10	0,13	0,13	0,87
1997-07-01	2,79	0,88	0,66	0,85	-0,62	0,77	-0,29	0,14	0,16	0,80
1997-10-01	2,68	1,07	0,52	0,79	-0,74	0,74	-0,21	0,09	0,14	0,86
1998-01-01	2,63	1,11	0,38	0,68	-0,76	0,72	0,18	-0,04	-0,02	0,84
1998-04-01	2,64	1,23	0,38	0,72	-0,90	0,69	0,02	-0,02	0,07	0,89
1998-07-01	2,57	1,19	0,21	0,56	-0,82	0,67	-0,08	-0,03	0,10	0,87
1998-10-01	2,50	0,94	-0,03	0,28	-0,88	0,64	-0,29	0,03	0,25	0,81
1999-01-01	2,40	1,11	0,12	0,46	-0,96	0,62	-0,24	0,09	0,34	0,90
1999-04-01	2,12	0,97	0,35	0,62	-0,98	0,58	-0,75	0,23	0,54	0,67
1999-07-01	2,26	0,98	0,51	0,75	-1,00	0,55	-0,85	0,34	0,69	0,57
1999-10-01	2,20	1,04	0,62	0,87	-1,10	0,51	-0,90	0,33	0,64	0,51
2000-01-01	2,23	1,34	0,76	1,09	-1,12	0,47	-0,83	0,29	0,60	0,45
2000-04-01	2,14	1,72	0,64	1,11	-1,18	0,44	-0,76	0,43	0,81	0,54
2000-07-01	2,19	1,65	0,51	0,97	-1,14	0,41	-0,71	0,60	1,07	0,58
2000-10-01	2,67	1,44	0,37	0,78	-1,20	0,39	-0,82	0,81	1,40	0,68
2001-01-01	2,34	0,73	0,14	0,35	-1,00	0,36	-0,66	0,90	1,61	0,68
2001-04-01	2,63	0,01	0,24	0,21	-0,90	0,35	-0,81	1,10	1,90	0,75
2001-07-01	2,79	0,26	0,11	0,01	-0,64	0,34	-0,19	1,04	1,84	0,81
2001-10-01	3,12	0,94	0,01	-0,29	-0,24	0,31	-1,23	1,32	2,14	0,70
2002-01-01	3,60	0,92	0,15	-0,17	-0,12	0,30	-0,47	1,13	1,95	0,64
2002-04-01	3,10	0,91	0,16	-0,16	-0,04	0,27	-0,34	1,06	1,85	0,51
2002-07-01	3,12	1,10	-0,20	-0,53	-0,10	0,25	-0,48	1,09	1,83	0,48
2002-10-01	2,81	1,19	-0,32	-0,66	-0,02	0,23	-0,45	1,07	1,72	0,42
2003-01-01	2,84	1,37	-0,36	-0,76	-0,02	0,20	-0,60	1,07	1,67	0,38
2003-04-01	3,03	1,29	-0,49	-0,84	0,15	0,17	-0,58	1,01	1,60	0,22
2003-07-01	2,80	1,70	-0,22	-0,74	0,15	0,13	-0,44	0,87	1,43	-0,08
2003-10-01	2,86	1,56	-0,20	-0,68	-0,04	0,10	-0,59	0,94	1,46	0,18
2004-01-01	2,72	1,43	-0,32	-0,74	-0,12	0,06	-0,75	0,94	1,44	-0,04
2004-04-01	2,76	1,22	-0,06	-0,45	-0,18	0,04	-0,59	0,87	1,31	-0,08
2004-07-01	2,52	0,76	-0,19	-0,41	-0,28	0,00	-0,76	0,97	1,35	-0,16
2004-10-01	2,65	0,30	-0,25	-0,31	-0,28	0,05	-0,72	0,96	1,19	-0,19
2005-01-01	2,49	0,28	-0,19	-0,07	-0,36	0,08	-1,38	1,26	1,35	-0,36
2005-04-01	2,41	0,67	-0,26	0,00	-0,48	0,12	-1,46	1,26	1,25	-0,45
2005-07-01	3,02	0,97	-0,23	0,12	-0,56	0,17	-1,59	1,32	1,12	-0,52
2005-10-01	3,04	1,13	-0,11	0,27	-0,56	0,19	-1,38	1,29	0,93	-0,68

Date	LCO %	10Y-2Y %	10Y %	2Y %	Urate %	PCE (billions)	PSR %	HD %	CD %	Drate %
2006-01-01	1,77	1,30	-0,07	0,36	-0,70	0,23	-1,08	1,31	0,89	-0,55
2006-04-01	1,92	1,18	0,15	0,51	-0,76	0,27	-1,15	1,37	0,68	-0,34
2006-07-01	2,20	1,31	0,07	0,49	-0,76	0,31	-1,33	1,47	0,59	-0,29
2006-10-01	2,14	1,40	-0,05	0,41	-0,88	0,33	-1,25	1,55	0,56	-0,32
2007-01-01	2,34	1,37	-0,03	0,42	-0,84	0,37	-1,20	1,56	0,53	-0,33
2007-04-01	2,32	1,22	0,05	0,44	-0,84	0,40	-1,18	1,63	0,59	-0,24
2007-07-01	2,45	0,87	0,00	0,28	-0,74	0,43	-1,33	1,74	0,67	0,06
2007-10-01	2,80	0,37	-0,21	-0,06	-0,66	0,47	-1,40	1,86	0,74	0,35
2008-01-01	2,95	0,63	-0,47	-0,61	-0,54	0,49	-1,26	1,79	0,72	0,48
2008-04-01	3,26	0,43	-0,38	-0,46	-0,34	0,53	-0,65	1,48	0,41	0,58
2008-07-01	3,70	0,48	-0,39	-0,49	0,06	0,53	-1,07	1,55	0,49	0,80
2008-10-01	4,28	1,08	-0,66	-0,92	0,59	0,46	-0,36	1,51	0,45	1,61
2009-01-01	4,76	0,86	-0,88	-1,04	1,44	0,44	-0,36	1,44	0,37	2,19
2009-04-01	5,58	1,41	-0,62	-0,99	2,06	0,43	-0,01	1,09	-0,01	2,45
2009-07-01	5,92	1,63	-0,54	-0,99	2,26	0,47	-0,61	0,96	-0,13	2,23
2009-10-01	5,75	1,74	-0,56	-1,05	2,44	0,49	-0,52	0,71	-0,44	2,07
2010-01-01	6,60	1,99	-0,45	-1,03	2,38	0,52	-0,44	0,45	-0,66	2,26
2010-04-01	6,56	1,78	-0,55	-1,05	2,26	0,55	-0,16	0,15	-0,99	1,58
2010-07-01	5,48	1,34	-0,86	-1,18	2,16	0,57	-0,16	-0,04	-1,19	1,22
2010-10-01	4,91	1,51	-0,82	-1,19	2,18	0,61	-0,22	-0,19	-1,21	0,68
2011-01-01	4,39	1,95	-0,56	-1,12	1,90	0,65	-0,01	-0,38	-1,15	0,39
2011-04-01	3,43	1,80	-0,68	-1,17	1,92	0,68	-0,11	-0,46	-1,18	0,16
2011-07-01	3,67	1,21	-1,02	-1,27	1,88	0,70	-0,05	-0,56	-1,24	-0,04
2011-10-01	2,97	0,80	-1,19	-1,28	1,66	0,72	-0,02	-0,68	-1,24	-0,13
2012-01-01	2,63	0,77	-1,19	-1,27	1,44	0,76	0,36	-0,91	-1,42	-0,33
2012-04-01	2,53	0,52	-1,28	-1,27	1,40	0,77	0,59	-1,02	-1,45	-0,46
2012-07-01	2,52	0,34	-1,37	-1,28	1,29	0,79	0,33	-0,97	-1,30	-0,55
2012-10-01	2,45	0,40	-1,34	-1,28	1,15	0,82	1,08	-1,26	-1,51	-0,76
2013-01-01	2,30	0,69	-1,23	-1,28	1,11	0,85	-0,40	-0,93	-1,00	-0,86
2013-04-01	2,15	0,73	-1,21	-1,28	0,99	0,86	-0,20	-1,01	-0,96	-0,98
2013-07-01	2,13	1,45	-0,90	-1,24	0,81	0,88	-0,19	-1,04	-0,92	-1,10
2013-10-01	2,09	1,54	-0,88	-1,26	0,63	0,92	-0,35	-1,04	-0,85	-1,14
2014-01-01	2,00	1,51	-0,87	-1,24	0,47	0,95	-0,09	-1,14	-0,85	-1,21
2014-04-01	2,00	1,29	-0,93	-1,22	0,19	1,00	0,04	-1,20	-0,82	-1,28
2014-07-01	1,88	1,04	-0,99	-1,18	0,11	1,04	0,08	-1,24	-0,76	-1,39
2014-10-01	1,82	0,75	-1,09	-1,18	-0,12	1,07	0,16	-1,26	-0,68	-1,56
2015-01-01	1,75	0,32	-1,22	-1,15	-0,22	1,08	0,36	-1,26	-0,63	-1,66
2015-04-01	1,74	0,54	-1,13	-1,15	-0,28	1,12	0,21	-1,22	-0,47	-1,70
2015-07-01	1,75	0,51	-1,11	-1,12	-0,48	1,15	0,15	-1,18	-0,33	-1,67

Date	LCO %	10Y-2Y %	10Y %	2Y %	Urate %	PCE (billions)	PSR %	HD %	CD %	Drate %
2015-10-01	1,77	0,30	-1,12	-1,06	-0,52	1,17	0,17	-1,23	-0,45	-1,66
2016-01-01	1,83	0,02	-1,25	-1,06	-0,60	1,20	0,21	-1,26	-0,47	-1,67
2016-04-01	1,82	0,13	-1,32	-1,09	-0,58	1,24	-0,01	-1,18	-0,28	-1,60
2016-07-01	1,88	0,30	-1,40	-1,10	-0,60	1,27	-0,04	-1,13	-0,14	-1,56
2016-10-01	2,11	0,04	-1,15	-1,00	-0,68	1,31	-0,04	-1,15	-0,06	-1,46
2017-01-01	2,19	0,13	-1,01	-0,91	-0,80	1,35	0,07	-1,20	-0,07	-1,41
2017-04-01	2,11	0,15	-1,09	-0,89	-0,90	1,38	0,21	-1,23	-0,06	-1,36
2017-07-01	2,20	0,25	-1,10	-0,86	-0,96	1,42	0,21	-1,24	-0,05	-1,28
2017-10-01	2,23	0,48	-1,04	-0,74	-1,06	1,48	0,03	-1,24	0,01	-1,36
2018-01-01	2,23	0,58	-0,87	-0,56	-1,12	1,53	0,22	-1,32	-0,07	-1,27
2018-04-01	2,23	0,76	-0,80	-0,44	-1,18	1,59	0,21	-1,36	-0,12	-1,28
2018-07-01	2,26	0,97	-0,80	-0,37	-1,29	1,62	0,24	-1,36	-0,08	-1,24
2018-10-01	2,25	0,99	-0,75	-0,32	-1,25	1,66	0,34	-1,34	-0,03	-1,24
2019-01-01	2,25	1,08	-0,92	-0,44	-1,23	1,67	0,59	-1,31	0,01	-1,18
2019-04-01	2,27	1,03	-1,06	-0,57	-1,35	1,73	0,18	-1,25	0,13	-1,14
2019-07-01	2,31	1,14	-1,30	-0,74	-1,39	1,77	0,09	-1,22	0,20	-1,17
2019-10-01	2,31	1,04	-1,30	-0,78	-1,39	1,81	0,17	-1,22	0,25	-1,26
2020-01-01	2,29	0,94	-1,49	-0,97	-1,27	1,75	1,01	-1,29	0,17	-0,98
2020-04-01	2,26	0,70	-1,79	-1,31	4,34	1,34	6,82	-2,19	-1,08	-1,63
2020-07-01	1,91	0,68	-1,80	-1,33	1,76	1,71	3,25	-1,86	-0,66	-1,92
2020-10-01	1,52	0,44	-1,71	-1,32	0,53	1,76	2,37	-1,69	-0,44	-1,85
2021-01-01	1,54	0,13	-1,50	-1,33	0,17	1,91	4,68	-2,72	-1,71	-2,11

(Adapted from: FRED, Federal Reserve Bank of St. Louis, US).

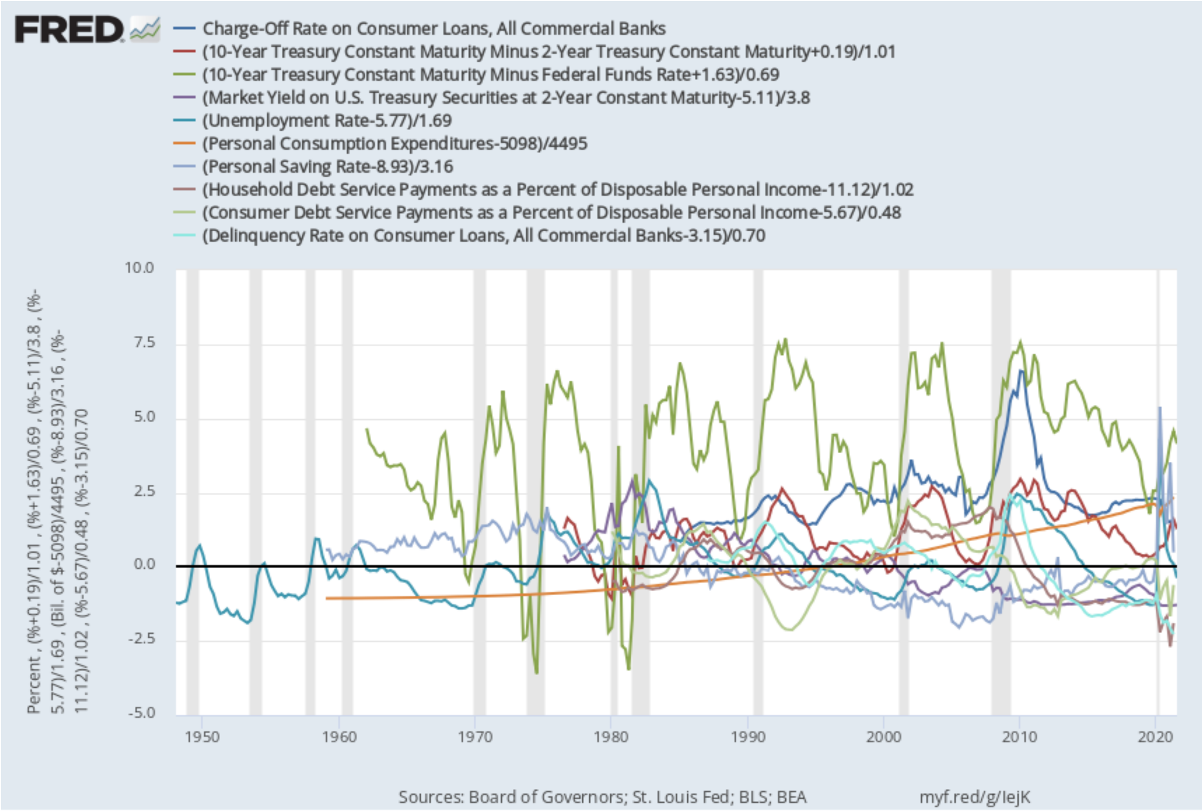


Figure 4.2: Line graph of the consumer Loan Charge Off Rate and nine independent variables.

(Adapted from: FRED, Federal Reserve Bank of St. Louis, US).

The line graphs shown in Figure 4.2 are examples of how the inbuilt FRED analysis tool was used for comparative analysis of the consumer Loan Charge Off Rate and nine (9) significant variables of Model US2. The independent variable data were standardised (Z-values) before the analysis was carried out. Most of the values are within two (2) standard deviations from their means.

4.2.5 Bivariable and Multivariable Regression and ANOVA Analyses

This section is about the higher order bivariable and multivariable regression analyses. To arrive at the list of eight (8) significant explanatory variables for credit risk model US1 from the initial 17 variables, first, bivariable regression analyses were performed to gather information on how each of the 17 independent variables individually relate with the Loan Charge Off Rate (see Annexure D). As explained in Section 3.3.3.1, in a regression analysis, the contribution of each variable is represented by the beta coefficient. The coefficient is the degree of change in the outcome variable for every unit of change in the explanatory variable.

The variables with significant coefficient values at the 95% confidence level (t-statistic of greater or equal to [≥] 1.96 in absolute terms or p-values of less than or equal to [≤] 0,05) would be deemed as significant contributors to the variance in the Loan Charge Off Rates. Therefore, explanatory variables with p-values of more than 0.05 and/or R– squared values equal to or less than 5% were deemed statistically insignificant and omitted (see Table 4.4). This resulted in the explanatory variables being reduced to 13. A multivariable regression analysis of the Loan Charge Off Rate on the 13 explanatory variables was done next, resulting in another five (5) explanatory variables being dropped as a result of their becoming insignificant in the combination (see Table 4.5). In the third round of the multivariable regression analysis, all eight (8) explanatory variables were found to be significant (t-statistic of ≥ 1.96 in absolute terms or p-values of ≤ 0,05). The results of this regression analysis are shown in Table 4.6. The eight (8) explanatory variables explain 91.2% of the variation in the dependent variable (Loan Charge Off Rate) on an adjusted R-square basis. As stated in Section 3.3.3 of the methodology chapter of this study, existing credit risk models explain 20-30% of the variation in LGD (Loan Loss Given Default).

Table 4.4 Summary of A Series of Correlations and Multivariable Regression Analysis of Loan Charge Off Rate and All Selected 17 US Independent Variables

	Key		**Bivariable/Multivariable regression and ANOVA analysis**
	Dependent variable (LCO)	(grey)	Round 1 (light blue)
	Sentiment explanatory variables	(yellow)	Round 2 (purple)
	Macro-economic explanatory variables	(orange)	Round 3 (green)
	Obligor explanatory variables	(navy)	

S/N	**Explanatory variable**	**Short name**	**Coefficient (sensitivity)**	**Standard error**	**T-stat**	**p-value**	**R-square**	**round 1**	**round 2**	**round 3**	**Finalists**
			Bivariable/Multiple regression and ANOVA analysis								
	Charge off rate	LCO									
1	Autos and L /Trucks	VEH	0,052	0,073	0,713	0,477	91,5%	include	include	exclude	exclude
2	Gold Price	GOLD	0	0,995	2,38	0	3,1%	include	exclude	exclude	exclude
3	Oil price (WTI)	OIL	-0,262	0,051	-5,139	0,000	91,2%	include	include	include	include
4	S&P Case-Shiller home price index	S&P CSPI	0,003	0,990	1,81	0,07	1,6%	include	exclude	exclude	exclude
5	10Y Inflation rate	INFL	-0,238	1,200	0,67	0,508	0,0%	include	exclude	exclude	exclude
6	10Y Treasury CM	T10Y	-1 257	537	-2,341	0,021	91,2%	include	include	include	include
7	10Y-2Y Treasury CM	10Y-2Y(Tilt)	3 320	1 419	2,341	0,021	91,2%	include	include	include	include
8	2Y Treasury CM	T2Y	-3 870	1 654	-2,341	0,021	91,2%	include	include	include	include
9	Consumer Price Index	CPI	-0,012	0,027	-0,439	0,662	91,5%	include	include	exclude	exclude
10	Gross Domestic Product	GDP	-3,142	1,962	-1,601	0,112	91,5%	include	include	exclude	exclude
11	Unemployment rate	UNRATE	0,484	0,057	8,538	0,000	91,2%	include	include	include	include
12	Personal cons. exp.	PCE	1,184	0,107	11,014	0,000	91,2%	include	include	include	include
13	Personal saving rate	PSR	-0,351	0,044	-7,884	0,000	91,2%	include	include	include	include
14	Household DSP / DPI	HD	-0,106	0,080	-1,323	0,188	91,5%	include	include	exclude	exclude
15	Consumer DSP / DPI	CD	0,101	0,067	1,507	0,134	91,5%	include	include	exclude	exclude
16	Delinquency rate	DRATE	0,869	0,041	21,334	0,000	91,2%	include	include	include	include
17	HNPOCCLL	HNPO	0	0,984	2,966	0,004	5,1%	include	exclude	exclude	exclude

(Data from: FRED, Federal Reserve Bank of St. Louis, US).

As shown in Table 4.4, credit losses (Charge Off Rates) were regressed on explanatory variables, and those that were found not to be significant were dropped. A final list of eight (8) remained (see the light green shadings in the last four columns). The red shadings denote variables dropped at each stage of the analysis.

The initial list of 17 independent explanatory variables was reduced to a final list of eight (8) significant variables. Nine (9) variables were eliminated on the basis of having p-values above 0,05 and/or R^2 values when individually regressed with LCO – of less than or equal to (≤) 5%. That it took only three (3) rounds of regression analyses to arrive at the final eight (8) significant explanatory variables is possibly mainly due to the good initial choices of the independent variables. The combination of eight (8) significant variables is not the only such combination. Technically, the number of models that can be built from various combinations of the Charge Off Rate and the 17 explanatory variables is large – 131,071. This was calculated and shown in Table 4.7 using the formula for calculation of unordered k-element combinations of *n* objects without repetition (Sheffield, no date; Ďuriš *et al.*, 2021). It is, therefore, plausible through careful study of correlation and regression analyses statistics to find other sets of variables that in combination are all significant. Using such a technique, a set of nine (9) variables – that were all significant in combination – similar to the one with eight (8) variables was found. In that set, Household Debt Service Payment as percentage of Disposable Personal Income (HD) and Consumer Debt Service Payment as percentage of Disposable Personal Income (CD) were included, while Oil Price (OIL) was excluded.

The two (2) consumer credit risk models – named model US1 (with eight (8) explanatory variables) and model US2 (with nine (9) explanatory variables) were back tested on their power to predict credit losses in the form of the Loan Charge Off Rate. The results of the back test for Model US1 are shown in Table 4.8, Figure 4.4, and Figure 4.5. The final regression analysis statistics for Model US2 are shown in Table 4.9, and its back test results are shown in Table 4.10, Figure 4.6 and Figure 4.7. The results show that model US1 is marginally better than model US2 – it has a better R-squared value (91.2% vs 90.5%) and estimation accuracy on average (85% vs 84%). The idea of back testing was to gain insight and confidence that a regression model is suitable and valid for use in credit risk estimation. The back testing used is termed in-sample since the period for which the coefficients and other statistics of the models are derived coincides with the period over which the Loan Charge Off Rate was estimated. If the coefficients and statistics are non-satisfactory, it shows in the back testing results. The researcher can alter the model – for example, by selecting a different combination of explanatory variables or including new ones and repeat the process of estimating the coefficients and back testing. This process (see Figure 4.3) can be repeated many times over until the modeller is comfortable that the choice of the regression model is fit for purpose.

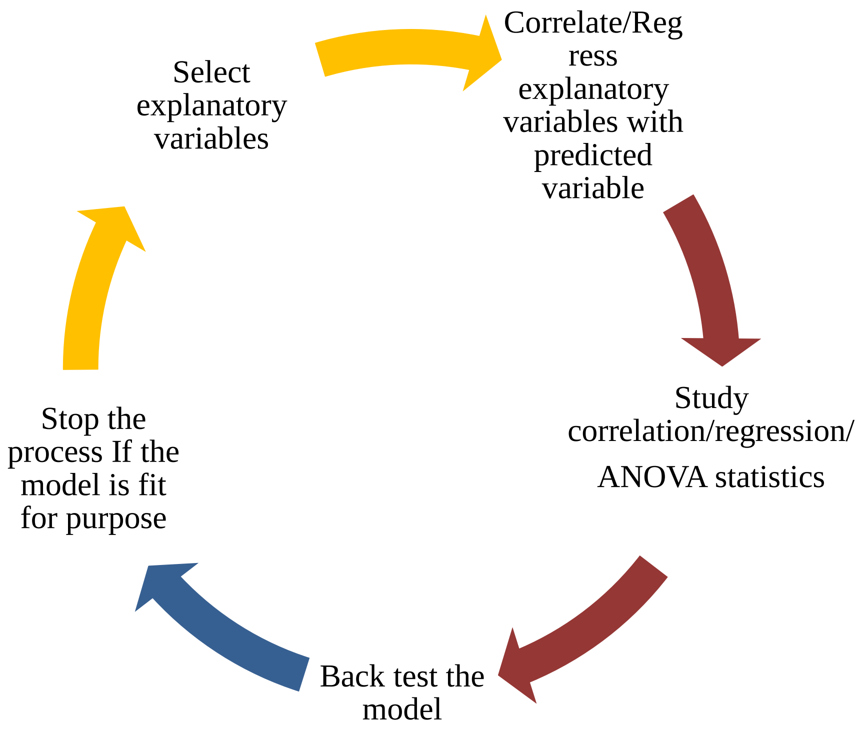


Figure 4.3: Process for Triangulation of Credit Risk Model

(Source: Own preparation based on Bolton (2009).

Shown in Figure 4.3 is the process for triangulation of a credit risk model – from the selection of explanatory variables through multiple analyses to back testing. In back testing, the actual Loan Charge Off Rates were compared with values estimated using the model. The triangulation revealed that linear multivariable regression is suitable and fit for estimating credit risk (as demonstrated by the over 90% R-square). Under the linear multivariable regression, the explanatory variables are significant (at 95% confidence level) as indicated by p-values that are less than 0.05 and, consequently T-stats equal to or greater than 1.96 in absolute terms. Combined, the variables explain over 90% of the variation in the independent variable (Loan Charge Off Rate).

Table 4.5: Multiple Regression Analysis of 13 Significant Explanatory Variables

Regression Statistics	
Multiple R	0,961
R Square	0,923
Adjusted R Square	0,915
Standard Error	0,290
Observations	137

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	13	124,645	9,588	113,627	0,000
Residual	123	10,379	0,084		
Total	136	135,024			

		Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
	Intercept	2,465	0,025	99,310	0,000	2,416	2,514
1	VEH	0,052	0,073	0,713	0,477	-0,092	0,196
2	OIL	-0,218	0,056	-3,919	0,000	-0,327	-0,108
3	T10Y-2Y (Tilt)	-1 200,100	529,225	-2,268	0,025	-2 247,669	-152,530
4	T10Y	3 169,443	1 397,597	2,268	0,025	402,986	5 935,900
5	T2Y	-3 694,417	1 629,200	-2,268	0,025	-6 919,319	-469,515
6	CPI	-0,012	0,027	-0,439	0,662	-0,066	0,042
7	GDP	-3,142	1,962	-1,601	0,112	-7,026	0,743
8	UNRATE	0,533	0,069	7,705	0,000	0,396	0,670
9	PCE	4,276	1,950	2,193	0,030	0,417	8,136
10	PSR	-0,327	0,055	-5,984	0,000	-0,436	-0,219
11	HD	-0,106	0,080	-1,323	0,188	-0,263	0,052
12	CD	0,101	0,067	1,507	0,134	-0,032	0,234
13	DRATE	0,904	0,060	14,953	0,000	0,784	1,024

(Data from: FRED, Federal Reserve Bank of St. Louis, US).

The data in Table 4.5 show that VEH, CPI, GDP, HD, and CD are insignificant in this combination. This means that the explanatory variables are possibly not significant, or what they explain in the variation of credit losses is explained to a larger extent by other variables.

Table 4.6: Regression Statistics and ANOVA for Eight (8) Significant Explanatory Variables

Regression Statistics	
Multiple R	0,958
R Square	0,917
Adjusted R Square	0,912
Standard Error	0,295
Observations	137

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	8	123,854	15,482	177,398	0,000
Residual	128	11,171	0,087		
Total	136	135,024			

		Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
	Intercept	2,46	0,025	97,65	0,000	2,41	2,51
1	OIL	-0,262	0,051	-5,13	0,000	-0,362	-0,161
2	T10Y-2Y (Tilt)	-1 257	537	-2,34	0,021	-2 320	-194
3	T10Y	3 320	1 418	2,34	0,021	513	6 127
4	T2Y	-3 870	1 653	-2,34	0,021	-7 142	-598
5	UNRATE	0,484	0,057	8,53	0,000	0,372	0,596
6	PCE	1,184	0,107	11,01	0,000	0,971	1,396
7	PSR	-0,351	0,044	-7,88	0,000	-0,439	-0,263
8	DRATE	0,869	0,041	21,33	0,000	0,788	0,950

(Data from: FRED, Federal Reserve Bank of St. Louis, US).

The eight (8) explanatory variables in Table 4.6 are significant as indicated by p-values that are less than 0.05 and consequently T-statistics of equal to or greater than 1.96. Combined, they explain 91.2% of the variation in the independent variable (Loan Charge Off Rate), and the resultant regression equation is Consumer Credit Risk Model US1.

Table 4.7: Number of Possible Models With 17 Explanatory Variables

Total no. of variables	No. of variables in the model	(n-k)	n!	k!(n-k)!	Number of potential models
n	k				n!/(k!(n-k)!)
17	17	0	355 687 428 096 000	355 687 428 096 000	1
17	16	1	355 687 428 096 000	20 922 789 888 000	17
17	15	2	355 687 428 096 000	2 615 348 736 000	136
17	14	3	355 687 428 096 000	523 069 747 200	680
17	13	4	355 687 428 096 000	149 448 499 200	2 380
17	12	5	355 687 428 096 000	57 480 192 000	6 188
17	11	6	355 687 428 096 000	28 740 096 000	12 376
17	10	7	355 687 428 096 000	18 289 152 000	19 448
17	9	8	355 687 428 096 000	14 631 321 600	24 310
17	8	9	355 687 428 096 000	14 631 321 600	24 310
17	7	10	355 687 428 096 000	18 289 152 000	19 448
17	6	11	355 687 428 096 000	28 740 096 000	12 376
17	5	12	355 687 428 096 000	57 480 192 000	6 188

17	4	13	355 687 428 096 000	149 448 499 200	2 380
17	3	14	355 687 428 096 000	523 069 747 200	680
17	2	15	355 687 428 096 000	2 615 348 736 000	136
17	1	16	355 687 428 096 000	20 922 789 888 000	17
Total number of potential models					**131 071**

(Source: Own preparation based on Sheffield (n.d) and Ďuriš et al. (2021)).

The outcomes of the calculations shown in Table 4.7 indicate that the number of possible models which can be constructed by regressing credit losses with any number of the explanatory variables varies from as low as one (1) when all the variables are in the model to as high as 24 310, when eight (8) or nine (9) explanatory variables are included in the model.

4.2.6 Back Testing Results for Model US1

Shown in Table 4.8, Figure 4.4, and Figure 4.5, are the results of back testing Consumer Credit Risk Model US1. An analysis of these tabular and graphical representations of the results reveals that the model generally overestimates the credit losses, and its accuracy varies from as low as 50% to as high as 99%, with the average accuracy being 85% over the period 1987 – 2021. The line representing the estimated impairments closely tracks that of actual impairments. However, it is mostly above it, meaning that it generally overestimates rather than underestimate the credit losses. Its performance is relatively uniform, with only 16 incidences of low estimation performances (less than 75%) spread over nine (9) years. Only year 1991 had three (3) incidences of underperformance. The other eight (8) years had one (1) or two (2) such incidences each.

Table 4.8: Back Test Results for Model US1

Date	Charge off rate (%)	Estimated charge off rate (%)	Accuracy (0-1)	Date	Charge off rate (%)	Estimated charge off rate (%)	Accuracy (0-1)	Date	Charge off rate (%)	Estimated charge off rate (%)	Accuracy (0-1)
1987-01-01	1,47	1,67	87%	1998-07-01	2,57	2,64	97%	2010-01-01	6,60	6,30	95%
1987-04-01	1,47	1,88	72%	1998-10-01	2,50	2,64	94%	2010-04-01	6,56	5,57	85%
1987-07-01	1,43	1,80	74%	1999-01-01	2,40	2,71	87%	2010-07-01	5,48	5,17	94%
1987-10-01	1,49	1,70	86%	1999-04-01	2,12	2,74	71%	2010-10-01	4,91	4,71	96%
1988-01-01	1,45	1,61	89%	1999-07-01	2,26	2,71	80%	2011-01-01	4,39	4,28	97%
1988-04-01	1,47	1,56	94%	1999-10-01	2,20	2,67	79%	2011-04-01	3,43	4,06	82%
1988-07-01	1,52	1,57	97%	2000-01-01	2,23	2,62	83%	2011-07-01	3,67	3,89	94%
1988-10-01	1,46	1,63	88%	2000-04-01	2,14	2,63	77%	2011-10-01	2,97	3,62	78%
1989-01-01	1,57	1,67	93%	2000-07-01	2,19	2,65	79%	2012-01-01	2,63	3,16	80%
1989-04-01	1,55	1,85	81%	2000-10-01	2,67	2,75	97%	2012-04-01	2,53	3,03	80%
1989-07-01	1,54	1,92	75%	2001-01-01	2,34	2,82	79%	2012-07-01	2,52	3,00	81%
1989-10-01	1,64	1,92	83%	2001-04-01	2,63	3,06	84%	2012-10-01	2,45	2,56	96%
1990-01-01	1,64	1,99	79%	2001-07-01	2,79	3,02	92%	2013-01-01	2,30	2,98	70%
1990-04-01	1,72	2,04	81%	2001-10-01	3,12	3,58	85%	2013-04-01	2,15	2,77	71%
1990-07-01	1,83	2,42	68%	2002-01-01	3,60	3,35	93%	2013-07-01	2,13	2,60	78%
1990-10-01	1,97	2,27	85%	2002-04-01	3,10	3,22	96%	2013-10-01	2,09	2,66	73%
1991-01-01	2,20	2,53	85%	2002-07-01	3,12	3,15	99%	2014-01-01	2,00	2,45	78%
1991-04-01	2,33	3,25	60%	2002-10-01	2,81	3,13	89%	2014-04-01	2,00	2,19	90%
1991-07-01	2,40	3,22	66%	2003-01-01	2,84	3,12	90%	2014-07-01	1,88	2,13	86%
1991-10-01	2,28	3,07	65%	2003-04-01	3,03	3,10	98%	2014-10-01	1,82	2,08	86%

Date	Charge off rate (%)	Estimated charge off rate (%)	Accuracy (0-1)	Date	Charge off rate (%)	Estimated charge off rate (%)	Accuracy (0-1)	Date	Charge off rate (%)	Estimated charge off rate (%)	Accuracy (0-1)
1992-01-01	2,39	2,99	75%	2003-07-01	2,80	2,90	96%	2015-01-01	1,75	2,06	82%
1992-04-01	2,27	2,88	73%	2003-10-01	2,86	3,12	91%	2015-04-01	1,74	2,04	83%
1992-07-01	2,10	2,76	68%	2004-01-01	2,72	2,92	93%	2015-07-01	1,75	2,14	78%
1992-10-01	2,07	2,58	75%	2004-04-01	2,76	2,86	97%	2015-10-01	1,77	2,17	77%
1993-01-01	1,84	2,41	69%	2004-07-01	2,52	2,75	91%	2016-01-01	1,83	2,18	81%
1993-04-01	1,82	2,33	72%	2004-10-01	2,65	2,70	98%	2016-04-01	1,82	2,24	77%
1993-07-01	1,72	2,11	78%	2005-01-01	2,49	2,75	89%	2016-07-01	1,88	2,30	77%
1993-10-01	1,56	1,83	83%	2005-04-01	2,41	2,64	90%	2016-10-01	2,11	2,43	85%
1994-01-01	1,49	1,83	77%	2005-07-01	3,02	2,54	84%	2017-01-01	2,19	2,43	89%
1994-04-01	1,42	1,63	85%	2005-10-01	3,04	2,41	79%	2017-04-01	2,11	2,42	85%
1994-07-01	1,41	1,57	89%	2006-01-01	1,77	2,37	66%	2017-07-01	2,20	2,49	87%
1994-10-01	1,47	1,53	96%	2006-04-01	1,92	2,57	66%	2017-10-01	2,23	2,46	90%
1995-01-01	1,44	1,53	94%	2006-07-01	2,20	2,71	77%	2018-01-01	2,23	2,46	90%
1995-04-01	1,63	1,77	91%	2006-10-01	2,14	2,69	75%	2018-04-01	2,23	2,45	90%
1995-07-01	1,82	2,02	89%	2007-01-01	2,34	2,75	83%	2018-07-01	2,26	2,45	92%
1995-10-01	1,99	2,37	81%	2007-04-01	2,32	2,81	79%	2018-10-01	2,25	2,58	85%
1996-01-01	2,14	2,19	98%	2007-07-01	2,45	3,12	72%	2019-01-01	2,25	2,57	86%
1996-04-01	2,26	2,48	90%	2007-10-01	2,80	3,32	82%	2019-04-01	2,27	2,69	82%
1996-07-01	2,33	2,57	89%	2008-01-01	2,95	3,37	86%	2019-07-01	2,31	2,70	83%
1996-10-01	2,40	2,68	88%	2008-04-01	3,26	3,16	97%	2019-10-01	2,31	2,64	86%
1997-01-01	2,56	2,80	91%	2008-07-01	3,70	3,75	99%	2020-01-01	2,29	2,63	85%
1997-04-01	2,77	2,75	99%	2008-10-01	4,28	4,88	86%	2020-04-01	2,26	2,36	96%

Date	Charge off rate (%)	Estimated charge off rate (%)	Accuracy (0-1)	Date	Charge off rate (%)	Estimated charge off rate (%)	Accuracy (0-1)	Date	Charge off rate (%)	Estimated charge off rate (%)	Accuracy (0-1)
1997-07-01	2,79	2,70	97%	2009-01-01	4,76	5,84	77%	2020-07-01	1,91	2,42	73%
1997-10-01	2,68	2,65	99%	2009-04-01	5,58	6,17	89%	2020-10-01	1,52	2,27	50%
1998-01-01	2,63	2,51	96%	2009-07-01	5,92	6,29	94%	2021-01-01	1,54	1,17	76%
1998-04-01	2,64	2,59	98%	2009-10-01	5,75	6,15	93%			Average	85%

(Data from: FRED, Federal Reserve Bank of St. Louis, US)

The data in Table 4.8 shows that Consumer Credit Risk Model US1 generally overestimates credit losses. Its accuracy varies from as low as 50% to as high as 99%. The average accuracy is 85% over the period 1987 – 2021.

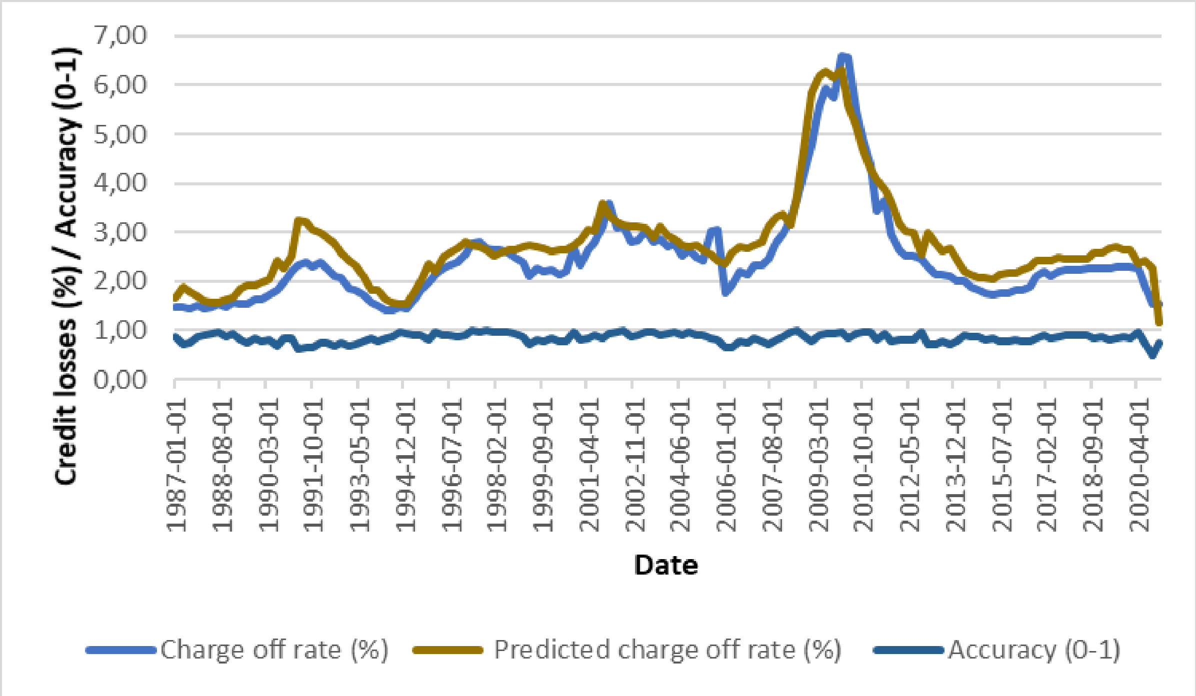


Figure 4.4: Back Testing Results for Model US1

(Data from: FRED, Federal Reserve Bank of St. Louis, US).

For Consumer Credit Risk Model US1, the line representing the estimated Charge Off Rate closely tracks that of the actual Charge Off Rate (Figure 4.4). However, it is generally above it, meaning that it overestimates rather than underestimate the credit losses. There was a significant decrease in credit losses during the Covid-19 period of 2020[8].

[8] The significant decrease in credit losses during the Covid-19 period of 2020 was due to cash grants and loan guarantees that were given to US citizens and corporations by their government to cushion them against the adverse effects of the pandemic (Borio, 2020).

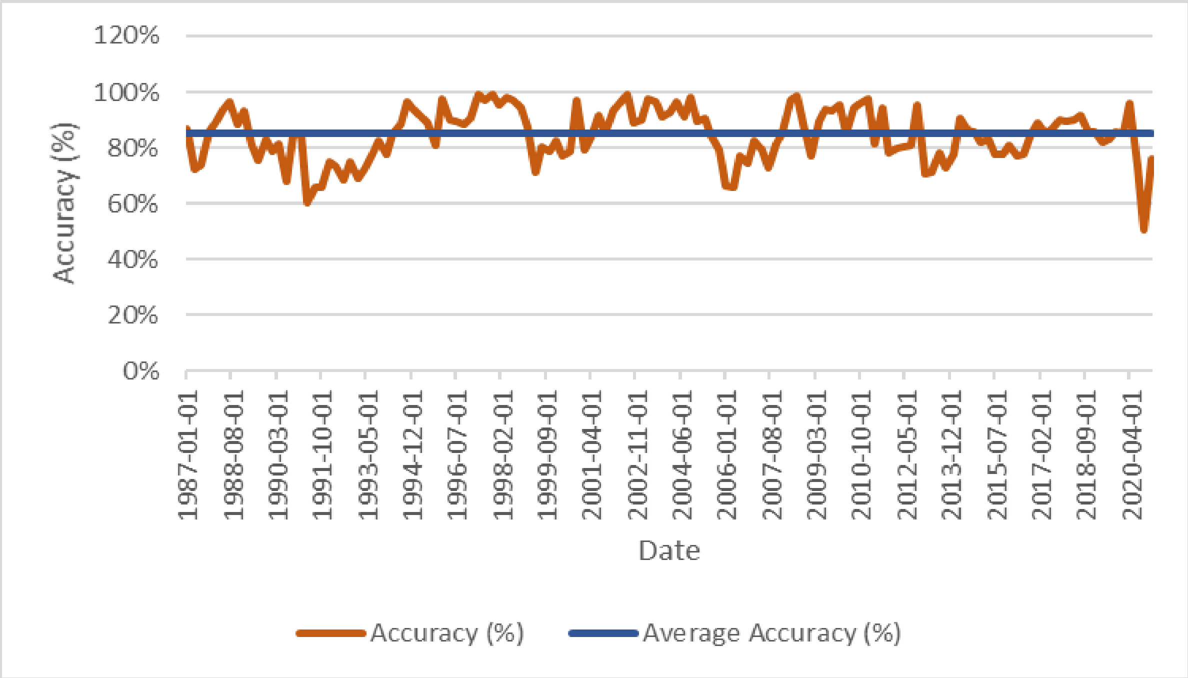


Figure 4.5: Back Testing Accuracy Results for Model US1

(Data from: FRED, Federal Reserve Bank of St. Louis, US).

For Consumer Credit Risk Model US1, its performance is relatively uniform, with only 16 incidences of low estimation performances (less than 75%) spread over nine (9) years. Only year 1991 had three (3) incidences of underperformance. The other eight (8) years had one (1) or two (2) incidences each (Figure 4.5).

Table 4.9: Regression Statistics and ANOVA For Nine (9) Significant Explanatory Variables

Regression Statistics	
Multiple R	0,955
R Square	0,912
Adjusted R Square	0,905
Standard Error	0,306
Observations	137

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	9	123,107	13,679	145,770	0,000
Residual	127	11,917	0,094		
Total	136	135,024			

		Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
	Intercept	2,465	0,026	94,174	0,000	2,413	2,516
1	T10Y-2Y (Tilt)	-1 269,799	557,241	-2,279	0,024	-2 372,479	-167,119
2	T10Y	3 353,474	1 471,586	2,279	0,024	441,471	6 265,476
3	T2Y	-3 909,058	1 715,448	-2,279	0,024	-7 303,620	-514,496
4	UNRATE	0,464	0,061	7,622	0,000	0,343	0,584
5	PCE	0,878	0,090	9,774	0,000	0,700	1,056
6	PSR	-0,285	0,050	-5,753	0,000	-0,383	-0,187
7	HD	-0,206	0,074	-2,784	0,006	-0,352	-0,060
8	CD	0,220	0,055	4,035	0,000	0,112	0,328
9	DRATE	0,918	0,046	19,819	0,000	0,826	1,010

(Data from: FRED, Federal Reserve Bank of St. Louis, US).

The nine (9) explanatory variables shown in

Table 4.9 are significant as indicated by p-values that are less than 0.05 and consequently T-statistics equal to or greater than 1.96. Combined, they explain over 90% of the variation in the independent variable (Loan Charge Off Rate), and the resulting regression equation is Consumer Credit Risk Model US2.

4.2.7 Back Testing Results for Model US2

Shown in Table 4.10, Figure 4.6 and Figure 4.7 are the results of back testing Consumer Credit Risk Model US2. An analysis of these tabular and graphical representations of the results reveals that the model generally overestimates the credit losses, and its accuracy varies from as low as 56% to as high as 100% with the average accuracy being 84% over the period 1987-2021. The line representing the estimated Charge Off Rates closely tracks that of actual Charge Off Rates. However, it is generally above it, meaning that it generally overestimates rather than underestimate the credit losses. Its incidences of low estimation performances (less than 75%) were almost twice as many as those of Model US1. It had two (2) years – 1987 and 2013 – with four (4) incidences each, one (1) year with three (3) incidences and 10 years having one (1) or two (2) incidences each.

Table 4.10: Back Testing Results for Model US2

Date	Charge off rate (%)	Estimated Charge off rate (%)	Accuracy (%)	Date	Charge off rate (%)	Estimated Charge off rate (%)	Accuracy (%)	Date	Charge off rate (%)	Estimated Charge off rate (%)	Accuracy (%)
1987-01-01	1,47	1,98	65%	1998-07-01	2,57	2,67	96%	2010-01-01	6,60	6,26	95%
1987-04-01	1,47	2,08	58%	1998-10-01	2,50	2,67	93%	2010-04-01	6,56	5,50	84%
1987-07-01	1,43	1,99	61%	1999-01-01	2,40	2,74	86%	2010-07-01	5,48	5,09	93%
1987-10-01	1,49	1,88	74%	1999-04-01	2,12	2,75	70%	2010-10-01	4,91	4,69	95%
1988-01-01	1,45	1,77	78%	1999-07-01	2,26	2,73	79%	2011-01-01	4,39	4,36	99%
1988-04-01	1,47	1,70	85%	1999-10-01	2,20	2,69	78%	2011-04-01	3,43	4,20	77%
1988-07-01	1,52	1,66	91%	2000-01-01	2,23	2,66	81%	2011-07-01	3,67	3,95	92%
1988-10-01	1,46	1,72	82%	2000-04-01	2,14	2,70	74%	2011-10-01	2,97	3,77	73%
1989-01-01	1,57	1,81	85%	2000-07-01	2,19	2,78	73%	2012-01-01	2,63	3,41	70%
1989-04-01	1,55	2,00	71%	2000-10-01	2,67	2,91	91%	2012-04-01	2,53	3,22	73%
1989-07-01	1,54	2,05	67%	2001-01-01	2,34	2,99	72%	2012-07-01	2,52	3,19	73%
1989-10-01	1,64	2,03	76%	2001-04-01	2,63	3,20	78%	2012-10-01	2,45	2,76	87%
1990-01-01	1,64	2,06	74%	2001-07-01	2,79	3,19	86%	2013-01-01	2,30	3,16	63%
1990-04-01	1,72	2,04	82%	2001-10-01	3,12	3,60	85%	2013-04-01	2,15	2,97	62%

Date	Charge off rate (%)	Estimated Charge off rate (%)	Accuracy (%)	Date	Charge off rate (%)	Estimated Charge off rate (%)	Accuracy (%)	Date	Charge off rate (%)	Estimated Charge off rate (%)	Accuracy (%)
1990-07-01	1,83	2,45	66%	2002-01-01	3,60	3,41	95%	2013-07-01	2,13	2,87	65%
1990-10-01	1,97	2,33	82%	2002-04-01	3,10	3,31	93%	2013-10-01	2,09	2,85	64%
1991-01-01	2,20	2,49	87%	2002-07-01	3,12	3,26	96%	2014-01-01	2,00	2,68	66%
1991-04-01	2,33	3,18	64%	2002-10-01	2,81	3,21	86%	2014-04-01	2,00	2,49	75%
1991-07-01	2,40	3,13	70%	2003-01-01	2,84	3,23	86%	2014-07-01	1,88	2,40	72%
1991-10-01	2,28	2,98	69%	2003-04-01	3,03	3,15	96%	2014-10-01	1,82	2,15	82%
1992-01-01	2,39	2,84	81%	2003-07-01	2,80	2,90	96%	2015-01-01	1,75	1,94	89%
1992-04-01	2,27	2,73	80%	2003-10-01	2,86	3,12	91%	2015-04-01	1,74	2,00	85%
1992-07-01	2,10	2,60	76%	2004-01-01	2,72	2,94	92%	2015-07-01	1,75	2,00	86%
1992-10-01	2,07	2,38	85%	2004-04-01	2,76	2,87	96%	2015-10-01	1,77	1,98	88%
1993-01-01	1,84	2,22	79%	2004-07-01	2,52	2,80	89%	2016-01-01	1,83	1,93	95%
1993-04-01	1,82	2,14	83%	2004-10-01	2,65	2,76	96%	2016-04-01	1,82	2,11	84%
1993-07-01	1,72	1,90	90%	2005-01-01	2,49	2,75	90%	2016-07-01	1,88	2,19	84%
1993-10-01	1,56	1,62	96%	2005-04-01	2,41	2,64	91%	2016-10-01	2,11	2,34	89%
1994-01-01	1,49	1,59	93%	2005-07-01	3,02	2,57	85%	2017-01-01	2,19	2,36	92%
1994-04-01	1,42	1,42	100%	2005-10-01	3,04	2,36	78%	2017-04-01	2,11	2,34	89%
1994-07-01	1,41	1,37	97%	2006-01-01	1,77	2,36	67%	2017-07-01	2,20	2,42	90%
1994-10-01	1,47	1,33	90%	2006-04-01	1,92	2,54	68%	2017-10-01	2,23	2,43	91%
1995-01-01	1,44	1,41	98%	2006-07-01	2,20	2,63	81%	2018-01-01	2,23	2,49	88%
1995-04-01	1,63	1,70	96%	2006-10-01	2,14	2,50	83%	2018-04-01	2,23	2,51	88%
1995-07-01	1,82	1,99	91%	2007-01-01	2,34	2,52	92%	2018-07-01	2,26	2,53	88%
1995-10-01	1,99	2,38	80%	2007-04-01	2,32	2,63	87%	2018-10-01	2,25	2,56	86%
1996-01-01	2,14	2,22	96%	2007-07-01	2,45	3,02	77%	2019-01-01	2,25	2,54	87%
1996-04-01	2,26	2,51	89%	2007-10-01	2,80	3,35	80%	2019-04-01	2,27	2,69	82%
1996-07-01	2,33	2,62	87%	2008-01-01	2,95	3,49	82%	2019-07-01	2,31	2,68	84%
1996-10-01	2,40	2,77	84%	2008-04-01	3,26	3,54	91%	2019-10-01	2,31	2,62	87%
1997-01-01	2,56	2,86	88%	2008-07-01	3,70	4,05	91%	2020-01-01	2,29	2,61	86%
1997-04-01	2,77	2,77	100%	2008-10-01	4,28	4,74	89%	2020-04-01	2,26	2,46	91%
1997-07-01	2,79	2,72	98%	2009-01-01	4,76	5,60	82%	2020-07-01	1,91	2,36	76%
1997-10-01	2,68	2,70	99%	2009-04-01	5,58	6,06	91%	2020-10-01	1,52	2,18	56%
1998-01-01	2,63	2,55	97%	2009-07-01	5,92	6,18	96%	2021-01-01	1,54	1,22	80%
1998-04-01	2,64	2,61	99%	2009-10-01	5,75	6,09	94%		Average		84%

(Data from: FRED, Federal Reserve Bank of St. Louis, US).

The data in Table 4.10 shows that Consumer Credit Risk Model US2 generally overestimates credit losses. Its accuracy varies from as low as 56% to as high as 100%. The average accuracy is 84% over the period 1987 – 2021.

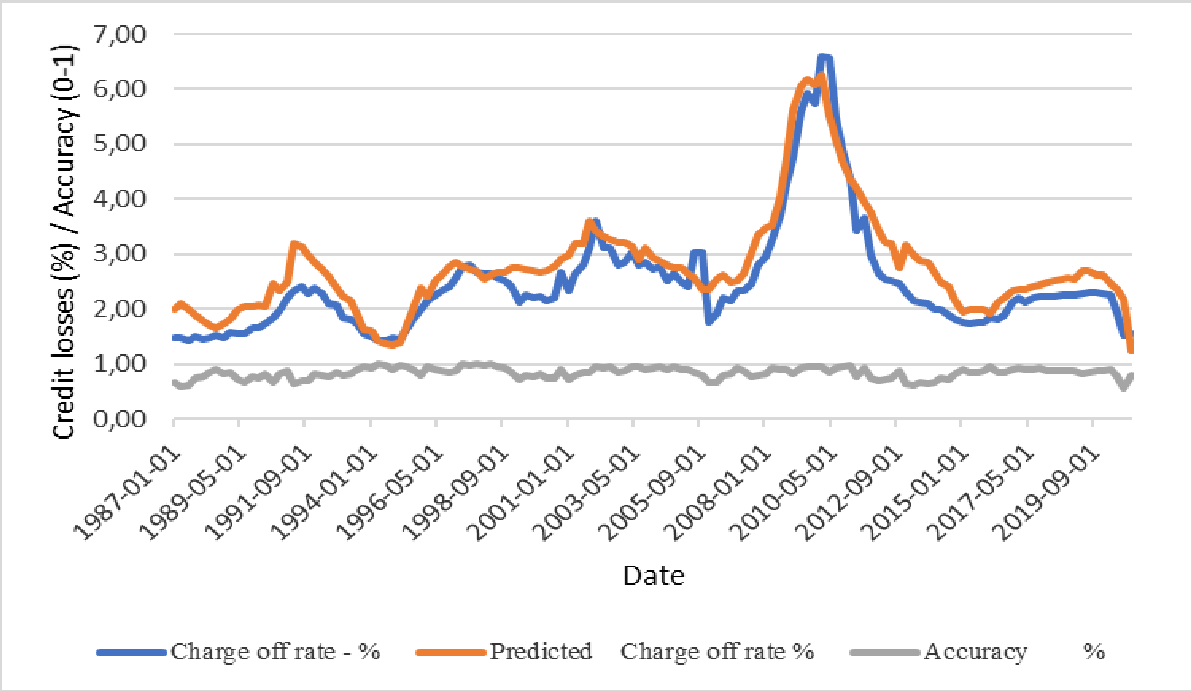


Figure 4.6: Back Testing Results for Model US2

(Data from: FRED, Federal Reserve Bank of St. Louis, US).

For Credit Risk Model US2, the line representing the estimated Charge Off Rate closely tracks that of the actual Charge Off Rate (Figure 4.6). However, it is generally above it, meaning that it generally overestimates rather than underestimate the credit losses. There was a significant decrease in credit losses during the Covid-19 period of 2020.

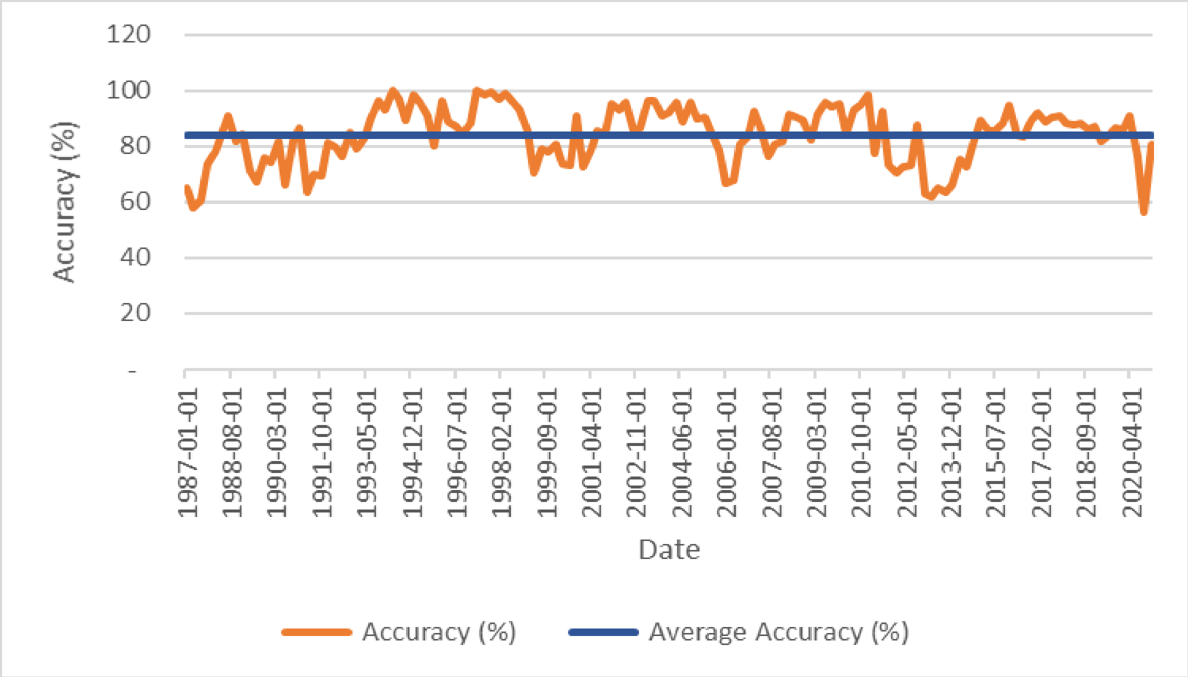


Figure 4.7: Back Testing Accuracy Results for Model US2

(Data from: FRED, Federal Reserve Bank of St. Louis, US).

For consumer credit risk model US2, its incidences of low estimation performances (less than 75%) were almost twice those of Model US1. It had two (2) years – 1987 and 2013 – with four (4) incidences each, one (1) year with three (3) incidences and 10 years having one (1) or two (2) incidences each (Figure 4.7).

4.2.8 Developed Market Consumer Credit Risk Models

For the developed US market, five (5) models resulting from correlation, bivariable regression and multivariable regressing analyses of the Loan Charge Off Rate (LCO) on the significant variables were considered; three (3) models that include the intercept or constant term, and two (2) models that do not include the intercept or constant term. For ease of reference, the models were named US1, US2, US3, US4 and US5 as shown in Table 4.11.

Table 4.11: Developed Market Consumer Credit Risk Models

S/N	**Explanatory Variables/ Analyses**	**Model US1**		**Model US2**		**Model US3**		**Model US4**		**Model US5**	
		Variable present	t Stat	Variable present	t Stat	Variable present	t Stat	Variable present	t Stat	Variable present	t Stat
3	Oil	√	-5,14	x		x		x		x	
7	Tilt	√	-2,34	√	-2,28	√	-0,27	√	0,00	√	5,98
6	T10Y	√	2,34	√	2,28	√	0,27	√	0,00	√	-2,86
8	T2Y	√	-2,34	√	-2,28	√	-0,27	√	0,00	x	
11	Unrate	√	8,54	√	7,62	√	0,91	√	2,12	x	
12	PCE	√	11,0	√	9,77	√	1,17	√	0,62	x	
13	PSR	√	7,88	√	-5,75	√	-0,69	√	-2,19	√	-2,79
14	HD	x		√	-2,78	√	-0,33	x		x	
15	CD	x		√	4,03	√	0,48	x		√	1,19
16	Drate	√	21,33	√	19,82	√	2,36	x		x	
N/A	Intercept	2.46		2.46						2.46	
N/A	R_a^2	91.2%		90.5%		6.5%		3.7%		33.1%	
N/A	Back Test Accuracy	85%		84%							

(Data from: FRED, Federal Reserve Bank of St. Louis, US).

Shown in Table 4.11 are five (5) models resulting from regressing the Charge Off Rate on all/some of the eight (8)/nine (9) significant explanatory variables. Those that remained significant at the confidence level of 95% are indicated by a t-statistic greater or equal to (≥) 1.96 in absolute terms while those that became insignificant have a t-statistic of less than (<) 1.96 in absolute terms in the new combination.

Models US1 and US2, with eight (8) and nine (9) significant variables, respectively have regression R-squared values of over 90%. Such high R-squared values typically imply that most (about 90%) of the variation in the Loan Charge Of Rate (LCO) is captured by the models. This is markedly higher than the well-published R-squared of 20 to 30 per cent range for LCD models (Baesens, 2015; Baesens, Rosch and Scheule, 2016). To establish how the variation may have arisen, further combinations of the explanatory variables and the predicted variable were studied.

In one of three such cases (see models US3, US4 and US5 in Table 4.11), selecting from the explanatory variables of models US1 and US2 which ones to include (generally high t-statistic, or low coefficient error variables) in the regression with the constant term, results in a goodness of fit of around 30% (see the R-squared value of model US5 in Table 4.11). Even though this R-squared value is characteristic of a poor regression model fit, it is on par with that found and published by other academics (Baesens, Rosch and Scheule, 2016; Barth *et al.*, 2018; Jin *et al.*, 2021). A closer look at the coefficients and t-statistics of Model US5 with the R-squared of 33.1%, reveals that the tilt variable is significant. The tilt is the difference between the 10-year (T10Y) and 2-year (T2Y) treasury yield. The tilt is important to financial institutions that borrow at the long 10-year treasury yield to fund at the shorter 2-year treasury yield. A multi-variable regression that incorporates both the tilt, and the 10-year (T10Y) may be subject to collinearity. Since the tilt variable remains significant (in Table 4.11, t-stat of 5.98 for Model US5 (selectively reduced variables)) when both the tilt and 10-year (T10Y) variables are included in the regression, this means that the multivariable regression adjusts in such a way that multi-collinearity problems are minimised. Another observation for Model US5 is that the CD variable became insignificant (its t-stat of 1.19 is less than 1.96) in the mix. Model US1 and model US2 – whose R-square values are significantly high at 90% range - on the other hand have all the explanatory variables being significant (all variables' t-statistics are greater than 1.96).

It is to be noted that Barth et al., (2018), in their study on the forecasting of Loan Charge Off Rates for two (2) small and two (2) large banks in the US market, did not seek to determine whether the variables that they used were significant – in combination - or not. Bolton (2009, p.53) points out that while it is essential to use significant variables in the model, there may be circumstances – like legal requirements – in which an explanatory variable is used even if it is not statistically significant. From the five (5) developed market credit risk models, it is observed that the two (2) models – US1 and US2 – with the high R-squared values, have all their variables in combination being significant (t-statistic of >1.96 in absolute terms).

Hence, it may be argued that reducing variables in such a way that all the variables in the model are significant in combination is important. It may therefore be concluded that a linear multivariable regression model with a constant intercept and significant variables is plausible for modelling credit risk.

For all consumer banks in the United States of America, it can be deduced that consumer credit risk models that would improve the estimation of the Charge Off Rate (LCO) or credit losses are:

$$\mathrm{LCO} = \beta_0 + \beta_1 \mathrm{OIL} + \beta_2 \mathrm{Tilt} + \beta_3 \mathrm{T10Y} + \beta_4 \mathrm{T2Y} + \beta_5 \mathrm{UNRATE} + \beta_6 \mathrm{PCE} + \beta_7 \mathrm{PSR} + \beta_8 \mathrm{DRATE} + \mathcal{E} \qquad \text{(Model US1)}$$

and,

$$\mathrm{LCO} = \beta_0 + \beta_1 \mathrm{Tilt} + \beta_2 \mathrm{T10Y} + \beta_3 \mathrm{T2Y} + \beta_4 \mathrm{UNRATE} + \beta_5 \mathrm{PCE} + \beta_6 \mathrm{PSR} + \beta_7 \mathrm{HD} + \beta_8 \mathrm{CD} + \beta_9 \mathrm{DRATE} + \mathcal{E} \qquad \text{(Model US2)}$$

where LCO is the Loan Charge Off Rate for the quarter, β_0 is the Charge Off Rate (intercept) when the values of all the explanatory variables are equal to zero. And $\beta_1 \dots \beta_9$ are the coefficients denoting the Charge Off Rate changes with respect to the explanatory variables in the quarter. $\mathcal{E}$ is the standard error that accounts for the variation in credit losses that the explanatory variables do not explain.

From Table 4.6 and Table 4.9, these figures are extracted to give the Consumer Credit Risk Models US1 and US2:

$$\mathrm{LCO} = 2{,}46 - 0{,}26 * \mathrm{OIL} - 1257{,}23 * \mathrm{Tilt} + 3320{,}41 * \mathrm{T10Y} - 3870{,}42 * \mathrm{T2Y} + 0{,}48 * \mathrm{UNRATE} + 1{,}18 * \mathrm{PCE} - 0{,}35 * \mathrm{PSR} + 0{,}87 * \mathrm{DRATE} + 0{,}30 \qquad \text{(Model US1)}$$

$$\mathrm{LCO} = 2{,}46 - 1269{,}80 * \mathrm{Tilt} + 3353{,}47 * \mathrm{T10Y} - 3909{,}06 * \mathrm{T2Y} + 0{,}46 * \mathrm{UNRATE} + 0{,}88 * \mathrm{PCE} - 0{,}28 * \mathrm{PSR} - 0{,}21 * \mathrm{HD} + 0{,}22 * \mathrm{CD} + 0{,}92 * \mathrm{DRATE} + 0{,}31 \qquad \text{(Model US2)}$$

Regression equations or Consumer Credit Risk Models US1 and US2 have statistically significant adjusted R-squared of 91.2% and 90.5%, respectively, and therefore statistically fit the credit losses well. Models US1 and US2 are similar, with the only difference being the number of explanatory variables each has. Model US2 has all the explanatory variables of Model US1, except the Oil Price, plus an additional two (Household Debt Service Payments as a percentage of Disposable Personal Income (HD), and Consumer Debt Service Payments as a percentage of Disposable Personal Income (CD)) and is fundamentally sound in that it resulted in the rate at which credit losses change with every unit change in the Personal Saving Rate (PSR), HD, and CD, being in the same range – 0,285, – 0,21, and 0,22 respectively (see Table 4.9). The full name of the consumer debt variable, according to the FRED, is the Seasonally Adjusted Quarterly Consumer Debt Service Payments as a percentage of Disposable Personal Income. Since the level of consumer debt service payments is one of the leading variables of the Loan Charge Off-Rates (DiGeorgia, 2001; Jareño and Negrut, 2016), a good credit risk regression model is expected to assign weight to this variable and PSR and HD in similar proportions. That the model also assigns a heavy weight of 0,88 to Personal Consumption Expenditure (PCE) and 0,92 to Delinquency Rate (DRATE) is to be expected. The ability of the borrowers to meet their debt obligations is directly dependent on how much of their personal income is spend and their delinquency rate is a leading indicator of the consumer credit losses.

Another facet of equation (Model) US2 is that all the significant variables other than the T10Y, UNRATE, PCE, CD, and DRATE are negatively weighted in the regression. This means that the T2Y treasury yield and the Personal Savings Rate (PSR) are countercyclical variables. From an economics perspective, when the Personal Saving Rate increases, the economy of the country is generally healthier, and as a result, the Loan Charge Off Rates generally decreases. Similarly, when the 2-year treasury yield increases, it means that related fixed-income market bond prices decrease. This means that fixed-income markets could be in distress, and with the pass-through in inflation to the consumer market, this could lead to an increase in the Loan Charge Off-Rate. The countercyclical response of the significant T2Y and the PSR variables in the regression equation (Model) US2 is therefore, fundamentally sound. For all practical purposes equation (Model) US2 is sound from a mathematical economics perspective. As aforementioned, Model US1 is similar to Model US2 and therefore, the same conclusions can be drawn with respect to the model.

4.3 EMERGING MARKET OF SOUTH AFRICA ANALYSES

In the previous section, five (5) models – US1, US2, US3, US4 and US5 built using data from the developed market of the US were presented. US5 model had a R-squared value that conforms to the academically published fit quality. However, one of its variables – Consumer Debt Service Payments as a percentage of Disposable Personal Income (CD) – was found to be insignificant (having a t-statistic (of 1.19) which is less than 1.96 in absolute terms). Models US3 and US4 had more variables falling in the insignificant class and, consequently poorer R-squared values. Parameters of Model US1 and Model US2 and sensibility were arguably economically sound. In this section, the thought leadership and techniques established in the developed market consumer credit risk models is leveraged to derive a credit risk model representative of the emerging market of South Africa. For ease of reference, the South African consumer credit risk models were named SA1, SA2, SA3 and SA4. Model SA1 is a result of regressing Impairments on nine (9) significant explanatory variables while Model SA2 is a refined form of Model SA1. The two models (SA1 and SA2) are described in Sections 4.3.2 and 4.3.3, respectively. Models SA3 and SA4 are additional models obtained through similar regression analyses of credit losses on lagging/leading explanatory variables as explained in Section 4.3.4.

4.3.1 Description of Dependent/Independent Variables

For the emerging market of South Africa consumer credit risk modelling, as stated in Section 3.3.2, 20 independent variables were selected. Impairments as a percentage of gross loans for South Africa were used as a proxy for realised credit losses (or the equivalent of Loan Charge Off Rates used in the developed market analyses of Section 4.2). These variables are shown and described in Table 4.12 for ease of reference.

Table 4.12: Description of Variables

S/N	Dependent / Independent variable	Short Name	Description	Units	Source
	Impairments %	I-mts	Advances Impairments to Gross Loans and Advances	Percent	WB/SARB
1	CPI growth %	CPIG	Consumer Price Index: Total All Items for South Africa	Growth Rate Previous Period	OECD
2	ZAR/$ Forex	FX	South Africa / U.S. Foreign Exchange Rate	South African Rand to One U.S. Dollar	FED
3	Interest TB %	ITB	Interest Rates, Government Securities, Treasury Bills for South Africa	Percent per Annum	IMF

S/N	Dependent / Independent variable	Short Name	Description	Units	Source
4	Interest IBR %	IBR	3-Month or 90-day Rates and Yields: Interbank Rates for South Africa	Percent	OECD
5	Interest 10Y T %	T10Y	Long-Term Government Bond Yields: 10-year: Main (Including Benchmark)	Percent	OECD
6	10 Y 3M	10Y3M	10Y – T3M Tilt of yield curve = funding costs = inflation expectation	Percent	FED
7	10 Y TB	10YTB	Tilt of yield curve = short term funding costs	Percent	FED
8	Oil Price US$	OIL	Crude Oil Prices: West Texas Intermediate (WTI) Cushing, Oklahoma	Dollars per Barrel	USEIA
9	Share price growth %	SPG	ALSI Total Share Prices for All Shares for South Africa	Growth Rate Previous Period	OECD
10	Share PI	SPI	ALSI Total Share Prices for All Shares for South Africa (pegged to 2015)	Index 2015=100	OECD
11	CPI	CPI	Consumer Price Index: All Items for South Africa (pegged to 2015)	Index 2015=100	OECD
12	P-car reg. index	VEH	Passenger Car Registrations in South Africa (pegged to 2015)	Index 2015=100	OECD
13	Car reg. growth rate %	VEHG	Sales: Retail trade: Car registration: Passenger cars for South Africa	Growth rate previous period	OECD
14	Disposable income/ Income	Income	Disposable income as a % of total income	Percent	OECD
15	Gold Price US$/ounce	GOLD	Producer Price Index by Industry: Gold Ore Mining: Gold Ores (pegged to 1985)	Index Jun 1985=100	USBLS
16	Unemployment Rate %	UNRATE	Unemployment Rate: Aged 15-64: All Persons for South Africa	Percent	OECD
17	Real GDP ZAR	GDP	Real Gross Domestic Product for South Africa	Domestic Currency	IMF
18	Res. Property PI 2010=100	RPPI	Real Residential Property Prices for South Africa (pegged to 2010)	Index 2010=100	BIS

S/N	Dependent / Independent variable	Short Name	Description	Units	Source
19	World Uncertainty Index	WUI (VOL)	World Uncertainty Index for South Africa	Index	BAF[9]
20	Constant Price GDP SA % Change	CPGDP	Constant Price Gross Domestic Product in South Africa	Percent Change from Year Ago	OECD

(Source: BIS, BAF, FED, IMF, OECD, SARB, USBLS, USEIA, and WB).

In Table 4.12, the 20 independent explanatory variables used in the emerging market of South Africa consumer credit risk modelling, are described together with the dependent variable (Impairments).

4.3.2 Consumer Credit Risk Model SA1

In this analysis, the Charge Off Rate (or realised losses) was represented by Impairments. The analysis started off with a population of 20 explanatory variables shown in Table 4.12. When correlation analysis between Impairments and the 20 explanatory variables was done, the results were the correlation matrix of Table 4.13. The table was divided into two parts for ease of reference. From the Table, it is observed that values, in absolute terms, for correlations between Impairments and CPI Inflation Growth, 10Y Treasury Yields, Car Registration Growth, and Unemployment Rate are low: -8%, 1%, 5% and 2% respectively. On this basis, these variables were therefore dropped. Correlations between Impairments and World Uncertainty Index, Real GDP, and Share Price Index are the three highest at -62%, -52%, and -49%, respectively. While the negative correlations between Impairments and Real GDP as well as Share Price Index is expected, the negative correlation with World Uncertainty Index (WUI) is opposite to expectation. Impairments are expected to decrease with an improving GDP outlook as well as improving share prices. The negative correlation between Impairments and WUI may imply that when credit consumers perceive that the credit market outlook is uncertain, they are more inclined to deleverage. The other observation from Table 4.13 is that several explanatory variables are strongly correlated (with correlations greater than 80%), and this may lead to multicollinearity – the notion of one variable in a multiple-step multivariable regression analysis, explaining most or all of the variation in the dependent variable that another variable that is strongly correlated with it explains, making one of the variables redundant and therefore missing in the regression model. Such a variable may therefore appear to be significant in one phase of a multi-step multivariable regression analysis and cease to be significant in subsequent analysis. This is discussed in detail in Section 4.3.2.1.

Table 4.13: Correlations Amongst Impairments and Explanatory Variables for South Africa

	S/N		1	2	3	4	5	6	7	8	9	10
S/N	**Dependent / Independent variable**	***Impairments %***	***CPI growth %***	***ZAR/$ Forex***	***Interest TB %***	***Interest IBR %***	***Interest 10Y T %***	***10 Y 3M %***	***10 Y TB %***	***Oil Price US$***	***Share price growth %***	***Share PI***
	Impairments %	100%										
1	CPI growth %	-8%	100%									
2	ZAR/$ Forex	-42%	-21%	100%								
3	Interest TB %	-42%	20%	-16%	100%							
4	Interest IBR %	-42%	20%	-17%	100%	100%						
5	Interest 10Y T %	1%	-1%	54%	14%	14%	100%					
6	10 Y 3M	41%	-20%	39%	-91%	-91%	29%	100%				
7	10 Y TB	41%	-20%	39%	-90%	-90%	30%	100%	100%			
8	Oil Price US$	14%	27%	-77%	5%	5%	-37%	-21%	-21%	100%		
9	Share price growth %	18%	-15%	-3%	-28%	-28%	-14%	21%	21%	-1%	100%	
10	Share PI	-49%	-17%	84%	-33%	-33%	29%	44%	44%	-48%	8%	100%
11	CPI	-30%	-24%	91%	-35%	-35%	45%	53%	53%	-62%	2%	93%
12	P-car reg. index	-41%	18%	-23%	-12%	-12%	-62%	-15%	-15%	46%	-6%	16%
13	Car reg. growth rate %	5%	-20%	17%	-13%	-12%	13%	18%	18%	-13%	9%	5%
14	Disposable income/Total income	33%	14%	-58%	27%	28%	-11%	-31%	-31%	40%	6%	-80%
15	Gold Price US$/ounce	31%	-14%	23%	-80%	-80%	5%	79%	80%	-3%	17%	35%
16	Unemployment Rate %	-2%	-8%	67%	-40%	-40%	42%	56%	56%	-47%	3%	75%
17	Real GDP ZAR	-52%	-24%	63%	-23%	-23%	4%	24%	24%	-33%	-10%	83%
18	Res. Property PI 2010=100	-44%	31%	-23%	70%	71%	4%	-67%	-66%	20%	-11%	-29%
19	World Uncertainty Index	-62%	1%	43%	16%	16%	11%	-10%	-10%	-28%	-1%	47%
20	Constant Price GDP SA % Change	-15%	25%	-60%	27%	27%	-53%	-49%	-49%	58%	-20%	-28%

	S/N	11	12	13	14	15	16	17	18	19	20
S/N	**Dependent / Independent variable**	***CPI***	***P-car reg. index***	***Car reg. growth rate %***	***Income/ Disposable income***	***Gold Price US$/ ounce***	***Unemployment Rate %***	***Real GDP ZAR***	***Res. Property PI 2010=100***	***World Uncertainty Index***	***Constant Price GDP SA % Change***
11	CPI	100%									
12	P-car reg. index	-11%	100%								
13	Car reg. growth rate %	12%	-29%	100%							
14	Income / Disposable income	-75%	-33%	19%	100%						
15	Gold Price US$/ounce	45%	4%	12%	-28%	100%					
16	Unemployment Rate %	83%	-7%	-9%	-72%	53%	100%				
17	Real GDP ZAR	75%	42%	-16%	-91%	24%	61%	100%			
18	Res. Property PI 2010=100	-43%	9%	-7%	42%	-63%	-41%	-32%	100%		
19	World Uncertainty Index	39%	25%	-1%	-49%	0%	21%	49%	11%	100%	
20	Constant Price GDP SA % Change	-48%	71%	-45%	-12%	-29%	-25%	9%	28%	7%	100%

(Data from FRED, Federal Reserve Bank of St. Louis, US, and the South African Reserve Bank).

4.3.2.1 Multivariable Regression Analyses

Multivariable regression analysis was done to ascertain how much the remaining sixteen explanatory variables explain variations in Impairments. The results of the analysis are shown in Table 4.14. P-values and t-statistic for the Treasury Bills Rates, Interbank Interest Rate and Disposable Income as a Percentage of Total Income are greater than 0,05 and less than 1.96 (in absolute terms), respectively. This means that these variables are not significant in this combination and were dropped. Regression analysis was similarly done a second, third and fourth time as shown in the results displayed in Table 4.15, Table 4.16 and Table 4.17, respectively. As a result of the second and third regression analyses, four explanatory variables – 10Y3M, 10YTB, SPI and CPI were dropped as they became insignificant in the combinations. In the final multivariable regression analysis, standardised values were used. Table 4.17 shows the final result in which all p-values and t-statistics are less than 0,05 and greater than 1,96 (in absolute terms), respectively. From this, Consumer Credit Risk Model SA1 was derived as follows:

$$LCO = \beta_0 + \beta_1 FX + \beta_2 OIL + \beta_3 SPG + \beta_4 VEH + \beta_5 GOLD + \beta_6 RGDP + \beta_7 RPI + \beta_8 WUI + \beta_9 CPGDP + \in$$

$$LCO = 4{,}05 - 0{,}41 * FX - 0{,}23 * OIL + 0{,}08 * SPG - 0{,}38 * VEH + 0{,}29 * GOLD - 0{,}28 * RGDP - 0{,}41 * RPI - 0{,}26 * WUI + 0{,}27 * CPGDP + 0{,}44$$

where LCO is the Loan Charge Off Rate (Impairments); FX is the USD/ZAR Foreign Exchange Rate; OIL is the crude Oil Price; SPG is the Share Price Growth; VEH is the Passenger Car Registrations; GOLD is the Gold Price; RDGP is the Real Gross Domestic Product; RPI is the Residential Price Index; WUI is the World Uncertainty Index; CGDP is the Constant Gross Domestic Product and ∈ is the Standard Error.

Table 4.14: Emerging Market of South Africa Multiple Regression Analysis: Impairments and 16 Explanatory Variables

Regression Statistics	
Multiple R	0,95
R Square	0,91
Adjusted R Square	0,90
Standard Error	0,32
Observations	159,00

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	16,00	154,16	9,64	98,42	0,00
Residual	143,00	14,93	0,10		
Total	159,00	169,09			

		Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*
	Intercept	19,39	3,25	5,97	0,00	12,96
2	ZAR/$ Forex	-0,20	0,04	-4,53	0,00	-0,28
3	Interest TB %	-0,18	0,36	-0,50	0,62	-0,90
4	Interest IBR %	0,23	0,36	0,62	0,53	-0,49
6	10 Y 3M	0,39	0,08	4,86	0,00	0,23
7	10 Y TB	-	-	65 535,00	#NUM!	-
8	Oil Price US$	-0,01	0,00	-2,71	#NUM!	-0,02
9	Share price growth %	0,03	0,01	3,40	0,00	0,01
10	Share PI	-0,03	0,01	-5,16	0,00	-0,05
11	CPI	0,04	0,01	4,02	0,00	0,02
12	P-car reg. index	-0,01	0,01	-2,04	0,04	-0,02
14	Income/ Disposable income	-0,05	0,05	-0,96	0,34	-0,15
15	Gold Price US$/ounce	-0,00	0,00	-1,97	0,05	-0,00
17	Real GDP ZAR	-0,00	0,00	-3,36	0,00	-0,00
18	Res. Property PI 2010=100	-0,06	0,02	-2,91	0,00	-0,10
19	World Uncertainty Index	-0,45	0,09	-4,76	0,00	-0,63
20	Constant Price GDP SA % Change	0,08	0,03	2,84	0,01	0,02

(Data from: FRED, Federal Reserve Bank of St. Louis, US, and the South African Reserve Bank).

In Table 4.14, P-values and t-statistics for Interest – TB, Interest – IBR and Disposable income/Total income (DI/TI) are greater than 0,05 and less than 1.96 (in absolute terms), respectively. These variables were subsequently dropped for being insignificant in this combination.

Table 4.15: Emerging Market of South Africa Multivariable Regression Analysis: Impairments and 13 Explanatory Variables

Regression Statistics	
Multiple R	0,95
R Square	0,91
Adjusted R Square	0,90
Standard Error	0,32
Observations	159

ANOVA

	df	***SS***	***MS***	***F***	***Significance F***
Regression	13	153,97	11,84	113,5	0,00
Residual	145	15,13	0,10		
Total	158	169,09			

		Coefficients	***Standard Error***	***t Stat***	***P-value***	***Lower 95%***	***Upper 95%***
S/N	Intercept	16,94	2,21	7,67	0,000	12,57	21,31
2	ZAR/$ Forex	-0,20	0,04	-5,23	0,000	-0,28	-0,13
6	10 Y 3M	0,18	0,36	0,51	0,613	-0,52	0,88
7	10 Y TB	0,20	0,36	0,55	0,586	-0,51	0,91
8	Oil Price US$	-0,01	0,00	-4,63	0,000	-0,01	-0,01
9	Share price growth %	0,03	0,01	3,41	0,001	0,01	0,04
10	Share PI	-0,03	0,01	-5,76	0,000	-0,04	-0,02
11	CPI	0,05	0,01	5,12	0,000	0,03	0,07
12	P-car reg. index	-0,01	0,00	-2,94	0,004	-0,02	-0,00
15	Gold Price US$/ounce	-0,00	0,00	-2,72	0,007	-0,00	-0,00
17	Real GDP ZAR	-0,00	0,00	-3,29	0,001	-0,00	-0,00
18	Res. Property PI 2010=100	-0,06	0,01	-4,12	0,000	-0,09	-0,03
19	World Uncertainty Index	-0,40	0,09	-4,60	0,000	-0,57	-0,23
20	Constant Price GDP SA % Change	0,11	0,02	5,01	0,000	0,06	0,15

(Data from: FRED, Federal Reserve Bank of St. Louis, US, and the South African Reserve Bank).

In Table 4.15, P-values and t-statistics for 10Y 3M and 10Y TB are greater than 0,05 and less than 1.96 (in absolute terms) respectively. These variables were subsequently dropped as this means that they are now insignificant.

Table 4.16: Emerging Market of South Africa Multivariable Regression Analysis: Impairments and 11 Explanatory Variables

	Regression Statistics						
	Multiple R	0,91					
	R Square	0,83					
	Adjusted R Square	0,82					
	Standard Error	0,44					
	Observations	159					
	ANOVA						
		df	***SS***	***MS***	***F***	***Significance F***	
	Regression	11	140,32	12,76	65,165	0,00	
	Residual	147	28,78	0,20			
	Total	158	169,09				
		Coefficient	***Standard Error***	***t Stat***	***P-value***	***Lower 95%***	***Upper 95%***
	Intercept	24,08	2,90	8,30	0,000	18,35	29,82
2	ZAR/$ Forex	-0,15	0,05	-2,86	0,005	-0,25	-0,05
8	Oil Price US$	-0,01	0,00	-3,27	0,001	-0,01	-0,00
9	Share price growth %	0,02	0,01	2,20	0,029	0,00	0,04
10	Share PI	-0,01	0,01	-1,15	0,251	-0,02	0,01
11	CPI	0,02	0,01	1,36	0,175	-0,01	0,04
12	P-car reg. index	-0,02	0,01	-3,63	0,000	-0,03	-0,01
15	Gold Price US$/ounce	0,00	0,00	4,24	0,000	0,00	0,00
17	Real GDP ZAR	-0,00	0,00	-2,86	0,005	-0,00	-0,00
18	Res. Property PI 2010=100	-0,12	0,02	-6,55	0,000	-0,16	-0,09
19	World Uncertainty Index	-0,66	0,11	-5,80	0,000	-0,88	-0,43
20	Constant Price GDP SA % Change	0,08	0,03	2,83	0,005	0,02	0,14

(Data from: FRED, Federal Reserve Bank of St. Louis, US, and the South African Reserve Bank).

In Table 4.16, P-values and t-statistics for Share PI and CPI are greater than 0,05 and less than 1.96 (in absolute terms) respectively. These variables were subsequently dropped as this means that they are in significant in this combination.

Table 4.17: Emerging Market of South Africa Multivariable Regression Analysis (Impairments and 9 Explanatory Variables)

Regression Statistics	
Multiple R	0,91
R Square	0,83
Adjusted R Square	0,82
Standard Error	0,44
Observations	159

ANOVA

	df	***SS***	***MS***	***F***	***Significance F***
Regression	9	139,95	15,55	79,50	0,00
Residual	149	29,14	0,20		
Total	158	169,09			

		Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*
	Intercept	25,09658	2,69	9,32	0,00	19,77
2	ZAR/$ Forex	-0,12896	0,04	-3,23	0,00	-0,21
8	Oil Price US$	-0,00934	0,00	-3,45	0,00	-0,01
9	Share price growth %	0,01955	0,01	2,03	0,04	0,00
12	P-car reg. index	-0,02517	0,00	-5,63	0,00	-0,03
15	Gold Price US$/ounce	0,00110	0,00	5,77	0,00	0,00
17	Real GDP ZAR	-0,00001	0,00	-2,88	0,00	-0,00
18	Res. Property PI 2010=100	-0,13125	0,02	-7,21	0,00	-0,17
19	World Uncertainty Index	-0,66163	0,11	-5,83	0,00	-0,89
20	Constant Price GDP SA % Change	0,08273	0,03	2,90	0,00	0,03

(Data from: FRED, Federal Reserve Bank of St. Louis, US, and the South African Reserve Bank).

In Table 4.17, P-values and t-statistics for all nine explanatory variables are less than 0,05 and greater than 1.96 (in absolute terms) respectively. This means that they are all significant in this combination and collectively explain 82 % of the variation in impairments (see adjusted R square).

4.3.2.2 Back Testing Results for Consumer Credit Risk Model SA1

Consumer Credit Risk Model SA1 was back tested, and the test results were as shown in Table 4.18, Figure 4.8 and Figure 4.9 . The actual monthly credit losses represented by Impairments and the corresponding credit loss values estimated by the model are shown in the second and third columns of Table 4.18, respectively. Shown on the fourth column are the calculated respective accuracies. The same information is shown in graphical representation in Figure 4.8 with “Accuracy” on a scale of 0 to 1 instead of 0 to 100 (per cent). In Figure 4.9, “Accuracy” is shown on its own on a scale of 0 to 100 (per cent). This model generally overestimated the credit losses with accuracies that varied from as low as 54% to as high as 100%, with an average of 86% over the period 1987 to 2001. The estimated credit losses closely tracked the actual credit losses. The low estimation performances (less than 75%) are concentrated mainly in 2008, 2017, and 2018 with no more than two (2) incidences in each of 2009, 2016, and 2020. The incidences of low estimation accuracy seem to have no particular pattern.

Table 4.18: Back Testing Results for Consumer Credit Risk Model SA1

DATE	Impairments %	Estimated Impairments	Accuracy %
2008-01-01	2,30	2,31	99
2008-02-01	2,30	2,64	85
2008-03-01	2,40	2,38	99
2008-04-01	2,40	3,21	66
2008-05-01	2,60	3,32	72
2008-06-01	2,70	3,09	86
2008-07-01	2,80	3,83	63
2008-08-01	3,00	4,07	64
2008-09-01	3,10	4,12	67
2008-10-01	3,42	4,03	82
2008-11-01	3,60	4,43	77
2008-12-01	3,42	4,99	54
2009-01-01	4,20	5,58	67
2009-02-01	4,60	5,59	78

DATE	Impairments %	Predicted Impairments	Accuracy %
2011-05-01	5,73	5,96	96
2011-06-01	5,56	5,95	93
2011-07-01	5,49	5,76	95
2011-08-01	5,29	5,73	92
2011-09-01	5,06	5,81	85
2011-10-01	4,90	5,39	90
2011-11-01	4,81	5,39	88
2011-12-01	4,69	5,16	90
2012-01-01	4,73	5,44	85
2012-02-01	4,71	5,44	85
2012-03-01	4,58	5,25	85
2012-04-01	4,64	5,19	88
2012-05-01	4,56	5,09	88
2012-06-01	4,47	5,20	84

DATE	Impairments %	Predicted Impairments	Accuracy %
2014-09-01	3,35	4,03	80
2014-10-01	3,35	3,50	96
2014-11-01	3,27	3,81	84
2014-12-01	3,28	3,72	86
2015-01-01	3,28	4,14	74
2015-02-01	3,23	4,21	70
2015-03-01	3,21	3,98	76
2015-04-01	3,22	3,91	79
2015-05-01	3,26	3,78	84
2015-06-01	3,31	3,76	86
2015-07-01	3,33	3,39	98
2015-08-01	3,27	3,38	97
2015-09-01	3,20	3,29	97
2015-10-01	3,24	3,55	90

DATE	Impairments %	Estimated Impairments %	Accuracy %
2018-01-01	3,08	4,48	55
2018-02-01	3,10	4,39	58
2018-03-01	3,24	4,38	65
2018-04-01	3,31	4,16	74
2018-05-01	3,39	3,98	83
2018-06-01	3,50	3,79	92
2018-07-01	3,60	3,75	96
2018-08-01	3,57	3,68	97
2018-09-01	3,66	3,49	95
2018-10-01	3,72	3,74	99
2018-11-01	3,74	4,09	91
2018-12-01	3,73	4,21	87
2019-01-01	3,85	4,09	94
2019-02-01	3,81	4,06	93

DATE	Impairments %	Estimated Impairments	Accuracy %
2009-03-01	4,80	5,53	85
2009-04-01	5,10	5,82	86
2009-05-01	5,40	5,83	92
2009-06-01	5,50	5,67	97
2009-07-01	5,60	5,83	96
2009-08-01	5,80	5,91	98
2009-09-01	5,90	5,91	100
2009-10-01	5,84	5,85	100
2009-11-01	5,94	5,86	99
2009-12-01	5,94	5,98	99
2010-01-01	5,86	5,83	99
2010-02-01	5,82	5,66	97
2010-03-01	5,88	5,82	99
2010-04-01	5,95	5,62	94
2010-05-01	5,91	5,44	92

DATE	Impairments %	Predicted Impairments	Accuracy %
2012-07-01	4,43	4,85	90
2012-08-01	4,36	4,87	88
2012-09-01	4,29	4,93	85
2012-10-01	4,28	5,26	77
2012-11-01	4,06	5,19	72
2012-12-01	4,04	5,32	68
2013-01-01	4,09	5,11	75
2013-02-01	4,06	5,09	75
2013-03-01	4,01	4,95	76
2013-04-01	4,01	4,46	89
2013-05-01	4,01	4,51	88
2013-06-01	3,94	4,21	93
2013-07-01	3,95	4,34	90
2013-08-01	3,83	4,46	84
2013-09-01	3,71	4,54	78

DATE	Impairments %	Predicted Impairments	Accuracy %
2015-11-01	3,08	3,18	97
2015-12-01	3,12	3,01	96
2016-01-01	3,08	2,88	93
2016-02-01	3,16	3,37	93
2016-03-01	3,38	3,59	94
2016-04-01	3,13	3,20	98
2016-05-01	3,17	3,04	96
2016-06-01	3,17	3,18	100
2016-07-01	3,15	3,73	82
2016-08-01	3,17	3,86	78
2016-09-01	3,24	3,78	83
2016-10-01	2,91	3,66	74
2016-11-01	2,85	3,73	69
2016-12-01	2,87	3,57	76
2017-01-01	2,89	3,65	74

DATE	Impairments %	Estimated Impairments %	Accuracy %
2019-03-01	3,77	3,87	97
2019-04-01	3,79	4,08	92
2019-05-01	3,83	3,92	98
2019-06-01	3,73	4,05	91
2019-07-01	3,73	3,92	95
2019-08-01	3,80	3,79	100
2019-09-01	3,83	3,90	98
2019-10-01	3,76	3,71	99
2019-11-01	3,79	3,83	99
2019-12-01	3,89	3,77	97
2020-01-01	3,98	4,91	77
2020-02-01	3,96	4,80	79
2020-03-01	4,04	5,01	76
2020-04-01	4,27	6,13	57
2020-05-01	4,59	5,38	83

DATE	Impairments %	Estimated Impairments	Accuracy %
2010-06-01	5,91	5,70	96
2010-07-01	5,84	5,82	100
2010-08-01	5,86	5,69	97
2010-09-01	5,88	6,10	96
2010-10-01	5,95	6,03	99
2010-11-01	5,81	5,93	98
2010-12-01	5,79	5,91	98
2011-01-01	5,82	5,68	98
2011-02-01	5,81	5,49	95
2011-03-01	5,78	5,42	94
2011-04-01	5,79	5,86	99

DATE	Impairments %	Predicted Impairments	Accuracy %
2013-10-01	3,69	4,16	87
2013-11-01	3,65	4,15	86
2013-12-01	3,64	3,98	91
2014-01-01	3,59	4,00	89
2014-02-01	3,57	3,96	89
2014-03-01	3,51	4,03	85
2014-04-01	3,54	4,04	86
2014-05-01	3,57	4,08	86
2014-06-01	3,42	3,88	86
2014-07-01	3,45	4,21	78
2014-08-01	3,42	4,12	80

DATE	Impairments %	Predicted Impairments	Accuracy %
2017-02-01	2,84	3,85	64
2017-03-01	2,84	4,06	57
2017-04-01	2,89	3,87	66
2017-05-01	2,90	3,85	67
2017-06-01	2,90	3,77	70
2017-07-01	2,85	3,74	69
2017-08-01	2,83	3,73	68
2017-09-01	2,77	3,68	67
2017-10-01	2,79	3,32	81
2017-11-01	2,81	3,22	85
2017-12-01	2,84	3,25	85

DATE	Impairments %	Estimated Impairments %	Accuracy %
2020-06-01	4,89	4,69	96
2020-07-01	5,01	5,11	98
2020-08-01	4,99	5,04	99
2020-09-01	5,03	4,96	99
2020-10-01	5,03	5,10	99
2020-11-01	5,03	5,07	99
2020-12-01	5,16	5,03	98
2021-01-01	5,24	5,56	94
2021-02-01	5,19	5,40	96
2021-03-01	5,12	5,06	99
Average			86

(Data from: FRED, Federal Reserve Bank of St. Louis, US, and the South African Reserve Bank).

Consumer Credit Risk Model SA1 generally overestimates credit losses. Its accuracy varies from as low as 54% to as high as 100%. The average accuracy is 86% over the period 2008 – 2021 (Table 4.18).

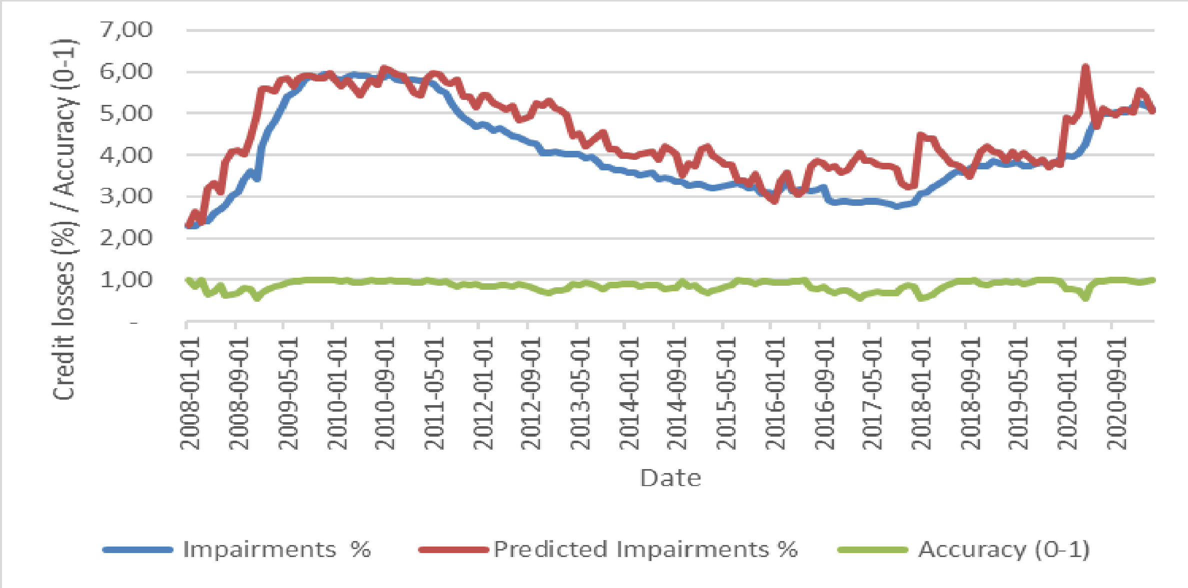


Figure 4.8: Back Testing Results for Consumer Credit Risk Model SA1

(Data from: FRED, Federal Reserve Bank of St. Louis, US, and the South African Reserve Bank).

As shown in Figure 4.8, for Consumer Credit Risk Model SA1, the line representing the estimated Impairments closely tracks that of actual Impairments. However, it is generally above it, meaning that it generally overestimates rather than underestimate the credit losses. There was a significant increase in credit losses during the Covid-19 period of 2020[10].

[10] There was a significant increase in credit losses during the Covid-19 period of 2020 due to lockdowns and other measures with adverse effects on the economy imposed by the government. The likelihood that bank borrowers would not be able to pay their loans consequently increased (Brouwer, Huttenhuis and ter Hoeven, 2021).

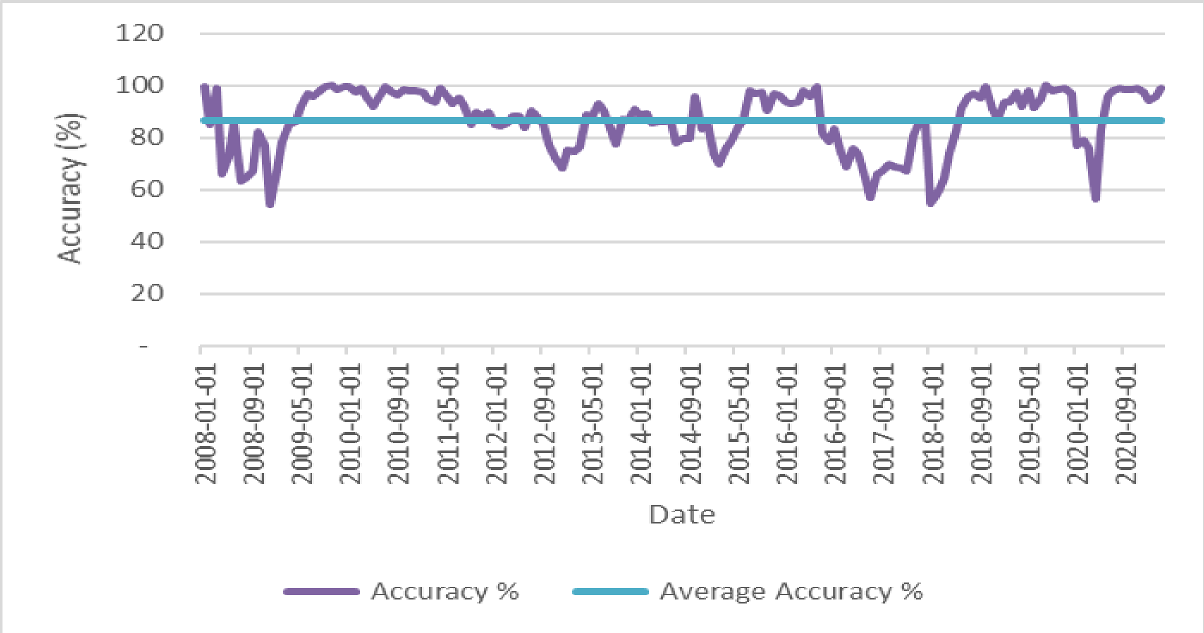


Figure 4.9: Back Testing Accuracy Results for Consumer Credit Risk Model SA1

(Data from: FRED, Federal Reserve Bank of St. Louis, US, and the South African Reserve Bank).

As shown in Figure 4.9, for Consumer Credit Risk Model SA1, the low estimation performances (less than 75%) are concentrated mainly in 2008, 2017, and 2018 with no more than two (2) incidences in each of 2009, 2016, and 2020.

4.3.3 Consumer Credit Risk Model SA2

The nine (9) significant variables of Consumer Credit Risk Model SA1 were given a lead of one (1) to 10 months, and the correlations between the corresponding Impairments and the led values were analysed. The results of the analysis are shown in Table 4.19. The values shaded in black show the highest correlations between Impairments and the nine (9) explanatory variables. Values that do not differ, with the shaded ones with more than ±3% are assumed to be the same. The shaded values are in the most suitable normal/lead periods. Regression analysis of Impairments on the normal/lead explanatory variables was done, and this led to another three (3) – Share Price Growth, Gold Price, and Real GDP – explanatory variables being dropped as they were no longer significant (see Table 4.20). The loss of significance may be due to multicollinearity – the notion of one variable explaining most of what another variable explains, rendering one of them redundant in the model.

Table 4.19: Consumer Credit Risk Model SA2: Correlations Between Impairments and Nine (9) Normal/Lead Explanatory Variables

	Status quo/Lead (months)	**Status quo**	**+1**	**+2**	**+3**	**+4**	**+5**	**+6**	**+7**	**+8**	**+9**	**+10**
S/N	**Dependent / Independent variable**	***Impairments***										
	Impairments %	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
2	ZAR/$ Forex	-42%	-42%	-42%	-42%	-42%	-43%	-43%	-44%	-45%	-46%	-47%
8	Oil Price US$	14%	11%	9%	7%	7%	7%	7%	8%	10%	11%	13%
9	Share price growth %	18%	17%	15%	11%	8%	6%	6%	1%	0%	-3%	-6%
12	P-car reg. index	-41%	-46%	-51%	-55%	-58%	-60%	-60%	-64%	-65%	-66%	-66%
15	Gold Price US$/ounce	31%	26%	22%	16%	11%	6%	6%	-6%	-12%	-17%	-22%
17	Real GDP ZAR	-52%	-55%	-58%	-61%	-63%	-65%	-65%	-67%	-68%	-68%	-68%
18	Res. Property PI 2010=100	-44%	-42%	-38%	-35%	-30%	-26%	-26%	-16%	-11%	-5%	1%
19	World Uncertainty Index	-62%	-63%	-64%	-65%	-65%	-64%	-64%	-61%	-59%	-58%	-55%
20	Constant Price GDP SA % Change	-15%	-18%	-20%	-20%	-20%	-20%	-20%	-20%	-19%	-17%	-15%

Source: Adopted from FRED, Federal Reserve Bank of St. Louis, US

Table 4.20: Results of Multivariable Regression Analysis of Impairments and Nine (9) Normal/Lead Explanatory Variables

Regression Statistics	
Multiple R	0,92
R Square	0,85
Adjusted R Square	0,84
Standard Error	0,40
Observations	151

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	9	127,60	14,18	89,21	0,00
Residual	141	22,41	0,16		
Total	150	150,01			

		Coefficients	***Standard Error***	***t Stat***	***P-value***	***Lower 95%***	***Upper 95%***
	Intercept	30,851	4,48	6,88	0,00	21,99	39,72
2	ZAR/$ Forex	-0,200	0,02	-8,52	0,00	-0,25	-0,15
8	Oil Price US$	-0,005	0,00	-2,11	0,04	-0,01	-0,00
9	Share price growth %	0,007	0,01	0,80	0,42	-0,01	0,02
12	P-car reg. index	-0,033	0,00	-8,21	0,00	-0,04	-0,02
15	Gold Price US$/ounce	0,000	0,00	0,33	0,75	-0,00	0,00
17	Real GDP ZAR	-0,000	0,00	-1,48	0,14	-0,00	0,00
18	Res. Property PI 2010=100	-0,194	0,04	-5,00	0,00	-0,27	-0,12
19	World Uncertainty Index	-0,332	0,11	-3,07	0,00	-0,55	-0,12
20	Constant Price GDP SA % Change	-0,047	0,01	-3,34	0,00	-0,07	-0,02

(Data from: FRED, Federal Reserve Bank of St. Louis, US, and the South African Reserve Bank).

As shown in Table 4.20, p-values and t-statistics for Share Price Growth, Gold Price and Real GDP are greater than 0,05 and less than 1.96 (in absolute terms), respectively. These variables were subsequently dropped as this means that they are not significant in this combination. The remaining six (6) explanatory variables – Foreign Exchange Rate (leading by eight (8) months, Oil Price (normal), Passenger Car Registration (leading by seven (7) months), Residential Property Price Index (normal), World Uncertainty Index or Volatility (normal), and Constant Price GDP Percentage Change (leading by two (2) months) – were next standardised. The standardisation was obtained by subtracting the mean of each variable data set from the value of the variable and dividing the outcome by the standard deviation of the relevant data set. The outcome of this analysis is shown in Table 4.21. This made it possible to do a comparative analysis of the credit losses (Impairments) and the explanatory variables and have better consumer credit risk model statistics (the sizes of the numbers for most regression statistics become smaller).

Table 4.21: Standardised Data for Six (6) Normal/Lead Explanatory Variables

S/N		2	8	12	18	19	20
DATE	**Impairments %**	**ZAR/US$ Forex**	**Oil Price US$**	**P-car reg. index**	**Res. Property PI2010=100**	**World Uncertainty Index**	**Constant Price GDP SA % Change**
2008-09-01	3,10	-1,21	1,56	0,49	0,70	0,20	0,60
2008-10-01	3,42	-1,00	0,38	0,76	0,60	0,25	0,60
2008-11-01	3,60	-0,90	0,45	0,52	0,60	0,25	0,60
2008-12-01	3,42	-0,97	1,14	0,00	0,60	0,25	0,08
2009-01-01	4,20	-1,02	1,12	0,11	0,42	0,38	0,08
2009-02-01	4,60	-0,91	1,23	0,15	0,42	0,38	0,08
2009-03-01	4,80	-1,02	0,85	0,29	0,42	0,38	0,43
2009-04-01	5,10	-1,00	0,78	0,42	0,15	0,30	0,43
2009-05-01	5,40	-0,87	0,37	0,71	0,15	0,30	0,43
2009-06-01	5,50	-0,34	0,08	0,80	0,15	0,30	0,82
2009-07-01	5,60	-0,23	0,16	1,05	0,01	0,34	0,82
2009-08-01	5,80	-0,29	0,14	1,47	0,01	0,34	0,82
2009-09-01	5,90	-0,30	0,07	1,50	0,01	0,34	0,82
2009-10-01	5,84	-0,28	0,34	1,47	0,13	0,39	0,82
2009-11-01	5,94	-0,28	0,44	1,57	0,13	0,39	0,82
2009-12-01	5,94	-0,59	0,29	1,50	0,13	0,39	0,51
2010-01-01	5,86	-0,78	0,45	1,32	0,22	0,31	0,51
2010-02-01	5,82	-0,88	0,37	1,26	0,22	0,31	0,51
2010-03-01	5,88	-0,91	0,57	1,30	0,22	0,31	0,18
2010-04-01	5,95	-0,91	0,71	1,31	0,21	0,26	0,18
2010-05-01	5,91	-1,05	0,26	1,24	0,21	0,26	0,18
2010-06-01	5,91	-1,05	0,32	1,12	0,21	0,26	0,45
2010-07-01	5,84	-1,05	0,37	1,32	0,14	0,30	0,45
2010-08-01	5,86	-1,06	0,38	0,85	0,14	0,30	0,45
2010-09-01	5,88	-1,06	0,32	0,76	0,14	0,30	0,68
2010-10-01	5,95	-1,00	0,60	0,69	0,13	0,34	0,68

S/N		2	8	12	18	19	20
DATE	**Impairments %**	**ZAR/US$ Forex**	**Oil Price US$**	**P-car reg. index**	**Res. Property PI2010=100**	**World Uncertainty Index**	**Constant Price GDP SA % Change**
2010-11-01	5,81	-1,08	0,71	0,51	0,13	0,34	0,68
2010-12-01	5,79	-1,10	0,91	0,23	0,13	0,34	0,79
2011-01-01	5,82	-1,00	0,92	0,54	0,07	0,13	0,79
2011-02-01	5,81	-1,01	0,89	0,48	0,07	0,13	0,79
2011-03-01	5,78	-1,04	1,50	0,11	0,07	0,13	0,74
2011-04-01	5,79	-1,12	1,79	0,68	0,01	0,34	0,74
2011-05-01	5,73	-1,17	1,42	0,31	0,01	0,34	0,74
2011-06-01	5,56	-1,24	1,22	0,13	0,01	0,34	0,71
2011-07-01	5,49	-1,22	1,26	0,20	-0,04	0,18	0,71
2011-08-01	5,29	-1,26	0,79	0,04	-0,04	0,18	0,71
2011-09-01	5,06	-1,23	0,76	0,36	-0,04	0,18	0,50
2011-10-01	4,90	-1,15	0,79	0,20	-0,04	0,11	0,50
2011-11-01	4,81	-1,24	1,26	0,11	-0,04	0,11	0,50
2011-12-01	4,69	-1,29	1,32	0,11	-0,04	0,11	0,42
2012-01-01	4,73	-1,25	1,39	0,05	-0,07	0,08	0,42
2012-02-01	4,71	-1,27	1,47	0,10	-0,07	0,08	0,42
2012-03-01	4,58	-1,27	1,64	0,37	-0,07	0,08	0,27
2012-04-01	4,64	-1,18	1,52	0,54	-0,08	0,08	0,27
2012-05-01	4,56	-1,03	1,15	0,59	-0,08	0,08	0,27
2012-06-01	4,47	-0,91	0,62	0,50	-0,08	0,08	0,36
2012-07-01	4,43	-0,85	0,86	0,67	-0,03	0,07	0,36
2012-08-01	4,36	-0,83	1,13	0,45	-0,03	0,07	0,36
2012-09-01	4,29	-0,89	1,14	0,71	-0,03	0,07	0,36
2012-10-01	4,28	-1,01	0,93	0,73	-0,03	0,23	0,36
2012-11-01	4,06	-1,02	0,80	0,63	-0,03	0,23	0,36
2012-12-01	4,04	-0,95	0,86	0,75	-0,03	0,23	0,27
2013-01-01	4,09	-0,85	1,15	0,81	-0,03	0,20	0,27
2013-02-01	4,06	-0,77	1,18	0,80	-0,03	0,20	0,27

S/N		2	8	12	18	19	20
DATE	**Impairments %**	**ZAR/US$ Forex**	**Oil Price US$**	**P-car reg. index**	**Res. Property PI2010=100**	**World Uncertainty Index**	**Constant Price GDP SA % Change**
2013-03-01	4,01	-0,81	1,08	0,89	-0,03	0,20	0,27
2013-04-01	4,01	-0,81	1,04	0,87	-0,03	0,13	0,27
2013-05-01	4,01	-0,81	1,14	0,87	-0,03	0,13	0,27
2013-06-01	3,94	-0,69	1,20	0,99	-0,03	0,13	0,32
2013-07-01	3,95	-0,64	1,58	0,68	-0,01	0,22	0,32
2013-08-01	3,83	-0,70	1,66	1,20	-0,01	0,22	0,32
2013-09-01	3,71	-0,64	1,65	0,96	-0,01	0,22	0,36
2013-10-01	3,69	-0,62	1,40	1,11	0,14	0,10	0,36
2013-11-01	3,65	-0,52	1,12	1,29	0,14	0,10	0,36
2013-12-01	3,64	-0,55	1,28	1,30	0,14	0,10	0,58
2014-01-01	3,59	-0,47	1,15	1,24	0,17	0,03	0,58
2014-02-01	3,57	-0,27	1,41	1,03	0,17	0,03	0,58
2014-03-01	3,51	-0,30	1,41	0,99	0,17	0,03	0,37
2014-04-01	3,54	-0,25	1,47	0,75	0,14	0,20	0,37
2014-05-01	3,57	-0,28	1,47	0,71	0,14	0,20	0,37
2014-06-01	3,42	-0,30	1,63	0,68	0,14	0,20	0,16
2014-07-01	3,45	-0,21	1,53	0,65	0,13	0,37	0,16
2014-08-01	3,42	-0,15	1,23	0,83	0,13	0,37	0,16
2014-09-01	3,35	0,01	1,09	0,75	0,13	0,37	0,20
2014-10-01	3,35	0,03	0,71	0,83	0,25	0,16	0,20
2014-11-01	3,27	-0,04	0,34	0,89	0,25	0,16	0,20
2014-12-01	3,28	-0,10	0,36	0,80	0,25	0,16	0,15
2015-01-01	3,28	-0,14	0,88	1,07	0,38	0,17	0,15
2015-02-01	3,23	-0,06	0,74	0,98	0,38	0,17	0,15
2015-03-01	3,21	-0,06	0,85	1,16	0,38	0,17	0,36
2015-04-01	3,22	-0,06	0,57	1,20	0,31	0,14	0,36
2015-05-01	3,26	0,04	0,36	1,02	0,31	0,14	0,36
2015-06-01	3,31	0,06	0,34	0,83	0,31	0,14	0,15

S/N		2	8	12	18	19	20
DATE	**Impairments %**	**ZAR/US$ Forex**	**Oil Price US$**	**P-car reg. index**	**Res. Property PI2010=100**	**World Uncertainty Index**	**Constant Price GDP SA % Change**
2015-07-01	3,33	0,07	0,72	1,00	0,28	0,12	0,15
2015-08-01	3,27	0,20	1,07	0,80	0,28	0,12	0,15
2015-09-01	3,20	0,22	0,95	0,77	0,28	0,12	0,02
2015-10-01	3,24	0,22	0,92	0,78	0,39	0,12	0,02
2015-11-01	3,08	0,38	1,08	0,75	0,39	0,12	0,02
2015-12-01	3,12	0,35	1,31	0,85	0,39	0,12	0,27
2016-01-01	3,08	0,35	1,54	0,56	0,35	0,13	0,27
2016-02-01	3,16	0,45	1,60	0,67	0,35	0,13	0,27
2016-03-01	3,38	0,50	1,29	0,60	0,35	0,13	0,44
2016-04-01	3,13	0,64	1,16	0,60	0,28	0,36	0,44
2016-05-01	3,17	0,87	0,90	0,50	0,28	0,36	0,44
2016-06-01	3,17	0,82	0,81	0,75	0,28	0,36	0,09
2016-07-01	3,15	1,03	0,99	0,76	0,22	0,16	0,09
2016-08-01	3,17	1,30	0,99	0,60	0,22	0,16	0,09
2016-09-01	3,24	1,71	0,97	0,16	0,22	0,16	0,03
2016-10-01	2,91	1,53	0,77	0,02	0,21	0,03	0,03
2016-11-01	2,85	1,41	0,95	0,16	0,21	0,03	0,03
2016-12-01	2,87	1,17	0,68	0,21	0,21	0,03	0,04
2017-01-01	2,89	1,40	0,65	0,03	0,17	0,07	0,04
2017-02-01	2,84	1,31	0,61	0,08	0,17	0,07	0,04
2017-03-01	2,84	1,11	0,79	0,15	0,17	0,07	0,01
2017-04-01	2,89	0,91	0,72	0,15	0,18	0,11	0,01
2017-05-01	2,90	0,99	0,83	0,03	0,18	0,11	0,01
2017-06-01	2,90	0,96	0,97	0,11	0,18	0,11	0,01
2017-07-01	2,85	0,96	0,90	0,02	0,16	0,23	0,01
2017-08-01	2,83	0,93	0,84	0,75	0,16	0,23	0,01
2017-09-01	2,77	0,84	0,77	0,26	0,16	0,23	0,12
2017-10-01	2,79	0,73	0,69	0,08	0,12	0,43	0,12

S/N		2	8	12	18	19	20
DATE	**Impairments %**	**ZAR/US$ Forex**	**Oil Price US$**	**P-car reg. index**	**Res. Property PI2010=100**	**World Uncertainty Index**	**Constant Price GDP SA % Change**
2017-11-01	2,81	0,65	0,48	0,13	0,12	0,43	0,12
2017-12-01	2,84	0,81	0,42	0,06	0,12	0,43	0,32
2018-01-01	3,08	0,75	0,17	0,06	0,11	0,14	0,32
2018-02-01	3,10	0,64	0,24	0,07	0,11	0,14	0,32
2018-03-01	3,24	0,72	0,22	0,20	0,11	0,14	0,15
2018-04-01	3,31	0,74	0,07	0,17	0,09	0,01	0,15
2018-05-01	3,39	0,72	0,09	0,18	0,09	0,01	0,15
2018-06-01	3,50	0,89	0,00	0,24	0,09	0,01	0,07
2018-07-01	3,60	1,00	0,14	0,06	0,06	0,04	0,07
2018-08-01	3,57	0,70	0,01	0,14	0,06	0,04	0,07
2018-09-01	3,66	0,42	0,11	0,15	0,06	0,04	0,08
2018-10-01	3,72	0,30	0,13	0,29	0,03	0,34	0,08
2018-11-01	3,74	0,31	0,46	0,21	0,03	0,34	0,08
2018-12-01	3,73	0,39	0,78	0,07	0,03	0,34	0,21
2019-01-01	3,85	0,53	0,70	0,17	0,05	0,09	0,21
2019-02-01	3,81	0,77	0,55	0,11	0,05	0,09	0,21
2019-03-01	3,77	0,78	0,41	0,13	0,05	0,09	0,24
2019-04-01	3,79	1,01	0,17	0,16	-0,01	0,04	0,24
2019-05-01	3,83	1,22	0,30	0,12	-0,01	0,04	0,24
2019-06-01	3,73	1,14	0,56	0,05	-0,01	0,04	0,00
2019-07-01	3,73	1,01	0,45	0,09	0,01	0,09	0,00
2019-08-01	3,80	1,06	0,55	0,15	0,01	0,09	0,00
2019-09-01	3,83	0,93	0,46	0,11	0,01	0,09	0,22
2019-10-01	3,76	0,92	0,59	0,02	-0,01	0,10	0,22
2019-11-01	3,79	1,10	0,46	0,09	-0,01	0,10	0,22
2019-12-01	3,89	1,03	0,34	0,10	-0,01	0,10	0,40
2020-01-01	3,98	1,11	0,44	0,17	-0,20	0,17	0,40
2020-02-01	3,96	1,17	0,74	0,15	-0,20	0,17	0,40

S/N		2	8	12	18	19	20
DATE	**Impairments %**	**ZAR/US$ Forex**	**Oil Price US$**	**P-car reg. index**	**Res. Property PI2010=100**	**World Uncertainty Index**	**Constant Price GDP SA % Change**
2020-03-01	4,04	0,99	1,65	0,13	-0,20	0,17	0,30
2020-04-01	4,27	1,34	2,19	0,03	-0,11	0,12	0,30
2020-05-01	4,59	1,24	1,68	0,06	-0,11	0,12	0,30
2020-06-01	4,89	1,27	1,26	0,15	-0,11	0,12	4,73
2020-07-01	5,01	1,23	1,16	0,06	-0,05	0,18	4,73
2020-08-01	4,99	1,11	1,09	0,30	-0,05	0,18	4,73
2020-09-01	5,03	1,11	1,20	0,12	-0,05	0,18	1,78
2020-10-01	5,03	1,30	1,21	1,52	0,01	0,36	1,78
2020-11-01	5,03	1,82	1,15	4,96	0,01	0,36	1,78
2020-12-01	5,16	2,41	0,89	3,37	0,01	0,36	1,34
2021-01-01	5,24	2,29	0,68	1,56	-0,09	0,30	1,34
2021-02-01	5,19	1,96	0,37	1,48	-0,09	0,30	1,34
2021-03-01	5,12	1,84	0,23	1,27	-0,09	0,30	0,95

(Data from: FRED, Federal Reserve Bank of St. Louis, US, and the Reserve Bank of South Africa).

The data of the variables in Table 4.21 was standardised for ease of comparative analysis and model construction. Each column of numbers has a positive and negative set whose average is zero (see Annexure I in which the corresponding non-standardised data is tabulated).

Next, multivariable regression analysis of Impairments on the six (6) explanatory variables was done, and the results – as seen in Table 4.22– show that all of them were significant as all the p-values are less than 0,05. From this, Consumer Credit Risk Model SA2 was derived as follows:

$$LC0_t = \beta_0 + \beta_1 FX_{t+8} + \beta_2 OIL_t + \beta_3 VEH_{t+7} + \beta_4 RPI_t + \beta_5 VOL_t + \beta_6 GDP_{t+2} + \in$$

$$LC0_t = 4{,}128 - 0{,}677 * FX_{t+8} - 0{,}098 * OIL_t - 0{,}569 * VEH_{t+7} - 0{,}287 * RPI_t - 0{,}138 * VOL_t - 0{,}170 * GDP_{t+2} + 0{,}40$$

LCO is the Loan Charge Off Rate (Impairments); FX is the USD/ZAR Foreign Exchange Rate; OIL is the Crude Oil Price; VEH is the Passenger Car Registrations; RPI is the Residential Price Index; WUI is the World Uncertainty Index that is a proxy for volatility (VOL), GDP is the Constant Gross Domestic Product and ∈ is the Standard Error.

Table 4.22 Consumer Credit Risk Model SA2: Regression analysis: Impairments and Normal/Lead 6 explanatory variables

Regression Statistics	
Multiple R	0,92
R Square	0,85
Adjusted R Square	0,84
Standard Error	0,40
Observations	151

ANOVA

	df	***SS***	***MS***	***F***	***Significance F***
Regression	6	127,10	21,18	133,1	0,00
Residual	144	22,91	0,16		
Total	150	150,01			

		Coefficients	***Standard Error***	***t Stat***	***P-value***	***Lower 95%***	***Upper 95%***
	Intercept	29,048	2,72	10,68	0,00	23,67	34,43
2	ZAR/$ Forex	0,222	0,02	-12,27	0,00	0,26	0,19
8	Oil Price US$	0,004	0,00	1,76	0,08	0,01	0,00
12	P-car reg. index	0,037	0,00	-14,39	0,00	0,04	0,03
18	Res. Property PI 2010=100	0,187	0,03	7,20	0,00	0,24	0,14
19	World Uncertainty Index	0,339	0,11	3,19	0,00	0,55	0,13
20	Constant Price GDP SA % Change	0,052	0,01	3,91	0,00	0,08	0,03

(Data from: the Federal Reserve Bank of St. Louis, US, and the Reserve Bank of South Africa)

As shown in Table 4.22, p-values and t-statistics for all six (6) explanatory variables are less than 0,05 and greater than 1.96 (in absolute terms), respectively. This means that they are all significant and collectively explain 84% of the variation in impairments (see adjusted R square).

4.3.3.1 Back Testing Results for Consumer Credit Risk Model SA2

Consumer Credit Risk Model SA2 was thereafter back tested, and the results of the test were as shown in Table 4.23, Figure 4.10 and Figure 4.11. In Table 4.23, the actual monthly credit losses represented by Impairments and the corresponding credit loss values estimated by the model are shown in the second and third columns of the table. Shown on the fourth column are the calculated respective accuracies. The same information is shown in graphical representation in Figure 4.10 with “Accuracy” on a scale of 0 to 1 instead of 0 to 100 (per cent). In Figure 4.11, “Accuracy” is shown on its own on a scale of 0 to 100 (per cent). This model generally overestimated the credit losses (similar to Model SA1) with accuracies that varied from as low as 54% to as high as 100% with an average of 88% over the period 1987 to 2001. The estimated credit losses closely tracked the actual credit losses. The low estimation performances (less than 75%) were concentrated mainly in 2017 and 2018, with one incidence each in 2008 and 2020. As was the case with Model SA1, the incidences of low performance appear not to have any particular pattern.

Table 4.23: Back Testing Results for Consumer Credit Risk Model SA2

DATE	Impairments %	Estimated Impairments %	Accuracy %	DATE	Impairments %	Estimated Impairments %	Accuracy %	DATE	Impairments %	Estimated Impairments %	Accuracy %
2008-09-01	3,10	3,87	75	2012-11-01	4,06	5,08	75	2017-01-01	2,89	3,46	80
2008-10-01	3,42	3,88	86	2012-12-01	4,04	4,96	77	2017-02-01	2,84	3,55	75
2008-11-01	3,60	4,05	88	2013-01-01	4,09	4,80	83	2017-03-01	2,84	3,75	68
2008-12-01	3,42	4,62	65	2013-02-01	4,06	4,75	83	2017-04-01	2,89	3,83	67
2009-01-01	4,20	5,13	78	2013-03-01	4,01	4,73	82	2017-05-01	2,90	3,67	73
2009-02-01	4,60	5,09	89	2013-04-01	4,01	4,69	83	2017-06-01	2,90	3,80	69
2009-03-01	4,80	5,32	89	2013-05-01	4,01	4,68	83	2017-07-01	2,85	3,69	70
2009-04-01	5,10	5,81	86	2013-06-01	3,94	4,51	86	2017-08-01	2,83	3,20	87
2009-05-01	5,40	5,86	91	2013-07-01	3,95	4,67	82	2017-09-01	2,77	3,55	72
2009-06-01	5,50	5,58	99	2013-08-01	3,83	4,37	86	2017-10-01	2,79	3,75	65
2009-07-01	5,60	5,97	93	2013-09-01	3,71	4,48	79	2017-11-01	2,81	3,82	64
2009-08-01	5,80	6,26	92	2013-10-01	3,69	4,03	91	2017-12-01	2,84	3,61	73
2009-09-01	5,90	6,29	93	2013-11-01	3,65	3,87	94	2018-01-01	3,08	3,98	71
2009-10-01	5,84	6,04	97	2013-12-01	3,64	3,83	95	2018-02-01	3,10	4,07	69
2009-11-01	5,94	6,10	97	2014-01-01	3,59	3,71	97	2018-03-01	3,24	3,96	78
2009-12-01	5,94	6,23	95	2014-02-01	3,57	3,68	97	2018-04-01	3,31	3,88	83
2010-01-01	5,86	6,01	97	2014-03-01	3,51	3,77	93	2018-05-01	3,39	3,87	86
2010-02-01	5,82	6,05	96	2014-04-01	3,54	4,06	85	2018-06-01	3,50	3,77	92
2010-03-01	5,88	5,93	99	2014-05-01	3,57	4,12	85	2018-07-01	3,60	3,88	92
2010-04-01	5,95	5,90	99	2014-06-01	3,42	4,18	78	2018-08-01	3,57	4,06	86
2010-05-01	5,91	6,00	98	2014-07-01	3,45	4,27	76	2018-09-01	3,66	4,24	84
2010-06-01	5,91	5,86	99	2014-08-01	3,42	4,15	79	2018-10-01	3,72	4,48	79
2010-07-01	5,84	6,15	95	2014-09-01	3,35	4,09	78	2018-11-01	3,74	4,87	70
2010-08-01	5,86	5,85	100	2014-10-01	3,35	3,71	89	2018-12-01	3,73	4,68	74
2010-09-01	5,88	5,75	98	2014-11-01	3,27	3,75	85	2019-01-01	3,85	4,18	92
2010-10-01	5,95	5,67	95	2014-12-01	3,28	3,94	80	2019-02-01	3,81	4,02	94

DATE	Impairments %	Estimated Impairments %	Accuracy %	DATE	Impairments %	Estimated Impairments %	Accuracy %	DATE	Impairments %	Estimated Impairments %	Accuracy %
2010-11-01	5,81	5,60	96	2015-01-01	3,28	3,62	90	2019-03-01	3,77	3,99	94
2010-12-01	5,79	5,39	93	2015-02-01	3,23	3,60	88	2019-04-01	3,79	3,98	95
2011-01-01	5,82	5,48	94	2015-03-01	3,21	3,46	92	2019-05-01	3,83	3,88	99
2011-02-01	5,81	5,45	94	2015-04-01	3,22	3,50	91	2019-06-01	3,73	3,96	94
2011-03-01	5,78	5,19	90	2015-05-01	3,26	3,53	92	2019-07-01	3,73	4,01	93
2011-04-01	5,79	5,84	99	2015-06-01	3,31	3,68	89	2019-08-01	3,80	4,02	94
2011-05-01	5,73	5,67	99	2015-07-01	3,33	3,45	96	2019-09-01	3,83	4,12	92
2011-06-01	5,56	5,62	99	2015-08-01	3,27	3,53	92	2019-10-01	3,76	4,09	91
2011-07-01	5,49	5,62	98	2015-09-01	3,20	3,56	89	2019-11-01	3,79	4,02	94
2011-08-01	5,29	5,60	94	2015-10-01	3,24	3,54	91	2019-12-01	3,89	4,10	95
2011-09-01	5,06	5,36	94	2015-11-01	3,08	3,46	88	2020-01-01	3,98	4,40	89
2011-10-01	4,90	5,36	91	2015-12-01	3,12	3,50	88	2020-02-01	3,96	4,60	84
2011-11-01	4,81	5,43	87	2016-01-01	3,08	3,59	83	2020-03-01	4,04	4,79	82
2011-12-01	4,69	5,63	80	2016-02-01	3,16	3,45	91	2020-04-01	4,27	4,29	99
2012-01-01	4,73	5,51	84	2016-03-01	3,38	3,47	97	2020-05-01	4,59	4,29	94
2012-02-01	4,71	5,48	84	2016-04-01	3,13	3,31	94	2020-06-01	4,89	5,28	92
2012-03-01	4,58	5,32	84	2016-05-01	3,17	3,18	100	2020-07-01	5,01	5,17	97
2012-04-01	4,64	5,18	88	2016-06-01	3,17	2,97	94	2020-08-01	4,99	5,41	92
2012-05-01	4,56	5,07	89	2016-07-01	3,15	3,09	98	2020-09-01	5,03	4,70	93
2012-06-01	4,47	5,08	86	2016-08-01	3,17	3,00	95	2020-10-01	5,03	5,49	91
2012-07-01	4,43	4,71	94	2016-09-01	3,24	2,98	92	2020-11-01	5,03	7,36	54
2012-08-01	4,36	4,82	89	2016-10-01	2,91	3,33	86	2020-12-01	5,16	5,79	88
2012-09-01	4,29	4,69	91	2016-11-01	2,85	3,31	84	2021-01-01	5,24	4,80	92
2012-10-01	4,28	5,00	83	2016-12-01	2,87	3,43	81	2021-02-01	5,19	4,95	95
								2021-03-01	5,12	4,81	94
								Average			88

(Data from: FRED, Federal Reserve Bank of St. Louis, US, and the South African Reserve Bank).

Consumer Credit Risk Model SA2 generally overestimates the credit losses. Its accuracy varies from as low as 54% to as high as 100%. The average accuracy is 88% over the period 2008 – 2021.

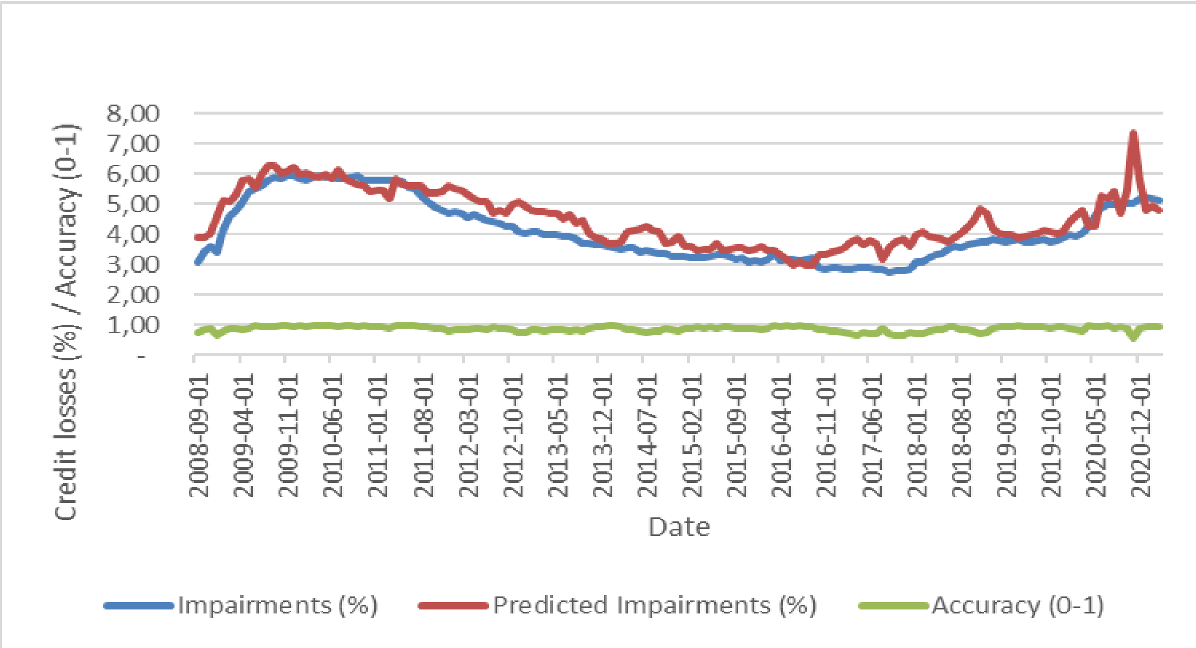


Figure 4.10: Back Testing Results for Consumer Credit Risk Model SA2

(Data from: FRED, Federal Reserve Bank of St. Louis, US, and the South African Reserve Bank).

For Consumer Credit Risk Model SA2, the line representing the estimated impairments closely tracks that of actual impairments. As was the case with Consumer Credit Risk Model SA1, it is generally above it, meaning that it generally overestimates rather than underestimate the credit losses. There was a significant increase in credit losses during the Covid-19 period of 2020.

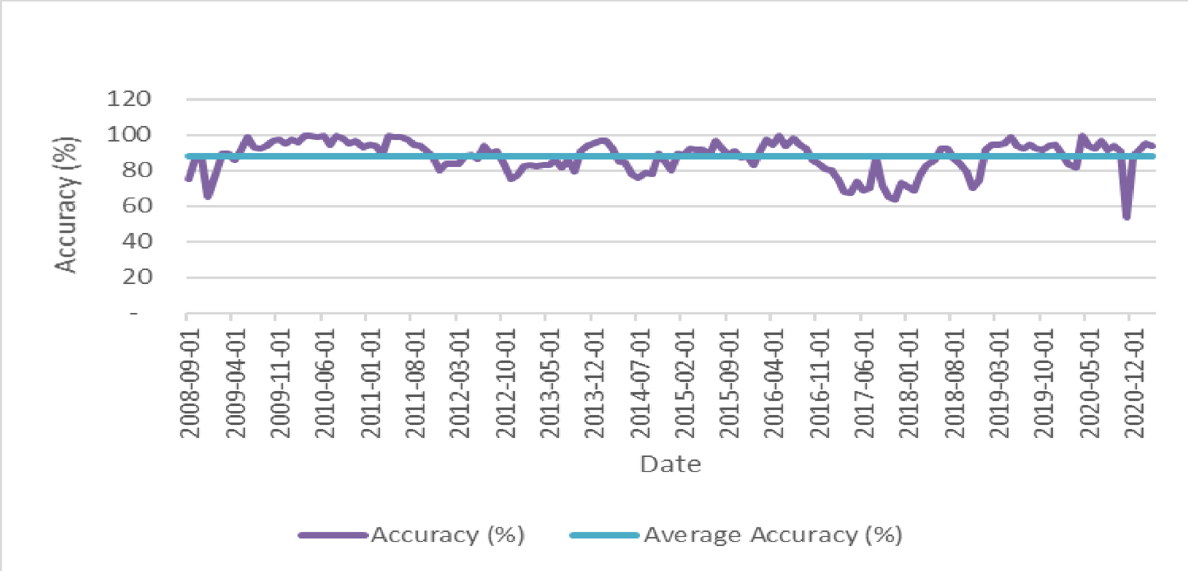


Figure 4.11: Back Testing Accuracy Results for Consumer Credit Risk Model SA2

(Data from: FRED, Federal Reserve Bank of St. Louis, US, and the South African Reserve Bank).

As shown in Figure 4.11, for Consumer Credit Risk Model SA2, the low estimation performances (less than 75%) are concentrated mainly in 2017 and 2018, with one incidence each in 2008 and 2020. The back test accuracy scores for the emerging market of South Africa consumers credit risk model SA2 (EMCCRM SA2) have a Gini coefficient of 0.06. A Gini coefficient measures the statistical dispersion of all the possible pairs of a data set from equality. When all the values of the data set are positive as is the case with the back test accuracy scores of EMCCRM SA2 , the Gini coefficient (G), also referred to as Gini ratio or index, is calculated using the formula, $G = \Sigma((2i\text{-}n\text{-}1)^*x_i) / n\Sigma x_i$, where x represent values of the data set ranked in ascending order and assigned a positional index i =1, to n, the total number of values (Fessel, 2020). The Gini coefficient of 0.06 implies that the scores are generally more equal than unequal i.e., the statistical dispersion about equality is small.

4.3.4 Consumer Credit Risk Models SA3 and SA4

Two (2) additional credit risk models, SA3 and SA4, were similarly derived from correlation analysis of a combination of lagged (-ve in math symbol subscript) and leading (+ve in math symbol subscript) explanatory variables. The two (2) models were obtained by scientifically reducing the variables (described in Section 4.3.1) and linearly regressing using a constant intercept (similar to that in Section 4.2.5). In Figure 4.12, the process steps used to reduce the variables to a more manageable set for regression analysis purposes are given.

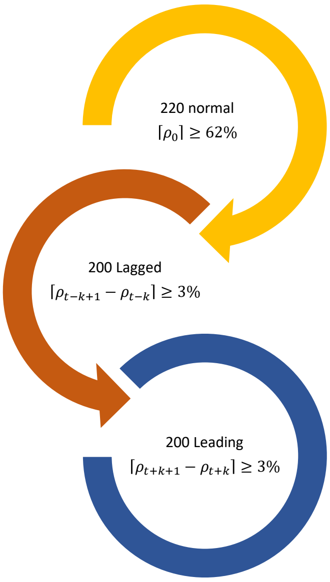


Figure 4.12: Variable Reduction Framework: Coincidental, Lagging, and Leading Variables

Following the process shown in Figure 4.12, coincidental or normal (no lags or leads) correlations (220 unique correlations) were first computed. A careful study of the correlations between the Impairments and each variable – see Annexures J and K – reveals that the number of variables is first reduced when a critical absolute correlation of more than 60% is chosen. Furthermore, it was also found that leading/lagging variables are further reduced if the analyses are confined to the lead (or lag) correlations that are higher than 60%. Finally, for the variables whose correlations with the estimated variable are higher than 60%, the lead or lag period used in the model is the one at which any further addition or subtraction of lead or lag periods (in months) does not change the correlation by more than 3%. As aforementioned, the scientific framework of the process used to reduce the variables selectively is shown in Figure 4.12. Using this framework resulted in two (2) emerging market of South Africa Consumer Credit Risk Models SA3 and SA4, each with only seven (7) terms (a constant intercept term and six (6) economic variables). The two (2) models were described as follows:

$$LC0_t = \beta_0 + \beta_1 FX_{t-10} + \beta_2 OIL_{t-9} + \beta_3 SPI_{t+10} + \beta_4 VEH_{t+6} + \beta_5 VOL_{t+3} + \beta_6 GDP_{t+3} \quad \text{(Model SA3)}$$

$$where, \beta_0 = 4{,}065;\ \beta_1 = -0{,}139;\ \beta_2 = -0{,}434;\ \beta_3 = -0{,}242;\ \beta_4 = 0{,}380;\ \beta_5 = -0{,}514; and\ \beta_6 = 0{,}437$$

and,

$$LCO_t = \beta_0 + \beta_1 FX_{t-10} + \beta_2 OIL_{t-9} + \beta_3 SPI_{t+10} + \beta_4 VEH_{t+6} + \beta_5 VOL_{t+3} + \beta_6 GDP_{t+6}$$

(Model SA4)

$$where, \beta_0 = 4{,}065;\ \beta_1 = -0{,}079;\ \beta_2 = -0{,}439;\ \beta_3 = -0{,}243;\ \beta_4 = 0{,}358;\ \beta_5 = -0{,}540;\ and\ \beta_6 = 0{,}393$$

where LCO is the Loan Charge Off Rate (Impairments); FX is the USD/ZAR Foreign Exchange Rate; OIL is the Crude Oil Price; SPI is the Share Price Index; VEH is the Passenger Car Registrations; WUI (or VOL) is the World Uncertainty Index; GDP is the Real Gross Domestic Product and ∈ is the Standard Error.

The above equations show that the Share Price Index (SPI) is a leading indicator of the Loan Charge Off Rates. This result is consistent with a Journal of Financial Economics study by Duffie et al., (2007), which also found the equity market index to be an essential factor in their credit default risk model. For purposes of comparing the model parameters and discussing or contrasting the economic factors, the regression beta coefficients and their t-Stats were tabulated in Table 4.24.

Table 4.24 Multivariable Regression Coefficients of Credit Risk Models SA3 and SA4 and their Significance

	Model SA3 (1Q leading GDP)		**Model SA4 (2Q leading GDP)**	
	Coefficient	**t-Stat**	**Coefficient**	**t-Stat**
Intercept	4,065	(118,97) ****	4,065	(118,40) ****
FX	-0,139	(-1,34) *	-0,079	(-0,75)
OIL	-0,434	(-9,81) ****	-0,439	(-9,81) ****
SPI	-0,242	(-5,40) ****	-0,243	(-5,37) ****
VEH	0,380	(5,54) ****	0,358	(5,43) ****
VOL	-0,514	(-4,31) ****	-0,540	(-4,27) ****
GDP	0,437	(3,54) ****	0,393	(3,35) ****
R^2	0.86		0.85	
Adj R^2	0.85		0.84	

(Data from: FRED, Federal Reserve Bank of St. Louis, US, and the South African Reserve Bank).****, ***, **, * indicates statistical significance at the 1, 5, 10 and 20 percent levels.

The Vol variable is the World Uncertainty Index, SPI represents the Alsi-40 index, VEH represents Passenger Car Registrations, FX is the South African Rand (Zar) price of the dollar, OIL is the dollar Crude Oil WTI Price, and GDP is the Real GDP of South Africa.

In Table 4.24, the main difference between the two (2) models is that Model SA3 has GDP led by three months while Model SA4 has it led by six (6) months. The two (2) models (SA3 and SA4) of consumer credit risk estimation in the emerging market of South Africa are so close that even the statistical significance of the coefficients appears very close. See the t-Stat columns in Table 4.24. The one exception occurs with the foreign exchange regression factor. Model SA3 has an FX coefficient (β_1 in expression (Model SA3)) at the 20% significance level, whereas Model SA4 does not have an FX coefficient at this significance. Moreover, when two (2) quarters of leading GDP are included in the model, it appears that the significance of the FX coefficient impact diminishes. Since the FX usually has an economic passthrough into GDP (via inflation), the reduction of the FX at the two (2) quarter-lead GDP could be due to the model indirectly compensating for the FX risk when incorporating two (2) quarter-leads of GDP. From an econometric perspective, the model is, therefore sensible.

Even though it is expected that there should be a strong negative correlation between the Loan Charge Off Rate and GDP in the multivariable regression results of Table 4.24, this is not the case.

Common to both Models SA3 and SA4 is the fact that three (3) of the six (6) factor coefficients are negative (See Table 4.24). This evidence is consistent with the bias towards negative correlations and supports the conjecture that the Loan Charge of Rate is pro-cyclical. Models SA3 and SA4 variables explain 85% and 84% on adjusted R^2 basis, of the variation in the credit losses, respectively. Their explanatory power is comparable to those of Models SA1 and SA2. The latter pair's adjusted R^2 were 82% and 84%, respectively. Their shortcoming is that the FX variable in the models has a t-statistic of less than 1.96 in absolute terms and, therefore, not significant at the 95% confidence level. Additionally, models with explanatory variables that have been lagged with respect to credit losses cannot be used to estimate credit losses as the variables occur after the corresponding losses. However, the two (2) models are useful when reviewing (what-if-scenario analysis) the movement of realised losses with changes in explanatory variables.

Table 4.25: Number of Possible Credit Risk Models with 20 Explanatory Variables

Total no. of variables	**No. of variables in the model**	**(n-k)**	**n!**	**k!(n-k)!**	**Number of potential models**
n	**k**				**n!/(k!(n-k)!)**
20	20	0	2 432 902 008 176 640 000	2 432 902 008 176 640 000	1
20	19	1	2 432 902 008 176 640 000	121 645 100 408 832 000	20
20	18	2	2 432 902 008 176 640 000	12 804 747 411 456 000	190
20	17	3	2 432 902 008 176 640 000	2 134 124 568 576 000	1 140
20	16	4	2 432 902 008 176 640 000	502 146 957 312 000	4 845
20	15	5	2 432 902 008 176 640 000	156 920 924 160 000	15 504
20	14	6	2 432 902 008 176 640 000	62 768 369 664 000	38 760
20	13	7	2 432 902 008 176 640 000	31 384 184 832 000	77 520
20	12	8	2 432 902 008 176 640 000	19 313 344 512 000	125 970
20	11	9	2 432 902 008 176 640 000	14 485 008 384 000	167 960
20	10	10	2 432 902 008 176 640 000	13 168 189 440 000	184 756
20	9	11	2 432 902 008 176 640 000	14 485 008 384 000	167 960
20	8	12	2 432 902 008 176 640 000	19 313 344 512 000	125 970
20	7	13	2 432 902 008 176 640 000	31 384 184 832 000	77 520
20	6	14	2 432 902 008 176 640 000	62 768 369 664 000	38 760
20	5	15	2 432 902 008 176 640 000	156 920 924 160 000	15 504
20	4	16	2 432 902 008 176 640 000	502 146 957 312 000	4 845
20	3	17	2 432 902 008 176 640 000	2 134 124 568 576 000	1 140
20	2	18	2 432 902 008 176 640 000	12 804 747 411 456 000	190
20	1	19	2 432 902 008 176 640 000	121 645 100 408 832 000	20
Total number of potential models					**1 048 575**

(Source: Own preparation based on Sheffield (n.d) and Ďuriš et al. (2021)).

The calculations shown in Table 4.25 indicate that the number of possible models that can be constructed by regressing credit losses with any number of the explanatory variables varies from as low as one (1) when all the 20 variables are in the model to as high as 184 756, when 10 explanatory variables are included in the model. The total number of consumer credit risk models that can theoretically be built from the 20 explanatory variables is 1 048 575.

4.3.5 Emerging Market of South Africa Consumer Credit Risk Models

As established in Table 4.25, the total number of possible models for the emerging market of South Africa that can be constructed by regressing credit losses (represented by Impairments) on the 20 explanatory variables is 1 048 575. In this research four models SA1, SA2, SA3 and SA4 were developed using emerging market of South Africa data and in line with the knowledge, insights and techniques gained in building the developed market of US credit risk models (see Table 4.26)

Table 4.26: Emerging Market of South Africa Consumer Credit Risk Models

S/N	**Explanatory Variables/ Analyses**	**Model SA1**		**Model SA2**		**Model SA3**		**Model SA4**	
		Variable present	t Stat	Variable present	t Stat	Variable present	t Stat	Variable present	t Stat
2	FX	√	-3.23	√	-12.27	√	-1.34	√	-0.75
8	Oil	√	-3.45	√	1.76	√	-9.81	√	-9.81
9	SPI	√	2.03	x		√	-5.4	√	-5.37
12	Veh	√	-5.63	√	-14.39	√	5.54	√	5.53
15	Gold	√	5.77	x		x		x	
17	RGDP	√	-2.88	x		x		x	
18	RPI	√	-7.21	√	7.2	x		x	
19	WUI	√	-5.83	√	3.19	√	-4.31	√	-4.27
20	CPGDP (GDP)	√	2.9	√	3.91	√	3.54	√	3.35
N/A	Intercept	25,10		29.05		4.07		4.07	
N/A	R_a^2	82%		84%		85%		84%	
N/A	Back Test Accuracy	86%		88%					

(Data from: FRED, Federal Reserve Bank of St. Louis, US).

From the analyses of the t statistics, adjusted R-squared values, and back test accuracy values shown in Table 4.26, two (2) key observations are made. First is that models SA3 and SA4 have some explanatory variables whose t statistic values are significantly lower than 1.96 in absolute terms. A value of the t statistic lower than 1.96 indicates that the explanatory variable is insignificant in the model at the 95% confidence level. And secondly, the adjusted R-squared (R_a^2) and back test accuracy values of model SA2 are marginally better than those of model SA1. This means that the Emerging Market of South Africa Consumer Credit Risk Model SA2 (EMCCRM SA2) is the best of the four models. It is plausible that in addition to the four models (SA1, SA2, SA3 and SA4) described in this research, more could be constructed using similar methods or other methods to find suitable combinations of explanatory variables that are all significant at a chosen confidence level.

4.4 CHAPTER SUMMARY

In this chapter, the researcher presented the analyses and results of the research. The analyses were about the credit risk modelling using the developed market of US data and its adaptation for the understanding and estimation of credit losses in the emerging market of South Africa. This encompassed the description of the dependent and explanatory variables, analyses of correlations amongst the variables, bivariable and multivariable regression and ANOVA analyses, and back testing of the models. The models appraised for statistical significance were Consumer Credit Risk Models (CCRM) US1, US2, US3, US4 and US5 for the developed market of US and SA1, SA2, SA3 and SA4 for the emerging market of South Africa. In the next chapter, the researcher outlined the summary, conclusions, and recommendations of the research.

CHAPTER 5: SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

The previous chapter was a report on the research analysis and results. In the report, the variables used in building the credit risk models were described, and the correlation, data standardisation, bivariable and multivariable regression, and ANOVA analyses were presented. The development of the proof-of-concept developed market consumer credit risk model, and the use of insights gained in the construction of the emerging market of South Africa consumer credit risk model were discussed. In this chapter a summary of the findings, the conclusions reached, and recommendations arising from the research are presented. In the summary, a discussion of the findings and how they relate to the research objectives, literature review and theoretical framework is presented. The conclusion is a summary of what the researcher did, what was found, and what the implications are. In the recommendation section, the limitations of the research and suggestions for further study are outlined.

5.1 RESEARCH SUMMARY

The following are the research findings and how they relate to the literature review and the theoretical framework.

5.1.1 Model Building Analyses Blocks

Amongst other insights, knowledge and understanding gained during the credit risk model building process was the model building analyses blocks framework depicted in Figure 4.1. The framework was used to reduce the explanatory variables by elimination and build the credit risk model. In hierarchical order of reducing number of explanatory variables, the analyses blocks were correlations, bivariable regressions, ANOVA and multivariable regressions.

5.1.2 Emerging Market of South Africa Consumer Credit Risk Model SA2

The best of each set of proof-of-concept consumer credit risk models based on the developed market of the US data and the consumer credit risk models constructed out of the emerging market of South Africa data performed better (see Table 4.6 and Table 4.22) than the equivalent Loss Given Default (LGD) models. LGD models' performance, as measured by R-squared (R^2) values ranges between 20 and 30 per cent (Baesens, 2015; Baesens, Rosch and Scheule, 2016). The best proof-of-concept Consumer Credit Risk Model US1 had a R^2 value of 91.2% (with an in-sample back test average accuracy of 84%), and the best emerging market of South Africa Consumer Credit Risk Model SA2 had a R^2 value of 84% (with an in-sample back test average accuracy of 88%). LGD models, as the name suggest, are based on relationships between expected losses incurred in the event of a default and explanatory variables, while this research's credit risk modelling is based on relationships between realised credit losses and explanatory variables. In the literature reviewed, while economic and obligor explanatory variables were used in some LGD models (Angel and Heitzmann, 2015; Jin *et al.*, 2021), there was no mention of a specific action to include (in combination) proxy variables for sentiment – defined in this research as participants' behaviour in disequilibrium situations (DiGeorgia, 2001; Soros, 2012; Jareño and Negrut, 2016; Marks, 2022) It is plausible that this is due to LGD credit risk modellers assuming that the Efficient Market Hypothesis holds in all situations and at all times as it states. In this research, the Efficient Market Hypothesis was taken as not holding in disequilibrium situations and hence the use of proxy sentiment variables to capture borrower behaviour in such situations. The two subtle differences – that of the basis of modelling relationships and positions with respect to the Efficient Market Hypothesis in modelling approaches – may explain the credit loss estimation performance differences.

5.1.3 Coupling of Variables that Capture Sentiment with Economic Variables and Obligor Characteristics

Variables classified under all the three categories of data – sentiment, economic, and obligor – were in the proof-of-concept Credit Risk Model US1. These were Oil Price (Oil), a proxy variable for sentiment, T10Y2Y (10-Year Treasury Constant Maturity Rate minus the 2-Year Treasury Constant Maturity Rate), and T2Y (2-Year Treasury Constant Maturity Rate), and Unrate (Unemployment Rate) falling under the economic variable category and PCE (Personal Consumption Expenditure), PSR (Personal Savings Rate) and Drate (Delinquency Rate) as proxy variables for obligor characteristics. However, in the Emerging Market Consumer Credit Risk Model SA2, the proxy variables for sentiment were dominant, with only one variable for the economic category and none for the obligor category. These were FX (Foreign Exchange Rate), Oil (Oil Price), VEH (Vehicle Sales), VOL (Volatility as measured by the World Uncertainty Index) and RPI (Residential Property Index) for sentiment, and GDP (Gross Domestic Product) for economic categories. In research Finding 5.1.2, it was pointed out that the developed market Credit Risk Model US1 had a better R^2 (of 91.2%) than the emerging market of South Africa Consumer Credit Risk Model SA2 (at 84%). However, it had a lower average estimation accuracy (of 84% compared to 88% for the emerging market consumer credit risk model) when back tested. The variation in the composition of the variables in the three categories within the models may explain this difference in performance. However, this may not be conclusive as only one variable had been selected within the obligor category for the emerging market credit risk model since data availability within it was limited compared to the developed market. Moreover, when first selected, half of the explanatory variables (10) for the emerging market were of the economic category. The developed market data covered a longer period (1987-2021) than emerging market data (2008-2021). A longer period is likely to have included more periods of economic expansions and contractions and may explain the better R^2 for the developed market explanatory variables and realised credit losses. From these observations, the researcher concluded that consumer credit risks are affected by market sentiment and including proxy variables for the sentiment (tone of the market) does improve the consumer credit risk model loss estimations. The analysis and results of the research support the notion that coupling proxy variables for sentiment together with a set of economic and obligor characteristic variables improve the estimation of a consumer credit risk model. The p-values, and consequently the t-statistics for the proxy variables for sentiment that were described in Section 4.3.1 – Oil Price, Car Registration (to represent car sales), and World Uncertainty Index (volatility) were all found to be significant. See p-values/t-statistics for these variables in Models SA1, SA2, SA3 and SA4 shown in Table 4.17, Table 4.22 and Table 4.24. The finding of sentiment variables affecting consumer credit risk modelling is consistent with other findings in literature (Duffie, Saita and Wang, 2007). Duffie, Saita and Wang (2007) found that financial market variables affect corporate defaults.

5.1.4 Data Gaps in the Emerging Market of South Africa

Availability of ideal data for credit risk modelling in the emerging market of South Africa was a limitation in this research. The researcher could not, for example, obtain data on the Loan Charge Off Rates (or realised credit losses, the dependent variable used in the proof-of-concept consumer credit risk model building for the developed market of the US) for the emerging market of South Africa from the South African Reserve Bank (SARB, 2020). An officer working at the SARB confirmed that the bank does not collect Loan Charge Off Rates data from the country's consumer banks, therefore, the researcher used Loan Impairment data instead. Data on Impairments for the emerging market of South Africa was available in both the data banks of the Reserve Bank of St. Louis, US, and the South African Reserve Bank. However, neither of the sets of data was complete. The researcher combined the two sets to obtain the data that was used in building the emerging market of South Africa consumer credit risk models. The Impairments data was also available only on a monthly frequency and the portion from the South African Reserve Bank had to be extracted on a day-by-day and month-by-month basis as it had not been organised for ease of use in research in the way of the developed market of the US. The data series for the emerging market of South Africa explanatory variables were all extracted from the data bank of the Reserve Bank of St. Louis, US, which in turn had sourced most of the data mainly from the OECD and a few data series each from the IMF, US Bureau of Labour Statistics, US Energy Information Administration, and the World Bank.

5.1.5 Multicollinearity and its Effects on Consumer Credit Risk Modelling

Multicollinearity is a situation in linear multiple regression analysis in which some of the explanatory variables are strongly (correlation factor of greater or equal to 80%) correlated with each other (Shrestha, 2020; Simpson and Weiner, 2022). In both the correlation analyses amongst the explanatory variables used in building the proof-of-concept developed market consumer credit risk models (see Table 4.2) and the emerging market of South Africa consumer credit risk models (see Table 4.13), correlations factors greater than or equal to 80% were observed as follows:

- For the developed market modelling, GDP and T10Y, GDP and T2Y, PCE and T10Y, PCE and T2Y and PCE and GDP had correlation factors of -94%, -84%, -94%, -85%, and 100%, respectively. This may explain the presence of the explanatory variable PCE, and the absence of GDP in the consumer credit risk model US1 as the variables are perfectly correlated (with a correlations factor of 100%). Their correlations with T10Y and T2Y are also similar in values. PCE and GDP had similar correlations with Loan Charge Off Rates of 21% and 23%, respectively. However, GDP was indicated as being insignificant in the first multivariable regression analysis with a p-value greater than 0.05. However, GDP was found to be a significant explanatory variable when regressed with Loan Charge Off Rates on its own. Multicollinearity may explain this discrepancy. Therefore, it may lead to the wrong conclusion being made about the importance of GDP in explaining credit losses. Nevertheless, elimination of GDP from the credit risk model, does not affect its predictive power as PCE is perfectly correlated with it. This is in line with findings in the literature reviewed on the effects of multicollinearity (Jayakumar, Sulthan and Studies, 2014; Mundfrom, DePoy Smith and Kay, 2018).
- For the emerging market of South Africa credit risk modelling, there were also perfectly and strongly correlated explanatory variables. IBR (Interbank Interest Rate) and ITB (Interest Rate on Treasury Bills), I10YTB (Interest Rate on 10Y Treasury Bills) and

10Y3MTB (interest on 10 -Year Treasury minus that on 3-Month Treasury Bills) , and CPI (Consumer Price Index) and SPI (Share Price Index) had correlation factors of 100%, 100%, and 93%, respectively. However, none of these variables ended up in the emerging market of South Africa Consumer Credit Risk Model SA2. The selection of second round inflation variables like FX may explain the absence of inflation variables like CPI in the model.

5.1.6 The Efficient Market Hypothesis

One of the two gaps identified in the literature in this study was that the Efficient Market Hypothesis may not hold in disequilibrium situations. The Efficient Market Hypothesis holds that the price of a financial asset reflects all available information about its fundamental value. This means that an investor cannot earn a higher return than the expected return without taking on additional risk. However, in disequilibrium situations, the price of a financial asset can be far higher or lower than the fundamental value. In situations where the price is far lower than the fundamental value, taking on less risk (by buying the asset at the lower price) results in higher returns when the price increases to fundamental value. The returns are reversed in situations where the asset price is far higher than the fundamental value – taking on more risk (by buying an overpriced asset) results in a lower return when the price normalises. Consumer credit can be securitised and traded in the financial markets through instruments such as mortgage-backed securities. Therefore, theoretically, in situations where the Efficient Market Hypothesis holds, the prices of such securities should reflect the fundamental values of the assets (mortgage loans) upon which the securities are based. Movements of these prices, therefore, reflect credit consumer behaviour – for example, the degree to which they are honouring their loan obligations. Economic factors also affect the behaviour of credit consumers. If they have a pessimistic view of the state of their economic future, they will react to mitigate it and vice versa. In this study, the researcher identified explanatory variables that capture credit consumer behaviour in equilibrium (economic and obligor variables) and disequilibrium situations (proxy variables for sentiment). The results of the study indicate that coupling together proxy variables for sentiment with sets of economic and obligor characteristic variables improves the performance of a consumer credit risk model.

5.1.7 The Vast Number of Possible Credit Risk Models and its Implications

It was established in the results and analysis section (see Table 4.7 and Table 4.25) of this study that the number of possible consumer credit risk models for the developed market of the United

States of America and the emerging market of South Africa that can be built from the respective 17 and 20 explanatory variables selected for these markets are 131 071 and 1 048 575 respectively. In this study, comparative, correlation, ANOVA, and bivariable and multivariable regression analyses were used to reduce the explanatory variables to a smaller and significant set. It is possible that other sets of explanatory variables, which explain the variations in credit losses as well as or better than those used in this research, could be identified using this research's methods, their variations or completely different methods. Moreover, other explanatory variables with as good or better correlations with credit losses may be identified.

5.2 RESEARCH CONCLUSIONS

This section is a summary of what the researcher did, what was found, and what the implications are. In this research, the importance of credit risk as having ramifications on the economy, businesses and society was established (Baesens, Rosch and Scheule, 2016; DW Documentary, 2017), and the problem of inaccurate credit risk models in the market was identified (Baesens, 2015; Baesens, Rosch and Scheule, 2016; Honohan, 2016).

5.2.1 Emerging Market of South Africa Consumer Credit Risk Model (EMCCRM) SA2

The solution for the problem was to build a credit risk model, especially for the emerging market of South Africa, where data is sparse and credit risk modelling not well established (Apanga, Appiah and Arthur, 2016), using novel explanatory variables and analysis techniques. While similar credit risk models have used economic and obligor characteristic variables (Angel and Heitzmann, 2015; Baesens, Rosch and Scheule, 2016), the researcher did not find any that had sought to identify variables to explain credit consumer behaviour in situations where the Efficient Market Hypothesis (on which credit risk model building is generally based) does not hold (Soros, 2014). This research used proxy variables for Sentiment (defined in this research as behaviour or reactions of consumers, in the financial markets in general and in the credit markets in particular, in disequilibrium situations) (Soros, 2012; Marks, 2022). The result was an emerging market of South Africa consumer credit risk model with an estimation performance that is relatively better than similar Loss Given Default models reported on in the literature reviewed. The Emerging Market of South Africa Consumer Credit Risk Model SA2 (EMCCRM SA2) had a high R^2 value of 85 per cent compared to similar market LGD models with R^2 values of 20 to 30 per cent. This is the final model that should be adopted/adapted or tested by other researchers.

5.2.2 Theoretical Bridge for the EMH Gap (Limitation)

The gap identified in this research as the limitation of the Efficient Market Hypothesis – in not holding in disequilibrium situations[11] (Ang, Goetzmann and Schaefer, 2010; Lawson, 2015; Rossi and Gunardi, 2018) – may also have been bridged with the use of proxy variables for sentiment to account for credit market participants behaviour in far-from-equilibrium or disequilibrium situations (DiGeorgia, 2001; Soros, 2014; Jareño and Negrut, 2016; Dupor *et al.*, 2020; Marks, 2022). This, together with the finding presented in Section 5.2.1, constitute the two major contributions of the research.

5.2.3 Research Aim, Questions and Objectives

The Emerging Market of South Africa Consumer Credit Risk Model SA2 produced estimated credit loss values, when back tested with in-sample data, that closely tracked the realised credit losses. Since, it had a high R^2 value of 85 per cent compared to similar market LGD models with R^2 values of 20 to 30 per cent, and credit loss estimation accuracy of 88%, the main aim of the research was achieved. The aim of this research was to develop a consumer credit risk model, using data from the developed market of the United States of America, that would improve the estimation of consumer credit losses, and use the insights gained to develop a similar model for the emerging market of South Africa. The use of proxy variables for sentiment (like OIL, VEH, FX, WUI (VOL)) and obligor characteristic (like PSR, DRATE, PCE) as well as the choice of economic variables (like GDP, UNRATE, T10Y2Y(Tilt)) and the transformation of some of the data (for example by converting them from a quarterly to a monthly frequency) also partially answered the question raised in the research on possible data gaps in the emerging market of South Africa. The availability of a sizeable proportion of the data used in this research at the data bank of the Reserve Bank of St. Louis, US, ameliorated the data situation in the emerging market.

---

[11] It has been observed for example that during periods of financial market busts, security prices may fall below intrinsic value, presenting a buying opportunity ((Buffet, 1984). This is contrary to the Efficient Market Hypothesis. In such situations, credit risk modelling using only traditional economic and/or obligor variables with the assumption that the EMH holds may be incomplete. What is required is the inclusion of additional explanatory variables specifically selected for their ability to capture borrower behaviour in such situations. This is why including Sentiment Variables as defined in this research works.

5.2.3.1 Adaptability of the Credit Risk Modelling Techniques

The development of the emerging market of South Africa Consumer Credit Risk Model SA2 through adaptation of the modelling techniques and insights obtained in building the Consumer Credit Risk Model US1 (and other models) using data from the developed market of the United States of America demonstrated the universal applicability, under certain conditions, of the methods used in this research. As stated by Baesens (2015), the most important condition is the availability and quality of relevant data for model development. Data gaps in the emerging market of South Africa were pointed out in Section 5.1.4. An online search for data from other emerging markets like Kenya that could be used in consumer credit risk modelling was not successful. This is in line with findings by other researchers on the difficulty of accessing data in emerging markets (Apanga, Appiah and Arthur, 2016). However, all three jurisdictions (the US, South Africa and Kenya) have dedicated statistics bureaus (SARB, 2020; FED, 2022; KNBS, 2022). The main difference between the developed market of the United States of America and the two emerging markets is that the former has collected and organised data, for ease of use in research and education, from its own jurisdiction and other jurisdictions around the world. The collection, storage, and organisation formats used in the Federal Reserve Economic Data (FRED) database could be adopted by emerging markets.

5.2.4 Significance and Implications of Findings

This research developed an Emerging Market of South Africa Consumer Credit Risk Model (EMCCRM) SA2 with relatively good estimation accuracy. This means that the model can be adopted/adapted for use by financial institutions in the emerging market of South Africa or be adapted for use in other emerging markets. For the financial institutions, the model's economic value will accrue from its ease of construction (assuming that quality data is available), low cost and the lesser likelihood of setting aside overestimated economic and regulatory capital buffers.

5.3 RESEARCH RECOMMENDATIONS

This section of the thesis is about the limitations of the study and suggestions for further research.

5.3.1 Development of Proprietary Data

This research was limited by unavailability of research data, especially, in the emerging market of South Africa. Research could be done in future with data prepared in collaboration with one or more of the South African banks with a long history of operating in South Africa. Such data could be prepared following the formats that the Federal Reserve bank of St. Louis, a branch of the US Federal Reserve bank system, has used. This would facilitate the adaptation of credit risk models explored to bank-specific consumer credit risk models. Such models would more accurately capture the particular characteristics of the customers of the relevant bank and the policies and procedures that contribute towards consumer behaviour.

5.3.2 Other Consumer Credit Risk Models

In this research, it was established that the number of possible models that can be built from the 20 explanatory variables used in creating the EMCCRM SA2 is vast (1 048 575). It is, therefore, plausible that there is a model or several models from this vast number with better predictive power than EMCCRM SA2. Further research could be done using the methods employed in this study, their variations or other methods to identify different combinations of the explanatory variables (or additional variables) that further improve the estimation accuracy of the resulting model(s).

5.4 CHAPTER SUMMARY

In this final chapter of the thesis, the researcher outlined the summary, conclusions, and recommendations of the research. In the summary, the findings and how they relate to the literature review and the theoretical framework were discussed. The conclusions are a summary of what was done, what was found and what the implications are. In the recommendations section, the limitations of the research and the suggestions for further research were outlined.

REFERENCES

Abduholiqovna, Q.G. (2021) ‘Epistemological Approaches and the Psychology of Knowledge’, *Emergent: Journal of Educational Discoveries and Lifelong Learning*, 2(4), pp. 50–53.

Adamko, P., Klieštik, T. and Birtus, M. (2014) ‘History of Credit Risk Models’, in Lee, G. (ed.) *2014 2nd International Conference On Economics and Social Science (ICESS 2014)*. Shenchen, China.: Information Engineering Research Institute, USA, pp. 148–163.

Adom, D., Hussein, E.K. and Agyem, J.A. (2018) ‘Theoretical and Conceptual Framework : Mandatory Ingredients of a Quality Research’, *International Journal of Science Research*, 7(January).

Allen, D.E. and Powella, R.J. (2011) *Credit Risk Measurement Methodologies, 19th International Congress on Modelling and Simulation*. Perth, Australia. Available at: http://mssanz.org.au/modsim2011/D6/allen4.pdf (Accessed: 28 June 2019).

Ang, A., Goetzmann, W.N. and Schaefer, S.M. (2010) *The Efficient Market Theory and Evidence: Implications for Active Investment Management, Foundations and Trends® in Finance*. Available at: https://www.nowpublishers.com/article/Details/FIN-034 (Accessed: 5 July 2019).

Angel, S. and Heitzmann, K. (2015) ‘Over-Indebtedness in Europe: The Relevance of Country-Level Variables for the Over-Indebtedness of Private Households’, *Journal of European Social Policy*, Vol. 25(3), pp. 331– 351. doi:10.1177/0958928715588711.

Antoniades, A. (2018) *Gazing into the Abyss of Indebted Society: The Social Power of Money and Debt, Political Studies Review*. doi:10.1177/1478929918757135 journals.sagepub.com/home/psrev.

Apanga, M.A.N., Appiah, K.O. and Arthur, J. (2016) ‘Credit Risk Management of Ghanaian Listed Banks’, *International Journal of Law and Management*, 58(2), pp. 162–178. doi:10.1108/IJLMA-04-2014-0033.

Athanassiou and Theodosopoulou (2015) ‘Reducing Over-Reliance on Credit Rating Agencies: State of Play and Challenges Ahead’, *Maastricht Journal of European and Comparative Law*, 22(5), pp. 656–677.

Bae, J.K. and Kim, J. (2015) *A Personal Credit Rating Prediction Model Using Data Mining* *in Smart Ubiquitous Environments*, *International Journal of Distributed Sensor Networks*. doi:10.1155/2015/179060.

Baesens, B. (2015) *State of the Art in Credit Risk Modelling, Morgan Mckinley*. Available at: https://www.youtube.com/watch?v=1vF3dbcr1c8 (Accessed: 24 July 2018).

Baesens, B., Rosch, D. and Scheule, H. (2016) *Credit Risk Analytics: Measurement, Applications and Examples in SAS, SAS Institute Inc.* Cary, North Carolina, USA: Wiley.

Barth, J.R. *et al.* (2018) 'Forecasting Net Charge-Off Rates of Banks: What Model Works Best?', *Quantitative Finance and Economics*, 2(3), pp. 554–589. doi:10.3934/qfe.2018.3.554.

BCBS (2017) *High-level Summary of Basel III Reforms*. Basel, Switzerland: Bank for International Settlements. Available at: www.bis.org.

BCBS (2020) *Counterparty Credit Risk Definition and Explanation*. Basel, Switzerland. Available at: https://www.bis.org/basel_framework/chapter/ CRE/51.htm?tldate= 20220103&inforce=20220101&published=20191215 (Accessed: 27 July 2021).

Beerbaum, D. and Ahmad, S. (2015) 'Credit Risk according to IFRS 9 : Significant increase in Credit Risk and implications for Financial Institutions', pp. 1–7.

Black, F. and Scholes, M. (1973) 'The Pricing of Options and Corporate Liabilities', *The Journal of Political Economy,* 81(3), pp. 637–654. Available at: http://www.jstor.org/stable/1831029.

Blblica Inc. (2011) 'Exodus; Leviticus; Deuteronomy', in Biblica Inc. (ed.) *The Holy Bible, New International Version*. 2011th edn. Grand Rapids, Michigan: Zondervan, p. 123;198;312.

Bluhm, C., Overbeck, L. and Wagner, C. (2003) *An Introduction to Credit Risk Modelling*. New York, New York: Chapman & Hall/CRC.

Bolton, C. (2009) *Logistic Regression and its Application in Credit Scoring*. University of Pretoria.

Borio, C. (2019) *On Money, Debt, Trust and Central Banking*. 763. Basel, Switzerland.

Borio, C. (2020) 'The Covid - 19 Economic Crisis : Dangerously Unique', *Business Economics,* 55(4), pp. 181–190. doi:10.1057/s11369-020-00184-2.

Botha, I., Booyens, M. and de Wet, S. (2015) 'Measuring the Effect of the National Credit Act on Indebtedness in South Africa', *Journal of Economic and Financial Sciences*, 8(April), pp.

83–104.

Brouwer, T., Huttenhuis, J. and ter Hoeven, R. (2021) ‘Empirical Results for Expected Credit Losses of G-Sibs During COVID-19. The Proof of The Pudding is in the Eating’, *Maandblad voor Accountancy en Bedrijfseconomie*, 95(11/12), pp. 381–396. doi:10.5117/mab.95.75980.

Buffet, W. (1984) ‘The Superinvestors of Graham and Doddsville’, *Graham and Doddsville* [Preprint]. Available at: https://www8.gsb.columbia.edu/sites/valueinvesting/files/files/Buffett1984.pdf.

Burton, D. (2012) ‘Credit Scoring, Risk, and Consumer Lendingscapes in Emerging Markets’, *Environment and Planning*, 44(1), pp. 111–124. doi:10.1068/a44150.

Chopra, A. and Bhilare, P. (2018) ‘Application of Ensemble Models in Credit Scoring Models’, *Business Perspectives and Research*, 6(2), pp. 1–12. doi:10.1177/2278533718765531.

Dalio, R. (2013) *How the Economic Machine Works*. United States of America: Youtube. doi:https://www.youtube.com/watch?v=PHe0bXAIuk0.

Delcey, T. (2019) ‘Samuelson vs Fama on the efficient market hypothesis: The point of view of expertise’, *OEconomia*, 9(1), pp. 37–58. doi:10.4000/oeconomia.5300.

Demirgüç-Kunt, A. *et al.* (2015) *The Global Findex Database 2014 Measuring Financial Inclusion around the World*. Washington, DC. Available at: http://www.worldbank.org/globalfindex.

Demirgüç-Kunt, A. *et al.* (2018) *The Global Findex Database 2017: Measuring Financial Inclusion and the Fintech Revolution, Report*. Washington, DC: World Bank. doi:10.1596/978-1-4648-1259-0.

DiGeorgia, J. (2001) *Understanding the ‘Official’ 1 7 Key Economic Indicators and How They Move the U.S. Stock Market*. 21st Century Investor Publishing, Inc.

Dore, T. and Mach, T. (2019) *Marketplace Lending and Consumer Credit Outcomes : Evidence from Prosper*. 2019–022. Washington, DC. doi:https://doi.org/10.17016/FEDS.2019.022.

Dudovskiy, J. (2018) *The Ultimate Guide to Writing a Dissertation in Business Studies: A Step by Step Assistance*. Available at: https://research-methodology.net/research-philosophy/epistomology/#:~:text=Epistemology has many branches that relate to business studies.

Dueñas-peña, A., Barbosa-Guerrero, L.M. and Rozo, J.J.P. (2022) ‘Approach to a Characterization of Credit’, *Journal of Language and Linguistic Studies*, 18(2), pp. 276–297. Available at: https//:www.jlls.org JOURNAL.

Duffie, D., Saita, L. and Wang, K. (2007) ‘Multi-Period Corporate Default Prediction with Stochastic Covariates’, *Journal of Financial Economics*, 83(2007), pp. 635–665. doi:10.1016/j.jfineco.2005.10.011.

Dupor, B. (2019) ‘Economic Synopses’, *Crisis*, 16(20), p. 1.25. Available at: https://files.stlouisfed.org/files/htdocs/publications/economic-synopses/2018/11/30/does-the-yield-curve-really-forecast-recession.pdf.

Dupor, B. *et al.* (2020) ‘The 2008 U.S. Auto Market Collapse’, p. 38 pages. Available at: https://facultystaff.richmond.edu/~smehkari/polk33_jan20.pdf.

Ďuriš, V. *et al.* (2021) ‘Teaching combinatorial principles using relations through the placemat method’, *Mathematics*, 9(15), pp. 1–13. doi:10.3390/math9151825.

DW Documentary (2017) *How the Rich Get Richer - Money in the World Economy*. Germany: DW.Com. Available at: https:www.youtube.com/channel/UCW39zufHfsuGgpLvikh298Q?sub_confirmation=1#.

Engelmann, B. (2021) ‘A Simple and Consistent Credit Risk Model for Basel II / III , IFRS 9 and Stress Testing when Loan Data History is Short’, pp. 35–37.

Eubank, E.E. (2012) ‘A Case Study of the Effects of Consumer Credit Upon the Family’, *The Annals of the American Academy of Political and Social Science*, (November), pp. 211–220. doi:10.1177/000221947400701012.

FASB (2016) *Understanding Costs and Profit*. Norwalk, Connecticut. Available at: chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.fasb.org/document/blob?fileName=Understanding_cost_and_benefits_credit_losses_(6-16).pdf.

FED (2022) *FRED (Economic Data)*. Available at: https://fred.stlouisfed.org/ (Accessed: 17 May 2022).

Fessel, K. (2020) *Measuring Statistical Dispersion with the Gini Coefficient*. Available at: http://kimberlyfessel.com/mathematics/applications/gini-use-cases/ (Accessed: 8 May 2023).

Fons Wijnhoven and Plant, O. (2019) ‘Sentiment Analysis and Google Trends Data for Predicting Car Sales’, in *Thirty Eighth International Conference on Information Systems, South*

Korea 2017. the International Conference on Information Systems (ICIS).

Frey, F. (2017) 'SPSS (Software)', *The International Encyclopedia of Communication Research Methods*, (November 2016), pp. 1–2. doi:10.1002/9781118901731.iecrm0237.

Gabeshi, K. (2022) 'The Impact of Credit Activity on the Economic Evolution of the Developed Economies', (1), pp. 185–190. Available at: chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.utgjiu.ro/revista/ec/pdf/2022-01/22_Gabeshi.pdf.

Gaffney, E. and McCann, F. (2017) 'Credit risk under IFRS 9 accounting reforms : an application to Irish mortgages', (October 2017), pp. 1–25.

Garrett, G. (2000) 'The Causes of Globalization', *Comparative Political Studies*, 33(6/7), pp. 941–991. doi:10.4324/9781315251950-17.

Gestel, D.I.T. Van and Baesens, P.D.B. (2009) *Credit Risk Management Basic Concepts: Financial Risk Components, Rating Analysis, Models, Economic And Regulatory Capital*. New York, New York: Oxford University Press.

Gouvêa, M.A. (2007) 'Credit Risk Analysis Applying Logistic Regression, Neural Networks and Genetic Algorithms Models', in *Production and Operations Management Society, 18th Annual Conference*. Dallas, Texas, pp. 1–49. Available at: https://www.pomsmeetings.org/ConfPapers/007/007-0210.pdf.

Hao, C., Alam, M. and Carling, K. (2010) 'Review of the Literature on Credit Risk Modelling: Development of the Past 10 Years', *Banks and Bank Systems*, 5(3). Available at: https://businessperspectives.org/images/pdf/applications/publishing/templates/article/assets/3437/BBS_en_2010_3_Hao.pdf.

Harper, D. (2008) *Unexpected Losses (UL) of Credit Asset*, *Youtube*. Available at: https://www.youtube.com/watch?v=6Ddn3oKIOu8 (Accessed: 8 August 2018).

Honohan, P. (2016) *Bank Failures : The Limitations of Risk Modeling*. Chicago,. doi:10.1142/9789814280488.

Hull, J.C. (2012) *Options, Futures, and other Derivatives*. 8th edn. Edited by Battista Donna, T. O'Brien, and E. Senra-Sargen. New Jersey: Prentice Hall (Pearson Education Inc.).

Hyman, L. (2012) 'The Politics of Consumer Debt: U.S. State Policy and the Rise of Investment in Consumer Credit, 1920-2008', *Annals of the American Academy of Political and*

Social Science, 644(November), pp. 40–49. doi:10.1177/0002716212452721.

IASB (2019) *IFRS 9: Financial Instruments*. IFRS Foundation. doi:10.1007/978-3-030-29485-4_9.

International Monetary Fund (2010) ‘The Uses and Abuses of Sovereign Credit Ratings’, *Global Financial Stability Report*, (October), pp. 85–122.

International Monetary Fund (2022) *IMF Data*. Available at: https://www.imf.org/en/Data (Accessed: 17 May 2022).

Isaacs, W. (2014) *The Theory of Risk and Return, Allan Gray*. Available at: https://www.allangray.co.za/latest-insights/personal-investing/the-theory-of-risk-and-return/ (Accessed: 6 November 2018).

Ishikawa, T. (2009) *How I Caused the Credit Crunch*. London: Icon Books Limited.

Ivashina, V., Laeven, L. and Moral-Benito, E. (2020) *Loan Types and the Bank Lending Channel*. 27056. Boston, Massachussetts. doi:http://www.nber.org/papers/w27056.

J.P. Morgan (1997) *Introduction to CreditMetrics: The Benchmark for Understanding Credit Risk*. New York. Available at: http://homepages.rpi.edu/~guptaa/MGMT4370.09/Data/CreditMetricsIntro.pdf.

Jareño, F. and Negrut, L. (2016) ‘US stock market and macroeconomic factors’, *Journal of Applied Business Research*, 32(1), pp. 325–340. doi:10.19030/jabr.v32i1.9541.

Jarrow, R.A. (2012) ‘The Role of ABSs, CDSs, and CDOs in the Credit Crisis and the Economy’, *Rethinking the Financial Crisis*, pp. 210–234. Available at: chrome-extenhttps://www.russellsage.org/sites/all/files/Rethinking-Finance/Jarrow ABS CDS CDO 2.pdf.

Jayakumar, D.S., Sulthan, A. and Studies, I. (2014) ‘Identification of Multicollinearity and it ’ s Effect in Model Selection’, 7(September). doi:10.1285/i20705948v7n1p153.

Jin, J.Y. *et al.* (2021) ‘Banks’ Loan Charge-Offs and Macro-Level Risk’, *Journal of Behavioral and Experimental Finance*, 32(2021), pp. 1–20. doi:10.1016/j.jbef.2021.100573.

Joseph, A.D. (2018) ‘Challenges in Determining Credit Valuation Adjustment (CVA)’. Woodmead, Johannesburg.

Joseph, A.D. and Kimetto, G. (2020) ‘Emerging Market Default Risk Charge Modelling

Framework’, in *7th International Conference on Business Management (ICOBM) 2020*. Online: Association of Management Development Institutions in Pakistan (AMDIP). Available at: https://amdip.org.pk/Events/7th-International-Conference-on-Business-Management-(ICOBM)-2020.aspx.

Klieštik, T. and Cúg, J. (2015) ‘Comparison of Selected Models of Credit Risk’, *Procedia Economics and Finance*, 23(December), pp. 356–361. doi:10.1016/s2212-5671(15)00452-9.

KNBS (2022) *Kenya National Bureau of Statistics*. Available at: https://www.knbs.or.ke/.

Koulouridi (2020) ‘The Unique Features of the Pandemic-Triggered Recession Have Led Banks to Move More Quickly to Build Real-Time Data and Analytics into their Credit-Decision Engines’, *Mckinsey & Company*. Available at: https://www.mckinsey.com/business-functions/risk-and-resilience/our-insights/managing-and-monitoring-credit-risk-after-the-covid-19-pandemic.

Kruger, J.W., Joseph, A.D. and Aphane, A. (2012) ‘Oil and Gold Price Volatility’, *SSRN Electronic Journal* [Preprint], (August 2018). doi:10.2139/ssrn.1935192.

Langley, P. (2008) ‘Financialization and the Consumer Credit Boom’, *Competition & Change*, 12(2), pp. 133–147. doi:10.1179/102452908X289794.

Lawson, T. (2015) ‘Soros Theory of Reflexivity: A Critical Comment’, *Revue de Philosophie Économique*, 14(January), pp. 29–48. Available at: https://www.cairn.info/revue-de-philosophie-economique-2013-1-page-29.htm#.

Leedy, P. and Ormrod, J. (2015) *Practical Research, Planning and Design*. 10th edn. Upper Saddle River, New Jersey: Prentice Hall (Pearson Education Inc.).

Liew, J. (2014) ‘The Great Divide over Market Efficiency’, (August 2007), pp. 1–17.

Lowenstein, R. (2011) *When Genius Failed: The Rise and Fall of Long-Term Capital Management*. New York: Random House Trade Paperbacks.

Malkiel, B.G. (2003) *The Efficient Market Hypothesis and its Critics*.

Marks, H. (2006) ‘Risk’. New Jersey, pp. 1–13. Available at: https://www.oaktreecapital.com/insights/ howard-marks-memos.

Marks, H. (2022) ‘Bull Market Rhymes’. New York, New York: Oak Tree Capital Management L.P.

McNeil, A., Frey, R. and Embrechts, P. (2005) *Quantitative Risk Management: Concepts, Techniques and Tools,*. Princeton, New Jersey: Princeton University Press.

Mendenhall, W. and Sincich, T. (2020) *A Second Course in Statistics - Regression Analysis*. 8th edn. Hoboken, New Jersey: Pearson Education Inc.

Mendez-Carbajo, D. (2013a) *Data Revisions with FRED, Reserve, Federal Louis, St*. doi:10.5860/choice.51-1007.

Mendez-Carbajo, D. (2013b) *Ethical Use of Data With FRED, Reserve, Federal Louis, St*. doi:10.5860/choice.51-1007.

Meng, Y. and Yin, C. (2019) 'Trust and the Cost of Debt Financing', *Journal of International Financial Markets, Institutions and Money* [Preprint]. doi:https://doi.org/10.1016/j.intfin.2018.11.009 Available.

Merton, R.C. (1974) 'On the Pricing of Corporate Debt: The Risk Structure of Interest Rates', *The Journal of Finance*, 29(2), pp. 449–470. Available at: http://www.people.hbs.edu/rmerton/ Pricing%25 20of%2520corporate%2520debt.pdf.

Mester, L.J. (2015) 'What is the Point of Credit Scoring ?', *Business Review*. Philadelphia, Pennsylvania: Federal Reserve Bank of Philadelphia, pp. 3–16.

Mohajan, H. (2017) 'Two Criteria for Good Measurements in Research : Validity and Reliability', (83458).

Mundfrom, D., DePoy Smith, M. and Kay, L. (2018) 'The Effect of Multicollinearity on Prediction in Regression Models', *General Linear Model Journal*, 44(1), pp. 24–28. doi:10.31523/glmj.044001.003.

National Credit Regulator (2022) *NCR Annual Report*. Available at: https://www.ncr.org.za/documents/NCR Annual Report 2017.pdf (Accessed: 17 May 2022).

NCR (2018) *Annual Report 2016/2017*. Available at: http://www.ncr.org.za/annual-reports.

Nehrebecka, N. (2021) 'Internal Credit Risk Models and Digital Transformation : What to Prepare for ? An Application to Poland', *European Research Studies Journal*, XXIV(3), pp. 719–736.

Oliver, R.M. and Hand, D. (2005) 'A Survey of the Issues in Consumer Credit Modelling Research', *Journal of the Operational Research Society*, 56(April 2018), pp. 1006–1015.

doi:10.1057/palgrave.jors.2602018.

Organisation for Economic Cooperation and Development (2022) *OECD Data*. Available at: https://data.oecd.org/ (Accessed: 17 May 2022).

Peng, C.Y.J., Lee, K.L. and Ingersoll, G.M. (2002) 'An introduction to Logistic Regression Analysis and Reporting', *Journal of Educational Research*, 96(1), pp. 3–14. doi:10.1080/00220670209598786.

Prorokowski, L. (2018) 'IFRS 9 in credit risk modelling', 49(August 2017), pp. 639–670.

Rattner, S. (2013) 'Who's Right on the Stock Market?', *New York Times*, 14 November. Available at: https://www.nytimes.com/2013/11/15/opinion/rattner-whos-right-on-the-stock-market.html.

Renke, S., Roestoff, M. and Haupt, F. (2007) 'The National Credit Act : New Parameters for the Granting of Credit in South Africa', *Obiter*, pp. 229–270.

Republic of South Africa (2006) 'National Credit Act No. 34 of 2005', *Government Gazette*, 489(28619), pp. 4–6. Available at: https://www.saflii.org/za/legis/num_act/nca2005152.pdf.

Rossi, M. and Gunardi, A. (2018) 'Efficient Market Hypothesis and Stock Market Anomalies_ Empirical Evidence in Four European Countries', *The Journal of Applied Business Research*, 34(183–191).

Russel, P.S. and Torbey, V.M. (2002) 'The Efficient Market Hypothesis on Trial', *A journal of applied topics in business and economics*, pp. 1–16. doi:10.2469/faj.v42.n2.58.

Sanchez, J. (2017) 'The Information Technology Revolution and the Unsecured Credit Market', *Economic Enquiry* [Preprint]. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1111/ecin.12519.

Sanders, E.. (1992) *Judaism: Practice and Belief (63BCE-66CE)*. Philadelphia, Pennsylvania: Trinity Press International.

SARB (2020) *Statistics*. Available at: https://www.resbank.co.za/en/home/what-we-do/statistics (Accessed: 17 May 2022).

Sauder, H.O.-M. and Penas, M.F. (2006) *Lending to Small Businesses : The Role of Loan Maturity in Addressing Information Problems*. Vancouver, Canada. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=509882.

Sgambati, S. (2022) ‘Who Owes? Class Struggle, Inequality and the Political Economy of Leverage in the Twenty-First Century’, *Finance and Society*, 8(1), pp. 1–21. doi:10.2218/finsoc.7115.

Sheffield, S. (no date) ‘Permutations and Combinations, Pascal’s Triangle, Learning to Count’. Boston, Massachusetts: Massachusetts Instititute of Technology. Available at: https://math.mit.edu/~sheffield/2018600/mergedlectures.pdf.

Shrestha, N. (2020) ‘Detecting Multicollinearity in Regression Analysis’, (June), pp. 1–5. doi:10.12691/ajams-8-2-1.

Simpson, J. and Weiner, E. (2022) *Oxford English Dictionary*. Edited by D. Salazar. London, United Kingdom: Oxford University Press. Available at: https://www.oed.com/.

Smit, L., Swart, B. and Niekerk, F. van (2003) ‘Credit Risk Models in the South African Context’, *Investment Analysts Journal*, January(57), pp. 41–46.

Soros, G. (1992) *The Alchemy of Finance*. New York, New York: John Wiley & Sons, Inc.

Soros, G. (2003) ‘The Bubble of American Supremacy’, *The Atlantic Online* [Preprint], (December). Available at: http://www.theatlantic.com/issues/2003/12/soros.htm.

Soros, G. (2012) *The Soros Lectures at the Central European University, European University Institute*. New York, New York: PublicAffairs™. Available at: https://www.georgesoros.com/wp-content/uploads/2017/ 10/thesoroslectures-human _uncertainty_principle-2017_10_05.pdf.

Soros, G. (2014) ‘Fallibility, Reflexivity, and the Human Uncertainty Principle’, *Journal of Economic Methodology*, January, pp. 1–30.

Swartz, N.P. (2010) ‘The Prevention of Over-Indebtedness: The Problem of Interest and the Islamic Response’, in *International Conference on Over-indebtedness and Credit Regulation*. Pretoria, South Africa, pp. 1–17.

Takawira, O. and Mwamba, M. (2022) ‘An Analysis of Sovereign Credit Ratings Using Random Forest’, *International Journal Of Economics And Finance Studies*, (April), pp. 29–87. doi:10.34109/ijefs. 202220002.

Thornton, S. (2022) ‘Karl Popper’, *Stanford Encyclopedia of Philosophy*, pp. 1–23.

Venditti, F. and Veronese, G.F. (2021) ‘Global Financial Markets and Oil Price Shocks in Real

Time', *SSRN Electronic Journal* [Preprint], (2472). doi:10.2139/ssrn.3701652.

Ventura, J. and Voth, H.-J. (2016) *Debt into Growth : How Sovereign Debt Accelerated the First Industrial Revolution*. 2180. Barcelona: The National Bureau of Economic Research. Available at: http://www.nber.org/papers/w21280.

Walliman, N. (2011) *Research Methods: The Basics*. New York, New York: Routledge. Available at: www.eBookstore.tandf.co.uk.

Wisse, M. and Roeland, J. (2022) 'Building Blocks for Developing a Research Question : The ABC-Model', *Teaching Theology & Religion*, (March), pp. 22–34. doi:10.1111/teth.12603.

World Bank (2022) *The Global Findex database*. Available at: https://globalfindex.worldbank.org/ (Accessed: 17 May 2022).

ANNEXURE A: RESEARCH OUTPUT


7th International Conference on Business Management

December 19-20, 2020 | Online

School of Business and Economics

AMDIP

Association of Management Development Institutions in Pakistan

Certificate of Presentation

This is to certify that **Angelo Joseph** presented online the paper titled

Emerging Market Default Risk Charge Modelling Framework

authored by

Angelo Joseph, and Geoffrey Kimetto

in the conference

Ijaz Yusuf
Chief Organizer
7th ICoBM

Memoona Zareen
Secretary
AMDIP

ANNEXURE B: CREDIT RATINGS: MOODY’S, FITCH, AND STANDARD & POOR’S

Sortable Table Key	Moody's	Fitch	S&P
Highest grade credit	Aaa	AAA	AAA
Very high grade credit	Aa1, Aa2, Aa3	AA+, AA, AA-	AA+, AA, AA-
High grade credit	A1, A2, A3	A+, A, A-	A+, A, A-
Good credit grade	Baa1, Baa2, Baa3, Baa4	BBB+, BBB, BBB-	BBB+, BBB, BBB-
Speculative grade credit	Ba1, Ba2, Ba3	BB+, BB, BB-	BB+, BB, BB-
Very speculative credit	B1, B2, B3	B+, B, B-	B+, B, B-
Substantial risks - In default	Caa1, Caa2, Caa3, Ca	CCC, CC, C, RD, D	CCC+, CCC, CCC-, CC, C, D

Source: http://s.wsj.net/media/EURODEBTRATINGS1106_key.png.

The three main credit rating agencies rate credit instruments, other investment securities, companies, and countries (Annexure B).

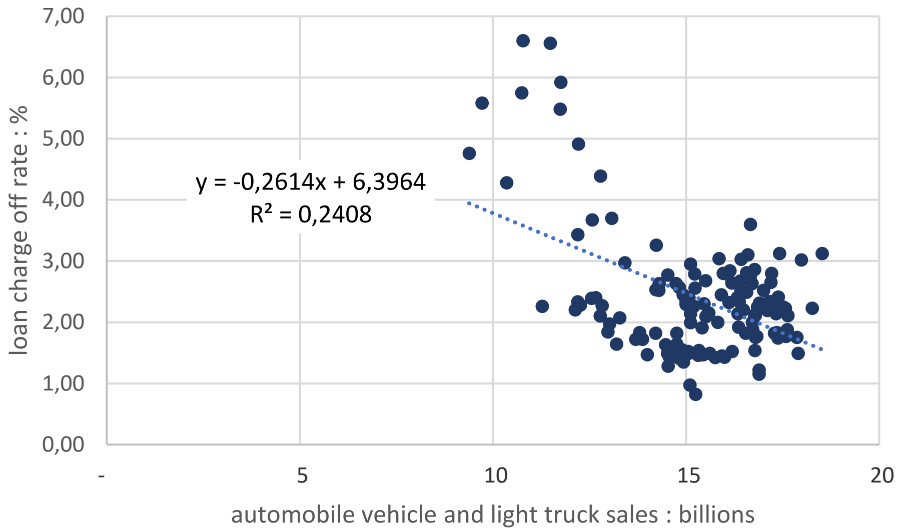
ANNEXURE C: S&P’s CREDIT RATING MIGRATION TABLE

Rating at year end		AAA	AA	A	BBB	BB	B	CCC	Default
Initial rating	AAA	90.81	8.33	0.68	0.06	0.12	0.00	0.00	0.00
	AA	0.70	90.65	7.79	0.64	0.06	0.14	0.02	0.00
	A	0.09	2.27	91.05	5.52	0.74	0.26	0.01	0.06
	BBB	0.02	0.33	5.95	86.93	5.30	1.17	1.12	0.18
	BB	0.03	0.14	0.67	7.73	80.53	8.84	1.00	1.06
	B	0.00	0.11	0.24	0.43	6.48	83.46	4.07	5.20
	CCC	0.22	0.00	0.22	1.30	2.38	11.24	64.86	19.79

Source: S&P’s CreditWeek (15 April 1996).

The matrix of Annexure C shows the probability of migrating from one S&P credit rating quality to another within one year.

ANNEXURE D: SCATTER DIAGRAMS AND BIVARIABLE REGRESSION ANALYSIS (US DATA)



Consumer Loan Charge Off Rate and Automobile Vehicle and Light Truck Sales

(Data from: FRED, Federal Reserve Bank of St. Louis, US).

Consumer Loan Charge Off Rate and Auto Vehicle and Light Truck Sales are negatively correlated.

Regression and ANOVA Analysis Of Consumer Loan Charge Off Rate and Automobile Vehicle and Light Truck Sales

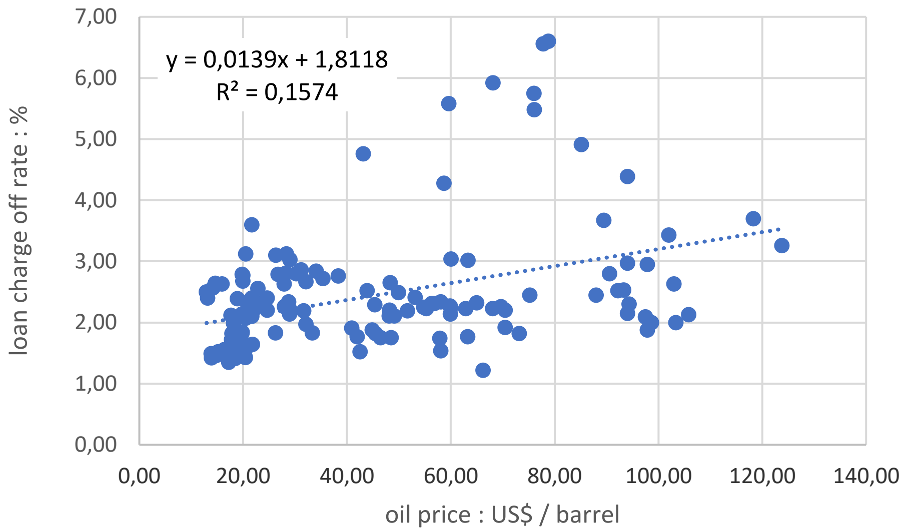
Regression Statistics	
Multiple R	0,491
R Square	0,241
Adjusted R Square	0,236
Standard Error	0,883
Observations	146

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	35,649	35,649	45,671	0,000
Residual	144	112,401	0,781		
Total	145	148,051			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95,0%*	*Upper 95,0%*
Intercept	6,396	0,597	10,707	0,000	5,216	7,577	5,216	7,577
autovehicle and light truck sales	- 0,261	0,039	- 6,758	0,000	- 0,338	- 0,185	- 0,338	- 0,185

The explanatory variable is significant as the p-value is <0.05. It explains 23.6% of the variation in credit losses.



Consumer Loan Charge Off Rate and WTI Oil Price

(Data from: FRED, Federal Reserve Bank of St. Louis, US).

Consumer Loan Charge Off Rate and WTI Oil Price are positively correlated.

Regression and ANOVA Analysis of Consumer Loan Charge Off Rate and WTI Oil Price

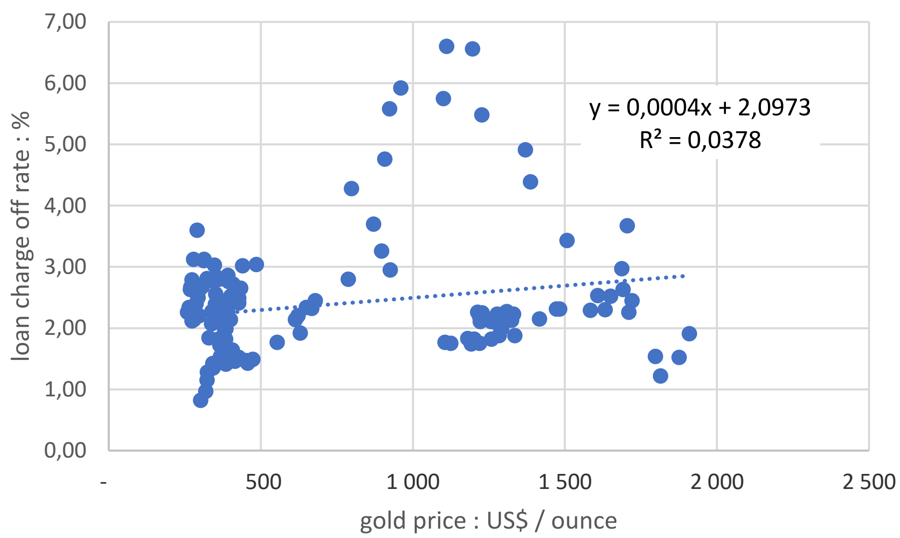
Regression Statistics	
Multiple R	0,397
R Square	0,157
Adjusted R Square	0,151
Standard Error	0,920
Observations	142

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	22,131	22,131	26,152	0,000
Residual	140	118,475	0,846		
Total	141	140,606			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95,0%*	*Upper 95,0%*
Intercept	1,812	0,143	12,676	0,000	1,529	2,094	1,529	2,094
Oil price - US$ / barrel	0,014	0,003	5,114	0,000	0,009	0,019	0,009	0,019

The explanatory variable is significant as the p-value is <0.05. It explains 15.1 % of the variation in credit losses.



Consumer Loan Charge Off Rate and Gold Price

(Data from: FRED, Federal Reserve Bank of St. Louis, US).

Consumer Loan Charge Off Rate and Gold Price are positively correlated.

Regression and ANOVA Analysis of Consumer Loan Charge Off Rate and Gold Price

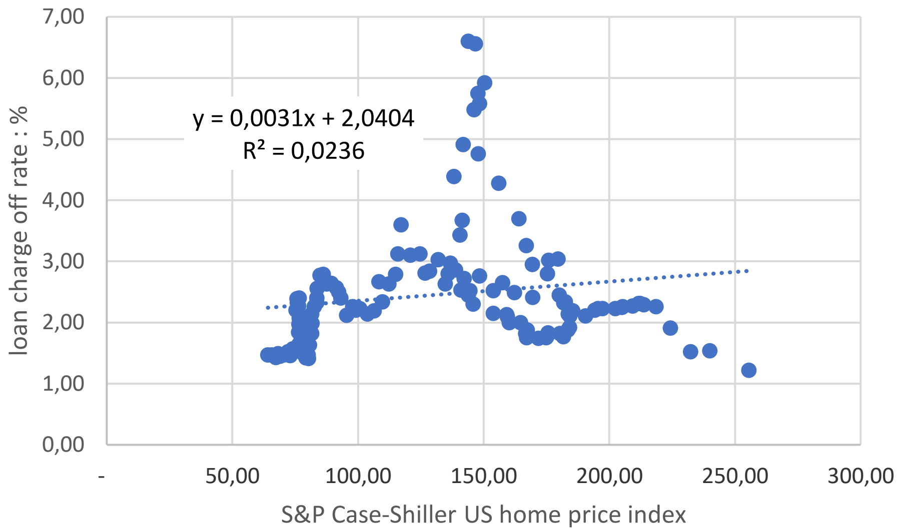
Regression Statistics	
Multiple R	0,194
R Square	0,038
Adjusted R Square	0,031
Standard Error	0,995
Observations	146

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	5,590	5,590	5,650	0,019
Residual	144	142,461	0,989		
Total	145	148,051			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95,0%*	*Upper 95,0%*
Intercept	2,097	0,148	14,179	0,000	1,805	2,390	1,805	2,390
Gold price - US$ / ounce	0,000	0,000	2,377	0,019	0,000	0,001	0,000	0,001

The explanatory variable is significant as the p-value is <0.05. It explains 3.1 % of the variation in credit losses.



Consumer Loan Charge Off Rate and The S&P Case-Shiller Home Price Index

(Data from: FRED, Federal Reserve Bank of St. Louis, US).

Consumer Loan Charge Off Rate and the S&P Case-Shiller Home Price Index appear to be positively correlated. On further examination, that correlation is found to be minimal and not significant. The p-value is > 0.05 at 0.072.

Regression and ANOVA Analysis Of Consumer Loan Charge Off Rate and The S&P Case-Shiller Home Price Index

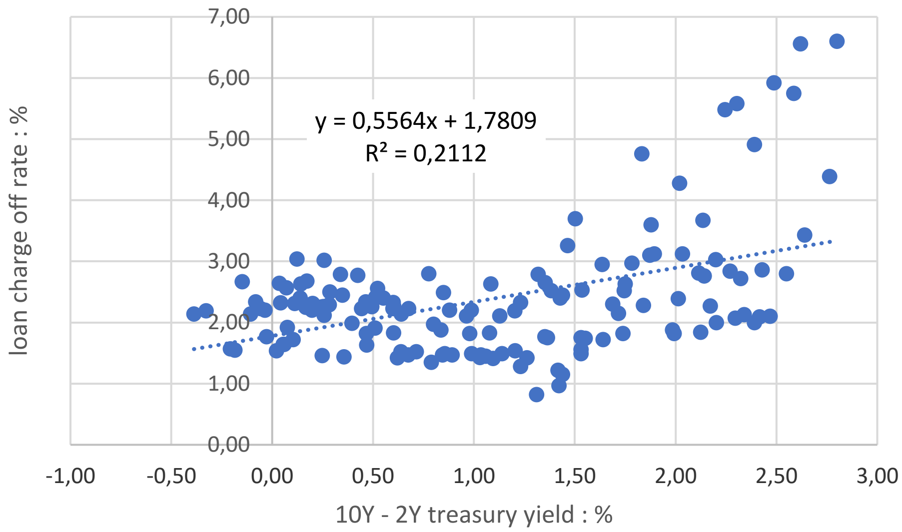
Regression Statistics	
Multiple R	0,154
R Square	0,024
Adjusted R Square	0,016
Standard Error	0,990
Observations	138

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	3,223	3,223	3,287	0,072
Residual	136	133,339	0,980		
Total	137	136,562			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95,0%*	*Upper 95,0%*
Intercept	2,040	0,244	8,361	0,000	1,558	2,523	1,558	2,523
CSUSPI	0,003	0,002	1,813	0,072	- 0,000	0,007	- 0,000	0,007

The explanatory variable is insignificant as the p-value is >0.05. It explains 1.6 % of the variation in credit losses.



Consumer Loan Charge Off Rate and 10Y – 2Y Treasury Yield

(Data from: FRED, Federal Reserve Bank of St. Louis, US).

Consumer Loan Charge Off Rate and 10Y-2Y Treasury Constant Maturity are positively correlated.

Regression and ANOVA Analysis of Consumer Loan Charge Off Rate and The 10Y-2Y Constant Maturity (Yield)

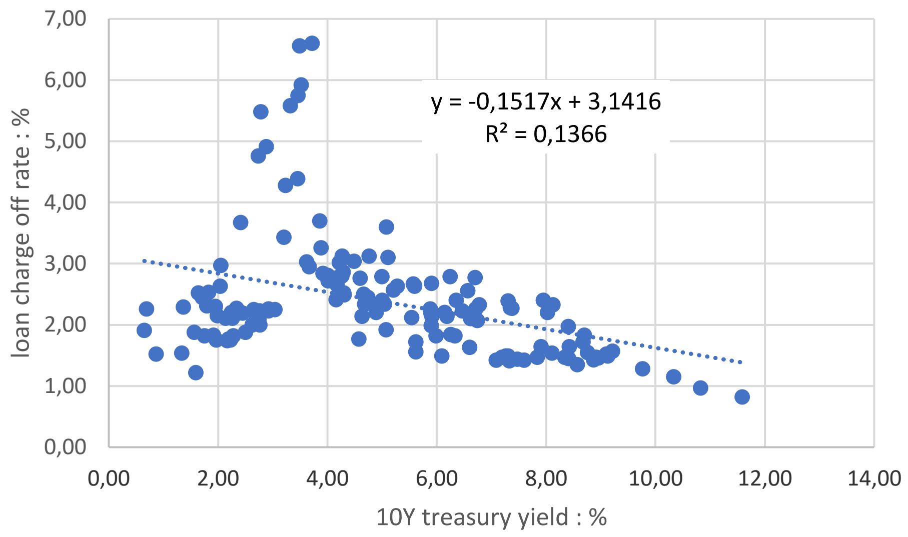
Regression Statistics	
Multiple R	0,460
R Square	0,211
Adjusted R Square	0,206
Standard Error	0,901
Observations	146

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	31,272	31,272	38,561	0,000
Residual	144	116,779	0,811		
Total	145	148,051			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95,0%*	*Upper 95,0%*
Intercept	1,781	0,123	14,466	0,000	1,538	2,024	1,538	2,024
10Y-2Y treasury yield	0,556	0,090	6,210	0,000	0,379	0,733	0,379	0,733

The explanatory variable is significant as the p-value is <0.05. It explains 20.6 % of the variation in credit losses.



Consumer Loan Charge Off Rate and 10Y Treasury Yield

(Data from: FRED, Federal Reserve Bank of St. Louis, US).

Consumer Loan Charge Off Rate and 10Ytreasury constant maturity are negatively correlated.

Regression and ANOVA Analysis of Consumer Loan Charge Off Rate and The 10Y Constant Maturity (Yield)

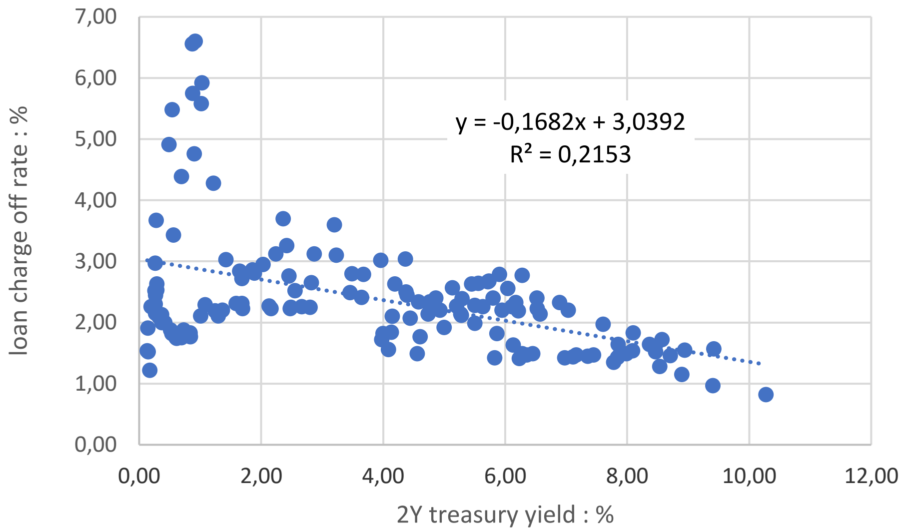
Regression Statistics	
Multiple R	0,370
R Square	0,137
Adjusted R Square	0,131
Standard Error	0,942
Observations	146

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	20,223	20,223	22,781	0,000
Residual	144	127,828	0,888		
Total	145	148,051			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95,0%	Upper 95,0%
Intercept	3,142	0,176	17,866	0,000	2,794	3,489	2,794	3,489
10Y treasury yield	- 0,152	0,032	- 4,773	0,000	- 0,215	- 0,089	- 0,215	- 0,089

The explanatory variable is significant as the p-value is <0.05. It explains 13.1 % of the variation in credit losses.



Consumer Loan Charge Off Rate and 2Y Treasury Yield

(Data from: FRED, Federal Reserve Bank of St. Louis, US).

Consumer Loan Charge Off Rate and 2Ytreasury Constant Maturity are negatively correlated.

Regression and ANOVA Analysis of Consumer Loan Charge Off Rate and the 2Y Constant Maturity (Yield)

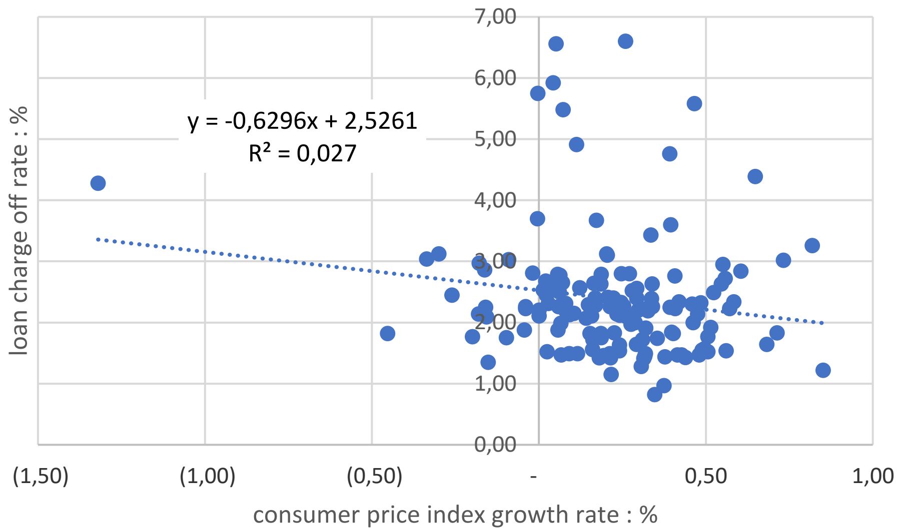
Regression Statistics	
Multiple R	0,464
R Square	0,215
Adjusted R Square	0,210
Standard Error	0,898
Observations	146

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	31,869	31,869	39,499	0,000
Residual	144	116,182	0,807		
Total	145	148,051			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95,0%*	*Upper 95,0%*
Intercept	3,039	0,127	23,865	0,000	2,788	3,291	2,788	3,291
2Y treasury yield	- 0,168	0,027	- 6,285	0,000	- 0,221	- 0,115	- 0,221	- 0,115

The explanatory variable is significant as the p-value is <0.05. It explains 21 % of the variation in credit losses.



Consumer Loan Charge Off Rate and Consumer Price Index Growth

(Data from: FRED, Federal Reserve Bank of St. Louis, US).

Consumer Loan Charge Off Rate and Consumer Price Index Growth Rate are positively correlated.

Regression and ANOVA Analysis of Consumer Loan Charge Off Rate and Consumer Price Index Growth Rate

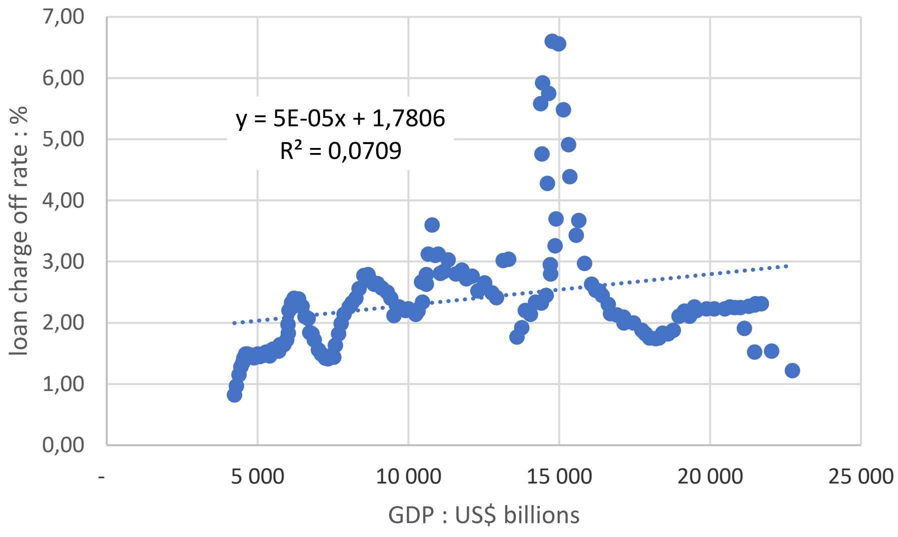
Regression Statistics	
Multiple R	0,164
R Square	0,027
Adjusted R Square	0,020
Standard Error	1,000
Observations	146

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	3,999	3,999	3,997	0,047
Residual	144	144,052	1,000		
Total	145	148,051			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95,0%*	*Upper 95,0%*
Intercept	2,526	0,107	23,527	0,000	2,314	2,738	2,314	2,738
Consumer price index growth rate	- 0,630	0,315	- 1,999	0,047	- 1,252	- 0,007	- 1,252	- 0,007

The explanatory variable is significant as the p-value is <0.05. It explains 2.0 % of the variation in credit losses.



Consumer Loan Charge Off Rate and GDP

(Data from: FRED, Federal Reserve Bank of St. Louis, US).

Consumer Loan Charge Off Rate and GDP are positively correlated.

Regression and ANOVA Analysis of Consumer Loan Charge Off Rate and GDP

Regression Statistics	
Multiple R	0,266
R Square	0,071
Adjusted R Square	0,064
Standard Error	0,977
Observations	146

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	10,503	10,503	10,996	0,001
Residual	144	137,548	0,955		
Total	145	148,051			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95,0%*	*Upper 95,0%*
Intercept	1,781	0,201	8,875	0,000	1,384	2,177	1,384	2,177
GDP	0,000	0,000	3,316	0,001	0,000	0,000	0,000	0,000

The explanatory variable is significant as the p-value is <0.05. It explains 6.4 % of the variation in credit losses.

Consumer Loan Charge Off Rate and Unemployment Rate

(Data from: FRED, Federal Reserve Bank of St. Louis, US).

Consumer Loan Charge Off Rate and the Unemployment Rate are positively correlated.

Regression and ANOVA Analysis Of Consumer Loan Charge Off Rate and the Unemployment Rate

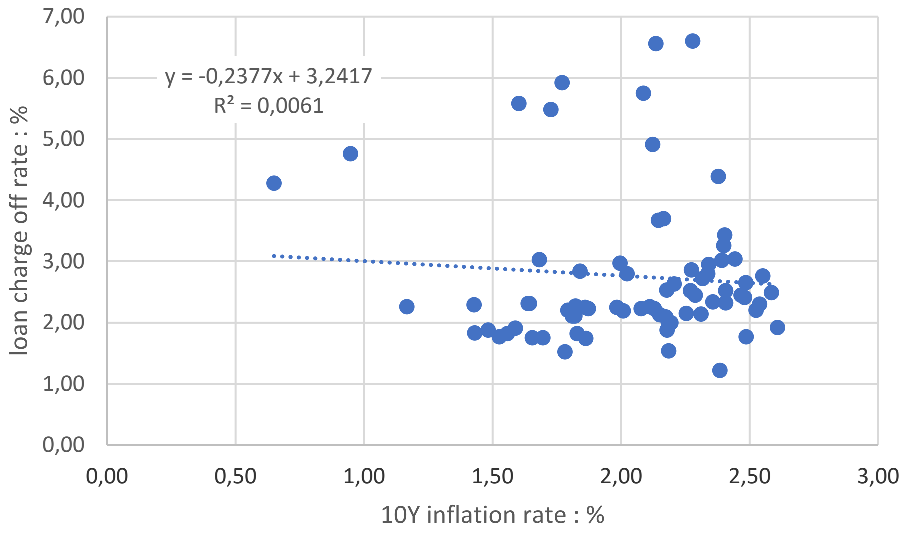
Regression Statistics	
Multiple R	0,426
R Square	0,181
Adjusted R Square	0,176
Standard Error	0,917
Observations	146

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	26,865	26,865	31,923	0,000
Residual	144	121,185	0,842		
Total	145	148,051			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95,0%*	*Upper 95,0%*
Intercept	0,811	0,289	2,803	0,006	0,239	1,383	0,239	1,383
Unemployment rate	0,265	0,047	5,650	0,000	0,172	0,357	0,172	0,357

The explanatory variable is significant as the p-value is <0.05. It explains 17.6 % of the variation in credit losses.



Consumer Loan Charge Off Rate and Inflation Rate

(Data from: FRED, Federal Reserve Bank of St. Louis, US).

Consumer Loan Charge Off Rate and The Inflation Rate are negatively correlated.

Regression and ANOVA Analysis of Consumer Loan Charge Off Rate and The Inflation Rate

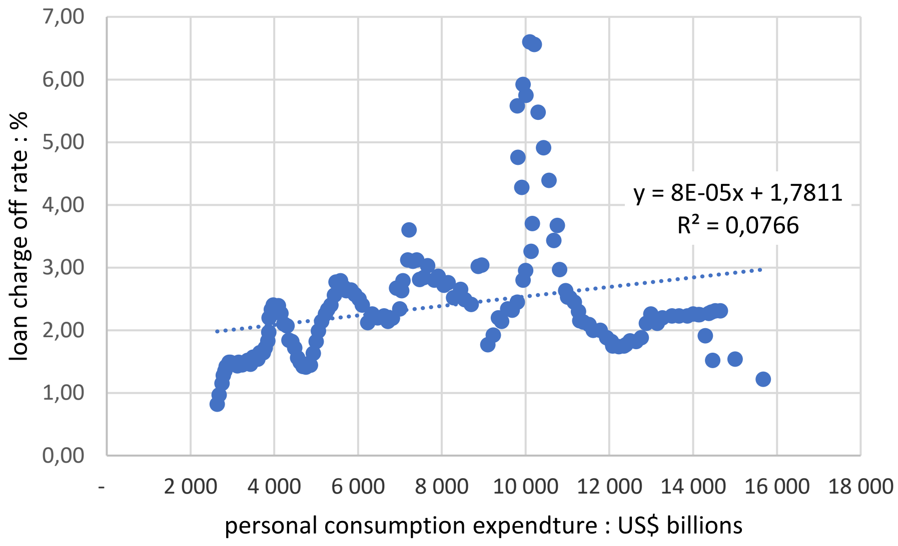
Regression Statistics	
Multiple R	0,078
R Square	0,006
Adjusted R Square	- 0,008
Standard Error	1,200
Observations	74

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0,638	0,638	0,443	0,508
Residual	72	103,641	1,439		
Total	73	104,279			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95,0%*	*Upper 95,0%*
Intercept	3,242	0,741	4,374	0,000	1,764	4,719	1,764	4,719
10Y Inflation rate	- 0,238	0,357	- 0,666	0,508	- 0,949	0,474	- 0,949	0,474

The explanatory variable is insignificant as the p-value is >0.05. It explains 0.8 % of the variation in credit losses.



Consumer Loan Charge Off Rate and Personal Consumption Expenditure

(Data from: FRED, Federal Reserve Bank of St. Louis, US).

Consumer Loan Charge Off Rate and Personal Consumption Expenditure are negatively correlated.

Regression and ANOVA Analysis Of Consumer Loan Charge Off Rate and Personal Consumption Expenditure

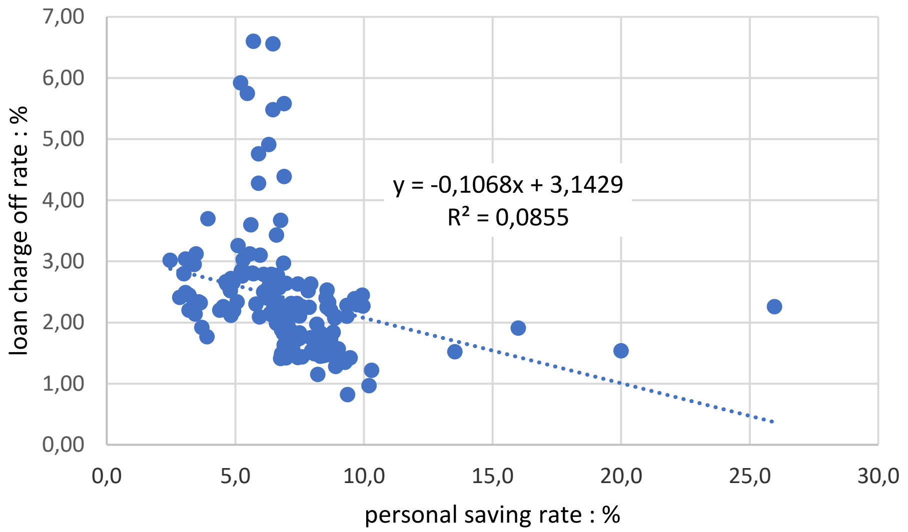
Regression Statistics	
Multiple R	0,277
R Square	0,077
Adjusted R Square	0,070
Standard Error	0,974
Observations	146

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	11,348	11,348	11,954	0,001
Residual	144	136,703	0,949		
Total	145	148,051			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95,0%*	*Upper 95,0%*
Intercept	1,781	0,194	9,202	0,000	1,399	2,164	1,399	2,164
PCE	0,000	0,000	3,457	0,001	0,000	0,000	0,000	0,000

The explanatory variable is significant as the p-value is <0.05. It explains 7 % of the variation in credit losses.



Consumer Loan Charge Off Rate and Personal Saving Rate

(Data from: FRED, Federal Reserve Bank of St. Louis, US).

Consumer Loan Charge Off Rate and Personal Saving Rate are negatively correlated.

Regression and ANOVA Analysis Of Consumer Loan Charge Off Rate and Personal Consumption Expenditure

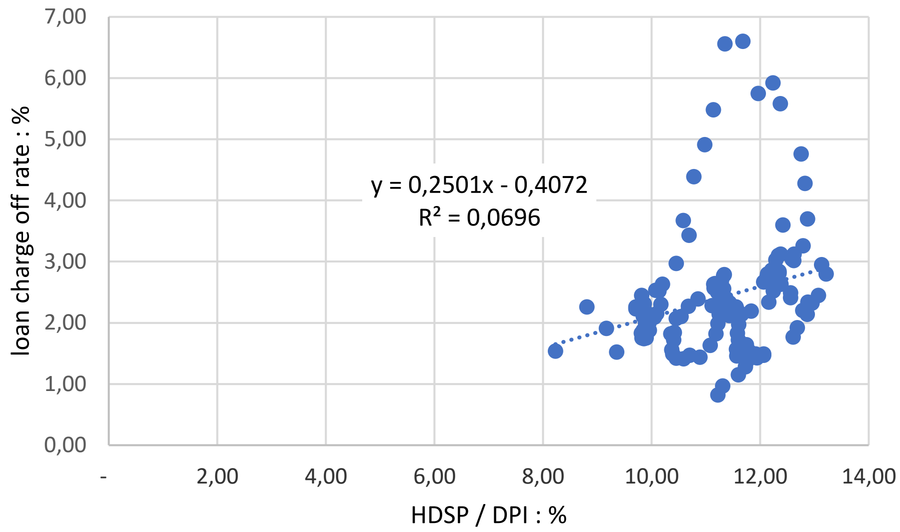
Regression Statistics	
Multiple R	0,292
R Square	0,085
Adjusted R Square	0,079
Standard Error	0,970
Observations	146

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	12,655	12,655	13,459	0,000
Residual	144	135,396	0,940		
Total	145	148,051			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95,0%*	*Upper 95,0%*
Intercept	3,143	0,221	14,253	0,000	2,707	3,579	2,707	3,579
PSAVER	- 0,107	0,029	- 3,669	0,000	- 0,164	- 0,049	- 0,164	- 0,049

The explanatory variable is significant as the p-value is <0.05. It explains 7.9 % of the variation in credit losses.



Consumer Loan Charge Off Rate and Household Debt Service Payment as a Percentage of Disposable Personal Income

(Data from: FRED, Federal Reserve Bank of St. Louis, US).

Consumer Loan Charge Off Rate and Household DSP/DPI are positively correlated.

Regression and ANOVA Analysis of Consumer Loan Charge Off Rate and Household DSP/DPI

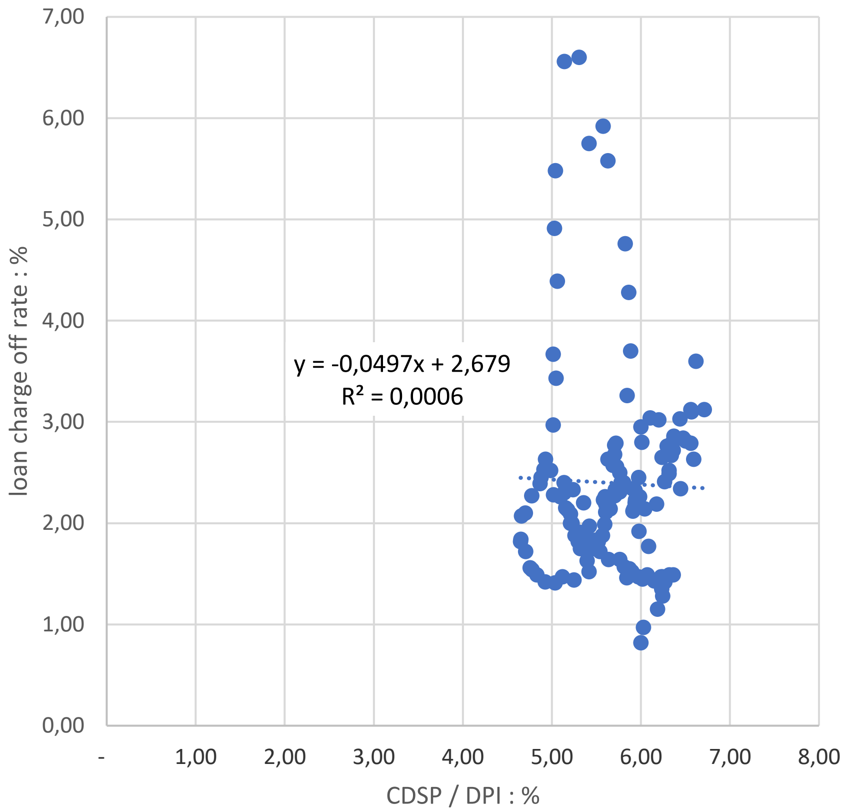
Regression Statistics	
Multiple R	0,264
R Square	0,070
Adjusted R Square	0,063
Standard Error	0,977
Observations	145

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	10,205	10,205	10,693	0,001
Residual	143	136,469	0,954		
Total	144	146,674			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95,0%*	*Upper 95,0%*
Intercept	- 0,407	0,861	- 0,473	0,637	- 2,110	1,296	- 2,110	1,296
HDSP / DPI	0,250	0,076	3,270	0,001	0,099	0,401	0,099	0,401

The explanatory variable is significant as the p-value is <0.05. It explains 6.3 % of the variation in credit losses.



Consumer Loan Charge Off Rate and Consumer Debt Service Payment as a Percentage of Disposable Personal Income

(Data from: FRED, Federal Reserve Bank of St. Louis, US).

Consumer Loan Charge Off Rate and Consumer DSP/DPI are negatively correlated.

Regression and ANOVA Analysis of Consumer Loan Charge Off Rate and Consumer DSP/DPI

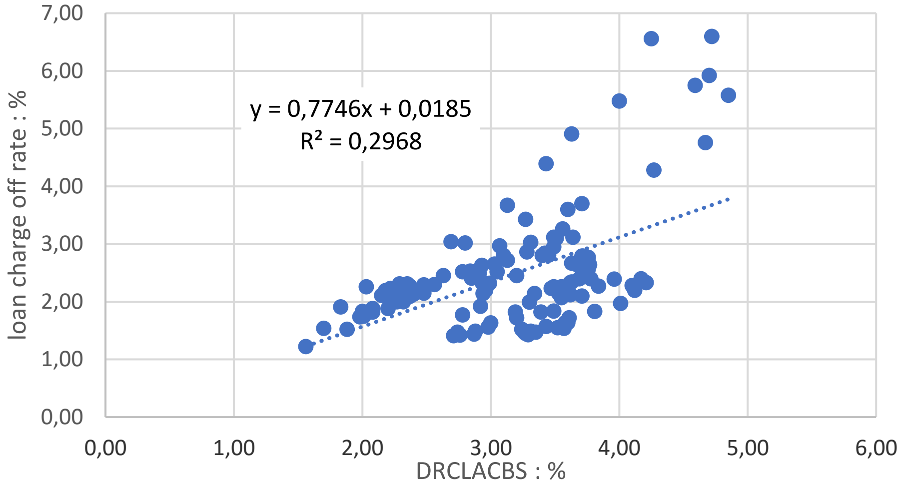
Regression Statistics	
Multiple R	0,025
R Square	0,001
Adjusted R Square	- 0,006
Standard Error	1,012
Observations	145

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0,091	0,091	0,089	0,766
Residual	143	146,583	1,025		
Total	144	146,674			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95,0%*	*Upper 95,0%*
Intercept	2,679	0,949	2,823	0,005	0,803	4,555	0,803	4,555
CDSP dv DPI	- 0,050	0,167	-0,298	0,766	- 0,379	0,280	- 0,379	0,280

The explanatory variable is insignificant as the p-value is >0.05. It explains 0.6 % of the variation in credit losses.



Consumer Loan Charge Off Rate and Delinquency Rate on Consumer Loans for all US Commercial Banks

(Data from: FRED, Federal Reserve Bank of St. Louis, US).

Consumer Loan Charge Off Rate and Delinquency Rate on consumer loans for all US commercial banks are positively correlated.

Regression and ANOVA Analysis of Consumer Loan Charge Off Rate and Consumer DSP/DPI

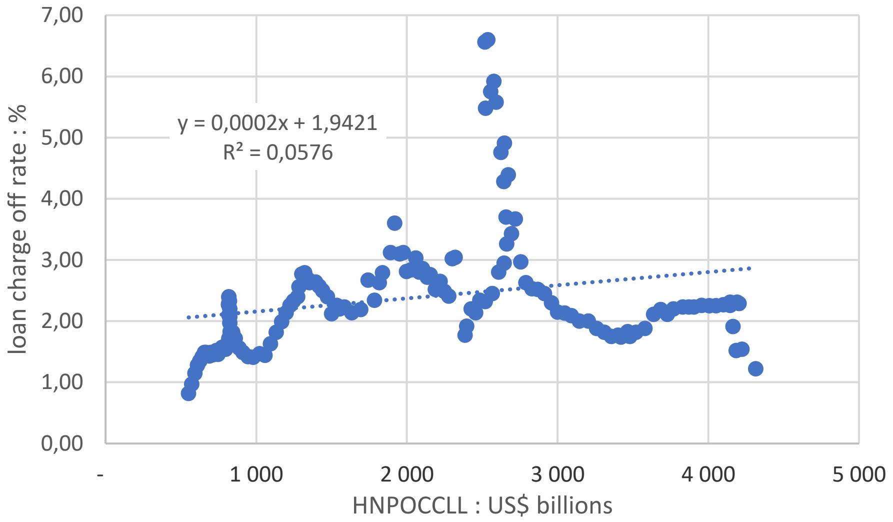
Regression Statistics	
Multiple R	0,545
R Square	0,297
Adjusted R Square	0,292
Standard Error	0,840
Observations	138

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	40,534	40,534	57,406	0,000
Residual	136	96,028	0,706		
Total	137	136,562			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95,0%*	*Upper 95,0%*
Intercept	0,019	0,330	0,056	0,955	- 0,633	0,670	- 0,633	0,670
DRCLACBS	0,775	0,102	7,577	0,000	0,572	0,977	0,572	0,977

The explanatory variable is significant as the p-value is <0.05. It explains 29.2 % of the variation in credit losses.



Consumer Loan Charge Off Rate and Household and Non-Profit Companies' Consumer Credit Liability Level

(Data from: FRED, Federal Reserve Bank of St. Louis, US).

Consumer Loan Charge Off Rate And Household and Non-Profit Companies' Consumer Credit Liability Level are positively correlated.

Regression and ANOVA Analysis Of Consumer Loan Charge Off Rate and Household and Non-Profit Companies' Consumer Credit Liability Level

Regression Statistics	
Multiple R	0,240
R Square	0,058
Adjusted R Square	0,051
Standard Error	0,984
Observations	146

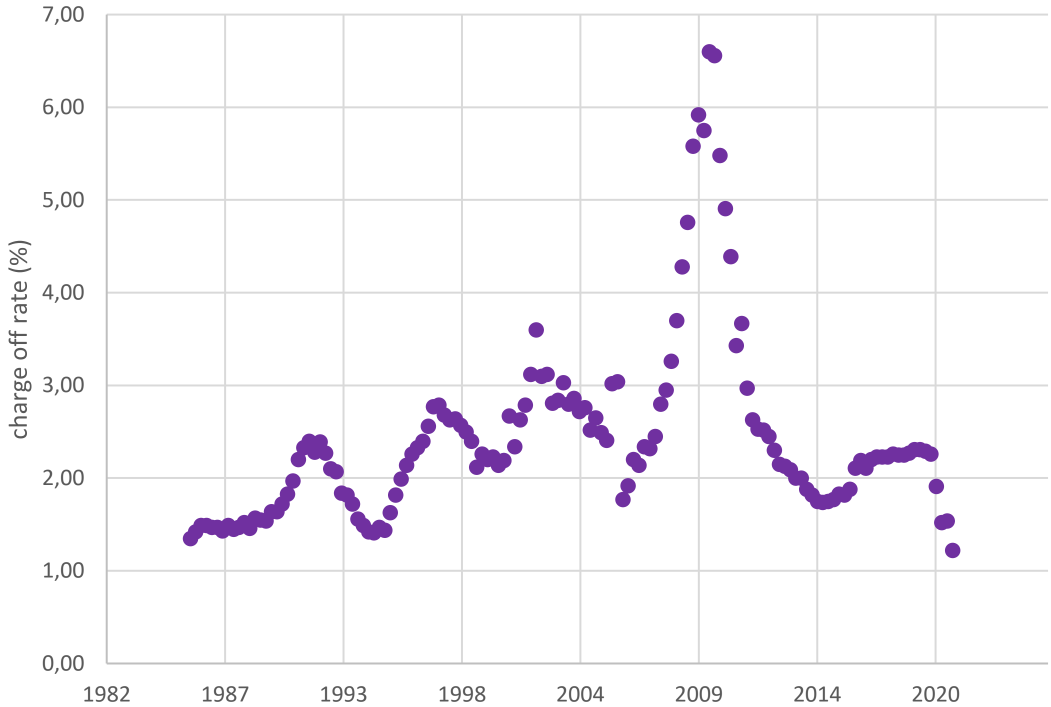
ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	8,522	8,522	8,795	0,004
Residual	144	139,528	0,969		
Total	145	148,051			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95,0%*	*Upper 95,0%*
Intercept	1,942	0,171	11,329	0,000	1,603	2,281	1,603	2,281
HNPOCCLL	0,000	0,000	2,966	0,004	0,000	0,000	0,000	0,000

The explanatory variable is significant as the p-value is <0.05. It explains 5.1 % of the variation in credit losses.

ANNEXURE E: CHARGE OFF RATE AS A MEASURE OF CREDIT RISK.



Charge Off Rate as a Measure Of Consumer Credit Risk

(Data from: FRED, Federal Reserve Bank of St. Louis, US).

The Loan Charge Off Rate exhibits noticeable cycles as it is by definition the loan losses a bank incurs during various business cycles.

ANNEXURE F: RESEARCH ETHICS CLEARANCE CERTIFICATE

University of South Africa PO Box 392, Unisa, 0003, South Africa
Cnr Janadel and Alexandra Avenues, Midrand, 1685, Tel: +27 11 652 0000 Fax: +27 11 652 0299
E-mail: sbl@unisa.ac.za Website: www.unisa.ac.za/sbl

SCHOOL OF BUSINESS LEADERSHIP
RESEARCH ETHICS REVIEW COMMITTEE (GSBL CRERC)

08 December 2020

Ref #: 2020_SBL_DBL_029_SD
Name of applicant: Mr GJ Kimetto
Student #: 64101304

Dear Mr Kimetto

Decision: Ethics Approval

Student: Mr GJ Kimetto, (gjkurgat@gmail.com , 082 790 0264)

Supervisor: Dr A Joseph, (angelo@quantanalyst.co.za , 083 753 3709)

Project Title: Emerging markets retail credit risk model.

Qualification: Doctor of Business Leadership (DBL)

Expiry Date: January 2023

Thank you for applying for research ethics clearance, SBL Research Ethics Review Committee reviewed your application in compliance with the Unisa Policy on Research Ethics.

Outcome of the SBL Research Committee:
Approval is granted for the duration of the Project

The application was reviewed in compliance with the Unisa Policy on Research Ethics by the SBL Research Ethics Review Committee on the 08/12/2020.

The proposed research may now commence with the proviso that:

1) The researcher will ensure that the research project adheres to the relevant guidelines set out in the Unisa Covid-19 position statement on research ethics attached
2) The researcher/s will ensure that the research project adheres to the values and principles expressed in the UNISA Policy on Research Ethics.

45 Building leaders who go beyond

SBL GRADUATE SCHOOL OF BUSINESS LEADERSHIP UNISA

University of South Africa, PO Box 392, Unisa, 0003, South Africa
Cnr Janadel and Alexandra Avenues, Midrand, 1685, Tel: +27 11 652 0000, Fax: +27 11 652 0299
E-mail: sbl@unisa.ac.za Website: www.unisa.ac.za/sbl

3) Any adverse circumstance arising in the undertaking of the research project that is relevant to the ethicality of the study, as well as changes in the methodology, should be communicated in writing to the SBL Research Ethics Review Committee.
4) An amended application could be requested if there are substantial changes from the existing proposal, especially if those changes affect any of the study-related risks for the research participants.
5) The researcher will ensure that the research project adheres to any applicable national legislation, professional codes of conduct, institutional guidelines and scientific standards relevant to the specific field of study.

Kind regards,

Prof R Ramphal
Chairperson: SBL Research Ethics Committee
011 – 652 0363 or ramphrr@unisa.ac.za

Prof P Msweli
Executive Dean: Graduate School of Business Leadership
011- 652 0256/mswelp@unisa.ac.za

ANNEXURE G: SAMPLE INDEPENDENT/DEPENDENT VARIABLES FOR A MORTGAGE

According to Baesens, Roesch, and Scheule (2016), independent variables, assuming a mortgage, which can be used in consumer credit risk modelling research are:

- Outstanding balance at observation time
- Loan-to-value ratio at observation time, in %
- Interest rate at observation time, in %
- House price index at observation time, base year = 100
- Gross domestic product (GDP) growth at observation time, in %
- Unemployment rate at observation time, in %
- Real estate type apartment = 1, otherwise = 0
- Investor borrower = 1, otherwise = 0
- Outstanding balance at origination time
- Credit score at origination time, in %
- Loan-to-value ratio at origination time, in %
- Interest rate at origination time, in %
- House price index at origination time, base year = 100
- Default observation at observation time
- Default (1) and nondefault (0) observation at observation time

Variables that were used in this research included some of the ones on this list.

ANNEXURE H: US DATA USED IN BUILDING THE MODELS

Date	LCO	T10Y2Y	T10Y	T2Y	UNRATE	PCE	PSR	HD	CD	DRATE
1987-01-01	1,47	0,84	7,20	6,35	6,6	2 984	8,9	11,93	6,23	3,35
1987-04-01	1,47	0,89	8,34	7,45	6,3	3 053	6,8	12,07	6,24	3,28
1987-07-01	1,43	1,03	8,87	7,84	6,0	3 117	7,4	11,94	6,15	3,29
1987-10-01	1,49	1,14	9,13	7,99	5,8	3 151	8,5	11,81	6,07	3,31
1988-01-01	1,45	1,06	8,41	7,35	5,7	3 232	8,3	11,75	6,02	3,27
1988-04-01	1,47	1,04	8,91	7,88	5,5	3 292	8,5	11,74	5,97	3,26
1988-07-01	1,52	0,64	9,10	8,47	5,5	3 362	8,6	11,69	5,90	3,24
1988-10-01	1,46	0,25	8,96	8,71	5,3	3 435	8,4	11,57	5,84	3,30
1989-01-01	1,57	-0,21	9,21	9,42	5,2	3 490	9,0	11,57	5,81	3,43
1989-04-01	1,55	-0,19	8,76	8,94	5,2	3 554	8,2	11,73	5,87	3,52
1989-07-01	1,54	0,02	8,11	8,09	5,2	3 609	8,0	11,79	5,85	3,57
1989-10-01	1,64	0,06	7,91	7,85	5,4	3 654	8,3	11,75	5,76	3,58
1990-01-01	1,64	0,06	8,42	8,36	5,3	3 738	8,3	11,65	5,64	3,60
1990-04-01	1,72	0,10	8,67	8,57	5,3	3 783	8,7	11,60	5,54	3,61
1990-07-01	1,83	0,60	8,70	8,10	5,7	3 847	8,3	11,58	5,48	3,81
1990-10-01	1,97	0,80	8,41	7,61	6,1	3 868	8,2	11,60	5,42	4,01
1991-01-01	2,20	0,99	8,02	7,04	6,6	3 874	8,7	11,58	5,36	4,12
1991-04-01	2,33	1,23	8,13	6,89	6,8	3 927	8,6	11,43	5,24	4,21
1991-07-01	2,40	1,43	7,95	6,52	6,9	3 973	8,5	11,32	5,14	4,17
1991-10-01	2,28	1,84	7,35	5,50	7,1	4 000	9,3	11,12	5,02	4,10
1992-01-01	2,39	2,01	7,31	5,29	7,4	4 100	9,6	10,86	4,87	3,96
1992-04-01	2,27	2,17	7,38	5,20	7,6	4 156	10,0	10,68	4,77	3,84
1992-07-01	2,10	2,47	6,62	4,15	7,6	4 227	9,3	10,54	4,70	3,71
1992-10-01	2,07	2,30	6,74	4,44	7,4	4 307	8,9	10,46	4,66	3,55
1993-01-01	1,84	2,13	6,26	4,13	7,1	4 350	8,8	10,42	4,65	3,49
1993-04-01	1,82	1,99	5,99	4,00	7,1	4 419	8,3	10,35	4,65	3,39
1993-07-01	1,72	1,64	5,62	3,97	6,8	4 487	7,4	10,41	4,71	3,20
1993-10-01	1,56	1,53	5,62	4,09	6,6	4 553	7,2	10,37	4,76	2,98
1994-01-01	1,49	1,53	6,09	4,56	6,6	4 621	6,8	10,39	4,83	2,88
1994-04-01	1,42	1,26	7,09	5,83	6,2	4 683	7,0	10,45	4,92	2,76
1994-07-01	1,41	1,09	7,33	6,23	6,0	4 753	6,8	10,59	5,04	2,71
1994-10-01	1,47	0,67	7,84	7,16	5,6	4 827	7,1	10,70	5,12	2,74
1995-01-01	1,44	0,36	7,47	7,11	5,5	4 862	7,6	10,89	5,25	2,87
1995-04-01	1,63	0,47	6,60	6,13	5,7	4 934	6,9	11,08	5,39	3,00
1995-07-01	1,82	0,47	6,33	5,86	5,7	4 999	6,9	11,19	5,52	3,19
1995-10-01	1,99	0,40	5,90	5,50	5,6	5 056	6,6	11,22	5,59	3,30
1996-01-01	2,14	0,64	5,91	5,27	5,5	5 131	6,7	11,24	5,65	3,34
1996-04-01	2,26	0,60	6,71	6,11	5,5	5 221	6,5	11,27	5,68	3,49
1996-07-01	2,33	0,60	6,78	6,18	5,3	5 275	6,7	11,32	5,71	3,62
1996-10-01	2,40	0,55	6,35	5,80	5,3	5 353	6,4	11,37	5,78	3,69
1997-01-01	2,56	0,52	6,57	6,04	5,2	5 433	6,3	11,33	5,73	3,75
1997-04-01	2,77	0,42	6,70	6,28	5,0	5 471	6,6	11,33	5,70	3,76
1997-07-01	2,79	0,34	6,24	5,90	4,9	5 579	6,1	11,34	5,72	3,71
1997-10-01	2,68	0,17	5,91	5,73	4,7	5 664	6,3	11,29	5,71	3,75
1998-01-01	2,63	0,14	5,59	5,45	4,6	5 721	7,4	11,14	5,63	3,74
1998-04-01	2,64	0,03	5,59	5,56	4,4	5 833	7,0	11,17	5,67	3,77
1998-07-01	2,57	0,07	5,21	5,14	4,5	5 927	6,7	11,15	5,69	3,76

Date	LCO	T10Y2Y	T10Y	T2Y	UNRATE	PCE	PSR	HD	CD	DRATE
1998-10-01	2,50	0,29	4,66	4,38	4,4	6 028	6,1	11,22	5,76	3,72
1999-01-01	2,40	0,14	5,00	4,86	4,3	6 102	6,2	11,28	5,81	3,78
1999-04-01	2,12	0,26	5,54	5,28	4,3	6 231	4,8	11,43	5,91	3,62
1999-07-01	2,26	0,25	5,88	5,63	4,2	6 335	4,5	11,56	5,98	3,55
1999-10-01	2,20	0,20	6,14	5,95	4,1	6 467	4,4	11,55	5,96	3,51
2000-01-01	2,23	-0,06	6,47	6,53	4,0	6 618	4,6	11,50	5,94	3,47
2000-04-01	2,14	-0,39	6,18	6,57	3,9	6 712	4,8	11,66	6,04	3,53
2000-07-01	2,19	-0,33	5,89	6,22	4,0	6 820	4,9	11,84	6,18	3,56
2000-10-01	2,67	-0,15	5,57	5,72	3,9	6 919	4,6	12,06	6,34	3,63
2001-01-01	2,34	0,46	5,04	4,58	4,2	6 995	5,1	12,16	6,45	3,63
2001-04-01	2,63	1,08	5,28	4,19	4,4	7 042	4,7	12,38	6,59	3,68
2001-07-01	2,79	1,32	5,00	3,68	4,8	7 070	6,4	12,32	6,56	3,72
2001-10-01	3,12	1,90	4,76	2,87	5,5	7 187	3,5	12,62	6,71	3,64
2002-01-01	3,60	1,88	5,08	3,20	5,7	7 218	5,6	12,42	6,62	3,60
2002-04-01	3,10	1,87	5,11	3,23	5,8	7 308	6,0	12,34	6,57	3,51
2002-07-01	3,12	2,03	4,27	2,24	5,7	7 397	5,6	12,38	6,56	3,49
2002-10-01	2,81	2,12	4,00	1,89	5,9	7 473	5,7	12,35	6,50	3,45
2003-01-01	2,84	2,27	3,92	1,65	5,9	7 567	5,2	12,35	6,48	3,42
2003-04-01	3,03	2,20	3,62	1,42	6,1	7 661	5,3	12,29	6,44	3,31
2003-07-01	2,80	2,55	4,23	1,68	6,1	7 821	5,7	12,14	6,36	3,10
2003-10-01	2,86	2,43	4,29	1,86	5,8	7 913	5,3	12,21	6,37	3,28
2004-01-01	2,72	2,32	4,01	1,69	5,7	8 049	4,8	12,21	6,36	3,13
2004-04-01	2,76	2,14	4,60	2,45	5,6	8 147	5,3	12,13	6,29	3,10
2004-07-01	2,52	1,74	4,30	2,56	5,4	8 283	4,8	12,24	6,32	3,05
2004-10-01	2,65	1,35	4,18	2,82	5,4	8 449	4,9	12,24	6,24	3,03
2005-01-01	2,49	0,85	4,30	3,45	5,3	8 552	3,1	12,56	6,32	2,91
2005-04-01	2,41	0,51	4,16	3,65	5,1	8 701	2,8	12,56	6,27	2,85
2005-07-01	3,02	0,26	4,22	3,96	5,0	8 868	2,5	12,62	6,20	2,80
2005-10-01	3,04	0,12	4,49	4,36	5,0	8 955	3,1	12,59	6,10	2,69
2006-01-01	1,77	-0,03	4,58	4,60	4,7	9 100	3,9	12,61	6,09	2,78
2006-04-01	1,92	0,07	5,07	5,00	4,6	9 228	3,7	12,68	5,98	2,92
2006-07-01	2,20	-0,04	4,89	4,93	4,6	9 354	3,2	12,79	5,93	2,96
2006-10-01	2,14	-0,11	4,63	4,74	4,4	9 427	3,4	12,87	5,92	2,94
2007-01-01	2,34	-0,08	4,68	4,76	4,5	9 572	3,6	12,88	5,90	2,93
2007-04-01	2,32	0,04	4,85	4,80	4,5	9 679	3,6	12,96	5,94	2,99
2007-07-01	2,45	0,35	4,74	4,39	4,7	9 798	3,2	13,08	5,98	3,20
2007-10-01	2,80	0,78	4,27	3,49	4,8	9 937	3,0	13,22	6,01	3,40
2008-01-01	2,95	1,64	3,67	2,03	5,0	10 004	3,4	13,14	6,00	3,49
2008-04-01	3,26	1,47	3,88	2,42	5,3	10 130	5,1	12,79	5,85	3,56
2008-07-01	3,70	1,50	3,86	2,36	6,0	10 159	3,9	12,87	5,88	3,71
2008-10-01	4,28	2,02	3,23	1,22	6,9	9 907	5,9	12,82	5,86	4,27
2009-01-01	4,76	1,83	2,74	0,90	8,3	9 815	5,9	12,76	5,82	4,67
2009-04-01	5,58	2,30	3,32	1,02	9,3	9 806	6,9	12,37	5,63	4,85
2009-07-01	5,92	2,49	3,52	1,03	9,6	9 939	5,2	12,24	5,57	4,70
2009-10-01	5,75	2,59	3,46	0,88	9,9	10 005	5,5	11,96	5,42	4,59
2010-01-01	6,60	2,80	3,72	0,92	9,8	10 102	5,7	11,68	5,30	4,72
2010-04-01	6,56	2,62	3,49	0,87	9,6	10 208	6,5	11,35	5,14	4,25
2010-07-01	5,48	2,25	2,78	0,54	9,5	10 301	6,5	11,14	5,04	4,00
2010-10-01	4,91	2,39	2,88	0,49	9,5	10 430	6,3	10,98	5,03	3,63

Date	LCO	T10Y2Y	T10Y	T2Y	UNRATE	PCE	PSR	HD	CD	DRATE
2011-01-01	4,39	2,76	3,46	0,69	9,0	10 558	6,9	10,78	5,06	3,43
2011-04-01	3,43	2,64	3,20	0,56	9,1	10 673	6,6	10,69	5,05	3,27
2011-07-01	3,67	2,14	2,41	0,28	9,0	10 755	6,8	10,58	5,02	3,13
2011-10-01	2,97	1,78	2,05	0,26	8,6	10 809	6,9	10,45	5,01	3,07
2012-01-01	2,63	1,75	2,04	0,29	8,3	10 959	7,9	10,20	4,93	2,93
2012-04-01	2,53	1,54	1,83	0,29	8,2	11 005	8,6	10,07	4,91	2,84
2012-07-01	2,52	1,38	1,64	0,26	8,0	11 059	7,8	10,14	4,99	2,78
2012-10-01	2,45	1,44	1,71	0,27	7,8	11 166	9,9	9,81	4,88	2,63
2013-01-01	2,30	1,69	1,95	0,26	7,7	11 266	5,8	10,17	5,14	2,56
2013-04-01	2,15	1,72	1,99	0,27	7,5	11 291	6,4	10,09	5,16	2,48
2013-07-01	2,13	2,34	2,71	0,37	7,2	11 379	6,4	10,05	5,18	2,40
2013-10-01	2,09	2,42	2,74	0,33	6,9	11 518	5,9	10,05	5,21	2,37
2014-01-01	2,00	2,39	2,77	0,37	6,7	11 618	6,7	9,95	5,21	2,32
2014-04-01	2,00	2,20	2,62	0,42	6,2	11 785	7,0	9,89	5,23	2,27
2014-07-01	1,88	1,98	2,50	0,52	6,1	11 934	7,1	9,84	5,26	2,20
2014-10-01	1,82	1,74	2,28	0,54	5,7	12 054	7,4	9,82	5,29	2,08
2015-01-01	1,75	1,36	1,97	0,60	5,5	12 084	7,9	9,82	5,32	2,01
2015-04-01	1,74	1,55	2,16	0,61	5,4	12 225	7,5	9,86	5,40	1,98
2015-07-01	1,75	1,53	2,22	0,69	5,1	12 348	7,3	9,90	5,47	2,00
2015-10-01	1,77	1,35	2,19	0,84	5,0	12 398	7,4	9,85	5,41	2,01
2016-01-01	1,83	1,08	1,91	0,84	4,9	12 495	7,5	9,81	5,40	2,00
2016-04-01	1,82	0,98	1,75	0,77	4,9	12 637	6,9	9,90	5,50	2,05
2016-07-01	1,88	0,84	1,56	0,73	4,9	12 759	6,8	9,96	5,57	2,08
2016-10-01	2,11	1,13	2,14	1,01	4,8	12 882	6,8	9,93	5,61	2,15
2017-01-01	2,19	1,20	2,45	1,24	4,6	13 046	7,1	9,88	5,60	2,18
2017-04-01	2,11	0,96	2,26	1,30	4,4	13 144	7,5	9,85	5,61	2,22
2017-07-01	2,20	0,88	2,24	1,36	4,3	13 268	7,5	9,83	5,61	2,27
2017-10-01	2,23	0,68	2,37	1,69	4,1	13 497	7,0	9,83	5,64	2,22
2018-01-01	2,23	0,60	2,76	2,16	4,0	13 667	7,5	9,75	5,60	2,28
2018-04-01	2,23	0,44	2,92	2,48	3,9	13 865	7,5	9,71	5,58	2,27
2018-07-01	2,26	0,26	2,92	2,67	3,8	14 003	7,6	9,71	5,60	2,30
2018-10-01	2,25	0,24	3,04	2,80	3,8	14 119	7,9	9,73	5,62	2,30
2019-01-01	2,25	0,17	2,65	2,49	3,9	14 156	8,6	9,76	5,64	2,34
2019-04-01	2,27	0,21	2,34	2,13	3,7	14 376	7,4	9,83	5,70	2,37
2019-07-01	2,31	0,11	1,80	1,69	3,6	14 530	7,2	9,86	5,74	2,35
2019-10-01	2,31	0,20	1,79	1,59	3,6	14 654	7,4	9,87	5,76	2,29
2020-01-01	2,29	0,28	1,37	1,08	3,8	14 439	9,7	9,79	5,72	2,48
2020-04-01	2,26	0,49	0,69	0,19	13,1	12 990	26,0	8,81	5,09	2,03
2020-07-01	1,91	0,51	0,65	0,14	8,8	14 294	16,0	9,17	5,31	1,83
2020-10-01	1,52	0,71	0,86	0,15	6,8	14 468	13,5	9,36	5,42	1,88
2021-01-01	1,54	1,20	1,34	0,13	6,2	15 005	20,0	8,23	4,78	1,70

(Adapted from: FRED, Federal Reserve Bank of St. Louis, US).

The data is for the dependent variable (loan charge off rate) and the nine (9) significant independent (explanatory) variables of model US2. There are 1370 data points spread over 34 years – from Q1 1987 to Q1 2021.

ANNEXURE I: SOUTH AFRICA DATA USED IN BUILDING THE MODELS

DATE	LCO	FX	OIL	SPG	VEH	GOLD	RGDP	RPPI	WUI (VOL)	CPGDP
2008-09-01	3,10	7,00	104,11	6,11	96,32	825	656 296	105	0,48	3,33
2008-10-01	3,4	7,66	76,61	18,02	101,02	813	679 423	104	0,37	3,33
2008-11-01	3,60	7,99	57,31	4,16	96,92	758	679 423	104	0,37	3,33
2008-12-01	3,42	7,76	41,12	5,72	87,81	820	679 423	104	0,37	1,31
2009-01-01	4,20	7,61	41,71	0,57	85,81	858	685 029	102	0,09	1,31
2009-02-01	4,60	7,94	39,09	4,89	85,11	940	685 029	102	0,09	1,31
2009-03-01	4,80	7,61	47,94	2,24	82,81	926	685 029	102	0,09	0,68
2009-04-01	5,10	7,67	49,65	5,25	80,41	893	687 854	100	0,25	0,68
2009-05-01	5,40	8,08	59,03	6,02	75,51	927	687 854	100	0,25	0,68
2009-06-01	5,50	9,78	69,64	3,26	73,91	948	687 854	100	0,25	2,22
2009-07-01	5,60	10,11	64,15	2,08	69,51	934	649 054	98	0,17	2,22
2009-08-01	5,80	9,92	71,05	7,43	62,11	950	649 054	98	0,17	2,22
2009-09-01	5,90	9,91	69,41	1,34	61,61	996	649 054	98	0,17	2,23
2009-10-01	5,84	9,98	75,72	2,97	62,21	1 044	661 877	100	0,08	2,23
2009-11-01	5,94	9,95	77,99	3,05	60,41	1 126	661 877	100	0,08	2,23
2009-12-01	5,94	8,96	74,47	2,34	61,71	1 135	661 877	100	0,08	1,01
2010-01-01	5,86	8,37	78,33	1,22	64,71	1 120	671 864	100	0,25	1,01
2010-02-01	5,82	8,03	76,39	3,56	65,81	1 096	671 864	100	0,25	1,01
2010-03-01	5,88	7,94	81,20	5,72	65,21	1 116	671 864	100	0,25	1,70
2010-04-01	5,95	7,94	84,29	2,99	65,01	1 148	684 146	100	0,35	1,70
2010-05-01	5,91	7,50	73,74	6,03	66,11	1 204	684 146	100	0,35	1,70
2010-06-01	5,91	7,49	75,34	0,46	68,31	1 232	684 146	100	0,35	2,75
2010-07-01	5,84	7,51	76,32	1,18	64,81	1 196	664 088	100	0,26	2,75
2010-08-01	5,86	7,48	76,60	0,35	72,91	1 213	664 088	100	0,26	2,75
2010-09-01	5,88	7,46	75,24	3,08	74,51	1 271	664 088	100	0,26	3,65
2010-10-01	5,95	7,67	81,89	5,28	75,81	1 343	682 169	100	0,18	3,65
2010-11-01	5,81	7,41	84,25	4,29	78,91	1 372	682 169	100	0,18	3,65
2010-12-01	5,79	7,34	89,15	1,68	83,71	1 394	682 169	100	0,18	4,06
2011-01-01	5,82	7,65	89,17	1,10	78,41	1 360	694 274	99	0,63	4,06
2011-02-01	5,81	7,64	88,58	1,38	79,51	1 371	694 274	99	0,63	4,06
2011-03-01	5,78	7,52	102,86	3,37	85,81	1 423	694 274	99	0,63	3,87
2011-04-01	5,79	7,29	109,53	3,32	75,91	1 474	707 477	98	0,18	3,87
2011-05-01	5,73	7,11	100,90	1,58	82,41	1 512	707 477	98	0,18	3,87
2011-06-01	5,56	6,91	96,26	2,26	85,61	1 528	707 477	98	0,18	3,76
2011-07-01	5,49	6,97	97,30	2,30	84,31	1 569	687 269	98	0,51	3,76
2011-08-01	5,29	6,82	86,33	7,01	87,11	1 760	687 269	98	0,51	3,76
2011-09-01	5,06	6,92	85,52	2,32	94,12	1 781	687 269	98	0,51	2,92
2011-10-01	4,90	7,18	86,32	2,00	91,32	1 668	705 429	98	0,66	2,92
2011-11-01	4,81	6,90	97,16	3,05	89,81	1 736	705 429	98	0,66	2,92
2011-12-01	4,69	6,72	98,56	0,72	85,81	1 653	705 429	98	0,66	2,61
2012-01-01	4,73	6,86	100,27	3,23	88,61	1 656	715 413	98	0,73	2,61
2012-02-01	4,71	6,79	102,20	2,62	89,61	1 743	715 413	98	0,73	2,61
2012-03-01	4,58	6,79	106,16	0,73	94,22	1 675	715 413	98	0,73	2,05
2012-04-01	4,64	7,09	103,32	0,04	97,22	1 649	730 146	98	0,72	2,05
2012-05-01	4,56	7,58	94,66	0,99	98,12	1 585	730 146	98	0,72	2,05
2012-06-01	4,47	7,95	82,30	0,85	96,62	1 596	730 146	98	0,72	2,38
2012-07-01	4,43	8,15	87,90	0,64	99,52	1 593	702 158	98	1,04	2,38

DATE	LCO	FX	OIL	SPG	VEH	GOLD	RGDP	RPPI	WUI (VOL)	CPGDP
2012-08-01	4,36	8,19	94,13	4,33	95,62	1 626	702 158	98	1,04	2,38
2012-09-01	4,29	8,00	94,51	0,90	100,12	1 742	702 158	98	1,04	2,38
2012-10-01	4,28	7,64	89,49	2,27	100,52	1 746	724 792	98	0,41	2,38
2012-11-01	4,06	7,61	86,53	2,30	98,82	1 724	724 792	98	0,41	2,38
2012-12-01	4,04	7,83	87,86	3,37	100,92	1 687	724 792	98	0,41	2,05
2013-01-01	4,09	8,15	94,76	3,65	101,92	1 672	730 226	98	0,48	2,05
2013-02-01	4,06	8,38	95,31	0,39	101,82	1 631	730 226	98	0,48	2,05
2013-03-01	4,01	8,25	92,94	0,10	103,32	1 591	730 226	98	0,48	2,06
2013-04-01	4,01	8,26	92,02	4,01	102,92	1 486	743 903	98	0,62	2,06
2013-05-01	4,01	8,26	94,51	5,21	103,02	1 416	743 903	98	0,62	2,06
2013-06-01	3,94	8,64	95,77	2,06	105,02	1 343	743 903	98	0,62	2,23
2013-07-01	3,95	8,80	104,67	1,19	99,72	1 284	717 484	98	0,42	2,23
2013-08-01	3,83	8,61	106,57	4,97	108,82	1 345	717 484	98	0,42	2,23
2013-09-01	3,71	8,80	106,29	2,70	104,52	1 348	717 484	98	0,42	2,39
2013-10-01	3,69	8,88	100,54	1,92	107,22	1 314	743 185	100	0,68	2,39
2013-11-01	3,65	9,19	93,86	1,67	110,32	1 277	743 185	100	0,68	2,39
2013-12-01	3,64	9,10	97,63	1,45	110,42	1 222	743 185	100	0,68	3,26
2014-01-01	3,59	9,35	94,62	3,64	109,52	1 243	745 651	100	0,83	3,26
2014-02-01	3,57	10,00	100,82	0,50	105,82	1 299	745 651	100	0,83	3,26
2014-03-01	3,51	9,91	100,80	1,87	105,12	1 337	745 651	100	0,83	2,43
2014-04-01	3,54	10,07	102,07	2,50	100,92	1 299	766 856	100	0,47	2,43
2014-05-01	3,57	9,96	102,18	2,05	100,12	1 289	766 856	100	0,47	2,43
2014-06-01	3,42	9,90	105,79	2,38	99,62	1 278	766 856	100	0,47	1,60
2014-07-01	3,45	10,20	103,59	2,21	99,12	1 313	732 211	100	0,12	1,60
2014-08-01	3,42	10,37	96,54	0,85	102,22	1 297	732 211	100	0,12	1,60
2014-09-01	3,35	10,89	93,21	0,63	100,92	1 241	732 211	100	0,12	1,78
2014-10-01	3,35	10,95	84,40	5,45	102,22	1 224	755 843	101	0,55	1,78
2014-11-01	3,27	10,74	75,79	4,33	103,32	1 176	755 843	101	0,55	1,78
2014-12-01	3,28	10,54	59,29	2,37	101,72	1 200	755 843	101	0,55	1,58
2015-01-01	3,28	10,41	47,22	0,83	106,52	1 249	759 758	102	0,54	1,58
2015-02-01	3,23	10,68	50,58	6,28	104,92	1 231	759 758	102	0,54	1,58
2015-03-01	3,21	10,66	47,82	0,22	108,02	1 181	759 758	102	0,54	2,41
2015-04-01	3,22	10,66	54,45	2,33	108,82	1 198	780 279	101	0,60	2,41
2015-05-01	3,26	10,99	59,27	0,32	105,62	1 198	780 279	101	0,60	2,41
2015-06-01	3,31	11,06	59,82	3,56	102,32	1 182	780 279	101	0,60	1,56
2015-07-01	3,33	11,09	50,90	0,14	105,32	1 132	750 792	101	1,16	1,56
2015-08-01	3,27	11,50	42,87	2,37	101,82	1 118	750 792	101	1,16	1,56
2015-09-01	3,20	11,55	45,48	1,49	101,32	1 125	750 792	101	1,16	0,91
2015-10-01	3,24	11,58	46,22	6,15	101,42	1 157	764 217	102	0,63	0,91
2015-11-01	3,08	12,09	42,44	0,83	100,92	1 088	764 217	102	0,63	0,91
2015-12-01	3,12	11,98	37,19	4,96	102,62	1 068	764 217	102	0,63	0,08
2016-01-01	3,08	11,97	31,68	3,60	97,52	1 096	765 471	102	1,18	0,08
2016-02-01	3,16	12,29	30,32	1,78	99,52	1 195	765 471	102	1,18	0,08
2016-03-01	3,38	12,46	37,55	6,83	98,22	1 246	765 471	102	1,18	0,74
2016-04-01	3,13	12,90	40,75	0,42	98,32	1 241	783 758	101	1,66	0,74
2016-05-01	3,17	13,64	46,71	0,16	96,52	1 260	783 758	101	1,66	0,74
2016-06-01	3,17	13,49	48,76	0,30	100,82	1 274	783 758	101	1,66	0,64
2016-07-01	3,15	14,14	44,65	0,28	101,02	1 337	747 421	100	1,25	0,64
2016-08-01	3,17	15,00	44,72	0,37	98,22	1 341	747 421	100	1,25	0,64

DATE	LCO	FX	OIL	SPG	VEH	GOLD	RGDP	RPPI	WUI (VOL)	CPGDP
2016-09-01	3,24	16,33	45,18	0,79	90,52	1 327	747 421	100	1,25	0,87
2016-10-01	2,91	15,76	49,78	1,86	87,41	1 269	768 076	100	0,96	0,87
2016-11-01	2,85	15,38	45,66	1,89	90,52	1 240	768 076	100	0,96	0,87
2016-12-01	2,87	14,59	51,97	0,84	91,52	1 152	768 076	100	0,96	0,84
2017-01-01	2,89	15,33	52,50	4,55	87,31	1 193	770 486	100	1,04	0,84
2017-02-01	2,84	15,05	53,47	0,06	86,41	1 233	770 486	100	1,04	0,84
2017-03-01	2,84	14,39	49,33	0,63	85,21	1 231	770 486	100	1,04	1,02
2017-04-01	2,89	13,78	51,06	2,24	85,11	1 267	790 482	100	1,13	1,02
2017-05-01	2,90	14,03	48,48	1,73	88,31	1 245	790 482	100	1,13	1,02
2017-06-01	2,90	13,93	45,18	4,28	85,91	1 261	790 482	100	1,13	0,97
2017-07-01	2,85	13,91	46,63	3,48	87,41	1 235	755 994	100	1,38	0,97
2017-08-01	2,83	13,84	48,04	4,38	100,92	1 282	755 994	100	1,38	0,97
2017-09-01	2,77	13,54	49,82	0,03	92,32	1 317	755 994	100	1,38	1,44
2017-10-01	2,79	13,19	51,58	3,47	86,41	1 281	780 325	100	1,82	1,44
2017-11-01	2,81	12,92	56,64	3,87	85,51	1 283	780 325	100	1,82	1,44
2017-12-01	2,84	13,45	57,88	2,66	86,81	1 266	780 325	100	1,82	2,23
2018-01-01	3,08	13,26	63,70	3,38	88,91	1 333	782 486	99	0,61	2,23
2018-02-01	3,10	12,89	62,23	4,06	89,11	1 334	782 486	99	0,61	2,23
2018-03-01	3,24	13,15	62,73	0,42	91,22	1 326	782 486	99	0,61	1,59
2018-04-01	3,31	13,21	66,25	1,73	90,82	1 334	801 180	99	0,89	1,59
2018-05-01	3,39	13,17	69,98	1,53	91,02	1 304	801 180	99	0,89	1,59
2018-06-01	3,50	13,70	67,87	0,57	91,92	1 282	801 180	99	0,89	0,72
2018-07-01	3,60	14,04	70,98	0,44	88,91	1 238	761 397	99	0,81	0,72
2018-08-01	3,57	13,09	68,06	1,50	90,22	1 202	761 397	99	0,81	0,72
2018-09-01	3,66	12,20	70,23	1,89	90,42	1 199	761 397	99	0,81	0,68
2018-10-01	3,72	11,82	70,75	6,79	92,92	1 215	781 144	99	0,17	0,68
2018-11-01	3,74	11,84	56,96	0,82	84,21	1 221	781 144	99	0,17	0,68
2018-12-01	3,73	12,11	49,52	1,54	89,11	1 250	781 144	99	0,17	0,17
2019-01-01	3,85	12,54	51,38	3,25	90,72	1 292	792 363	99	1,09	0,17
2019-02-01	3,81	13,31	54,95	2,81	89,81	1 320	792 363	99	1,09	0,17
2019-03-01	3,77	13,37	58,15	2,18	90,12	1 302	792 363	99	1,09	0,05
2019-04-01	3,79	14,09	63,86	4,34	90,62	1 288	809 635	98	0,81	0,05
2019-05-01	3,83	14,75	60,83	3,56	89,91	1 283	809 635	98	0,81	0,05
2019-06-01	3,73	14,49	54,66	3,17	88,61	1 358	809 635	98	0,81	1,00
2019-07-01	3,73	14,09	57,35	0,74	86,21	1 415	761 633	98	1,08	1,00
2019-08-01	3,80	14,26	54,81	5,08	85,21	1 497	761 633	98	1,08	1,00
2019-09-01	3,83	13,82	56,95	1,82	85,91	1 510	761 633	98	1,08	0,14
2019-10-01	3,76	13,81	53,96	0,90	88,11	1 495	788 307	98	1,10	0,14
2019-11-01	3,79	14,39	57,03	2,36	86,21	1 472	788 307	98	1,10	0,14
2019-12-01	3,89	14,15	59,88	0,29	86,01	1 480	788 307	98	1,10	0,57
2020-01-01	3,98	14,42	57,52	1,68	90,72	1 561	793 431	96	0,54	0,57
2020-02-01	3,96	14,59	50,54	1,49	85,21	1 600	793 431	96	0,54	0,57
2020-03-01	4,04	14,04	29,21	19,89	85,51	1 594	793 431	96	0,54	0,20
2020-04-01	4,27	15,16	16,55	6,20	88,31	1 680	805 965	97	0,64	0,20
2020-05-01	4,59	14,83	28,56	4,84	88,91	1 716	805 965	97	0,64	0,20
2020-06-01	4,89	14,91	38,31	6,55	85,11	1 734	805 965	97	0,64	17,47
2020-07-01	5,01	14,79	40,71	3,67	86,81	1 841	764 788	98	0,52	17,47
2020-08-01	4,99	14,40	42,34	1,78	82,51	1 971	764 788	98	0,52	17,47
2020-09-01	5,03	14,41	39,63	3,23	85,71	1 923	764 788	98	0,52	5,94

DATE	LCO	FX	OIL	SPG	VEH	GOLD	RGDP	RPPI	WUI (VOL)	CPGDP
2020-10-01	5,03	15,02	39,40	0,66	61,31	1 903	648 135	98	0,14	5,94
2020-11-01	5,03	16,68	40,94	3,87	1,40	1 870	648 135	98	0,14	5,94
2020-12-01	5,16	18,57	47,02	4,61	29,00	1 854	648 135	98	0,14	4,25
2021-01-01	5,24	18,18	52,00	6,50	60,61	1 870	744 590	98	0,27	4,25
2021-02-01	5,19	17,13	59,04	4,18	62,01	1 814	744 590	98	0,27	4,25
2021-03-01	5,12	16,74	62,33	2,34	65,71	1 722	744 590	98	0,27	2,73

(Adapted from: FRED, Federal Reserve Bank of St. Louis, US, and the South African Reserve Bank).

The data is for the dependent variable (Impairments) and the nine (9) significant independent (explanatory) variables. There are 1590 data points spread over 13 years – from January 2008 to March 2021.

ANNEXURE J: CORRELATIONS OF IMPAIRMENTS AND PREDICTOR LEAD VARIABLES FOR SOUTH AFRICA

	Leading (months) / Variables	**Status quo**	**+1**	**+2**	**+3**	**+4**	**+5**	**+6**	**+7**	**+8**	**+9**	**+10**
	Impairments %	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
1	CPI growth %	-8%	-8%	-8%	-8%	-7%	-7%	-7%	-5%	-4%	-5%	-2%
2	ZAR/$ Forex	-42%	-42%	-42%	-42%	-42%	-43%	-43%	-44%	-45%	-46%	-47%
3	Interest TB %	-42%	-37%	-31%	-24%	-17%	-9%	-9%	8%	16%	24%	31%
4	Interest IBR %	-42%	-37%	-31%	-24%	-17%	-9%	-9%	7%	16%	24%	31%
5	Interest 10Y T %	1%	3%	5%	8%	10%	11%	11%	15%	17%	20%	22%
6	10 Y 3M	41%	37%	32%	27%	21%	14%	14%	-1%	-9%	-17%	-24%
7	10 Y TB	41%	37%	32%	27%	21%	14%	14%	-1%	-9%	-16%	-24%

	Leading (months) / Variables	Status quo	+1	+2	+3	+4	+5	+6	+7	+8	+9	+10
8	Oil Price US$	14%	11%	9%	7%	7%	7%	7%	8%	10%	11%	13%
9	Share price growth %	18%	17%	15%	11%	8%	6%	6%	1%	0%	-3%	-6%
10	Share PI	-49%	-53%	-57%	-60%	-63%	-65%	-65%	-68%	-69%	-70%	-70%
11	CPI	-30%	-33%	-35%	-37%	-39%	-41%	-41%	-45%	-46%	-47%	-48%
12	P-car reg. index	-41%	-46%	-51%	-55%	-58%	-60%	-60%	-64%	-65%	-66%	-66%
13	Car reg. growth rate %	5%	7%	8%	8%	8%	8%	8%	9%	9%	9%	8%
14	Income / Disposable income	33%	37%	41%	44%	47%	50%	50%	56%	58%	60%	60%
15	Gold Price US$/ounce	31%	26%	22%	16%	11%	6%	6%	-6%	-12%	-17%	-22%
16	Unemployment Rate %	-2%	-6%	-10%	-14%	-18%	-22%	-22%	-29%	-32%	-35%	-34%

	Leading (months) / Variables	Status quo	+1	+2	+3	+4	+5	+6	+7	+8	+9	+10
17	Real GDP ZAR	-52%	-55%	-58%	-61%	-63%	-65%	-65%	-67%	-68%	-68%	-68%
18	Res. Property PI 2010=100	-44%	-42%	-38%	-35%	-30%	-26%	-26%	-16%	-11%	-5%	1%
19	World Uncertainty Index	-62%	-63%	-64%	-65%	-65%	-64%	-64%	-61%	-59%	-58%	-55%
20	Constant Price GDP SA % Change	-15%	-18%	-20%	-20%	-20%	-20%	-20%	-20%	-19%	-17%	-15%

(Data from: FRED, Federal Reserve Bank of St. Louis, US, and the South African Reserve Bank).

In combination with multivariable regression analyses, these correlations were used to determine the significant variables used in Credit Risk Models SA1 and SA2.

ANNEXURE K: CORRELATIONS OF IMPAIRMENTS AND PREDICTOR LAGGED VARIABLES FOR SOUTH AFRICA

	Lagging (months) / Variable	**-1**	**-2**	**-3**	**--4**	**-5**	**-6**	**-7**	**-8**	**-9**	**-10**
	Impairments	0,99	0,98	0,96	0,93	0,90	0,87	0,82	0,78	0,73	0,68
1	CPI growth	-0,05	-0,07	-0,06	-0,05	-0,03	0,01	0,04	0,06	0,07	0,07
2	ZAR/$ Forex	-0,46	-0,48	-0,51	-0,54	-0,56	-0,58	-0,61	-0,63	-0,64	-0,65
3	Interest TB	-0,45	-0,47	-0,49	-0,50	-0,51	-0,52	-0,52	-0,53	-0,53	-0,54
4	Interest IBR	-0,45	-0,47	-0,49	-0,50	-0,51	-0,52	-0,52	-0,53	-0,53	-0,53
5	Interest 10Y T	-0,03	-0,06	-0,09	-0,12	-0,14	-0,15	-0,17	-0,20	-0,23	-0,26
6	10 Y 3M	0,42	0,43	0,43	0,43	0,42	0,41	0,40	0,38	0,36	0,33
7	10 Y – TB	0,42	0,43	0,43	0,43	0,42	0,41	0,40	0,38	0,36	0,33
8	Oil Price US$	0,18	0,21	0,26	0,31	0,37	0,44	0,52	0,58	0,63	0,65
9	Share price growth	0,16	0,19	0,20	0,21	0,22	0,21	0,18	0,15	0,13	0,07

	Lagging (months) / Variable	**-1**	**-2**	**-3**	**--4**	**-5**	**-6**	**-7**	**-8**	**-9**	**-10**
10	Share PI	-0,49	-0,49	-0,49	-0,49	-0,48	-0,47	-0,47	-0,48	-0,48	-0,49
11	CPI	-0,32	-0,34	-0,36	-0,38	-0,40	-0,42	-0,44	-0,46	-0,47	-0,49
12	P-car reg. index	-0,34	-0,28	-0,21	-0,15	-0,09	-0,03	0,02	0,07	0,12	0,17
13	Car reg. growth rate	0,03	0,01	0,00	0,00	0,00	-0,01	-0,01	-0,01	-0,01	-0,02
14	Income/ Disposable income	0,32	0,31	0,31	0,31	0,31	0,31	0,32	0,32	0,33	0,34
15	Gold Price US$/ounce	0,33	0,35	0,38	0,40	0,41	0,43	0,45	0,46	0,47	0,48
16	Unemployment Rate	-0,02	-0,02	-0,03	-0,05	-0,06	-0,08	-0,11	-0,13	-0,16	-0,20
17	Real GDP ZAR	-0,50	-0,49	-0,49	-0,48	-0,47	-0,46	-0,46	-0,45	-0,44	-0,42
18	Res. Property PI 2010=100	-0,45	-0,46	-0,47	-0,47	-0,47	-0,47	-0,46	-0,45	-0,44	-0,43
19	World Uncertainty Index	-0,59	-0,57	-0,54	-0,51	-0,48	-0,45	-0,42	-0,39	-0,36	-0,32
20	Constant Price GDP SA Change	-0,09	-0,04	0,01	0,06	0,11	0,17	0,21	0,25	0,29	0,32

	Lagging (months) / Variable	-1	-2	-3	--4	-5	-6	-7	-8	-9	-10
	Min	-0,59	-0,57	-0,54	-0,54	-0,56	-0,58	-0,61	-0,63	-0,64	-0,65
	Max	0,42	0,43	0,43	0,43	0,42	0,44	0,52	0,58	0,63	0,65
	Average	-0,12	-0,12	-0,11	-0,10	-0,10	-0,09	-0,08	-0,08	-0,08	-0,08

(Data from: FRED, Federal Reserve Bank of St. Louis, US, and the South African Reserve Bank).

Independent variables that have a correlation of ≥ 60% with losses (Impairments) are used in the Credit Risk Models SA3 and SA4.

ANNEXURE L: CONSENT TO SUBMIT THESIS FOR EXAMINATION

Confidential

The Programme Administrator, DBL,
Graduate School of Business Leadership,
P. O. Box 392,
UNISA 0003.

CONSENT TO SUBMIT THESIS FOR EXAMINATION

Consent is hereby given to:

Student Name: Geoffrey John Kimetto

Student Number: 64101304

to submit his thesis, entitled, *Adapting a Developed Market Credit Risk Model for the Understanding and Estimation of Consumer Credit Losses in South Africa*, in its final form.

I have considered the originality software checking report obtained by the candidate and confirm that the thesis meets an acceptable standard of originality.

Supervisor: Dr. Angelo D Joseph **Date**: 17 November 2022

The student acknowledges that sufficient feedback was provided by the supervisor and that he took the responsibility to attend to the feedback in a way that satisfies the requirements for a research thesis on the DBL level.

Student: Geoffrey John Kimetto **Date**: 17 November 2022

Please follow the instructions for submission from the Executive Dean, College Of Graduate Studies, sent to you by Mrs. Elizabeth Goosen.

ANNEXURE M: PROOFREADING AND COPY EDITING CERTIFICATE

Emend

JILL STEVENSON

Copyediting and proofreading

May 2023

To whom it may concern

Re: Proofreading and copyediting of a research report entitled:

Adapting a Developed Market Credit Risk Model for the Understanding and Estimation of Consumer Credit Losses in South Africa

Prepared by: Student: Geoffrey John Kimetto, Student Number: 64101304

I, Jill Diane Stevenson, hereby confirm that the changes made to the above thesis were to ensure consistency of grammar and language (concord, spelling, punctuation) and to the conformity of format (headings, indexing, citations and references [Cite them right 11th ed. Harvard).

No other changes were made to the body of work submitted by the author (conclusions, recommendations, data, factual reporting or commentary).

Yours faithfully

J D Stevenson

Jill Stevenson
Certified Copyeditor and Proofreader

Cell: 0833092927
Email: Jilldiane18@gmail.com

ANNEXURE N: TURNITIN ORIGINALITY REPORT SUMMARY

Submit for 64101304 Kimetto Final

ORIGINALITY REPORT

11%	7%	8%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	**Submitted to University of South Africa** Student Paper	1%
2	**etonline-digitallibrary.com** Internet Source	<1%
3	**Kim, Jiseob. "Household's optimal mortgage and unsecured loan default decision", Journal of Macroeconomics, 2015.** Publication	<1%
4	**davegiles.blogspot.co.uk** Internet Source	<1%
5	**hdl.handle.net** Internet Source	<1%
6	**"Risk Management and Shareholders' Value in Banking", Wiley, 2012** Publication	<1%
7	**kupdf.net** Internet Source	<1%
8	**Submitted to The Robert Gordon University** Student Paper	<1%

turnitin

Digital Receipt

This receipt acknowledges that Turnitin received your paper. Below you will find the receipt information regarding your submission.

The first page of your submissions is displayed below.

Submission author:	Geoffrey John Kimetto
Assignment title:	COMPLETE proposal/thesis for examination
Submission title:	SUbmit for 64101304 Kimetto Final
File name:	Chapter_1-5_V1.121_2023.05.28.pdf
File size:	4.28M
Page count:	203
Word count:	64,298
Character count:	291,018
Submission date:	03-Jul-2023 01:34PM (UTC+0200)
Submission ID:	2125986932