The effect of the financial crisis on credit scoring in the retail credit market in South Africa

by

Jaco van der Walt

Dissertation submitted to The School of Economics
of the North-West University (Potchefstroom Campus)
in partial fulfilment of the requirements for the degree of
Magister Commercii (Risk Management)

Supervisor: Prof G. van Vuuren

Potchefstroom

November 2011

Acknowledgements

The research in this document led me to find my passion for risk management and especially quantitative analytics. I would like to thank my two supervisors for their contributions, patience and guidance. Prof Styger, your experience and calm demeanour was always shoulder for me to lean on. Though you could not see through the research with me, I pray for your health every day.

A huge thank you to Prof Gary van Vuuren for stepping in and guiding me to the finish line. You have been a fount of information and knowledge and an all-round great guy to work with! I promise to focus on writing in a less "folksy" manner when attempting research papers in the future.

To my family and friends, especially the love of my life, Therina, thank you all for your support, sacrifice and continuous prayers. I would have fallen by the wayside a long time ago if you weren't there to help me keep focused!

And then above all, I would like to thank my Heavenly Father for the opportunity and ability to complete this study. All that I am, all that I have done and all that I have achieved have been given me from Your loving hand.

Abstract

This study follows a three-pronged approach to investigate the effects of the global financial crisis on the South African retail credit market (using Woolworths as subject). These three prongs, or areas, include a literature study, step-by-step credit scoring guide and an application of this guide in an empirical study. To achieve this goal, credit scoring was selected as the quantitative tool to illustrate these effects. Two different periods were chosen to supply a snapshot of the retail credit industry, namely the retail credit situation *before* and *during* the global financial crisis.

To correctly define and understand the mechanics affecting South Africa's retail credit industry, a literature review was conducted to investigate the global financial crisis, the South African retail credit market and credit scoring itself. The literature investigation explains the global financial crisis and identifies some of the primary drivers behind it. These drivers included the US housing bubble, the introduction of subprime loans and the securitisation of these loans (mortgage backed securities). The study found that these drivers, especially the securitisation of subprime loans, were the vehicle used to enable the crisis to spread globally.

The ultimate goal of the study was to provide the individual, and companies, with an understanding of the global financial crisis' effects on the consumer specifically through their credit worthiness and retail credit behaviour. Through the use of credit scoring, the study found that at least one retailer (Woolworths) in the retail industry was affected. Woolworths placed a stronger emphasis on reducing their credit exposure whilst consumers were steadily increasing their facility utilisation.

Keywords: *Global financial crisis, financial crisis, consumer credit, credit scoring, scorecard development*

Opsomming

Die studie volg ’n drieledige benadering om die effekte van die finansiële krisis op die kleinhandel-kredietbedryf van Suid Afrika te ondersoek (met Woolworths as onderwerp). Die drie lede, of gebiede, sluit ‘n literatuur studie, ‘n stap-vir-stap kredietgraderings gids en die toepassing van hierdie gids in 'n empiriese studie. Om hierdie doel te bereik, is kredietgradering as 'n kwantitatiewe instrument gekies om die effekte (van die krisis) te illustreer. Twee verskillende tydperke is gekies om die situasie van die kleinhandel-kredietbedryf voor en tydens die finansiële krisis te voorsien.

Om die werkings van die kleinhandel-kredietbedryf in Suid Afrika korrek te defineer en te verstaan, is 'n literatuurstudie onderneem om die wêreldwye finansiële krisis, Suid-Afrika se kleinhandel-kredietbedryf en kredietgradering opsigself te ondersoek. Die literatuur studie ondersoek en verduidelik die finansiële krisis en identifiseer sommige van die primêre dryfvere. Hierdie dryfvere sluit die VSA eiendomskrisis, die bekendstelling van subprima-lenings en die sekuritering van hierdie lenings (verband-gesteunde sekuriteite) in. Die dryfvere, veral die sekuritering van subprima lenings, was gebruik as ’n voertuig om die krisis wêreldwyd te versprei.

Die uiteindelike doel van die studie was om die individu en maatskappye, met 'n begrip van die finansiële krisis en die uitwerking op die verbruiker spesifiek deur middel van hul kredietwaardigheid en kleinhandel krediet gedrag te voorsien. Deur gebruik te maak van kredietgradering, het die studie bevind dat ten minste een handelaar (Woolworths) in die kleinhandel bedryf geraak is. Woolworths het sterker klem geplaas op die vermindering van hul krediet-blootstelling, terwyl verbruikers stadig maar seker hul fasiliteitgebruik verhoog het.

Sleutelwoorde: *Wêreldwye finansiële krisis, finansiële krisis, verbruikers krediet, kredietpuntetellings, telkaart ontwikkeling*

Table of Contents

1. Chapter 1: Introduction 1

1.1 Background 1
1.2 Problem Statement 3
1.3 Objectives 3
1.4 Motivation 4
1.5 Research Methods 5
1.6 Study Outline 6

2. Chapter 2: The global financial crisis 7

2.1 Introduction 7
2.2 The origin of the financial crisis 9

2.2.1 The US housing bubble 10
2.2.2 Emergence of the Subprime Market 20
2.2.3 Securitisation of Subprime Loans 24

2.3 The Effects of the crisis on the global economy 33

2.3.1 Introduction: A Global View 33
2.3.2 Impact of the financial crisis on Europe 35
2.3.3 Impact of the crisis on Asia 38
2.3.4 Impact of the crisis on South America 41
2.3.5 Impact of the crisis on Africa 45

2.4 Conclusion 47

3. Chapter 3: Consumer credit and credit scoring 51

3.1 Consumer credit in South Africa 51
3.2 Credit scoring 56

3.2.1 Historical development of credit scoring 56
3.2.2 Understanding the credit scoring process 58

3.3 Developing a credit risk scorecard 64

3.3.1 The people and process 64
3.3.2 Scorecard development process: Stage 1 (Planning) 68
3.3.3 Scorecard development process: Stage 2 (Data review) 71
3.3.4 Scorecard development process: Stage 3 (Creating a database) 76
3.3.5 Scorecard development process: Stage 4 (Data exploration) 83
3.3.6 Scorecard development process: Stage 5 (Scorecard development) 90

4. Woolworths Credit Scorecards 105

4.1 Empirical study introduction 105
4.2 Credit Scorecard Development (Pre-Crisis) 107

4.2.1 Data exploration 108
4.2.2 Cleansing the database 110
4.2.3 Binning 113

4.2.4 Identifying 'Good/Bad' Indicators 119
4.2.5 Regression Analysis and Data Normalisation 121
4.3 Credit Scorecard Development (Post-Crisis) 125
4.3.1 Explanation of significant variables 126
4.3.2 Scorecard Report – Plug&Score (Lite)® 127
4.4 Credit scorecard results and conclusion 127
4.5 Data analysis (Woolworths ISC Data) 129
4.5.1 Customer base 129
4.5.2 Defaults 130
5. Chapter 5: Conclusion 135
5.1 The credit milieu 135
5.2 Future research possibilities 138
5.3 Final assertion 139
6. References 141

List of Annexes

A1: Regression coefficients for XYZ Bank example 151
A2: Data normalisation for XYZ Bank example 153
B1: Statistics and outliers (June 2006 data) 155
B2: SAS Binning Code 161
B3: Basic Statistics and Outliers (July 2009 Data) 162
B4: Risk segmentation (2006 data) 168
B5: Risk segmentation (2009 data) 169
B6: Good vs. bad distribution (2006 data) 170
B7: Good vs. bad distribution (2009 data) 171

List of Figures

Figure 2.1: US House Prices 11
Figure 2.2: US Population growth 11
Figure 2.3: US homeownership rate 13
Figure 2.4: The US mortgage interest rate 16
Figure 2.5: Average Credit Score (FICO) in the US 23
Figure 2.6: Movement of the Mortgage Backed Security 27
Figure 2.7: Bangladeshi export numbers 39
Figure 2.8: Nepali export numbers 39
Figure 2.9: South American Exchange rates 44
Figure 2.10: Post-Crisis versus Pre-Crisis Growth (sub-Saharan Africa) 46
Figure 2.11: World and South African GDP (US$ billions) 48
Figure 3.1: Gross credit granted per quarter 52
Figure 3.2: Consumers with impaired records 53
Figure 3.3: Job losses by Limit versus Severe impact (LSM) 54
Figure 3.4: Performance definition 73
Figure 3.5: Bad rate 75
Figure 3.6: Scatter-plot of age 87
Figure 3.7: SAS Proc Univariate Code 88
Figure 3.8: Logistic regression code 97
Figure 3.9: Data Normalisation Code 98
Figure 4.1: Data cleansing code 111
Figure 4.2: Code to remove Outliers 113
Figure 4.3: Histogram of Open Balance (Bins) 114
Figure 4.4: Histogram of Credit Limit (Bins) 115
Figure 4.5: Histogram of Purchase Value (Bins) 116
Figure 4.6: Histogram of Payments Value (Bins) 117
Figure 4.7: Histogram of Current Due Value (Bins) 118
Figure 4.8: Histogram of Bureau Score (Bins) 119
Figure 4.9: Sample Dataset Procedure 120
Figure 4.10: Gini coefficient and Kolmogorov-Smirnov Test (2006/06 scorecard) 124
Figure 4.11: Gini coefficient and Kolmogorov-Smirnov test (2009/07 scorecard) 127

Figure 4.12: Approval vs bad rate (2006 data) 128
Figure 4.13: Approval vs bad rate (2009 data) 129
Figure 4.14: Woolworths ISC accounts..... 130
Figure 4.15: Percentage defaults each month 131
Figure 4.16: Number of new defaults each month 131
Figure 4.17: Purchases, payments and defaults..... 132
Figure 4.18: End balance versus credit limit (exposure) 133
Figure 5.1: Credit facilities at a "current" position 136

List of Tables

Table 2.1: Trends in down-payment percentages 17
Table 2.2: Total Originations: Consolidated and Growth 21
Table 2.3: Top Subprime Mortgage Originators 25
Table 2.4: Top Subprime MBS Issuers 25
Table 2.5: Performance of GSAMP 2006-NC2) 29
Table 2.6: China's Economy in 2007 - 2009 41
Table 2.7: South American Economies: Real GDP, Consumer Prices, and Current Account Balance 43
Table 3.1: Decision Tree 62
Table 3.2: Use for Credit Scoring 64
Table 3.3: Sample cohort analysis 74
Table 3.4: Moments of age variable 88
Table 3.5: General statistics of age variable 89
Table 3.6: Quantiles of age variable 89
Table 3.7: Extreme observations of age variable 89
Table 3.8: Variable list (XYZ Bank – personal loan database) 91
Table 3.9: Age variable 92
Table 3.10: Gender variable* 92
Table 3.11: PTI variable 92
Table 3.12: Marital status variable* 92
Table 3.13: Last work record variable 93
Table 3.14 Children variable 93
Table 3.15: Cards variable* 93
Table 3.16: Credit history variable* 93
Table 3.17: Home ownership variable* 93
Table 3.18: Time at branch variable 93
Table 3.19: Transformed data 94
Table 3.20: Good/bad distribution 94
Table 3.21: XYZ Bank logistic regression coefficients (age variable) 98
Table 3.22: Score table (age, cards and children variables) 100
Table 3.23: Scorecard example (Mr X) 100
Table 4.1: Variable dictionary of Woolworths ISC data 108

Table 4.2: Variable to be used in Scorecard 110
Table 4.3: Basic Statistics (Scorecard Variables) 112
Table 4.4: Range for Outliers (Scorecard Variables) 112
Table 4.5: Variable: Open Balance (Bins) 114
Table 4.6: Variable: Credit Limit (Bins) 115
Table 4.7: Variable: Value of Purchases (Bins) 116
Table 4.8: Variable: Value of Payments (Bins) 117
Table 4.9: Variable: Current Due Value (Bins) 118
Table 4.10: Variable: Bureau Score (Bins) 118
Table 4.11: Sample Dataset Format (Extract) 121
Table 4.12: Transformed Sample Dataset Format (Extract) 121
Table 4.13: Scorecard (2006/06 Data) 122
Table 4.14: Significant Contributors (Variables and Categories – 2006/06 Data) 123
Table 4.15: Gini coefficient (classification quality) 124
Table 4.16: Scorecard (2006/06 data) 125
Table 4.17: Purchase Variable (2006 vs 2009 Data) 126
Table 4.18: Significant Contributors (Variables and Categories – 2006/06 Data) 126
Table 5.1: Research Possibilities 139

Chapter 1

Introduction

1.1 Background

The word "credit" is derived from the Greek word "credo", which means *to believe*. This describes the philosophy behind what economists today (November 2011) view as *credit*. Cash transactions occur with a central belief that the payment will be made (Ferguson, 2008). The world is still recovering from the greatest financial crisis1 since the great depression of the 1920s, which not only adversely affected financial institutions, but also the general public. An immediate public effect, stemming from the crisis, has been the collapse in cash-out borrowing from home equity. In the US this figure tumbled from an estimated US$700 billion in 2005 to US$100 billion in 2007 (Zuckerman, 2007). This phenomenon has affected all credit-related areas, and it is consumers who are left bearing the brunt.

In effect, the credit crisis has substantially increased the cost of credit and global consumers are paying for the US sub-prime debacle in more ways than one. Although South Africa is currently experiencing the lowest interest rates since 1973 (Liberta, 2011), the crisis initially caused a wide-spread increase in interest rates that particularly negatively affected mortgages, but also those individuals with high debt-to-income ratios (Mills, 2008). Consumers spend more on fuel, food and utilities than ever before. This leads to extensive credit card debt, forcing consumers to fall behind on mortgage payments (Moneyweb, 2011:1). The increase in consumer prices is partially due to the decrease in business loans as higher funding costs are transferred to the public via product prices (Mills, 2008).

---

1 The 'credit crisis' which began in 2008 and has arguably not yet run its course.

When the crisis began, South Africa was under strong economic fundamentals due to a protracted period of economic expansion (Zini, 2008). The period of rapid economic growth has highlighted some vulnerable areas, such as unemployment, inequality, poverty and crime that have plagued the country in recent years (Zini, 2008). The public reacted to the financial crisis in different ways, either increasing their amount of individual debt or recommitting to reduce their credit obligations. South African consumer savings display a downward trend and they could be depleted in ten years if the trend continues (Cameron, 2009).

In 2008, Neville Chester from Coronation Fund Managers argued that South Africa avoided a large part of the crisis 'through large measures of good luck' and 'careful planning' (Donohoe, 2008). The stability in the banking sector can be largely attributed to the careful observation and regulation by the Supervisory Department of the South African Reserve Bank (Donohoe, 2008). The biggest threat to the South African economy is therefore not the credit crisis itself, but subsequent secondary effects. Exports of manufactured products such as motor vehicles, which have allowed large companies to expand their production within South Africa and so help to curb the domestic demand, have decreased dramatically (Anon, 2008). The new NCA (National Credit Act), which was implemented in 2007, has made the South African public much more aware of certain facts about credit and applying for it. The tightened regulations for companies who provide credit are placing emphasis on the importance of attracting creditworthy customers and the methods of doing so (Wizard, 2007).

In the wake of the financial crisis the concept of credit, especially credit ratings, have become an important issue in the economic milieu. Ratings from major agencies are important because potential investors view them as high quality credit assessment tools (Hunt, 2008). Rating agencies have existed for many years, but since the Mexican economic crisis of 1994 – 1995 much emphasis has been placed on these agencies to anticipate events rather than to react to them (Larrian, *et al,* 1997). The advent of new rating methodologies and the increase in computer power, as well as the availability of more data, has enabled these

agencies to effectively rate the creditworthiness of countries, investments and individuals alike, using a method known as credit scoring (Mester, 1997).

Credit scoring, a statistical method used to predict whether a client will default on future loan obligations, have been used by corporates and banks to analyse credit and whether clients should be granted loans or credit since the 1950s (Mester, 1997). Credit scoring is now of the utmost importance: in the current (2011) financial turmoil consumers are still very vulnerable to increases in prices and it adversely effects their ability to apply for credit (FinMark Trust, 2011). Credit scorecards consist of a group of characteristics, statistically determined to be predictive in separating good from bad customers (Siddiqi, 2006). This method has become increasingly effective and has spread from financial institutions to businesses that provide credit to their customers (Altman & Saunders, 1997:1721).

The world is awash with news and information regarding how the credit crunch will affect countries, exchange rates, investments and other financial derivatives. The largest contributor of information and data on these effects is the International Monetary Fund (IMF), which releases the World Economic Outlook annually (IMF, 2010). The public in South Africa should take note and be aware of the problems this crisis can create and the effects it may have on their lives in general (Kadazi, 2011:246). The common financial factor between the public and the retail sector is the credit that these companies provide. Credit risk has become a subject of great importance, especially in the banking sector and regarding residential property (home) loans, but the retail credit sector has also been affected.

1.2 Problem Statement

How has the financial (credit) crisis affected the South African retail credit market from a credit scoring perspective?

1.3 Objectives

This dissertation aims to investigate the effect that the financial crisis has had on the South African public, specifically through their credit worthiness and the

changes that have occurred in the retail credit sector over the period 2005 – 2008. The study will be applied to a selected company which caters for the sample of consumers' credit ratings which will be used in the investigation. This will enable the public, as well as the companies affected, to understand the effects of the crisis on their financial ability and their customers, respectively.

Secondary objectives include the critical evaluation of the characteristics of the data retrieved. This includes basic data exploration, such as income levels and demographic information, and more intricate statistical methods which will establish problem areas when applying for credit.

In addition, this dissertation will explore the method of credit scoring and will provide a step-by-step guide in building, understanding and interpreting a score card. This will enable businesses which provide credit facilities to be able to evaluate their needs for in-house credit scoring and using external companies as an alternative.

1.4 Motivation

This dissertation will be of value not only to the consumer applying for credit, but also to the companies (in the retail sector) that provides the credit. Credit lenders need to adhere to the NCA's regulations and undertake the necessary credit checks before approving individuals for credit. This must be applied across the board whether it is vehicle, furniture or retail credit (Wizard, 2007).

Companies benefit from using credit scoring techniques in several ways. Firstly these techniques decrease the time needed for credit approval considerably, since all the characteristics of the applicant are incorporated in a structured statistical model and do not require individual investigation. Secondly this scoring method significantly increases the objectivity of the credit approval process. This objectivity helps companies ensure they are applying the same underwriting criteria to all potential credit applicants regardless of race, gender, or other factors prohibited by law from being used in credit decisions (Mester, 1997).

Conservative buying behaviour, stemming from job losses and the fluctuations in the economy has severely affected the retail market in South Africa. Retail trade figures have declined since the emergence of the financial crisis and Statistics South Africa reported a decrease of 4.9% in March 2009 and 6.7% in April 2009 (BuaNews, 2009). Consumers grew increasingly cautious about spending since 2008 (Kazadi, 2011:231), but how the crisis affected the credit worthiness of the public will be explored in this dissertation.

1.5 Research Methods

The dissertation will explore three different study areas: a literature study, step-by-step guide to credit scoring and an empirical credit data study.

Data collection will focus on the credit applicants of a popular retail store, Woolworths, over the period of 2005 – 2008. The data will be divided to create two scorecards, one with a snapshot of data from June 2006 and another with data from April 2009. The characteristic data will be collected via Alistair Purvis, the Chief Information Security Officer for Woolworths Financial Services. An agreement was made with Woolworths not to convey any personal information of their clients within this study. The decision of only using a single retailer was made due to the availability of data with the characteristics required for credit scoring as well as the cost of obtaining such data.

The literature study will be focused on the financial crisis, the origin, causes, spread and implications it has on the world economy. It will provide an overview of the effects the crisis currently has on South Africa's economy as a whole and then specifically credit risk. The literature will also continue to explain how the public is affected by the changes in the economy. These factors will then be combined to evaluate the effect they have on the public's ability to apply for credit.

The step-by-step guide to credit scoring will provide a detailed explanation of how to implement the technique of credit scoring in a business. It will explain the development of a scorecard according to different stages, such as preliminaries and planning, data review, database creation, implementation and interpretation.

The guide will explain the mathematical and statistical formulae as well as the programming code that will be used in the development of the scorecard.

The two scorecards will then be developed with the data collected via Woolworths Financial Services. An empirical study will be conducted to determine the differences between the two scorecards. An interpretation will then be made to explain the differences between the two developed scorecards. The data will be also be statistically evaluated, to determine the significant characteristics that are taken into account when approving or declining a credit application.

1.6 Study Outline

The dissertation comprises five chapters:

Chapter 1: An introduction to the research at hand.

Chapter 2: A literature review, which includes the explanation of the financial crisis and how it began, evolved and spread globally. The effects this crisis has had on South Africa's economy and credit risk specifically whilst incorporating different factors which have influenced the retail credit market in South Africa.

Chapter 3: This chapter will be the guide to effective credit scoring. Each aspect of the scoring method will be examined and described. This guide – aimed at companies already providing credit facilities – will explain the credit scoring methodology.

Chapter 4: In this chapter the implementation and development of the scorecard will be discussed. Further data exploration and the characteristic regression will also be conducted and the results analysed.

Chapter 5: This chapter concludes the work and provides an interpretation, conclusions and recommendations for future work. Interesting facts and findings that the study has produced are analysed and discussed.

Chapter 2

The global financial crisis

2.1 Introduction

Although it may not be as evident and visible as the Great Depression of the 1920s, the global credit crisis may have had the worst effect on the world's economy since the great depression in 1929 (Zumbrun, 2010). Everyone active in the financial economy has felt the effects of this phenomenon, whether it is through tightening credit policies, declines in vehicle sales or the decline in small business due to lenders that are struggling with commercial real estate mortgages (Zumbrun, 2010).

Frederic S. Mishken's "Anatomy of a Financial Crisis" defines this phenomenon as follows:

> *"A financial crisis is a disruption to financial markets in which adverse selection and moral hazard problems become much worse, so that financial markets are unable to efficiently channel funds to those who have the most productive investment opportunities. As a result, a financial crisis can drive the economy away from equilibrium with high output in which financial markets perform well to one in which output declines sharply." (Mishken, 1991)*

South Africa has, to a large extent, not been affected as badly as some countries with stronger ties to the US financial markets (Suzman, 2011). This is evident in the bankruptcies of Lehman Brothers in the US in September 2008 (Baker, 2011), as well as Northern Rock and Bradford and Bingley (B&B) in the UK in 2009. These banks were provided with government assistance in order to repay clients. These banks predominantly used savers' deposits to fund lending (BBC, 2008). Northern Rock and B&B attempted to expand their mortgage books by borrowing short-term funds from the money markets enable them to lend more to potential home buyers (BBC, 2008). At that time global interest rates were low by historical standards, so investors invested their money in US money markets (the future

origin of the crisis). South African banks and the South African Reserve Bank steered clear of the worst problems that would have negatively affected the country's financial wellbeing its ability to remaining profitable (SARB, 2008).

It is unlikely that any country (including South Africa) will escape the on-going effects of the crisis. In 2009, the IMF reported the worst world economic growth in decades with the world economy shrinking by 1.4% (IMF, 2011). Although the South African banking system escaped most of the direct impact, their global economic environment has come under severe stress, with little capital available for investment. In the wake of the financial crisis it is critical for the world community to agree on new and higher standards of global governance, especially when considering financial stability. This issue is currently (2011) being addressed with the introduction of Basel III banking regulation (BIS, 2011). The global economy requires greater transparency and accountability – especially due to the deeply integrated nature of the world economy today (SAinfo, 2009).

The first signs of recovery from the crisis only began towards the end of 2009 as the global economy began to slowly, systematically recover (Mucha, 2009). The origin country of the global crisis, the US, felt the pressure of a multitude of different problems affecting their economy. The US experienced the following challenges during the first six months of 2009:

- US$787 billion government spending and stimulus injected into the economy (Longley, 2009)
- 14.7 million Americans were unemployed in 2009, the highest unemployment number since 1948 (Shellhammer, 2009).
- High amounts of defaults on short-term debt obligations in California (Shellhammer, 2009).
- US$1.3 trillion of American wealth was eroded in first quarter of 2009 (Shellhammer, 2009).
- Car sales dropped to 30 year lows in 2009 (Shellhammer, 2009).

Thus the first quarter of 2009 indicated that the US economy had been steadily declining, decreasing by 6.1% after a 6.3% decline in the fourth quarter of 2008

(Rappeport, 2009). The trillions of dollars spent to help out banks and other institutions and to stimulate the economy slowed and halted the decline, thus enabling recovery in 2010. Current IMF (2011) evidence indicates that the recovery, mooted in 2010, is only the beginning of a possible double-dip recession which continues to threaten the global economy.

The question is: Where did it all go wrong? What is the reason for this global financial phenomenon that the global economy experienced in 2008? Was there a single cause or multiple causes compounding into the eventual financial crisis? Was the outbreak of the crisis inevitable? The financial crisis has affected all relevant parties, including individual consumers (Kazadi, 2011:244). Understanding the cause of the financial crisis may shed some light on the impact it had (and will have) on South African consumers.

2.2 The origin of the financial crisis

The precise origin of the global financial crisis is not known, but what is known is that there were several causes which interacted catastrophically to result in the crisis. The economic shocks of this crisis still affects many countries' social fabric, not just in the US but also other countries around the world (Shiller, 2008: 9). History proves the necessity of economic policies for preserving the social fabric. After World War I, committees of European countries made a mistake by enforcing one particular economic policy: the Treaty of Versailles (Shiller, 2008:9). This treaty, which acted as a closing document to the First World War, imposed on Germany fines far beyond the country's ability to pay. This treaty was heavily opposed by economist John Maynard Keynes. He subsequently resigned from the British delegation and wrote the book *The Economic Consequences of the Peace*, wherein he predicted that the treaty would lead to the disaster for Europe. Never the less, Keynes' views were disregarded and in the end the treaty was upheld. Germany was never able to pay the financial penalties imposed. The treaty caused considerable resentment and was an important factor that led to World War II (Shiller, 2008:10). The financial crisis of 2008 may not have as direct a

cause as that of World War II, but several man-made, and entirely avoidable, factors contributed to its eruption. Some of the contributing factors included:

- the US housing bubble,
- subprime loans and
- mortgage backed securities.

The next section will examine the rise and fall of the US housing bubble.

2.2.1 The US housing bubble

A primary cause of the subprime crisis was the US housing bubble that began to burst in 2006. Baker (2005:1) warned of the significant negative impact this would have on the US economy. From 1950 – 1995 house prices in the US had been steadily rising at approximately the same rate as other goods and services, such as cars, fuel, groceries and healthcare. As seen in Figure 2.1, US house prices increased by about 83% (even though adjusted for inflation) between 1998 and 2006. This significant increase caused housing bubble wealth to be in excess of US$5 trillion. This mass amount of property wealth was measured to be the difference between the actual price growth and that of their historical trend, which was to increase by the same percentage as inflation (Baker, 2005:1). The increase in home prices could not be explained by factors such as population growth and rising incomes of the people in the US, because the increase in both factors did not differ from the trend of the years before (Baker, 2005:2). In fact, the significant post war population growth (also known as the 'baby boom' of the 1950s and 1960s) saw a greater increase population growth than the late 1990s (as seen in Figure 2.2).

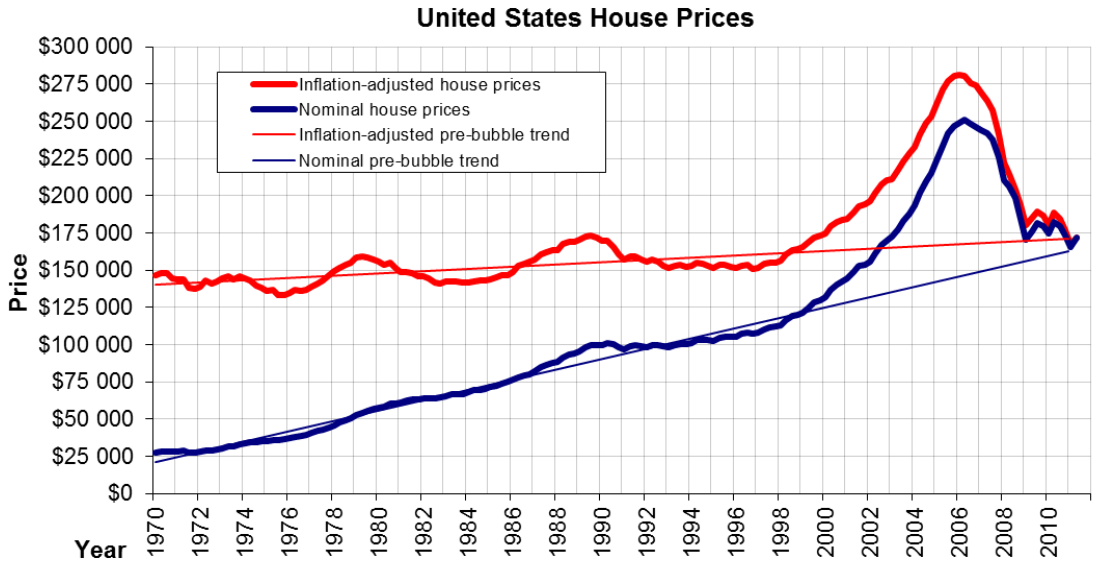


Figure 2.1: US House Prices

Source: Freddie Mac, 2011

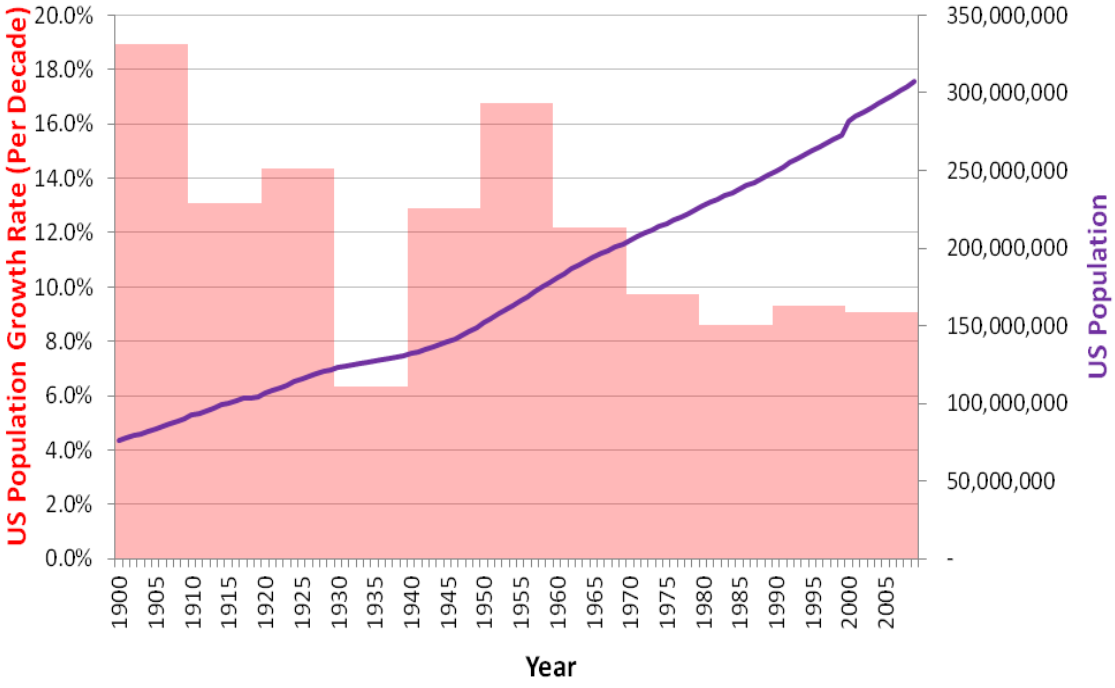


Figure 2.2: US Population growth

Source: US Census, 2010

House prices and rents are directly dependent, and although rent prices also increased alongside housing prices since the 1990s, they have *de*creased since 2003. Although not all of the states in the US were subject to the significant increases in home prices, the areas most affected by price increases included economic influential[2] areas in the US such as Washington DC and most of the East Coast north from the capital, the Pacific Coast region and many areas in between (Baker, 2005:2).

The main reason for the increase in home prices was the promotion of home ownership by the US government. According to the US Census, US home ownership rates rose in 1997 to 2005, in all regions, racial groups and income groups, between 65.7% and 68.6% (Figure 2.3). Most governments encourage home ownership as it boosts the capital flow in the economy, but the subprime crisis in the US is an example of the dangers of over-promoting ownership. Research conducted by Garriga, Gavin and Schlagenhauf in 2006 pointed out that the promotion of home ownership in the US happened through different channels.

[2] Economic influential areas: US cities that as a whole contribute most toward economic growth in that country.

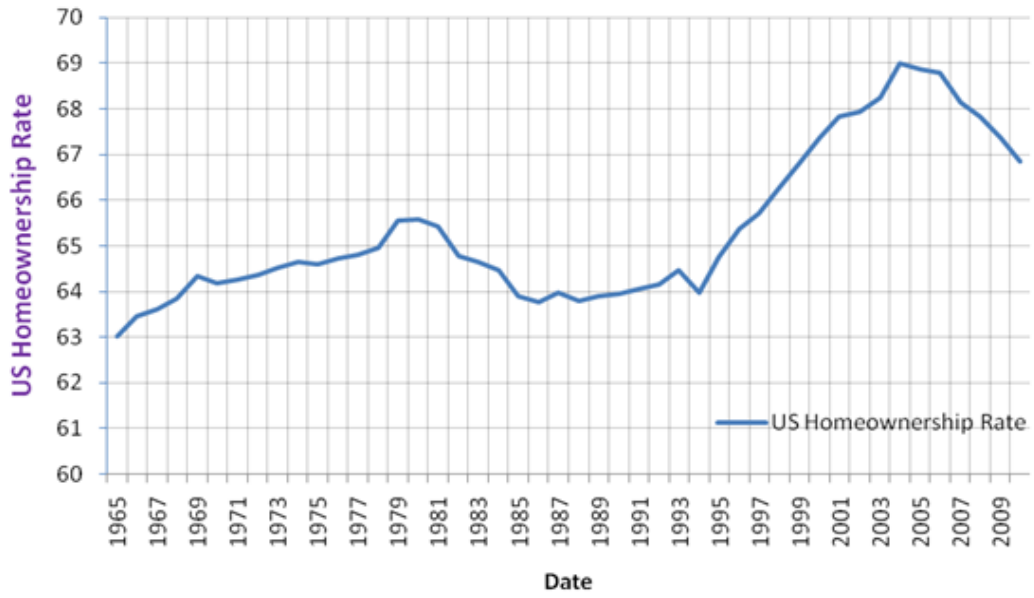


Figure 2.3: US homeownership rate

Source: US Census Bureau, 2011

2.2.1.1 Tax Law

The sixteenth amendment to the US constitution, which was passed in 1913, enabled a program that in a way supports homeownership through mortgage interest deduction. The program stated that homeowners who specify their tax returns were able to deduct interest payments on their home mortgage loans from their taxable income. Up until the 1986 tax reform, all interest payments on consumer loans could also be deducted from tax. This tax reform increased the value of mortgage interest deduction, because homeowners could substitute other types of debt for their mortgage debt. A calculation of the average marginal subsidy rate, done by the TAXISM program of the National bureau of Economic Research, showed that an average taxpayer would deduct 23 cents from their payable tax for the last dollar of mortgage interest paid. The tax law was thus a contributing factor to the increase in home ownership rates (Garriga, Gavin, Schlagenhauf, 2006).

2.2.1.2 The Secondary Mortgage market

The US government created two agencies to increase funds available to help finance mortgages during the Great Depression. The Federal Housing Administration (FHA), established in 1934, insured the public of long-term fixed-rate mortgages. In 1938 the Federal National Mortgage Association (Fannie Mae) was established to purchase FHA-insured mortgages. This acted as the origin of a secondary market for home mortgages. A rapid rise in homeownership after World War II, with homeownership rate rising from 40 – 64% by 1995, was argued by Colton (2003) that both the VA (Veteran's Administration, which helped war-veterans to purchase homes) and FHA was the primary cause for their sharp rise in home ownership.

A restructuring by Congress of the Fannie Mae into a government sponsored enterprise (GSE) in 1968 led to its new name of Government National Mortgage Association (Ginnie Mae), which took over the functions of its predecessor. Ginnie Mae was given authorisation to guarantee principal and interest payments on its securities. These securities was backed by VA and FHA loans thus enabling Ginnie Mae to offer guaranteed mortgage-backed securities in 1970 (Garriga et al., 2006).

2.2.1.3 Affordable Housing Programs

In the US the Department of Housing and Urban Development (HUD) established three affordable housing programs namely:

- Home Ownership Zone (HOZ)
- Home Investment Partnership Program (HOME) and
- Self-Help Homeownership Opportunity Program (SHOP).

The HOZ program enabled communities to regain vacant and deteriorated housing and other property, which led to an increase in homeownership and promoted economic growth by creating neighbourhoods of new family homes. These newly established suburbs came to be known as home ownership zones. In 1996 US$30 million subsidy was provided for the HOZ program and in 1997 a total of US$12.7

million was set aside (Garriga et al., 2006). Both subsidies were used to help 3400 new and refurbished homes in different cities of the US.

Another initiative, HOME, became a very important funding source for HUD's home ownership objectives. Grants were provided by HOME to local and state governments to help increase the homeownership rate under primarily low-income and minority households (Garriga et al., 2006). The HOME funds are utilised in different ways, such as:

- homebuyer programs
- rehabilitation of owner-occupied units
- rental housing development, and
- tenant-based rental assistance.

According to a study undertaken by HUD, it was found that US$3.1 billion in HOME funds helped 270 000 low-income households to buy homes between 1992 and 2002 (Turnham, Herbert, Nolden, Feins, Bonjouni, 2004).

SHOP helped non-profit organisations to buy home sites and also assist the development or improvement of the infrastructure required. These sites were then used to set up volunteer-based homeownership programs for lower-income households. An example of an entity benefiting from the SHOP program is Habitat for Humanity which received funding of US$25 million in 2003 (Garriga et al., 2006).This particular program targeted individuals and families who were willing to contribute their own time and effort into personal home improvement.

2.2.1.4 US Housing Costs

The interest rate associated with a mortgage is an important factor when considering the cost of a home. Figure 2.4 indicates the average mortgage rate on loans used to buy existing (2011) homes in the US. The interest rate dropped from above 16% in 1981 to 6% in 2005. The decrease in interest rates allowed more US citizens to buy a home instead of renting one (Garriga et al., 2006).

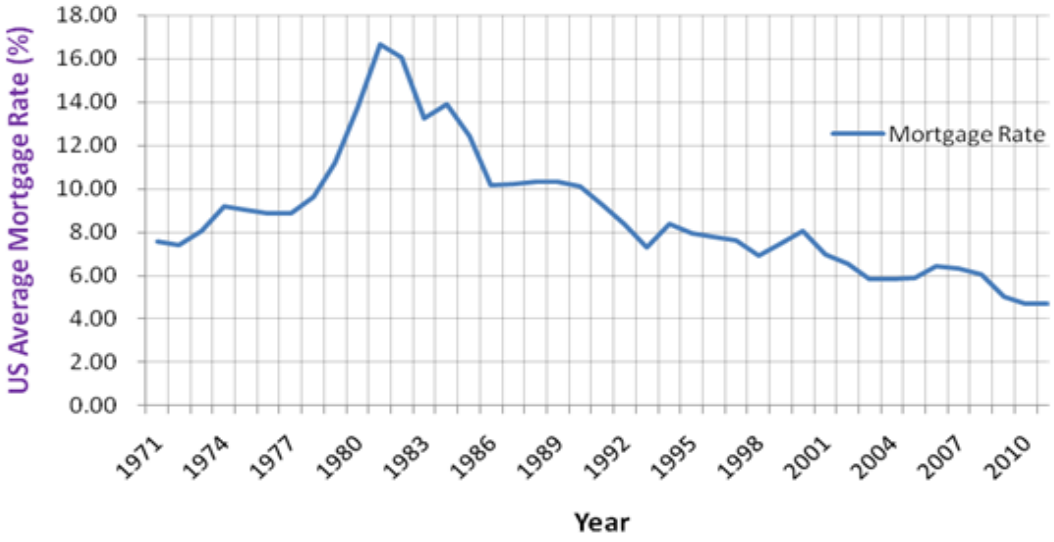


Figure 2.4: The US mortgage interest rate

Source: Freddie Mac, 2011

2.2.1.5 Financial Innovations

In the past, the US mortgage market was also affected by the FHA's introduction of mortgage insurance. This step encouraged the use of fixed-rate mortgages over a fixed 30-year period as well as a minimum down-payment. With financial innovations such as sub-prime lending, younger and low income households were able to buy homes with little or no down-payment. Chambers, Garriga, and Schlagenhauf (2005) found that innovations affecting the overall size of the down-payment are the most important factor when explaining the rise in homeownership.

A lower down-payment should in effect have an impact on the distribution of homeownership and thus the overall homeownership rate. Table 2.1 illustrates the average down-payment as a percentage of the loan size for first-time buyers as well as repeat buyers. There is a drop in the down-payment percentage between 1995 and 1999 due to the additional capital from the previous property sale. The decline in down-payment percentage could be explained by the increased use of private mortgage insurance (PMI). Fannie Mae and Freddie Mac regulations stated that a mortgage requires insurance when the loan-to-value ratio exceeds

80%. As a result, borrowers were able to buy a more expensive home than they previously could afford (Garriga et al., 2006).

Table 2.1: Trends in down-payment percentages

		Type of Loan	
Year	**Type of Buyer**	**FHA (%)**	**Other Loans (%)**
1995	First-time	21.6	29.8
	Repeat	22.0	33.3
	Total	23.2	33.5
1999	First-time	13.8	22.1
	Repeat	16.7	24.3
	Total	16.0	25.7
2001	First-time	16.3	24.1
	Repeat	26.5	28.5
	Total	22.6	27.0

Source: American Housing Surveys (1995,1999, 2001), Garriga et al. (2006)

2.2.1.6 Government Changes

In the middle of 2003 the US Federal Reserve decreased its repo rate to 1% (Shiller, 2008:40). That change led to the most rapid increase in home prices during the housing bubble. More important was the fact that the real (inflation-adjusted) federal funds rate remained negative for 31 months during 2002 and 2005. Shiller (2008:40) mentions the fact that this period of loose monetary policy should not be viewed as an external cause of the real estate bubble. The monetary policy, for both the Federal Bank and other central banks worldwide, was driven by the economic environment caused by the stock market bubble burst in the mid-1990s.

The impact of this loose monetary policy was then amplified by a large number of adjustable-rate mortgages (ARM's) issued after 2000, this included subprime mortgages. Because of the overall trend increasing homeowners, individuals became more involved in the real estate market to make their own profits. This led to a significant increase in ARM's financing was needed (Shiller, 2008:42).

With these methods the government in the US provided a wide range of consumers the opportunity to own a home. This made owning a home much more appealing and, as the demand for homes grew, the house prices increased in step, building the bubble year on year.

2.2.1.7 The bursting of the housing bubble

National leaders in the US conveniently turned a blind eye to the increase of home prices, as pride and the superiority of the capital system has always been approached with fervour. During the period of the US housing bubble, many authorities and political figures strongly denied or avoided the problem. Greenspan observed: "I would tell audiences that we were facing not a bubble but froth—lots of small local bubbles that never grew to a scale that could threaten the health of the overall economy" (Greenspan, 2007). President Bush also refrained from entering into conversation about the housing boom in his public pronouncements, instead focussing on successes. Shiller (2008:34) also mentions the comment made by Bernanke in 2005:

> *"House prices have risen by nearly 25% over the past two years. Although speculative activity has increased in some areas, at a national level these price increases largely reflect strong economic fundamentals, including robust growth in jobs and incomes, low mortgage rates, steady rates of household formation, and factors that limit the expansion of housing supply in some areas."*

These individuals were certainly aware of the possibility of the housing bubble, but the specific bubbles (stock market bubble of the 1990s and the housing bubble that followed) were, according to them, not significant enough to prompt any specific policy changes (Shiller, 2008:35).

The collapse began when these subprime mortgage loans started to default. Thus home foreclosures grew in number and the small subprime lender organisations went out of business. The primary reason for this happening was the increase of interest rates. Under Alan Greenspan the US were experiencing historic lows in interest rates between June 2003 and June 2004. These low rates encouraged people to buy homes more expensive than they could normally afford, but paying the same mortgage cost. Thereafter interest rates started to rise once again to battle inflation increases. The rise in interest rates severely affected subprime borrowers, and according to Tam (2009) in 2009, subprime home foreclosures were 24.1%, 25.7% and 27.2% in California, Florida and Nevada respectively. When the house buying trend had passed, the large amount of houses that was built during the boom time caused prices to sharply decrease. With this paradigm shift in the property market, investors and speculators were scared of holding on to real estate that would lose them money in the future. The house buying trend halted and people started selling off their property. The double supply of new homes for sale and old homes being sold only helped to drive property prices even lower (Tam, 2009).

The trouble quickly spread to Wall Street's investment banks as well as different hedge funds. The decrease in property values compounded the problems as defaulted mortgages were now worth much less to the banks and thus their investors. These investment banks were depended on borrowed money to purchase long term securities. These banks (for example Bear Sterns) were heavily invested in collateralised debt obligations (CDOs), which all became relatively worthless in a short amount of time (Ember, 2009).

Subprime loans were the primary reason for the housing bubble burst as it was responsible for significantly decreasing the home values in the US. The next section explains what subprime loans are and how they originated.

2.2.2 Emergence of the Subprime Market

Another financial innovation originating in the US mortgage market was the subprime market. This market made use of risk-based pricing to enable homebuyers who have poor credit ratings. For instance, the potential buyers who would fail a credit history check of the prime mortgage market would then look towards the subprime market to attain their necessary finance. The two major benefits to this type of lending are the increased numbers of homeowners as well as the opportunity for these owners to create their own wealth. It came at a price though, because of the complex nature of subprime lending it can either be of great benefit or risk to a company that extends credit. The promise of subprime lending is that it provides the opportunity of buying a home to individuals who either did not qualify or were discriminated against a bond in the past. In fact subprime lending is mostly found in suburbs with weaker economic conditions and a high concentration of minorities (Gruenstein & Zhai, 2005). However, because poor credit history is usually tied to default on previous loans and delinquent payments, the interest rates of subprime loans are much higher than normal rates (Chomsisengphet & Pennington-Cross, 2006).

The US only started to take notice of subprime loans in the mid-1990s. The growth of subprime lending in the decade prior to the financial crisis was quite significant. Table 2.2 indicates that total subprime (as stated by Inside B&C Lending) loans increased from US$65 billion in 1995 to US$332 billion in 2003 (Chomsisengphet et al., 2006).

Table 2.2: Total Originations: Consolidated and Growth

Year	Total B&C originations (US$ bn)	Top 25 B&C originations (US$ bn)	Top 25 market share of B&C	Total originations (US$ bn)	B&C market share of total
1995	65.0	25.5	39.3%	639.4	10.2%
1996	96.8	45.3	46.8%	785.3	12.3%
1997	124.5	75.1	60.3%	859.1	14.5%
1998	150.0	94.3	62.9%	1,450.0	10.3%
1999	160.0	105.6	66.0%	1,310.0	12.2%
2000	138.0	102.2	74.1%	1,048.0	13.2%
2001	173.3	126.8	73.2%	2,058.0	8.4%
2002	213.0	187.6	88.1%	2,680.0	7.9%
2003	332.0	310.1	93.4%	3,760.0	8.8%

Source: Inside B&C Lending, Chomsisengphet et al. (2006)

The subprime growth market's acceleration was in a big part due to it becoming legal practice in 1980 (Chomsisengphet et al., 2006)**.**This was documented and made possible by the Depository Institutions Deregulation and Monetary Control Act (DIDMCA). The act enabled financial institutions to increase their rates and fees. Then in 1982 the balloon payment was introduced. These two specific laws allowed for the emergence of the subprime market, but it was the introduction of the Tax Reform Act (TRA) in 1986 which allowed the subprime market to become an incredibly profitable industry (Chomsisengphet et al., 2006)**.** By prohibiting the deduction of interest on consumer loans, the TRA eventually caused an increase in the demand for mortgages. This was because the act allowed interest deductions on mortgages for a primary residence as well as one additional home. Due to low and steadily declining interest rates, cash-out refinancing[3] became a useful tool for homeowners to properly determine their home's value.

[3] The term *cash-out refinancing* indicates that a new loan is larger than the previous loan and the borrower will receive the difference in cash.

More than half of all subprime loan organisations in the early 2000s were for cash-out refinancing (Chomsisengphet et al., 2006)**.**

During the 1990s subprime mortgages were relatively new and seemed profitable at first glance, but the actual performance of these loans in the long run were unknown. By 1997 the arrear amounts and defaulted loans were already above expected levels. According to Temkin, Johnson, and Levy (2002) many lenders had under-priced subprime mortgages in the growing competitive mortgage market during the mid-1990s.

With the focus set on subprime borrowers, many banks tried to successfully identify these individuals in order to take appropriate action. The 2001 Interagency Expanded Guidance for Subprime Lending Programs (Ashcraft & Schuermann, 2008:14) describes the subprime borrower as an individual who displays different types of characteristics pertaining to credit risk, which includes one or more of the following:

- has two or more 30-day delinquencies in the past year or one or more 60-day delinquencies in the past two years,
- has undergone judgment, foreclosure, repossession, or charge-off in the past year,
- has been declared bankrupt in the past five years,
- has a relatively high default probability as estimated by a credit risk bureau to have a FICO score of 660 or less (depending on collateral and product). Or any other risk bureau with similar default probability likelihood, and
- has a debt service-to-income ratio of 50 or higher (Ashcraft & Schuermann, 2008:14).

On average in the US, adjustable rate mortgage (ARM) borrowers have a lower FICO[4] score than fixed rate mortgage (FRM) borrowers (Figure 2.5). To give an example, in 2003 the average FICO score of ARM's were 52 points less than that of the FRM's (623 versus 675).

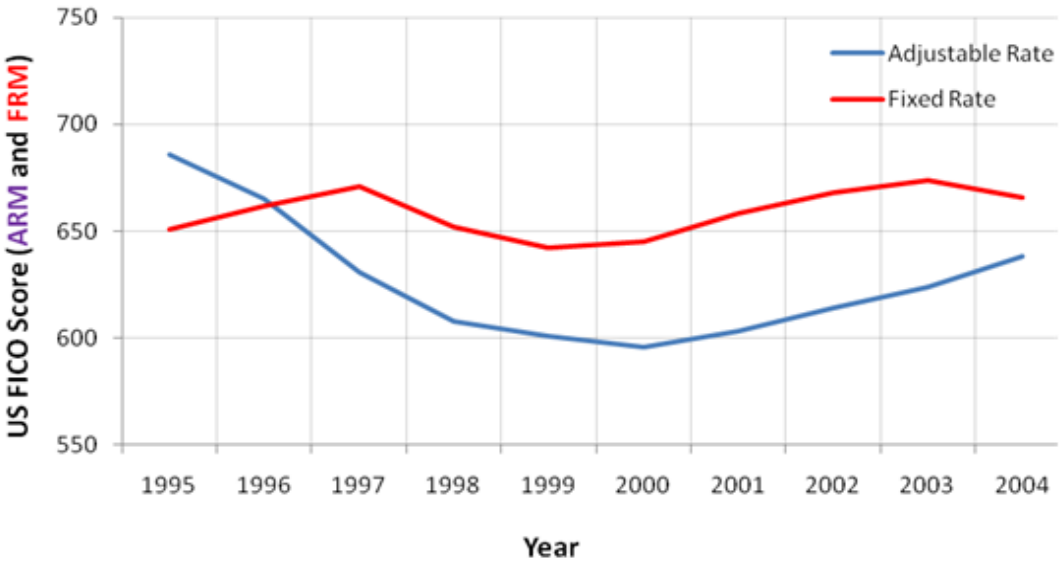


Figure 2.5: Average Credit Score (FICO) in the US

Source: Loan Performance ABS securities data base of subprime loans, Chomsisengphet & Pennington-Cross, 2006

Average credit scores decreased during the 1990s, but since 2000 they started to increase. Thus subprime lenders continued to expand during the 1990s and so extending credit to less credit-worthy borrowers (Chomsisengphet et al., 2006).

The default of subprime mortgages would have only affected the credit institutions and banks who gave them, thus confining the subprime crisis to the US. But these banks made use of mortgage backed securities (MBS) to sell these subprime loan-books to investors from all around the world. These mortgage backed securities thus served as a financial vehicle to change the subprime crisis to the global

[4] A credit score, developed by Fair Isaac Corporation to determine (calculate) a consumer's credit worthiness or credit risk.

financial crisis. The next section looks at and investigates mortgage backed securities.

2.2.3 Securitisation of Subprime Loans

Banks in the US wanted to hedge their risks and attract investors at the same time. Mortgage backed securities (MBS) were developed to achieve this goal. Investors from all over the world invested in the banks and financial institutions in the US and in doing so took on the risk of subprime loan books as well.

2.2.3.1 Mortgage Backed Securities

In essence mortgage backed are debt obligations that represents claim to cash flows on a big pool of mortgage loans. This is most commonly applicable to residential property. The securities were also grouped in one of the top two ratings as determined by an accredited credit rating agency (Investopedia, 2011). These mortgage loans originate in banks and other mortgage loan providers. The top mortgage loan providers (before the global financial crisis) are shown in Table 2.3 below. After these loans were issued by the companies, they were sold to different entities (Table 2.4) and are pooled together in categories. The previously mentioned entities then issue securities that represent claims on the principal and interest payments to be made by borrowers on the loans of the subprime pool, a process is called securitisation.

Table 2.3: Top Subprime Mortgage Originators

Rank	Lender	2006		2005	
		$US bn	Share (%)	$US bn	Share (%)
1	HSBC	52.8	8.8%	58.6	-9.9%
2	New Century Financial	51.6	8.6%	52.7	-2.1%
3	Countrywide	40.6	6.8%	44.6	-9.1%
4	CitiGroup	38.0	6.3%	20.5	85.5%
5	WMC Mortgage	33.2	5.5%	31.8	4.3%
6	Fremont	32.3	5.4%	36.2	-10.9%
7	Ameriquest Mortgage	29.5	4.9%	75.6	-61.0%
8	Option One	28.8	4.8%	40.3	-28.6%
9	Wells Fargo	27.9	4.6%	30.3	-8.1%
10	First Franklin	27.7	4.6%	29.3	-5.7%
	Top 25	543.2	90.5%	604.9	-10.2%
	Total	600.0	100.0%	664.0	-9.8%

Source: Inside Mortgage Finance, 2007

Table 2.4: Top Subprime MBS Issuers

Rank	Lender	2006		2005	
		$US bn	Share (%)	$US bn	Share (%)
1	Countrywide	38.5	8.6%	38.1	1.1%
2	New Century	33.9	7.6%	32.4	4.8%
3	Option One	31.1	7.0%	27.2	15.1%
4	Fremont	29.8	6.6%	19.4	53.9%
5	Washington Mutual	28.8	6.4%	18.5	65.1%
6	First Franklin	28.3	6.3%	19.4	45.7%
7	Residential Funding Corp	25.9	5.8%	28.7	-9.5%
8	Lehman Brothers	24.4	5.4%	35.3	-30.7%
9	WMC Mortgage	21.6	4.8%	19.6	10.5%
10	Ameriquest	21.4	4.8%	54.2	-60.5%
	Top 25	427.6	95.3%	417.6	2.4%
	Total	448.6	100.0%	508.0	-11.7%

Source: Inside Mortgage Finance, 2007

2.2.3.2 New Century Financial Example

Ashcraft & Schuermann (2008) studied the securitisation of 3,949 subprime loans with a combined principal balance of US$881 million. These subprime loans were originated by New Century Financial in the second quarter of 2006. This specific study is of special interest because it clearly illustrates how subprime loans, from a severely underperforming book, came to be issued. The example also discusses the structure and ultimately the sell to investors. What follows is a summary of the actual study.

In both 2004 and 2006 New Century Financial was the second largest lender of subprime mortgages in the US, providing subprime loans to the value of US$51.6 billion (Inside Mortgage Finance, 2007). Due to the automated internet-based loan submission and the integrated pre-approval system FastQual, mortgages increased at a compound annual growth rate of 59% during the period of 2000 and 2004. New Century Financial started to struggle with early payment defaults at the start of 2007 and finally filed for bankruptcy in April of 2007 (Ashcraft & Schuermann, 2008:13).

Goldman Sachs originally bought the subprime loans, originated by New Century Financial. They in turn sold the loans to a bankruptcy-remote special purpose vehicle called GSAMP TRUST 2006-NC2 (Figure 2.6).

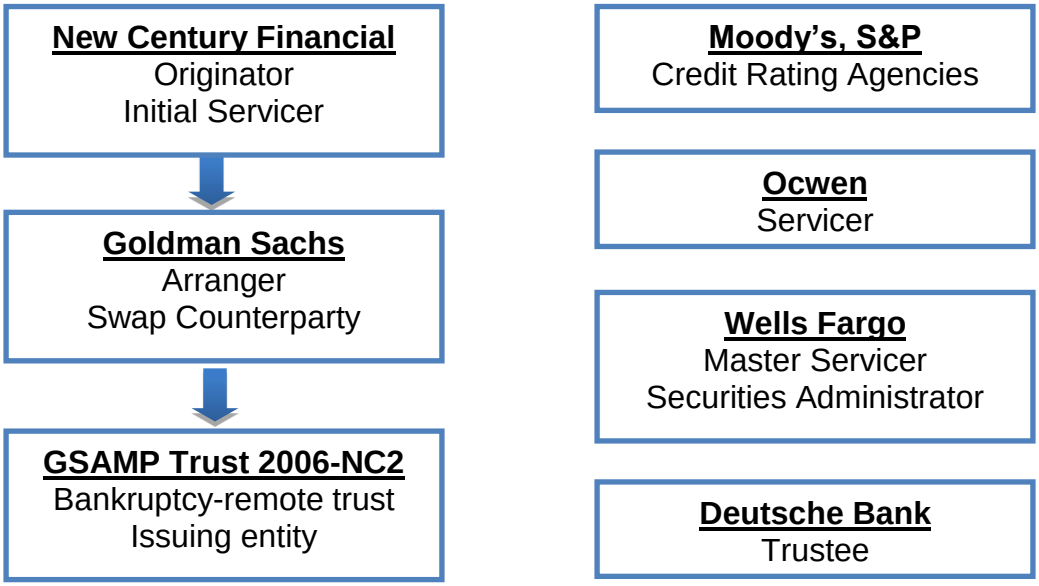


Figure 2.**6:** **Movement of the Mortgage Backed Security**

Source: Ashcraft & Schuermann, 2008

To be able to purchase these subprime loans the trust issued asset-backed securities. New Century was responsible for these loans in the beginning but with the creation of the loans were moved to Ocwen Loan Servicing. Ocwen received a fee of 50 basis points (US$4.4 million) annually to take care of the subprime loans (Ashcraft & Schuermann, 2008:14).

The pool of subprime mortgage loans that New Century used as collateral in the securitisation provided the following statistical summary:

- first-line loans accounted for 98.7% and the remaining partition is second-line home equity loans,
- 43.3% of the loans were purchase loans, given to the borrower to purchase a property. The remaining portion was used for cash-out refinancing,
- 90.7% of the mortgagors claimed to occupy the property as their primary residence and the remaining mortgagors were claimed to be investors,
- 73.4% of the loans were for single-family homes. The rest were separated into multi-family dwellings or condos,
- the two main states in the securitisation were California and Florida with 38.0% and 10.5% respectively,

- the average borrower in the pool had a FICO score of 626. The proportion was as follows: 31.4% had a FICO score below 600, 51.9% between 600 and 660, and 16.7% above 660 and
- the ratio of total debt service of the borrower to gross income (income before taxes) was 41.78% (Ashcraft & Schuermann, 2008:15).

These statistics indicated that the primary purpose of most of the above mentioned homes was not to buy a new home but to refinance an existing home (Ashcraft & Schuermann, 2008:15). Secondly, while almost none of the subprime lenders had a FICO score of above 660 in this pool, they were more aggressively underwritten than those with a lower score. Also, loans to borrowers with high FICO scores tend to be much larger were less likely to be owner occupied as well (Ashcraft & Schuermann, 2008:15).

The most important issue facing the subprime market in 2005/2006 was the impact of payment reset[5] on the monthly payments of borrowers (Ashcraft & Schuermann, 2008:21). A study undertaken by Cagan (2007) on mortgage payment resets attempted to explain what portion of resetting loans will end up in foreclosure. Cagan (2007) suggested that in an environment with no home price appreciation and 100% employment, 12% of subprime loans will end up in foreclosure due to reset.

Since this study was undertaken in 2007, market conditions have deteriorated dramatically and the originations of subprime loans have all but disappeared. These conditions have made Cagan's assumptions about equity risk look quite optimistic even under worst case stress scenarios. This is also true of the author's assumption that reset risk is constant across credit as a whole. Subprime lenders were thus be more likely to foreclose due to reset, resulting in the estimates of foreclosure found in the research reports of US investment banks to be inaccurate (Ashcraft & Schuermann, 2008:23).

[5] Payment reset: This happens when a bank recalculates the rates issued toward loans to adapt to the current economic stance of the country.

The payment resets thus had a significant impact to the defaults of subprime loans, many of which ended up in foreclosure. This thus also had a negative impact on the mortgage backed securities investors bought in the US as in the case with the New Century Financial example.

Table 2.5 depicts the performance of the GSAMP 2006-NC2 deal in August 2007. Column one to three report on mortgage deals in the pool that are 30-days, 60-days and 90-days delinquent. Loans in foreclosure are documented in column four and column five shows loans where the bank has the property title. Column six reports actual cumulative losses and the final column indicate the fraction of original loans that remain in the pool.

Table 2.5: Performance of GSAMP 2006-NC2)

Date (2007)	30d	60d	90d	Foreclosure	Bankruptcy	REO	Cum loss	CPR	Principal
Aug	6.32	3.39	1.70	7.60	0.90	3.66	0.25	20.35	72.48
Jul	5.77	3.47	1.31	7.31	1.03	3.15	0.20	20.77	73.90
Jun	5.61	3.09	1.43	6.92	0.70	2.63	0.10	25.26	75.38
May	4.91	3.34	1.38	6.48	0.78	1.83	0.08	19.18	77.26
Apr	4.68	3.38	1.16	6.77	0.50	0.72	0.04	15.71	78.68
Mar	4.74	2.77	1.12	6.76	0.38	0.21	0.02	19.03	79.84
Feb	4.79	2.59	0.96	6.00	0.37	0.03	0.00	23.08	81.29
Jan	4.58	2.85	0.88	5.04	0.36	0.00	0.00	28.54	83.12

Source: Cagan, 2007

UBS (2007a) developed an approach to use actual deal performance in the calculation of lifetime loss estimates. The author reports that around 70% of loans within a 60-day, 90-day and bankruptcy categories will eventually default on their mortgages in a four month period. This assumption was made using historical data on loans which were in an environment of low home price appreciation (< 5%) (Ashcraft & Schuermann, 2008:24).

Default numbers "in the pipeline"[6] for remaining loans in the next four months was constructed as follows:

$$\boldsymbol{Pipeline\ Default} = \mathbf{0.7} \times (\boldsymbol{60d} + \boldsymbol{90d} + \boldsymbol{bankruptcy}) + (\boldsymbol{foreclosure} + \boldsymbol{real\ estate\ owned}) \qquad \textit{(2.1)}$$

where:

- 60 d: Payment of mortgage loans 60 days past due
- 90 d: Payment of mortgage loans 90 days past due
- Bankruptcy: Customers in bankruptcy after mortgage foreclosure
- Foreclosure: Mortgage loans terminated and property repossessed
- Real estate owned: Loans in which the bank has title to the property

Thus the pipeline default for the GSAMP 2006-NC2 is calculated as 15.5%. It suggests that this fraction of loans still residing in the pool is subject to default. The total default calculation is done by combining the above measure with the remaining fraction of loans in the mortgage pool, the actual cumulative losses to date as well as an assumption about the different severities of losses. In UBS (2007a) an LGD (loss given default) of 37% is assumed.

$$\boldsymbol{Total\ default} = \boldsymbol{pipeline\ default} \times (\boldsymbol{fraction\ of\ loans\ remaining}) + \left(\frac{\boldsymbol{Cum\ loss}}{\boldsymbol{Loss\ severity}}\right) \qquad \textit{(2.2)}$$

where:

- pipeline default: calculated in Equation 2.1.
- fraction of loans remaining: fraction of original loans remaining in pool.
- Cum loss: actual cumulative losses.

[6] Pipeline: the mortgage payments that are 60 days, then 90 days behind and finally end up in foreclosure. Thus "Pipeline Default" refers to the amount of loans that are in the process towards foreclosure.

Thus applying this calculation to the GSAMP 2006-NC2, a total default of 11.9% is calculated. This fraction would then have defaulted in four months.

Lastly, UBS (2007a) paper used historical data to calculate the total number of defaults across the lifespan of the mortgage backed security. It specifically looked at a certain mapping that is constructed between the weighted-average loan age and the fraction of lifetime default. For example, an average structured deal realises about 33% of its defaults by month 13, 59% of the defaults by month 23, 75% by month 35 and 100% by month 60.

$$\boldsymbol{Projected\ cumulative\ default} = \left(\frac{\boldsymbol{Total\ Default}}{\boldsymbol{Default\ timing\ factor}}\right) \qquad \textit{(2.3)}$$

where:

- Total Default: default numbers at a specific point in time.
- Default timing factor: numbered factor of the specific month.

Application of this knowledge to the New Century example, where the pool of mortgage loans originated in May 2006, calculates the average loan to about 16 months old at the end of August 2007. The default timing factor for 20 months calculated to 51.2%. It was used since default predictions were calculated four months in the future. This then suggests that the projected cumulative default for this mortgage pool is 23.2%. Assuming a loss given default at 37%, results for the expected life time loss of this mortgage pool is 8.6%.

The backward looking nature of the above approach and the fact that it excludes payment reset are potential weaknesses when reviewing the approach. When looking at payment reset, losses are likely to be more than the historical curve used above. This implies that the number of actual life time losses observed up till now, is going to be too small, resulting in the predicted life time losses which are too low. UBS (2007b) addressed this problem by developing a "shut-down" model, which takes into account the problem of borrowers refinancing to mitigate payment reset. In this "shut-down" model UBS incorporates a loss severity of 45%. These assumptions imply that a more conservative view on losses would be to scale those from the loss projection model above by a factor of two. This would provide

a lifetime loss rate of 17.2% for the New Century mortgage pool (Ashcraft & Schuermann, 2008:25). The above examples show that the securitisation of mortgage loans initially seemed to be a good idea, but after careful analysis was put forth, that assumption was contradicted.

2.2.3.3 The Tipping Point

The culmination of all the above mentioned events including subprime lending, the creation of special vehicles for these mortgages and the burst of the housing bubble led to the financial crisis. It originated in the US and led to the collapse of large financial institutions, the assistance provided to banks by national governments and downturns in stock markets around the world (Baily, Litan & Johnson, 2008). Across the globe economies slowed during this period as credit tightened and international trade declined. Credit experts condoned credit rating agencies and investors who failed to accurately price the risk involved with mortgage-related financial products. In addition to this governments did not adjust their regulatory practices to address the new millennium's financial markets and central banks had to respond with an array of counter measures including bailouts, fiscal stimulus and expansion of monetary policy (UNENF, 2010).

The three subjects discussed in this section; housing bubble, subprime mortgages and mortgage backed securities, were the main drivers behind the global financial crisis. The mortgage backed securities served as a vehicle to spread the financial problem, created in combination by the housing bubble burst and subprime lending, to the rest of the world (Baily Litan & Johnson, 2008). Countries who were heavily invested in the US felt the effects of the crisis more than those who were financially insulated against them. The biggest effect though, was not the direct impact of the crisis, but the second round effects it carried. These second round effects such as the drop in exports (due to a country's financial strain) had and will have a lasting effect on many countries around the world. The next section details some of the effects the global financial crisis had on different regions of the world (United Nations, 2009).

2.3 The Effects of the crisis on the global economy

2.3.1 Introduction: A Global View

Each country had experienced the global financial crisis in different ways. It depended a lot on the type of financial exposure the country had to the US, or the exposure their primary economic partners had. Because of continuous integration around the globe, there was no way that such a crash in the world's leading nation would not affect the rest of the globe. Financial ties and investments in US banks spelled chaos for all investors around the financial world. As these companies were affected they in return affected the economy in that country, leading to considerable stock market crashes (Topçu, 2009:2). These impacts filtered down to the individual citizens as their finances and cost of living had to carry the burden.

In January 2009 the International Monetary Fund (IMF) released an update pertaining to its World Economic Outlook (IMF, 2009). The update significantly downgraded expected growth for 2009 due to the trying global economic conditions. The estimates of global GDP dropped from 2.2% in November 2008 to 0.5% in the updated outlook. The actual figure dropped to an even lower number as the world's economic growth contracted by 0.5% in 2009 (IMF, 2011). Trouble engulfed the financial world as businesses and consumers experienced a very difficult period with retrenchments, bankruptcies and all round low confidence (Topçu, 2009:2).

The global economic community now faced a financial crisis that originated in the US. Thus the subprime mortgage crisis which became more apparent during 2007 and 2008 can be characterised by constricted reserve of liquidity in the global credit markets and banking systems. A considerable slow-down in the economies of developed countries has seen a halt in development as well as affecting their financial markets. As house prices dropped to losses of around US$2.4 trillion in as short as eight months before July 2008, it led to a crash in the US financial markets and thus in turn other developed and developing countries worldwide (Lin, 2008).

Stock markets crashed globally as developed and developing countries saw drops of around 50-75% after their recent (before crisis) stock-market highs (Cali, Massa, Velde. 2009). The US lost around US$16.2 trillion in 2009 and a number of investment banks collapsed or had to be rescued. The IMF began to support countries like Hungary, Iceland and the Ukraine. Interest rates were cut worldwide in October 2008 as a response to the crisis and some had fallen further as the world tried to react against the global financial crisis (Cali et al., 2009).

As a result of the previously discussed increase in subprime mortgages and the coinciding increase in real estate prices led to global shortage in liquidity within the financial markets. This in turn led to a severe stock market crash affecting institutions across the board. These institutions included insurers, reinsurers, pension funds and hedge funds. This also led to currency turmoil in many countries, which affected developing countries, the most (Topçu, 2009:4).

Global economic growth was not the only aspect to suffer negatively in the financial crisis. According to Topçu (2009:5) the crisis also affected the world in the following ways:

- **The decline in world trade and investment levels:** The IMF reported that global trade volumes will decrease by an estimated 2.8% in 2009, which is in stark contrast to the expected 2.0% growth reported in November 2008. The actual figure dwarfed all IMF expectations as world trade volumes decreased by 10.9% (IMF, 2011).
- **An increase in global unemployment:** January 2009 saw an unprecedented 4.8 million unemployment claims in the US, the highest claim count since the US started keeping records in 1967. The uncertainty caused a great unease and loss of confidence, which led to a decrease in consumer spending around the world. (Topçu, 2009:5). The drop in spending led to:
- **A drop in price inflation and demand for goods and services:** headline inflation dropped to below 1% in February 2009 for advanced economies (IMF, 2009). This in essence eased the situation for consumers but could

also have encouraged them to hold back on spending as prices were perceived to be falling.

- **Financial strains on governments:** This was due to lower tax revenues generated from companies and employment as well as higher pay-outs for unemployment and on-going financial aid. Government spending increased worldwide as the UK government, for instance, the budget deficit was almost £90 billion in 2009 (Seager, 2009).
- **Pressure on Government bodies:** Pressure increased on governments and bodies such as the IMF to take measures to combat the problems. In January 2009 the US, for instance, passed a bill for a US$787 billion bailout package in January 2009 (New York Times, 2010).
- **Social and Political unrest on the rise:** In both developing and developed countries, the public started to assign blame for their economic woes. In the UK, for instance, strikes broke out in January 2009 over foreign labour issues (Topçu, 2009:6).

2.3.2 Impact of the financial crisis on Europe

The growth and development of social, economic and even political areas in Europe has undergone significant growth and change over the last twenty years, but the impact of the global financial crisis has put much of this forward momentum under pressure. According to Jelašić (2009) many central and eastern European countries (CEE) were unaware or underestimated the effects of the global financial crisis due to the fact that their economy is built on borrowing, foreign investment and, for some members, funds from Brussels (Jelašić, 2009). This economic model, especially using excessive borrowing has placed many countries in serious default problems since 2010 continuing through to today. These models thus need a complete overhaul. In the wake of the global financial crisis, even as many countries face sovereign default in 2011, problems cannot be resolved by only placing faith in help from the World Bank, IMF or European Union. These entities should however aid the troubling countries in the areas of health, education and pension structures of the troubled European countries.

Another problem that regulators within the CEE countries faced and are still facing in 2011 is that they have limited tools at their disposal to act proactively. This is because most of the banks are owned by financial institutions which have headquarters in the EU (Jelašić, 2009).

The European Central Bank and the Bank of England cut their interest rates to record lows on the 5th of March 2009 in reply to the economic slowdown across Europe and the European Union (Perry & Walker, 2009). Official numbers confirmed that in 2008 the European Union's gross domestic product decreased to 0.8%. The worst growth under the global financial crisis in Europe was experienced in 2009 when the economy contracted by 4% (CIA, 2011). Collapsing global trade has had serious implications for export dependent countries like Germany which is the world's leading exporter and fourth-biggest economy. During the global financial crisis, it has also been among the more conservative countries. German leaders were against spending and lending huge sums to fend off recession, such as the previously stated US$787 billion bailout package for the US, and warned that it will backfire in the long run. This statement holds particular value as Standard & Poor's downgraded the AAA rating of the US to AA+ in August 2011. This decision came after the US raised their debt ceiling up from US$ 14 trillion (Detrixhe, 2011). But by October 2008, however, German officials accepted that they will have to introduce support schemes (Perry & Walker, 2009). With global customers curbing their spending, BMW's and other luxury German exports were finding it difficult to make the trip abroad. The country eventually posted a shrink of 4.7% in GDP for 2009 (CIA, 2011), pushing authorities to consider large infusions of money to assist struggling European countries (Whitlock, 2009).

The pressure was evident across Europe as France, another big economic entity, announced on the 21st of January 2009 that the government added a US$13.6 billion cash injection into the struggling domestic banking sector. This announcement in conjunction with other state aid brought the total intervention amount of the French financial-services industry to US$27.2 billion between 2008 and 2009 (Scott, 2009).

By that stage, only Société Générale, the second largest bank in France, had agreed to a recapitalisation of a further US$2.2 billion. The top French bank, BNP Paribas, was down 30% over the same period between 2008 and 2009 (Scott, 2009).

With economic pressure surfacing everywhere Euro-zone unemployment increased to 8.2% in January of 2009 as companies had to dismiss countless workers, this added more pressure to the already struggling governments in terms of unemployment funding (Perry & Walker, 2009).

These problems, experienced as a direct result of the financial crisis, also led to the mismanagement of the Greek economy. The Greek government deceived world governing bodies by hiding elements on their national budgets, which led to the Greek Sovereign debt crisis (De Grauwe, 2010). The ousting of the Greek government led to scrutiny of other European countries and gave rise to the PIIGS (Portugal, Italy, Ireland, Greece, and Spain) sovereign debt crisis. The financial markets and European Union authorities were also responsible for letting the crisis degenerate into a systemic across Europe. It was essential to stop the Greek debt crisis as soon as possible as the default of the Greek government would adversely affect government bond markets across the Eurozone. On 11 May 2010 Europe agreed to a €500 billion rescue package to stop the panic of the Greek sovereign debt crisis spreading any further. This rescue package consisted of €440 billion from the European Union members as well as €60 billion from the existing European Union budget for helping the economies struggling with sovereign debt (Watts, 2010). A further €250 billion bailout package was offered by the IMF (Watts, 2010). The substantial bailout package came after discussion and deliberation at an emergency summit in Brussels. After the summit, the European Union's commissioner for economic and monetary affairs indicated that the size of the fund shows that the IMF would defend the Euro at all costs (Watts, 2010).

Being but a temporary solution to Greece's debt woes, a strategic solution has to be found. This solution will spark major changes in the Greek economic and budgetary policies. It will also aim to address the problem that the monetary union is not embedded in a political union throughout the Eurozone: the ultimate goal

would be political unification De Grauwe (2010). The proposed structural and policy changes have an important quality of being signals of a determination of the member countries to build act proactively in building the political future. Such signals make it clear that members are determined to preserve and build the European Union (De Grauwe, 2010).

2.3.3 Impact of the crisis on Asia

Most of Asia, especially South Asia, was still recovering from the food and fuel price crisis when the global financial crisis arrived. The food and fuel crisis already had a significant effect on income losses which ranged from 34% in the Maldives to 8% in Bangladesh (World Bank, 2009). South Asia, by and large, already faced numerous macroeconomic difficulties in countries such as Pakistan, Sri Lanka, and the Maldives. This was due to fragile political and social environments and the global financial crisis then put further strain on the already vulnerable countries. Although India responded very well to the food and fuel price crisis it was significantly impacted by the global financial crisis because of their strong relations larger to world economies (World Bank, 2009). Smaller countries like Bangladesh, Nepal, and Bhutan have been largely unaffected by the financial crisis because of the small nature of their financial markets. The World Bank (2009) predicted that the second wave of effects of the crisis could affect these countries in an adverse manner with export earnings and tourism receipts experiencing slowdown. This though proved not to be the case as Bangladesh exports rose over the period 2008 to 2010 and Nepal's exports remained stable (as seen in Figure 2.7 and Figure 2.8).

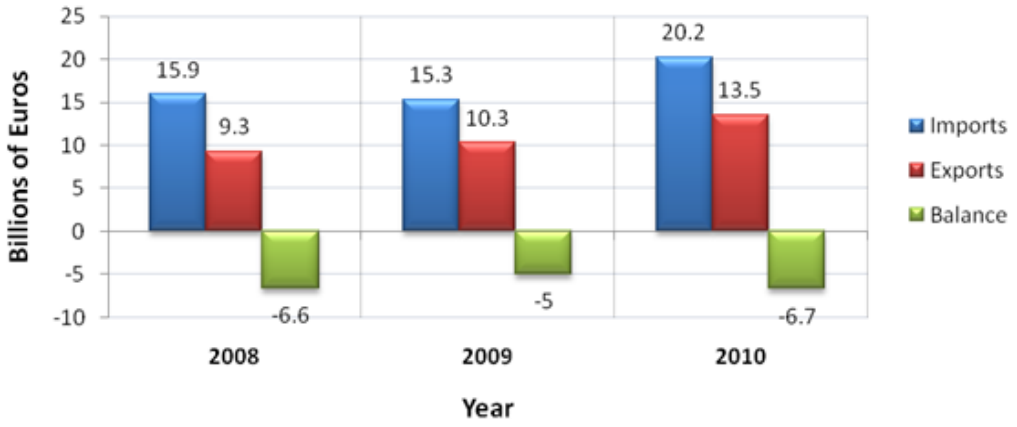


Figure 2.7: Bangladeshi export numbers

Source: IMF (Direction of Trade Statistics - DoTS)

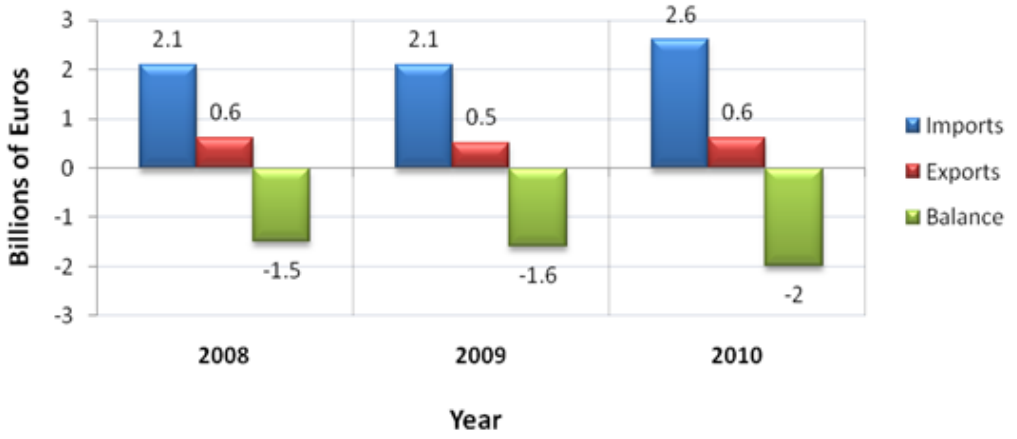


Figure 2.8: Nepali export numbers

Source: IMF (Direction of Trade Statistics - DoTS)

South Asia's largest economy, India, came under considerable strain with the effects of the global financial crisis. The crisis had negative effects on various areas in the country including drop in the stock market, net capital outflows and a decrease in foreign investment (World Bank, 2009). These effects combined to cause a fast declining exchange rate and a big increase in short term interest rates. The second round effects of the crisis have also taken its toll on the country, as a slowdown in domestic demand and exports slowed the Asian economy. Between the months October and November in 2008 exports declined.

There was an estimated loss of 500 000 jobs in India during same period. The GDP growth was estimated by the World Bank to drop to 7% for 2008, but India's GDP actually dropped by a larger margin to 4.9% in 2008 (World Bank, 2011).

The Indian government employed measures to counteract the effects of the global financial crisis by stabilising the exchange rate. Foreign exchange reserves were also stabilised around US$250 billion for 2009 (World Bank, 2009). The government still keeps a close eye to changes in the monetary and fiscal measures to keep the financial sector stabilised. They made an effort to ensure that the country possesses adequate liquidity, and try to stimulate domestic demand. The country did have a rise in non-performing loans, but as a whole the financial sector has developed a risk centric culture. The drop in inflation rates, due to a decline in global fuel and commodity prices, has created an opportunity for monetary easing and the introduction of a stimulus package. (World Bank, 2009).

In the Middle-East, the initial forecast for growth was estimated in 2009 to be 3.9%, this figure indicated that the area would be one of the most resilient to the crisis effects. However, oil dependant countries such as Saudi-Arabia entered recession in January 2009 with a drop in GDP growth from 4.2% in 2008 to 0.6% in 2009 (World Bank, 2011).

Most economists had the opinion, prior to the financial crisis, that China would be able to detach itself from the West when a global crisis hits. This observation was proven as unfounded as China also felt the effects of the global financial crisis. This was especially true for the southern China, due to their focus on exports and light industry (Schmidt, 2009).

Schmidt (2009) reports that in this area thousands of companies have had to dismiss tens of thousands of people as most have gone bankrupt or decreased significantly in size. The financial sector underwent a dramatic decrease as two thirds of the stock market was wiped out from the period October 2007 to November 2009 (China Stock market index dropped from 6000 to 2000) (Trading Economics, 2011).

This rapid decrease in the Chinese stock market index was mainly due to big Western banks and companies selling their shares in the Chinese financial sector. This enabled them to raise capital to help counter the crisis (Schmidt, 2009).

The crisis has also severely affected to the CIC (Chinese Investment corporation), which is China's developing sovereign wealth fund. The CIC adopted a strategy of investing heavily in Western companies, which did not pay off (Schmidt, 2009). China's gross domestic product growth rate fell by 4.6 % in 2008. Exports also decrease by US$ 228 billion from 2008 to 2009. On the import side, numbers show a decrease of US$ 126 billion in from 2008 to 2009 (See Table 2.6 below).

Table 2.6: China's Economy in 2007 - 2009

	2007	2008	2009
GDP (billion CNY)	26581	31404	34051
GDP Growth Rate (%)	14.2%	9.6%	9.1%
Unemployment (%)	4.0%	4.2%	4.3%
Inflation Rate (%)	4.8%	5.9%	-0.7%
Exports (billion US$) Imports (billion US$) Trade Surplus (billion US$)	1218 956 262	1430 1132 298	1202 1006 196
Current Account Balance (billion US$)	371.8	426.1	261.1

Sources: Statistical Bureau of the Peoples' Republic of China, CIA World Factbook, World Bank

2.3.4 Impact of the crisis on South America

South America has had its fair share of financial unrest in the past, the debt crisis of the early 80s which re-emerged in 2002. The financial impact of the global financial crisis had a lesser effect in South America in comparison to previous crises (Jara, Moreno, Tovar, 2009). This according to Jara et al (2009) was primarily due to overall improved and stable bond markets as well as net balance sheet positions in these countries.

As with many countries and regions across the globe, effects on South America intensified significantly with the bankruptcy of the Lehman Brothers in September 2008 (Jara et al., 2009). The downturn in the international banking system tied with the decrease in investor risk appetite resulted in a decreased demand for emerging market assets. This created another sharp depreciation in the currencies of emerging market economies (EMEs). Some countries entered swap agreements to stabilise their reserve positions, the largest of these was the US Federal Reserve Bank deal, which saw Brazil, Mexico, Singapore and South Korea enter into temporary currency arrangements for the provision of dollar liquidity (Boorman, 2009). There are a few elements that separate the global financial crisis from previous South American crises. First, and most importantly, the crisis found its origin in far more advanced economies rather than in South America itself. Secondly, the fact that South American countries had a much lower public external debt base gave their respected governments the ability to more effectively stabilise the private markets. And thirdly, this crisis was a new type of issue. It was not a macro economic issue faced in the past but more focussed on banking and financial stability (Jara et al., 2009).

At the beginning of the crisis in August 2007, South America had experienced an average real GDP growth rate of around 5.7% (Table 2.7). As the global financial crisis matured South America's real GDP growth slipped down to 4.2% in 2009 and the combined economies contracted by 1.5% in 2009 (IMF, 2009).

Table 2.7: South American Economies: Real GDP, Consumer Prices, and Current Account Balance

	Real GDP				Consumer prices (1)				Current acc bal (2)			
	07	08	09	10	07	08	09	10	07	08	09	10
Western Hemisphere	**5.7**	**4.2**	**-1.5**	**1.6**	**5.4**	**7.9**	**6.6**	**6.2**	**0.4**	**-0.7**	**-2.2**	**-1.6**
South America and Mexico (3)	**5.7**	**4.2**	**-1.6**	**1.6**	**5.3**	**7.7**	**6.7**	**6.3**	**0.7**	**-0.3**	**-1.9**	**-1.3**
Argentina (4)	8.7	7.0	-1.5	0.7	8.8	8.6	6.7	7.3	1.6	1.4	1.0	1.8
Brazil	5.7	5.1	-1.3	2.2	3.6	5.7	4.8	4.0	0.1	-1.8	-1.8	-1.8
Chile	4.7	3.2	0.1	3.0	4.4	8.7	2.9	3.5	4.4	-2.0	-4.8	-5.0
Colombia	7.5	2.5	0.0	1.3	5.5	7.0	5.4	4.0	-2.8	-2.8	-3.9	-3.3
Ecuador	2.5	5.3	-2.0	1.0	2.3	8.4	4.0	3.0	2.3	2.4	-3.5	-2.3
Mexico	3.3	1.9	-3.7	1.0	4.0	5.1	4.8	3.4	-0.8	-1.4	-2.5	-2.2
Peru	8.9	9.8	3.5	4.5	1.8	5.8	4.1	2.5	1.4	-3.3	-3.3	-3.2
Uruguay	7.6	8.9	1.3	2.0	8.1	7.9	7.0	6.7	-0.8	-3.6	-1.7	-2.4
Venezuela	8.4	4.8	-2.2	-0.5	18.7	30.4	36.4	43.5	8.8	12.3	-0.4	4.1
Central America	6.9	4.3	1.1	1.8	6.8	11.2	5.9	5.5	-7.0	-9.2	-6.1	-7.1
The Caribbean	5.8	3.0	-0.2	1.5	6.7	11.9	4.0	5.8	-1.5	-2.8	-5.1	-4.1

1. Movements in consumer prices are shown as annual averages.
2. % of GDP.
3. Includes Bolivia and Paraguay.
4. Private analysts estimate consumer price index (CPI) inflation has been considerably higher.

Source: IMF, World Economic Outlook, 2009

Consumer prices had a brief spike in 2008 but their increase remained stable at 6.7% and 7.6% for 2009 and 2010 respectively. Gross capital inflows began to recede substantially from early 2008 and over the year as a whole, which led to a decline in South America's gross portfolio inflows (Jara et al., 2009). According to Jara et al (2009) the reason for this decline in gross capital inflows was an increase in international risk aversion. This risk aversion saw financial institutions in developed countries trying to increase liquidity, stabilise their balance sheets and battle against high currency volatility. With an overall decrease in capital flows, the downward movement of commodity prices and the severe loss of confidence following the bankruptcy of the Lehman Brothers, the currencies across South America experienced severe depreciations. This also increased the cost of external financing of the area (Jara et al., 2009).

Figure 2.9 indicates a major change in South American exchange rates in September 2008. This was due to the declared bankruptcy of Lehman Brothers (indicated by the red line on Figure 2.10) on 15 September 2008. The event was the largest trigger of financial crisis in the South American region.

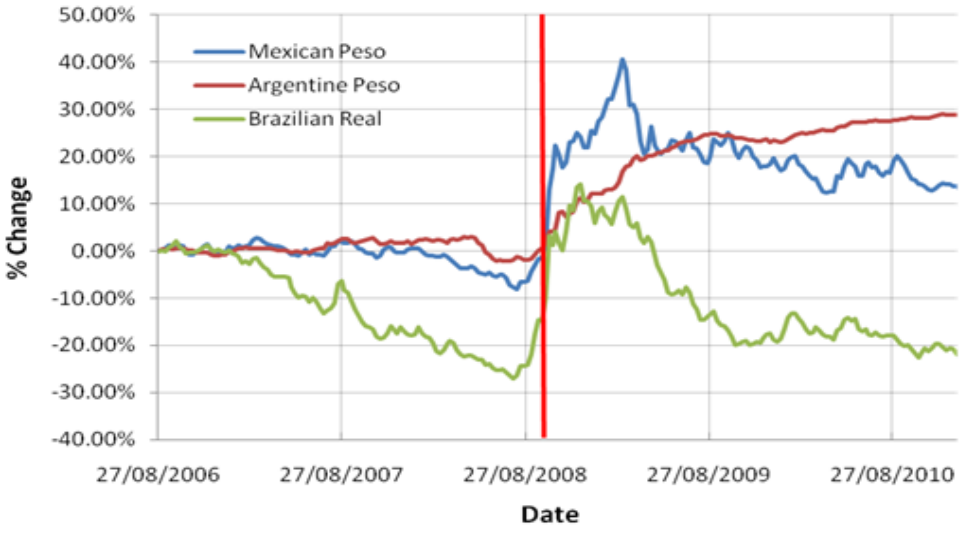


Figure 2.9: South American Exchange rates

Source: OANDA, 2011

Although South America was affected by the global financial crisis, the impact to most of the region's domestic financial markets has so far been less severe in comparison to that of other developing regions. Jara et al (2009) contributes this fact to the foreign asset accumulation by residents in some countries as well as the contributions made in the past to the domestic debt markets across the South American region.

2.3.5 Impact of the crisis on Africa

Following with the trend of developing regions around the world, Africa was expected to be relatively unaffected by the effects of the global financial crisis because of various different reasons. These reasons included factors like the relatively small foreign investment (of African countries in developed industrial countries) which in effect would provide weak financial links to Africa (IMF, 2009). But the second round effects have affected and will continue to influence the economies of a developing African market in terms of investing and import and export trade. The United Nations Conference on Trade and Development (UNCTAD) estimated a 9.2% decrease in overall exports for Africa in 2009 (UNCTAD, 2009). The actual figure was a decrease of 15% as every corner of Africa experienced a significant decrease in trade (WTO, 2010).

There was a general feeling that the short-term effects on many African countries were likely to be mitigated by the fact that most of them are relatively decoupled from the global financial system (United Nations, 2009). Moreover, this was because of the simplistic and conservative banking system found in many African countries. Many African banks stick to the rigorous standardised approach set forth by the Basel II accord. These banks also maintain prudent financial regulations, controls on foreign exchange and were very rarely exposed to subprime loans. These factors have shielded the continent's financial structures from the full effects of the crisis (United Nations, 2009).

This did not mean that Africa was unaffected by the financial crisis. The medium and long-term effects are the ones affecting the developing region the most, as millions of Africans were pushed over the poverty line in 2008 and 2009. The IMF's World Economic Outlook report in April 2010 stated that there was a decline in real GDP for Sub-Saharan Africa from 5.5% in 2008 to 2.1% in 2009 (IMF, 2010).

The tightening in global credit has also led to a strong decrease in private investment flows and bank financing which hampers the banks' capital flows as well as trade financing. A decline in demand for African exports in 2009 also contributed to a sharp fall in commodity prices and trade figures (Balchin, 2009).

The IMF stated in their "World Economic Outlook" of 2010 that the current account deficit for the entire Sub-Saharan Africa widened by 3.3% of GDP to reach 2.1%. Sub-Saharan GDP declined to 2.1% from 6.9% in 2007 and overall inflation increased to 10.6% from 7.1% in 2007 (Seen in Figure 2.10 below). Furthermore, many African countries face the increasing pressure of debt repayment as the developed countries also tighten their policies and lending activities and focus more on recovering on their loans (Balchin, 2009).

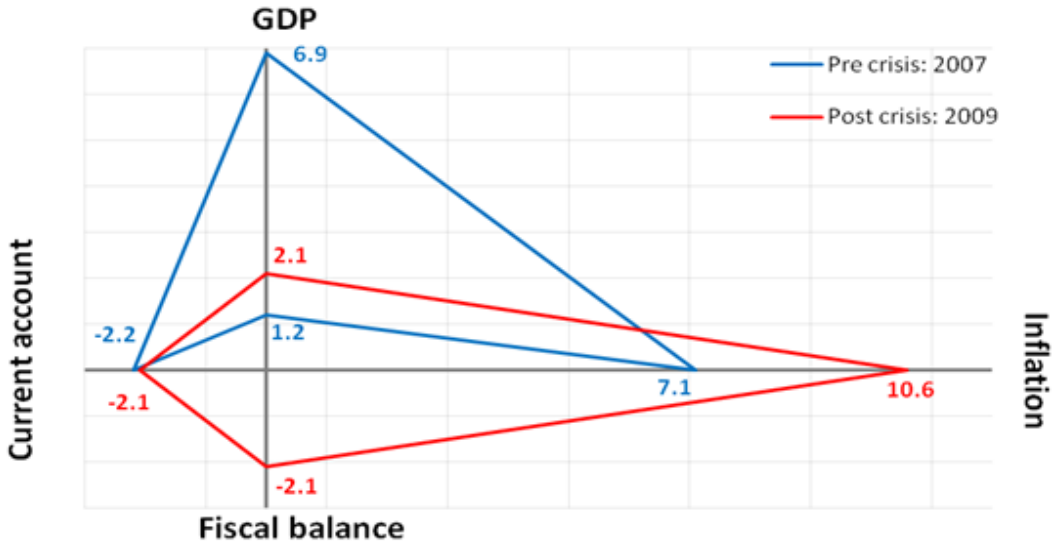


Figure 2.10: Post-Crisis versus Pre-Crisis Growth (sub-Saharan Africa)

Source: (IMF, 2009) (IMF,2010)

In response, Africa has to focus on building towards growth over the long term. African countries and especially South Africa has adopted a myriad of different steps to effectively mitigate the impact of the global financial crisis. The most notable of these steps were to include interest rate cuts, increased liquidity for banks and firms (with Basel III on the horizon), recapitalisation of financial institutions, labour and trade policy changes as well as focussing on regulatory reforms (Balchin, 2009). Each country applied these measures differently according to need, vulnerability and fiscal space (United Nations, 2009). Most though, set up task forces to investigate and advise government on the financial crisis. The biggest issue facing most African countries is the lack of financing to enable them to employ these mitigating factors. Africa will thus have to rely on the international community to lend support to slowly enable emerging countries in Africa to become more economically independent (United Nations, 2009).

2.4 Conclusion

The world and its economy faced and are still facing numerous challenges against the effects, whether primary or secondary, of the financial crisis. Numerous different countries, emerging and developed, have implemented steps to fight against the adverse effects of the financial crisis. Of course every region and country is unique in its approach to combating the crisis, as economic circumstances vary quite significantly across the globe.

Figure 2.11 indicates the world GDP at purchasing power parity (PPP), which gives the sum value of all goods and services produced in the world valued at prices prevailing in the US. The GDP at purchasing power parity is a measure used when investigating and comparing per capita welfare, living conditions and resource use across countries. The graph shows a drop at 2008 and 2009 after steady growth in the previous years. This gives a clear indication that not only did GDP fall globally, but also the means to purchase goods and services declined worldwide.

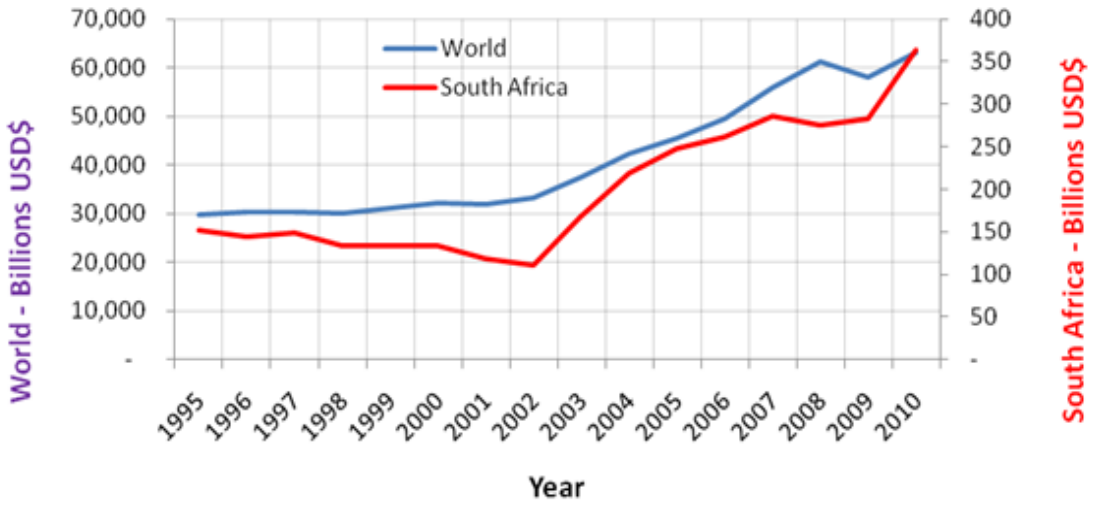


Figure 2.11: World and South African GDP (US$ billions)

Source: World Bank, 2011

Whilst the world is slowly putting the global financial crisis to bed, new financial and social pressures are pushing the world economy towards a possible double-dip. With the democracy movement in Egypt, Libya and Syria, Middle East governments have to focus on economic as well as social pressures. But these events in the Middle East are overshadowed economically by the growing issue surrounding sovereign debt.

The global financial crisis has put the world's financial affairs under a microscope. So much so that the current sovereign debt crisis has come to the fore. The IMF has extended loans to Greece, Portugal and many other European countries struggle with debt (Rooney, 2011). Ireland, for example has a debt to GDP ratio of well over a 100%, up from a relatively low 25% before the global financial crisis (Harrison, 2011). Even the world's largest economy, the US, came near to defaulting on its debt in August 2011. The government was forced to intervene and raise the debt ceiling by US$2.4 trillion to stop a potential US default (MarketWatch, 2011).

Due to strict policy and fiscal discipline shown by the government over recent years, South Africa's debt to GDP is only 35% in 2011 (Cohen, 2011). South Africa though will do well to address the catalising factor of Greece's debt woes, government corruption. Through years of concealing corruption and actual debt, Greece fell into their current sovereign debt woes (Cohen, 2011).

South Africa's strict fiscal measures and the reserve bank sticking to tight exchange controls, led to financial institutions to not, by and large, purchase the financial assets which were reconstructed by Wall Street at the height of the global financial crisis. This provided some protection for the country against most of the direct effects. Though, SA's GDP did not remain untouched by the effects of the crisis, as indicated in Figure 2.11 above (Baxter, 2009).

As countries around the world slashed interest rates, increased inflation, introduced stimulus packages and were seeking out every opportunity to shield against the crisis, the individual at the bottom of the chain was affected in almost every aspect of their daily lives. The effects had the most prominent impact on each individual's salary or wage at the end of the month, with increase in commodity prices, fuel, taxes etc.

The financial crisis has put the whole world on the defensive, and that includes not only the governments, policy makers and economists, but also the everyday consumer. People around the world are tracking their credit and financial situation on a much more structured and detailed level than before, focussing on rebuilding savings and debt repayments (ArticlesBase, 2010). There has been a worldwide decline in consumer credit during the financial crisis and this can be seen primarily in the mortgage and vehicle markets (ACEA, 2009). Countries have adopted stricter credit approval regulations which resulted in a lower rate of credit approvals. This is evident in South Africa as well, with the introduction of the National Credit Act implemented on the 1st of June 2007 (Nedbank, 2007).

The next chapter explores the effects of the financial crisis on South Africa's consumer credit market, focusing more on the everyday credit such as clothing and food (from retailers such as Woolworths) than primary consumer credit such as home loans and vehicle credit finance. Thereafter the investigation will continue into the development and build of basic credit scorecards used by credit institutions (including retailers) in South Africa to process applications for store credit. Once guided through the development and building method of these scorecards, actual scorecards will be constructed (in Chapter 4) and analysed to determine the effects and facts of the financial crisis on consumer credit in South Africa.

Chapter 3

Consumer credit and credit scoring

3.1 Consumer credit in South Africa

The financial ability of the public is interlinked with every economic decision the government makes and in South Africa one such decision was the National Credit Act (NCA), fully implemented in 2007 (NCA, 2006). The NCA in general protects the public of being granted credit recklessly. South Africa, previous to the introduction of the new credit act was granting consumer credit at a steadily growing pace. A survey conducted by Feasibility, an economic research organisation, revealed an increase of 32% active accounts between 2002 and 2008 (Feasibility, 2009). They also observed that since June 2007 there was a notable decline in the granting of credit, pointing to stricter affordability assessment regulations (Feasibility, 2009). Since the NCA's inception, credit providers reported an increase in credit rejection rates as well as a decreased rate in new accounts. According to the National Credit Regulator (NCR, 2011) there was a 50% decline in consumer credit since the fourth quarter of 2007 (R102bn) and the second quarter of 2009 (R50.9bn) as shown in Figure 3.1. The most significant impact came in mortgages and vehicle credit.

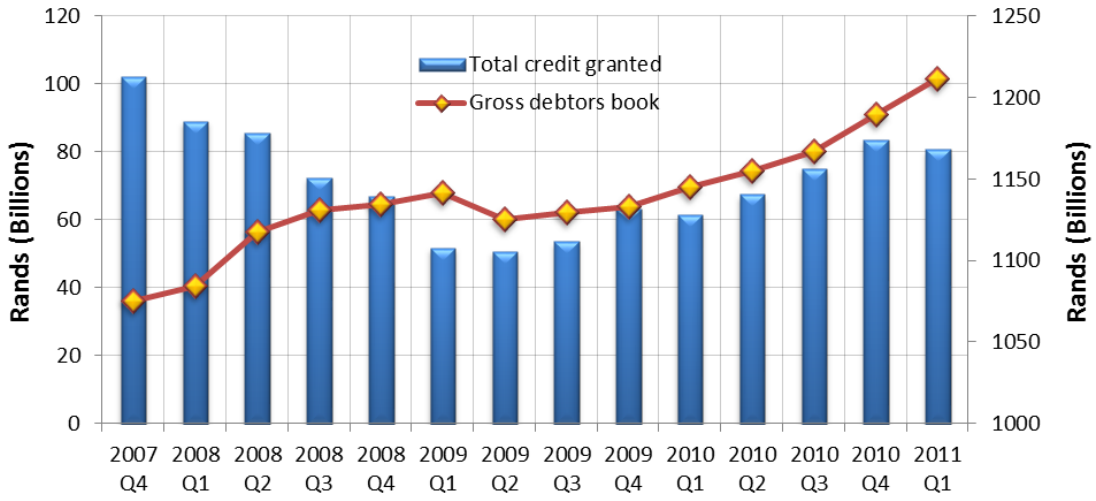


Figure 3.1: Gross credit granted per quarter

Source: NCR, 2011

The global financial crisis saw a significant decline in consumer credit in South Africa. There has also been an increase in the amount of impaired accounts since 2007 (NCR, 2009). The NCR analysed the credit performance of 17.8 million consumers, and concluded that there was a move from 36% to 42% impaired accounts between 2007 and 2009 (Figure 3.2 below). Of these investigated accounts 50 000 people were added to the impaired category per month and 1.7 million in total fell into arrears.

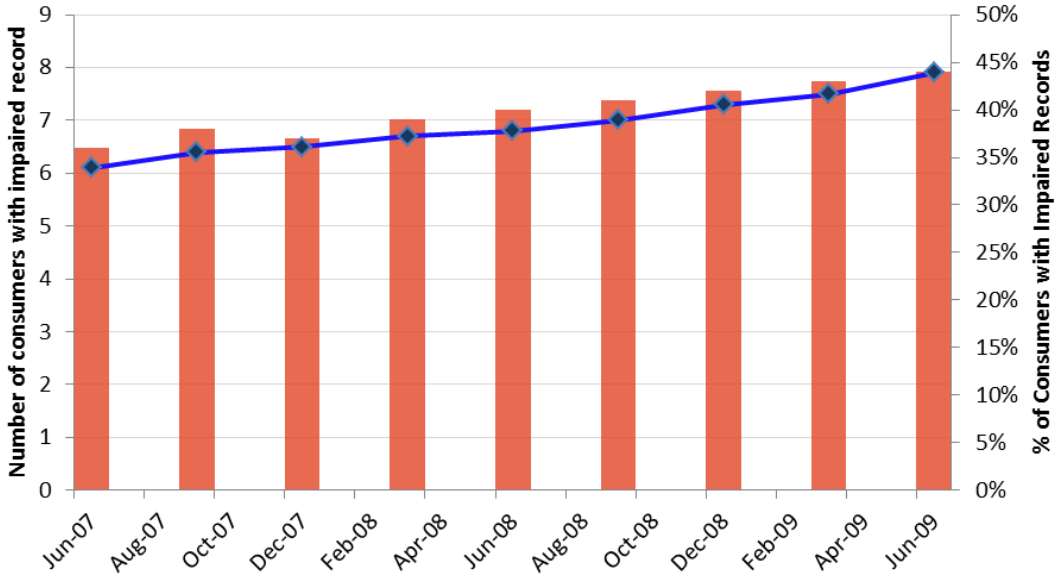


Figure 3.2: Consumers with impaired records

Source: NCR, 2009

A factor contributing to the increase of debt in South Africa is obviously the recent financial crisis (NCR, 2009). Individuals lost their jobs and means of financing their credit during the crisis. The study conducted for the NCR by Feasibility in 2009 divided individuals into the following groups:

- **Group 1:** The informal sector, domestic workers, social grant recipients & agricultural sector, earning less than R1,800 per month.
- **Group 2:** Entry level workers in the public and private sector, earning R1,800 to R6,000 per month.
- **Group 3:** The middle income sector, earning R6,000 to R17,000 per month.
- **Group 4:** The high income sector, earning more than R17,000 per month.

Feasibility found that that the group most affected by job losses, were individuals who fell into Group 2, followed by Group 3. Developed by the South African Advertising and Research Foundation (SAARF), the grouping used by Feasibility above has become the most widely used segmentation tool in South Africa.

This method of grouping is known as LSM grouping: a means of segmenting the South African market that cuts across race, gender, age or any other variable used to categorise people. Instead, it groups people according to their living standards.

The NCR researchers tested the different LSM groups under limited and severe impact scenarios to determine a ballpark figure of job losses in South Africa in 2009 (Figure 3.3).

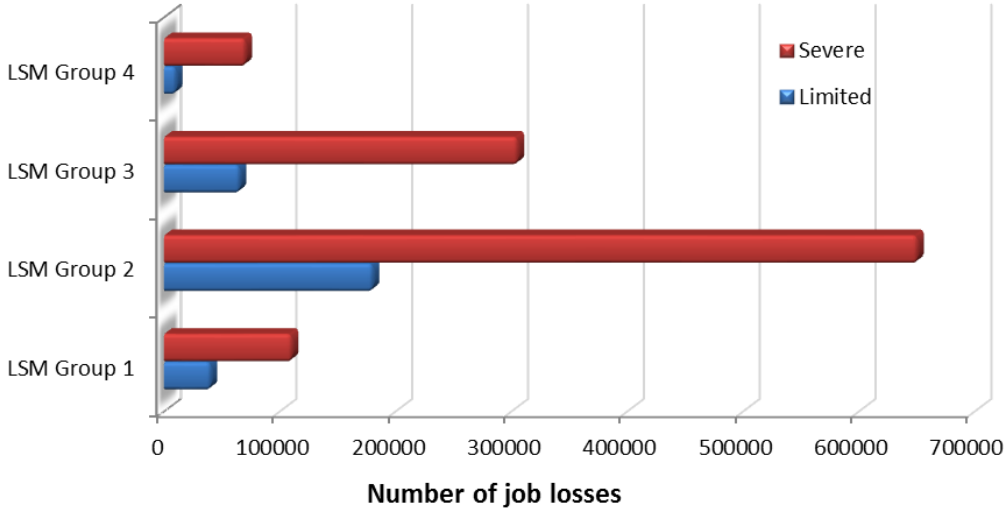


Figure 3.3: Job losses by Limit versus Severe impact (LSM)

Source: Feasibility, 2009

Figure 3.3 indicates that the highest risk group is those with moderate income where mortgage and vehicle repayments were the greatest threats. These credit facilities take longer to adjust and the impact of loss in asset value if repossessed also raises concern. There are several implications that a household faces when burdened with debt:

- Negative social consequences. There is set process to resolve an individual's financial crisis to help rehabilitate and become debt free (NCR, 2009),
- Without some type of debt relief, the individual's household income is primarily used to pay debts. Because of this, basic household needs are not

met and any social welfare assistance also ends up paying debts. School fees and municipal fees are neglected (NCR, 2009), and

- Ultimately, households turn to debt counsellors for assistance. Debt counsellors in South Africa struggle with years of backlog in the demand of their assistance. The financial crisis caused unemployment and reduced other income such as overtime, sales commissions, year-end bonuses and so forth. The opposite is also true, employees of credit providing companies were also under pressure, achieving sales targets and collection targets (NCR, 2009).

Thus the introduction of the NCA in 2007 addressed the loose credit regulations in the country at its core, the actual application for credit. The NCA was brought in to prevent consumers from falling into the trap of irreversible debt (Wizard, 2007).The new act aimed to address a few key areas in the lending of credit. Firstly, people are to be made more aware of credit, which allows the consumer to make the correct decisions before applying for credit. Now, when credit is given, the lender has to clearly state what interest rate charges is going to be (Wizard, 2007). The applicant needs to know upfront what payments the will have to make and what the total addition to the original purchase price will be. Secondly, credit lenders must now adhere to the NCR's credit regulations and make sure to do customer credit checks before approving credit (Wizard, 2007). This is applicable to mortgages, vehicle finance, retail as well as micro-lenders.

Thirdly, credit lenders may not, as they would often do in the past, increase an individual's credit limit without you giving them authority to do so (Wizard, 2007). And lastly, the credit bureaus need to ensure that the personal information they keep on individuals with a credit history are correct and remain accurate (Wizard, 2007). Credit Scoring, thus, has become an integral part of the credit application process. Once an application form has been submitted, for instance a consumer credit example (such as a Woolworths in-store account), the company would verify your credit standing through a credit bureau. This is accomplished by using credit scorecards.

3.2 Credit scoring

3.2.1 Historical development of credit scoring

The act of lending money is a concept that has been used from the very beginning and it is unlikely to change. A credit lender needs to find and attract potential debtors and then estimate their capacity and willingness to pay. History shows the use of credit over 4 000 years ago as found on a Babylonian tablet with the inscription:

> *"Two shekels of silver have been borrowed by Mas-Schamach, the son of Adradimeni, from the sun priestess Amat-Schamach, the daughter of WaradEnlil. He will pay the sun goddess interest. At the time of the harvest he will pay back the sum and the interest upon it." (Lewis, 1992).*

Banking started in the Middle Ages and up to the beginning of the twentieth century, banker's evaluation of risk from potential debtors, in terms of their ability and willingness to pay, has remained relatively the same (Marquez, 2008:3). There has always been a certain vague understanding of the professional and decisive banker. These individuals possessed judgment that could easily and accurately determine the risk in a potential debtor. This paradigm of thinking required the banker to be deeply informed and possess knowledge surrounding their client base and industry. This enabled them to know to whom they could lend money, how much he could lend, the type and size of the required security as well as the method of payment (Marquez, 2008:3). Since the mid-twentieth century to the present day, this perception of banking has undergone significant changes. In recent years emphasis has shifted to how risks are identified, mitigated and ultimately managed. This is especially true in the credit risk industry (Marquez, 2008:3).

With the steady expansion in the total numbers of the middle-class over the last century, money lenders realised there was a rapidly growing market for smaller loans with the added features of diversification and protection in the large numbers of the portfolio.

These big portfolios, producing low risk, made taking out a loan much more attractive to a potential client. This was the origin of commercial banks and pawn brokers. The idea of consumer credit became popular when cars (or automobiles) were introduced to the consumer. This product introduction accelerated the fast growing market of financing through credit. Because of this significant growth rate, the consumer quickly became just another number of debtors owing money to the bank. Since it was now impossible for a banker to access a single client's risk only based on intuition and secrets of the trade, structure and standardisation became an urgent requirement to guide the individual debtor's loan or other credit facility through a standard process and management framework. The banks and lenders urgently needed some kind of automatic classification of the quality of their debtors (Marquez, 2008:10).

The act of credit scoring became the first set approach to answer the problem of assessing credit risk in a mathematical, scientific and most importantly an automated way. This directly addressed the problem of processing large amounts of loan applications. During the 1930s some mail order companies were already overwhelmed by the large amount of credit applications (Marquez, 2008:10). These companies devised scoring systems to try and overcome the inconsistencies in results between different analysts when granting credit. The companies started by identifying certain attributes of good and bad debtors which was used to measure, grade and ultimately score (a simple sum of components) the application of a certain individual. In later years, with the invention of automatic calculators and later on computers with powerful calculation and processing power, the scoring process became easier and more automated. The scientific grounding of credit scoring was a technique called discriminant analysis (Fisher, 1936). This technique was used to differentiate groups in a population through quantifiable characteristics. This method is used when the common characteristics of a particular group are unobservable (Fisher, 1936).

The next step for credit scoring was to combine automation and discriminant analysis to eliminate the need for manual intervention from scoring analysts. In 1941, Durand recognised that the discriminant analysis approach could be used

as way to distinguish between good and bad loans (Durand, 1941). This automated approach had difficulty proving itself as an effective replacement for the knowledge built up in deciding the credit worthiness of a potential customer. Credit scoring was initially seen as an additional chore as many analysts would first approve or decline an application and then fill in the scorecard to reflect that decision. Perseverance did ultimately prevail and more credit scoring systems were developed over the years. In the initial scoring process, both selected variables and assigned scores were mainly based on judgment, but as the application of the methods were combined with uniform scoring criteria. This provided the credit origination with a process that provided consistency and predictability (Marquez, 2008:11).

The first credit scoring adaptation for business was developed in 1958 by the Fair Isaac Corporation (Estes, 2006). The company was the first to mathematically derive a probability that the borrower would make their payments on time.

3.2.2 Understanding the credit scoring process

3.2.2.1 Overview

Credit scoring has rapidly become an integral part of granting any sort of credit in the financial sector. It is increasingly being used by banks and credit providers to help make any type of commercial loan decisions (Estes, 2006). Thus it is important for individuals to understand their credit score as well as to take steps to ensure an improved score. An improved score enables individuals to secure the financing necessary to achieve their personal as well as business goals (Estes, 2006).

Modern credit scoring is designed to be a quick and objective process. The application of personal credit scoring is normally predicted on five categories:

- the borrower and his/her payment history,
- the amount of outstanding debt,
- the length of the credit history contained in the credit bureau,

- the number of credit inquiries on the report and
- the types of credit present.

The calculation result of these categories, also known as the individual's "credit score", is commonly known as a FICO[7] score in the US (Estes, 2006).

3.2.2.2 The scorecard

The increasing need for a similar scoring process, outside of the banking industry, was answered by Fair Isaac in 1995 when they developed business credit scoring (Estes, 2006). This adaptation used a set of shared data scorecards developed from 5,000 loan files from 17 banks in the US (Estes, 2006). When implementing business credit scoring outside a bank, the entity must carefully adjust the scoring scales, variables and weights to accurately reflect the business and its strategy. The outcome of these models can vary, but in general a high score represents a good risk to the bank or business and thus more favourable rates for the potential customer (Estes, 2006).

Even though the uses of credit scores in consumer and business lending seem similar, their calculation is determined in different ways. Years in a business, sales volumes, industry type and current liabilities are just a few variables that can be used to calculate a credit score within a business environment. Some banks apply a weighting (according to risk) to the same information. Although the credit score is a powerful decision making tool, banks do also take into account other factors when deciding on whether or not to grant credit (Estes, 2006).

3.2.2.3 The scoring system

The framework which underlies credit scoring is based on an ever expanding decision tree. Certain relative questions are formulated to distinguish between different types of customers.

[7] Named after the Fair Isaac Corporation.

These customers can be grouped into three primary groups:

Good Account: An account that is paid on time in each subsequent payment period.

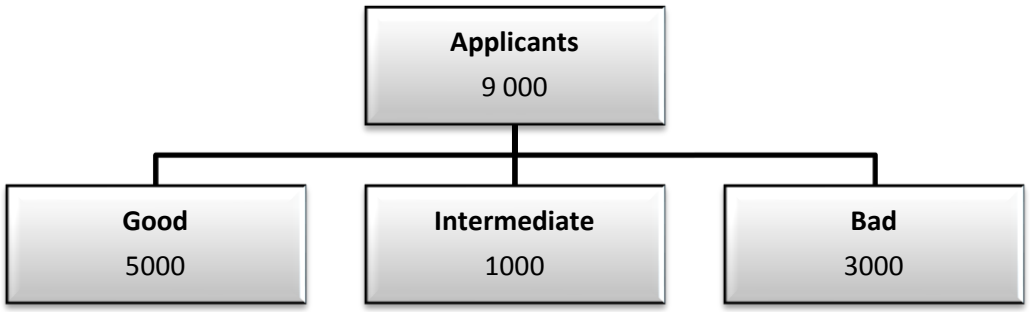
Bad Account: The organisation would normally decide on when an account is seen as bad, but the norm is 90 days delinquent.

Intermediate Account: Does not fall into either good not bad account categories as there is insufficient performance history.

The following example explains how the scoring decision tree functions:

Scenario 1:

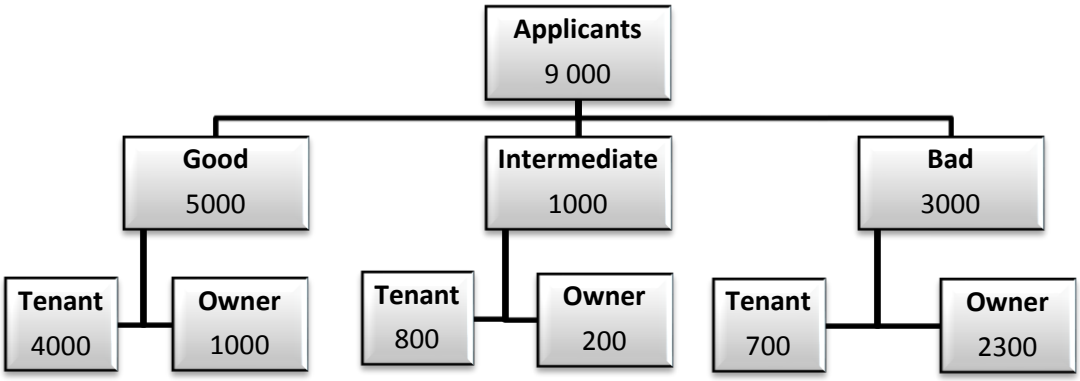
- Company's credit risk appetite is a **bad rate of 15%.**
- Assume no scoring system exists, thus everyone that applies is accepted.
- After some time, the population distribution looks as follows:



Total Bad Rate[8] = (3000)/(5000 + 1000 + 3000) = **33.3%,** thus the company will make a loss.

[8] Bad rate refers to the number of "bad" (defaulted) accounts that are included in the selection: $\frac{Number\ of\ Bad\ Accounts}{Number\ of\ Good\ \&\ Bad\ Accounts} \times 100$ (3.1)

One question is asked: "**Are you a home owner or a tenant?"**



$$\boldsymbol{Owner\ Bad\ Rate} = \frac{\mathbf{700}}{\mathbf{4000 + 800 + 700}} = \mathbf{12.7\%}$$

$$\boldsymbol{Tenant\ Bad\ Rate} = \frac{\mathbf{2300}}{\mathbf{1000 + 200 + 2300}} = \mathbf{65.7\%}$$

Thus, if the company accepts all owners and rejects all tenants, they will make a profit. To actively compete in their market the company wants to increase their market share. To achieve this they need an acceptance rate of 70%. If only owners will be accepted, the acceptance rate will only be:

$$\frac{\mathbf{4000 + 800 + 700}}{\mathbf{9}}\mathbf{000} = \mathbf{61.1\%}$$

Thus the company will have to accept some tenants; some of them are good, but not all. To identify the good accounts, the company asks another question. "Time with employer?"

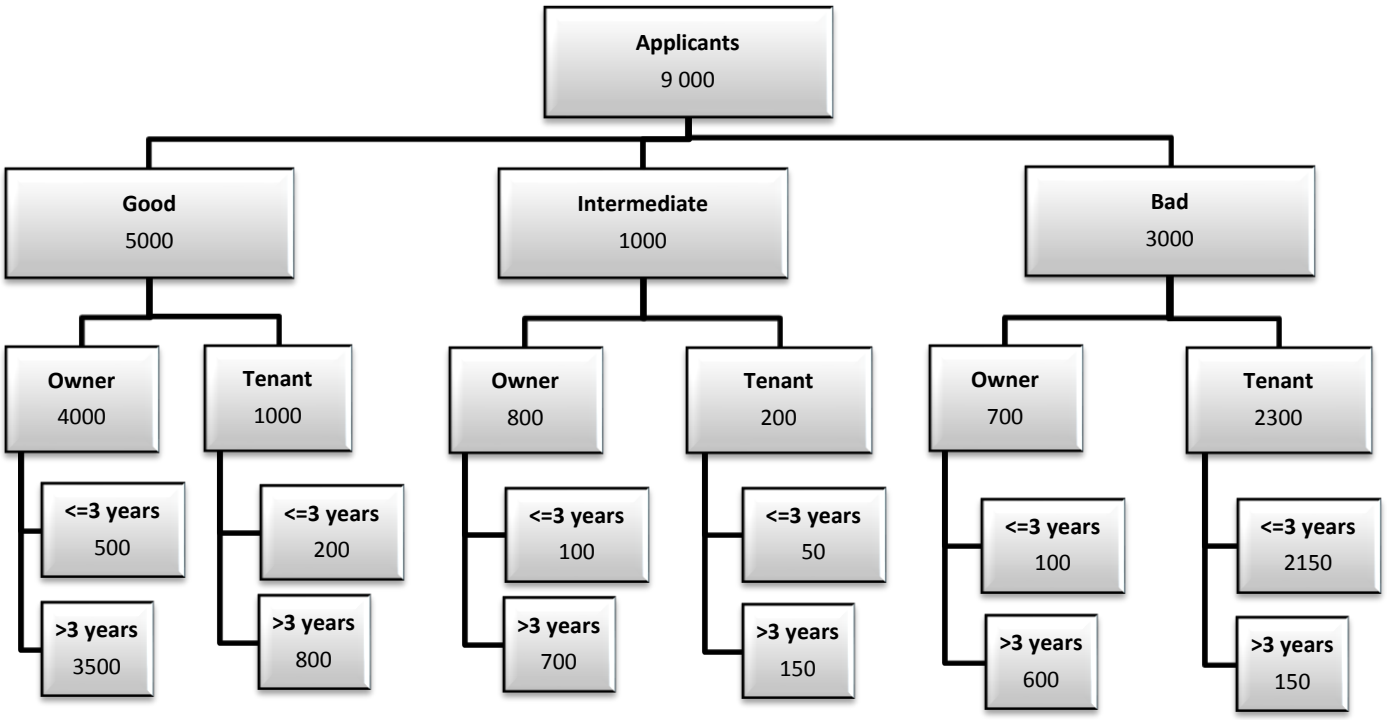


Table 3.1: Decision Tree

Decision 1: Owner vs. Tenant	Decision 2: Time with employer	Bad rate	Outcome (to company)
Owner	>3 years	12.5%	Acceptable
	<= 3years	14.3%	Acceptable
Tenant	>3 years	13.6%	Acceptable
	<= 3years	89.6%	Unacceptable

The company's total bad rate would then be: $\frac{12.5\% + 14.3\% + 13.6\%}{3} = 13.5\%$, *which is below the company's risk appetite of 15%.*

The total acceptance rate is: $\frac{4000 + 800 + 800 + 150 + 700 + 150}{9000} = 73.3\%$, *which is above the 70% required by the company.*

According to the above described process a company can build an effective credit scorecard by using their available data. There are many uses for credit scoring as well as factors that affect every scorecard.

3.2.2.4 Factors affecting a credit score

Many credit providers will look at the individual's personal credit record as an indicator of their ability as well as their willingness to repay a said loan. Credit applicants are advised to know their personal credit profile as errors can occur, especially when several inquiries for credit are made within a short period of time. Banks usually auto-decline potential borrowers who have unpaid tax hens or court judgments. Other factors such as liens, collections and legal actions can have a negative effect on the applicant's overall credit score, while individuals who promptly pay their debt obligations have an improved credit score. Inquiries for credit also negatively affect a credit score.

Another issue is how much credit card debt or lines of credit the applicant has compared to the credit that they have available. Thus a large amount of open balance revolving credit accounts can result in a lower credit score than a structured combination of mortgage, instalment and revolving credit balances.

3.2.2.5 What credit scoring can and cannot do

Credit scoring cannot predict default on a facility; rather it predicts the probability or likelihood of a "bad" outcome. Usually this will be some average of or total number of days in arrears at which the associated costs makes the facility unprofitable. The credit scoring system should also not be responsible for the approval or rejection of a loan application; instead the credit score must serve as an additional variable when the loan is reviewed. Lastly, credit scoring is not meant to increase acceptance numbers for credit facilities. It promotes efficiency and consistency while maintaining or reducing historical default rates.

Table 3.2: Use for Credit Scoring

UNABLE	ABLE
Predict individual defaults	Predict probability of 'bad' outcome as defined by company
Solely accept or reject applications	Automatically identify very good and very bad applications. Allows time to be allocated to borderline cases.
Increase acceptance rates	Increase profitability of companies (banks, small business lenders and credit providers) by reducing time spent to collect and workout credit facilities.

Banks have been successfully implementing credit scoring for a number of years in making decisions on consumer loans for credit cards, personal lines of credit as well as automotive loans. The same concept can be applied in almost every credit providing environment, such as personal retail credit cards (see Chapter 4).

3.3 Developing a credit risk scorecard

This section will give a high level outlook to the development of a basic scorecard. The section focuses primarily on the methods described by Siddiqi (2006), an American credit risk scorecard specialist, as well as simple industry related knowledge. The following aspects are important in the planning and development of an effective credit scorecard:

- the people and process and
- the scorecard development process.

3.3.1 The people and process

The development of a credit risk scorecard is collaboration between the Operational staff, IT and data mining individuals within the company. This is to ensure the creation of better scorecard and also to ensure that the solution is consistent to the strategic direction the company is taking. Credit scoring development also creates the opportunity for knowledge transfer between the collaborating departments. Developing scorecards in isolation, i.e. a single

department, will lead to such problems as: including out-dated characteristics, would be legally unadvisable, would ultimately be unable to implement in the operations of the company. Thus the level of involvement of a variety of staff is of significant importance, although they would be required at different stages in the development process.

The following roles are needed in the development of a successful scorecard.

3.3.1.1 Scorecard developer

This individual will perform the actual statistical analyses needed to develop the scorecard. He or she has to be a subject matter expert in performing data mining as well as statistical analyses, particularly those relating to predictive modelling. The will be responsible for implementing the scorecard with the use of previous risk modelling experience.

This person ensures that:

- the data are collected according to specific characteristics,
- all data abnormalities and outliers are taken into account and
- the credit scorecard development is statistically valid.

3.3.1.2 Portfolio risk manager/credit scoring manager

This individual is responsible for the management of the company's credit portfolio and will also become the owner of the credit scorecard (once implemented). He or she is a subject matter expert on the implementation of risk strategies using scores. The individual also has an in-depth knowledge of the processes and policies of the company as well as the risk profile of the company's loan book. They must use their knowledge of the application process to help develop and there after apply the credit risk scorecard.

This person ensures that they:

- help develop scorecard strategies and gauge the impacts of those strategies on their customer base,

- consider affects the scorecard will have on the business itself,
- give direction to scorecard developers when looking at direction and data and
- consider historical changes in the market, and adjust performance numbers accordingly.

3.3.1.3 Product managers

These managers are responsible for the products the company provides to their customers. These individuals have experience in implementing different product-marketing strategies. They know their target market well as well as the company's current client base. The development of future products is their responsibility.

These people ensure that they:

- offer insight surrounding the company's client base when scorecard characteristics are chosen as well as when impacts of strategies are gauged,
- coordinate the design of the new application forms to capture client information required for the scorecard and
- involve themselves in marketing to ensure the company reaches target markets that are in line with their scorecard segmentation[9].

3.3.1.4 Operational managers

These managers are responsible for departments such as Collections, Application Processing and Claims. Thus the development of scorecards will affect these departments as these managers have direct contact with the customers and in-depth knowledge of all customer related issues.

[9] Scorecard segmentation: the sub-populations that the scorecard will take into account. A scorecard will function differently for different segmentations in the market.

These people ensure:

- that they inform scorecard developers of issues pertaining to data collection and interpretation by front-end staff and
- that different departments can provide insight into factors that are predictive of negative behaviour, which helps the selection of characteristics[10] for the scorecard.

3.3.1.5 Project manager

This individual is responsible for the overall management and administration of the scorecard development project. Their responsibilities include creating a project plan, timelines, integration and development processes as well as the management of resources. The person must have a clear understanding of all relevant areas involved within the company.

3.3.1.6 IT managers (Information Technology)

These individuals are responsible for the direct management of all hardware as well as software products used throughout the company. They also, if the responsibility is assigned, manage the data warehouses of the company. They have in-depth knowledge of all the products involved in the development of the scorecards as well as the data needed.

They ensure that they:

- inform the scorecard developers on issues pertaining to data as well as coding and implementation and
- need to advise on the capability of the software the scorecards are being built on.

[10] These characteristics refer to different information collected from the client at application such as income level, address and age.

3.3.1.7 Enterprise risk managers

These managers are responsible for the management of the company's operational and financial risk at corporate level. The usually help with the oversight of the risk function, allocating capital and hedging. They have an understanding of all processes and policies relating to risk management, impacts on capital allocation and actuarial practices.

These people ensure that they:

- advise on new strategies that impact the company's risk profile,
- adjust capital allocation with the increase or decrease of risk levels and
- ensure that strategies comply with all risk policies and guidelines.

3.3.1.8 Legal staff

There are different laws and regulations regarding the collection of personal information to assess creditworthiness. Some information may be against these regulations and it is seen as good practice to submit the characteristics, required for the scorecard calculation, to the Legal department within the company. The legal staff will advise on whether any of these characteristics are contravening existing laws and regulations.

3.3.2 Scorecard development process: Stage 1 (Planning)

Scorecard projects do not start with the collection of data. To develop a successful scorecard, proper planning and analytical work is essential. The planning includes the reason behind the implementation of credit scoring, the individuals and participants required for the development and implementation of the scorecards and lastly assigning these individuals requirements and tasks.

3.3.2.1 Create a business plan

The first step in any project is to identify and prioritise the company's objectives for the specific project. In businesses where management has not already added

credit scoring to its forward looking agenda, an informative short presentation on the topic, advantages and requirements of credit scoring will at least add the value of exposing management to what is a growing practice in most credit providing institutions.

The creation of a business plan will provide a comparable measure to the company when priorities have to be placed on competing issues. It will also provide a laid out linear path for the project's key deliverables such as how post development validation is done and how to determine the "best" scorecard where more than one has been developed. The business plan will also clarify the actual role of credit scoring within the company. Different credit providers use scorecards in different ways, e.g. a credit card company will use it as a primary decision making tool whereas a bank (mortgage loan) will use it as one of several equally significant methods to gauge a client's credit worthiness (Caire & Kossmann, 2003).

The business plan should also determine whether the scorecard will be developed:

- in-house (by the company and its employees) or
- by external vendors.

The decision rests with management but is supplemented by the existence of sufficient data. Where sufficient data are available the decision depends on:

- the availability of resources,
- the expertise in developing scorecards for a specific product,
- the time frames for internal versus external development and
- the cost of acquiring scorecards to developing them internally (Siddiqi, 2006).

If sufficient data do not exist or the accessibility of the data is questionable the decision company can acquire generic scorecards from external vendors or the data may be obtainable from credit bureaus.

In some cases the company is unable to use a statistically developed scorecard. This could be due to:

- scarce data,
- the business volume of the product does not justify the cost in developing a statistical scorecard and
- the launch of a specific product does not allow for enough time to implement statistical scoring (Siddiqi, 2006).

When a company faces these circumstances they can opt to develop a "judgment-based adjudication model" also known as an "expert model". The development of such a model involves the selection of a group of characteristics judged to be predictive of risk. These characteristics are then assigned with ranges and points. These ranges and points are collected through expert judgment in conjunction with company strategy and policy. Although the expert model lacks a statistical foundation, it provides a consistent and objective decision making tool. The expert model should be constructed with participation from individuals in marketing, credit (adjudication), risk management as well as other relevant departments.

3.3.2.2 Project plan

The project plan includes a defined project scope as well as the timelines for the different key deliverables and project as a whole. The project plan will address the different approaches assigned to each deliverable as well as the implementation strategy for the scorecards. The foreseeable risks and contingencies should also be noted. The project plan will also ensure continuity between the development process and the post development process. This will allow an easy transition between the development team and the testing and implementing teams. The project plan will ensure that all stakeholders are aware of what needs to be done, who needs to do it and by when it should be completed.

A main section within the project plan should be the identification of risks. These risks could potentially affect the quality of the scorecard and should be identified at the preliminary stage. Where necessary, these risks should be accompanied by

mitigation plans. The most common risks associated with developing scorecards include:

- the data are unavailable or insufficient,
- the data are of poor quality and unreliable,
- there exist delays in the accessing of data,
- the data may not be predictive,
- the scorecard analytics cannot be handled by operational systems,
- there had been recent changes indirection and priorities and
- there exist legal and operational issues (Siddiqi, 2006).

The stakeholders are also identified within the project plan. They assemble a project team which represents all the affected departments (see Section 3.2). A management steering committee has to be formed to guide the design, development and implementation of the credit scoring system throughout the project cycle. The steering committee will approve major steps in the project each time they convene for a meeting. Each department's representative steering committee member will have the chance to express their views on the project's steps and objectives (Caire & Kossmann, 2003).

3.3.3 Scorecard development process: Stage 2 (Data review)

3.3.3.1 Data availability and quality

The key to reliable scorecards is accurate data. The data will form the statistical foundation of the scorecard, affecting variables, groups, calculations and ultimately the credit decision. Developing scorecards requires an acceptable number of "bad" as well as "good" data entries. The data required varies according to requirements, but a rule of thumb for application scorecard development is to acquire 2,000 "good" accounts and 2000 "bad" accounts. If the company decides to perform reject inference another 2000 "reject" accounts is also needed.

3.3.3.2 Definitions

Application Scorecards: The data are selected for each scorecard from a group of approved accounts that were opened in a specific time frame.

Behavioural Scorecards: The data are selected from a group of accounts that were current at a specific point in time. The data can also be at a specific delinquency status.

The organisation must undertake an analysis of their data needs as well as the data they have available. If internal data are available, they have to be reliable and easily accessible. The organisation may decide to develop scorecards based only on internal data, or they may choose to add additional data from external sources such as credit bureaus. Many organisations have a strong paper-based database and can decide to capture the data electronically to use in the scorecard. This creates an organisation-focused database that will create a more accurate scorecard. To be able to successfully define project parameters for the scorecard, data must be in an electronic database format.

3.3.3.3 Project parameters

This exercise serves a double purpose in not only defining the project parameters but also provides an insight into the organisation through data.

Certain types of accounts need to be excluded when taking a sample for the developmental process. The accounts to be used in development are those that the organisation would score during a business-as-usual credit granting process. Thus accounts that have abnormal performances must be excluded from the sample. Examples of these accounts are:

- accounts with fraud events,
- staff accounts at the organisation,
- VIP accounts,
- pre-approved accounts,

- accounts belonging to underage individuals and
- accounts that have been voluntary cancelled within the performance window.

To identify other accounts that should be excluded from the sample the organisation only has to look at whether or not the accounts could suffer from a bias issue.

Scorecards are developed under the assumption that past performance will forecast future performance. Thus the organisation will use previously opened accounts to predict the performance of newly opened accounts in the future. After the performance window has been identified, the accounts have to be monitored to identify "good" and "bad" accounts. Thus the performance window data along with their good/bad classification forms the development sample (Figure 3.4).

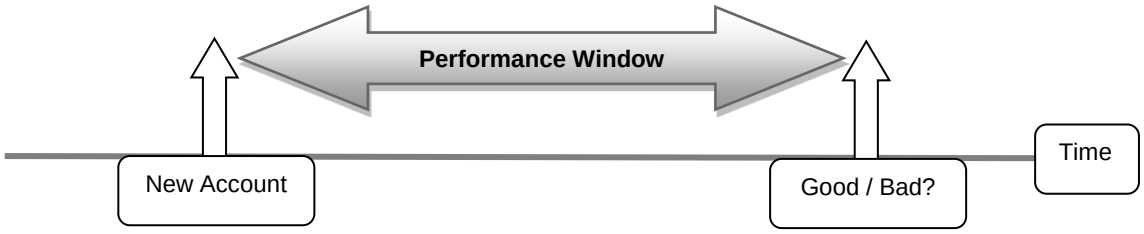


Figure 3.4: Performance definition

Source: (Siddiqi, 2006)

A simple way for any organisation to identify performance and sample windows is to study payment default performance of the selected portfolio. The analysed data are then used to plot the development of the identified "bad" cases over time.

Table 3.3: Sample cohort analysis

Bad = 90 days					Time on Books (months)				
2003	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Jan	0.00 %	0.44 %	0.87 %	1.40 %	2.40 %	2.80 %	3.20 %	3.60 %	**4.10 %**
Feb	0.00 %	0.37 %	0.88 %	1.70 %	2.30 %	2.70 %	3.30 %	**3.50 %**	
Mar	0.00 %	0.42 %	0.92 %	1.86 %	2.80 %	3.00 %	**3.60 %**		
Apr	0.00 %	0.65 %	1.20 %	1.90 %	2.85 %	**3.05 %**			
May	0.00 %	0.10 %	0.80 %	1.20 %	**2.20 %**				
Jun	0.00 %	0.14 %	0.79 %	**1.50 %**					
Jul	0.00 %	0.23 %	**0.88 %**						
Aug	0.00 %	**0.16 %**							
Sep	**0.00 %**								

Source: (Siddiqi, 2006)

The figures given in Table 3.3 refer to the percentage of accounts (in the portfolio) that had become delinquent after a certain number of months. In the 3rd line, for example, after being customers for six months there were 3.00% of the accounts that were 90 days past due (bad accounts).

Table 3.3 indicates a plot of bad rate month by month for a period of 14 months. The data are from a typical credit card portfolio. The bad rate initially develops quite rapidly in the first few months but then stabilises when the portfolio age reaches the 12 month mark. The development sample is then chosen where the bad rate has stabilised (where the curve in Figure 3.5 begins to level off).

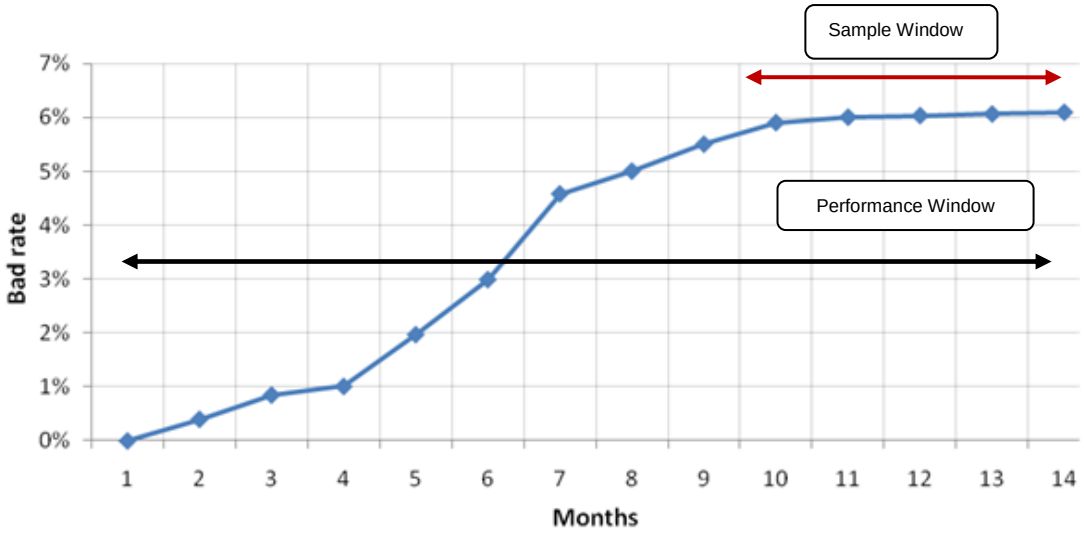


Figure 3.5: Bad rate

Source: (Siddiqi, 2006)

When identifying the development sample it is important to note the amount of data used. The best selection would be data from 6-17 months. The minimum criterion (6 months) would only just qualify as enough data to extract a development sample as less than six months would typically not show performance. By taking more than a year's worth of data a proper analysis can be done to eliminate abnormalities such as seasonality (Mays, 2001).

It is important to properly identify performance in a portfolio of accounts. The three primary performance groups are: "Good", "Bad" and "Intermediate". The definition of these groups has been stated but it is also important to identify the accounts within these groups correctly.

Bad Accounts: These accounts are usually tied to delinquency dates such as 30/60 or 90 days but would also include factors as bankruptcy and fraud events. A choice of delinquency dates will ultimately reflect on the accuracy of the scorecard. If the "bad" definition is decided to be 90 days in arrears a strong scorecard would be built but there may be a problem with sufficient data. If 30 days in arrears indicates a "bad" client then data would not be an issue but the accuracy of the scorecard will decline. Thus

the decision must be in line with the needs and objectives of the organisation. Commercial banks use a 90 day arrears status as definition of "bad" accounts according to the regulations of Basel II (Caire, Barton, de Zubiria, Alexiev, Dyer, Bundred & Brislin, 2006).

Good Accounts: Once a decision on what constitutes a "bad" account has been made and the "bad" accounts have been identified, the same analysis can be applied to identify "good" accounts. These accounts must also adhere to the organisation's objectives and are usually obvious to identify. Some characteristics of these accounts are:

- the accounts are never delinquent,
- the accounts are profitable (positive net present value),
- there have been no claims from the account,
- the account has never been bankrupt and
- no fraud events have occurred on the account.

A good account needs to retain its status as "good" throughout the performance window whereas a "bad" account only has to be defined as "bad" once throughout the performance window (Siddiqi, 2006).

Intermediate Accounts: These accounts are those that do not entirely fit the "bad" account definition or the "good" account definition. This is primarily because of insufficient performance history needed to classify the account.

3.3.4 Scorecard development process: Stage 3 (Creating a database)

After identifying the parameters of the data in stage 2, creation of the database can start. The data in the scorecard will contain a set of characteristics and a target variable.

3.3.4.1 Sample specification

As seen in the previous stages, certain specifications around the development of a scorecard should be decided and documented. These specifications are:

- the definition of "good" and "bad" accounts,
- the determination of portfolio bad rates and approval rates,
- the determination of the sample and performance windows,
- the determination of the definition of excluded accounts,
- the characteristics of the selected data.

The data in an organisation's database can be classified into two separate segments. The first is "records" or "accounts" which contains details for individual cases. The second segment is "fields" or "characteristics", containing individual pieces of information for each account. These two dimensions make up the rows (accounts) and columns (characteristics) of the spread sheet database. The characteristics in the database have assigned attributes, for example, "gender" would be the characteristic and "male" or "female" would be the attributes. Characteristics are made up of different data types, where numbers and text would be the most prominent (Anderson, 2007).

Thus the selection of the characteristics to be used in a credit scorecard should be carefully planned and each characteristic or variable carefully considered. There are different factors to help identify the best set of characteristics, they are:

- **The reliability and robustness of the characteristic.**

  There exists the possibility that some of the values of characteristics will be susceptible to manipulation. For example "income" which is usually a characteristic with low predictive power at best and totally unreliable at worst.

- **The predictive power of the characteristic.**

  Most predictive variables are usually derived from prior collective experience, whether it was in the same organisation or in a similar institution. Consultation with data collectors or front office staff will usually help to determine the expected predictive power of a variable (characteristic).

- **Interpretability of the characteristic.**

  Some characteristics tend to be subjective of nature when interpreted. For example, "occupation" can lead data collectors and developers of a credit scorecard to unconsciously disadvantage an individual based on subjectivity.

- **The ease of collecting the data.**

  The data capturing forms should strive to be easily understood and motivate the applicant to give accurate and sufficient data. Whenever possible the forms should refrain from using optional fields.

- **Future availability of the characteristic.**

  When selecting characteristics, the developers must be certain that these characteristics will be available for collection in the future. They must also determine if the characteristic will be relevant within the scorecard for the foreseeable future.

- **Legal issues surrounding characteristics.**

  Each selected characteristic should be reviewed to determine if they are in any way legally suspicious. Characteristics which could raise political questions should be excluded from the dataset.

- **Creation of ratios using characteristics.**

  The scorecard developers must ensure that the ratios they create, by dividing characteristics with other characteristics, are justified and easy to explain. Though some ratios can be exceptionally predictive, they must make business sense to justify their use.

- **Changes within the competitive environment.**

  Characteristics which may not be of much importance when creating the scorecard should still be considered based on their relevance in the future. Thus the organisation should consider future as well as past trends, with help from the credit bureaus (Siddiqi, 2006).

3.3.4.2 Sampling

The act of selecting a sample primarily focuses on the amount of good, bad and intermediate accounts needed for sampling. According to Siddiqi (2006), around 2000 of each would be regarded as sufficient. This method is called oversampling[11] and would still be subject to adjustments afterwards to ensure realistic forecasts. Choosing a large sample also reduces the effect of multicolinearity[12] and assists the results of linear regression to be statistically significant.

A second sampling method, known as "Proportional Sampling"[13], may also be used as long as there are enough cases of "goods" and "bads" for statistical validity. For example, if a loan portfolio has a 5% bad rate, then the sample should include 95 000 "good" and 5 000 "bad". Due to the sample already reflecting actual

[11] Oversampling method balances data set by increasing the number of minority class objects (examples).
[12] Multicolinearity is a statistical phenomenon where two or more predictor variables in a multiple regression model are highly correlated.
[13] Proportional Sampling is when the population is divided into subpopulations (strata) and random samples are taken of each stratum.

probabilities, there are no adjustments to be made after the scorecard is developed.

Methods to determine sample sizes:

$$n = \left(\frac{z\sigma}{d} \right)^2 \qquad (3.2)$$

where:

- z = z-statistic for the required level of confidence for example 1.96 for a 95% confidence level,
- σ = the population standard deviation and
- d = half the width of the desired interval (Kutner *et al*, 2005).

Consider a company that desires 95% confidence that the population mean of their average creditors balance is within a US$500 of their estimate. Assume σ is estimated at US$5 000, then, applying Equation 3.3:

$$n = \left(\frac{1.96 \cdot 5000}{500} \right)^2 \approx 384 \qquad (3.3)$$

In this instance the company requires a sample of approximately 384. If the company is able to specify the desired proportion of accounts in their target population as more or less than the proportion they estimate from their sample, then Equation 3.4 provides a more appropriate sample size.

$$n = \frac{z^2 - \rho \cdot (1 - \rho)}{d^2} \qquad (3.4)$$

where:

- z = z-statistic for the required level of confidence for example 1.96 for a 95% confidence level.
- ρ = the proportion to be achieved.

- d = half the width of the desired confidence level (e.g. a d% of the proportion estimated from the sample) (Kutner *et al*, 2005).

For example, if the company assumes the worst case scenario (which is 50/50) and a proportion of 4%, then:

$$n = \frac{1.96^2 - 0.5 \cdot (1 - 0.5)}{0.04} \approx 600 \qquad (3.5)$$

3.3.4.3 Examining the data

It is likely that the company will use their collated database not only for developing credit scorecards but also for different analysis required by the business. Thus the data must be checked. The following points are worth noting when examining data:

i) Non-segmented Dataset

If a company plans to develop more than one scorecard (different products, portfolios or business segments) it must create a separate dataset for each segment as well as a non-segmented dataset. This non-segmented dataset is used to identify any additional advantage or "lift"[14] gained from using segmented scorecards.

ii) Check for random and representative data

When looking at the selected accounts, they should be a random and representative group of the specific segment chosen for scorecard development (i.e. the data or accounts should be a representation of the population that will be scored in the future). A dataset skewed to represent more of a certain region or group within any one of the variables will not provide an accurate score as some accounts will seem to have a lesser risk relative to others.

[14] The incremental increase in sales and revenue of the company can be measured by looking at the difference between a segmented and non-segmented scorecard. This is called the "lift" factor.

iii) Quirks within the data

An important part of data mining is to check for data quirks, especially just before and after the sample window used. The most common data quirks are changes in the capturing methods of the data, for example the code for writing off accounts changed from "W" to "W/O". Other quirks may include the position of data within the database, for example the account balance was captured in column 30 and now it is captured in column 22 (where columns indicate a spread sheet format). Usually companies do not keep a record of these data changes and the individual conducting the data examination must rely on the memory of the database custodians. Some companies do however keep a data change log which will describe the changes and reasons behind them in full (Siddiqi, 2006).

3.3.4.4 Data adjustments

The most popular method of acquiring a sample is the "oversampling" method. The drawback of using this method is that the samples are often not entirely representative of the population. To address this problem the sample needs to be adjusted to take prior probabilities into account.

To achieve this, a method known as "factoring" is applied. This method statistically adjusts the development sample counts so that the good/bad rate of the sample more accurately reflects the good/bad rate of the population. The adjusting of a sample acquired through oversampling produces realistic forecasts and also provides insights into account performance during attribute groupings. This gives the company a valuable business advantage (Siddiqi, 2006).

The adjustment step should be done before grouping occurs. This will give a better understanding of the relationship between the variables and the target. This also provides the scorecard developer with a means of checking the groupings for required performance.

There are two common methods used to adjust samples:

i) Offset Method

The standard logarithm function used in regression of joint sampling is:

$$\boldsymbol{Logit}\,(\boldsymbol{\rho_i}) = \boldsymbol{\beta_0} + \boldsymbol{\beta_1 X_1} + \cdots + \boldsymbol{\beta_k X_k} \qquad \textit{(3.6)}$$

where:

- ρ_i reflects the posterior probability
- β_k are the regression coefficients
- X_k denotes the respective variables (Siddiqi, 2006).

ii) Oversampling Method

The oversampling method includes an offset $\left[ln\left(\frac{\rho_1\pi_0}{\rho_0\pi_1}\right)\right]$ included in the logit-function.

The new logit function or pseudo-model is:

$$\boldsymbol{Logit}(\boldsymbol{\rho_i}) = \boldsymbol{ln}\left(\frac{\boldsymbol{\rho_1\pi_0}}{\boldsymbol{\rho_0\pi_1}}\right) + \boldsymbol{\beta_0} + \boldsymbol{\beta_1 X_1} + \cdots + \boldsymbol{\beta_k X_k} \qquad \textit{(3.7)}$$

where:

- ρ_1 and ρ_0 are the proportion of target classes within the sample and
- π_1 and π_0 are the proportion of target classes within the population as a whole (Siddiqi, 2006).

3.3.5 Scorecard development process: Stage 4 (Data exploration)

After the database has been finalised the scorecard developer should have data that include the following:

- characteristics of the client portfolio and

- a target variable, which would be the event of default (as defined by the company).

There are various methods that can be used to develop an effective credit scorecard from this constructed dataset. It is up to the scorecard developer and company to decide on the method and how they want to quantify the relationship between the target variable, good/bad performance, and the data characteristics. This section will focus on model development using grouped attributes and the application of linear regression with understandable examples. The section will also have a high-level look at the alternative and more complex method of logistic regression.

3.3.5.1 Data exploration

It is seen as good practice to do an exploration of the dataset before any modelling is done. This is very relevant for large data sets, especially datasets where the capturing of the values were not controlled or done in a focused manner. Using simple statistics such as distribution of values, calculating mean/median values, identifying missing values and identifying outliers can greatly enhance the effectiveness of the scorecard (Datamology, 2010).

Missing data as well as outliers are the two most common data irregularities found when exploring data. These may be fields that were either captured incorrectly, were not available or discontinued by the company in the past. It is a prerequisite for linear and logistic regression that the data are complete (Siddiqi, 2006).

There are many different ways of dealing with missing values. The most used methods are:

i) Casewise deletion

Casewise deletion excludes all examples (cases) that have missing data in at least one of the selected variables. This method is the most commonly used approach when dealing with missing data, but it does not guarantee correct answers. The

problem with this approach is that the absolute deletion of missing values will likely result in a very small statistically insignificant dataset (Burke, 2009).

ii) Pairwise deletion

Another option is to use pairwise deletion. Instead of deleting all the accounts with any missing data, the scorecard developer uses only the accounts or data points which are complete for each calculation. Using this approach, different calculations in an analysis may be based on different sample sizes. This approach may be appropriate when working with a large sample, there are only a few missing accounts or the variables are not highly correlated. This approach however is more time intensive because of the larger samples and more calculations (Burke, 2009).

iii) Mean substitution

Mean substitution basically replaces all the missing data in a certain variable by the mean value for that same variable. The approach of mean substitution has its own disadvantages even though it appears to provide a complete dataset. The variability in the data set is continuously decreased in proportion to the amount of data points substituted with a mean value. This leads to a decrease in the spread or dispersion of the data. This approach is usually avoided when working with regression models as it interferes with the correlation between the individual variables.

There exist many data mining software packages, such as SAS Enterprise Miner, that has built in algorithms to impute missing data. These software packages and algorithms include mean and median replacement techniques as well as tree-based imputation. These methods all consider all the values of the other characteristics to determine or calculate the values of the missing data (SAS, 2003).

3.3.5.2 Outliers

Outliers are data that fall outside the normal range of a specific variable. For example, a distribution of age for a certain population may fall in the range of 18 – 55 years of age. The data indicate a few values of 101, 129, 99 and so forth. While these values, in some cases, may reflect accurate numbers, they are most likely to be the result of inaccurate data capturing. These values will have a negative effect on the regression models when they are taken into account (High, 2000).

Developing different techniques to identify outliers as well as understanding how they impact data analysis is important parts of a thorough data exploration, especially when statistical techniques are applied to deal with the outliers. For example, if outliers exist within a dataset, any statistical test based on sample means and variances can provide distorted and inaccurate results (High, 2000).

Other problems created by outliers include:

- creating a bias or distorting estimates,
- creating a distortion in the statistical significance of a certain variable (statistical significance, or the lack thereof, can be caused by a few or even a single outlier) and
- drawing incorrect conclusions (Allows for incorrect conclusions to be made if outliers are not accounted for) (High, 2000).

There are different ways to detect outliers within a dataset. The most common are:

i) Visual

If available, check the distribution of each variable on its own. Checking the distribution should be the first and easiest step to identify outliers within a category or variable. Using visual aids such a dot-plot or scatter-plot are very effective methods to examine how severe any outlying values actually are. Another option is to use a box-plot, since it requires no estimate for mean or standard deviation, nor uses assumptions for the distribution.

Univariate tests[15] are also used to identify outliers; however, many of these tests only check for a single outlier. They also make assumptions regarding the distribution of the variable, which in some cases are not accurate (for example, they assume a normal distribution when the variable actually has a very skew distribution). Additionally univariate test often require a mean as well as standard deviation estimate for calculation. Outliers can greatly affect the values of a mean or standard deviation (High, 2000).

In Figure 3.6 follows an example of identifying outliers using a scatter-plot. The red encircled dots are identified as outliers.

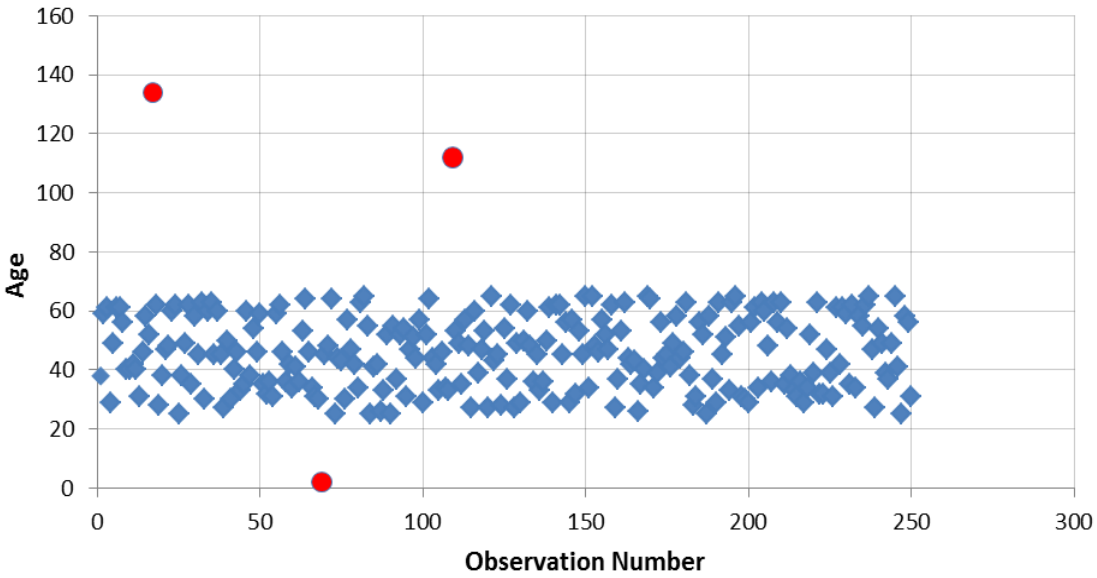


Figure 3.6: Scatter-plot of age

Source: Dummy data generated in Microsoft Excel

ii) IQR computation

Another approach is to calculate the inter-quartile-range[16] (IQR) and then use a multiple of it as a number that defines what values can be considered as outliers.

---

[15] A univariate test is an analysis carried out with the description of a single variable and its attributes of the applicable unit of analysis.

[16] The interquartile range (IQR), also called the midspread or middle fifty, is a measure of statistical dispersion, being equal to the difference between the upper and lower quartiles.

The box-plot approach uses this technique to effectively identify outliers, especially when working with large datasets that consists out of continuous data.

One way to implement an IQR computation is to use PROC UNIVARIATE with SAS (Statistical Analysis Software) and save the order statistics available with its OUTPUT, or HTML statement.

iii) SAS (Statistical Analysis Software)

SAS is a global business analytics company which focuses on providing their customers with statistically backed software to help with critical business decisions. SAS Foundation (or Base SAS) will be the only product used throughout the document.

3.3.5.3 Example

An analysis was done on a dataset containing 350 age data points. Using the PROC UNIVARIATE function in SAS (Figure 3.7), a summary of the data was determined including the moments and quantiles.

```
ods html file="C:\Users\jvdwalt\Desktop\Masters\SAS\Age.html";
Proc univariate data = Age;
var age;
run;
```

Figure 3.7: SAS Proc Univariate Code

Table 3.4: Moments of age variable

The SAS System			
The UNIVARIATE Procedure, Variable: AGE (AGE)			
Moments			
N	350	Sum Weights	350
Mean	42.04	Sum Observations	14714
Std Deviation	14.9	Variance	222.8
Skewness	0.8	Kurtosis	4.0
Uncorrected SS	696300	Corrected SS	77723.4
Coeff Variation	35.5	Std Error Mean	0.8

Source: Dummy data generated in SAS

Table 3.5: General statistics of age variable

General statistical measures			
Location		**Variability**	
Mean	42.0	Std deviation	14.9
Median	42.0	Variance	222.7
Mode	21.0	Range	132.0
		IQR	25.0

Source: Dummy data generated in SAS

Table 3.6: Quantiles of age variable

Quantiles (Definition 5)											
Quantile	100% Max	99 %	95 %	90 %	75% Q3	50% median	25% Q1	10 %	5 %	1 %	0% Min
Estimate	134	65	63	60	54	42	29	23	21	20	2

Source: Dummy data generated in SAS

Table 3.7: Extreme observations of age variable

Extreme Observations			
Lowest		**Highest**	
Value	**Obs**	**Value**	**Obs**
2	139	65	185
3	212	65	215
20	348	65	310
20	240	112	80
20	192	134	15

Source: Dummy data generated in SAS

To identify outliers within the AGE variable, three values are important. They are the first quartile (Q1), third quartile (Q3) and the inter-quartile range (IQR). The equations used are:

$$\textit{Potential Outliers: between} \quad \boldsymbol{q1 - (1.5 * iqr)\ and\ q3 + (1.5 * iqr)} \qquad \textit{(3.8)}$$

$$\textit{Definite Outliers: between} \quad \boldsymbol{q1 - (3 * iqr)\ and\ q3 + (3 * iqr)} \qquad \textit{(3.9)}$$

When Equations 3.8 and 3.9 are applied to the age data-set (values found in Table 3.6) the calculated outliers are given by Table 3.7. The values found on the boundaries of the age variable are also included in the extreme observation

output, they are the potential outliers and a decision can be made by the user (High, 2000).

3.3.6 Scorecard development process: Stage 5 (Scorecard development)

This stage of the scorecard process is the development of the scorecard, the preparation in terms of people, datasets and data exploration have been done in the previous three stages.

To illustrate the construction phase of a credit scorecard, Stage 4 will make use of dummy data to emulate a debtor's book of a small company, XYZ Bank[17]. The development of a scorecard for XYZ Bank will follow these steps:

i) identify the data to be used in building the scorecard,
ii) class the variables into intervals (Binning),
iii) determine the Good and Bad accounts,
iv) apply logistic regression to calculate variable coefficients,
v) normalise the output data to represent a score (out of 1000) and
vi) provide the results of the developed scorecard.

3.3.6.1 Data identification

The data used in this example are dummy data used to illustrate how to approach the development of an actual scorecard. The debtor's book acquired from XYZ Bank contains 2600 different accounts and each account has 10 variables which hold information of each client (For a personal loan product). The variables are a mixture of demographic and financial information. These variables are shown in Table 3.8.

[17] XYZ Bank is fictional, used for the example only. The data, however, are from a personal loan book of a small bank in the US (acquired from the Plug-n-Score website).

Table 3.8: Variable list (XYZ Bank – personal loan database)

Variable	Description
Age	The relevant age of each account holder at time of application.
Gender	The specific gender of the account holder.
PTI	"Percentage of taxed income". The percentage of tax the account holder pays on his/her income at time of application.
Marital status	The relationship status of the account holder at time of application.
Last worked record	The number of years the account holder has been employed at his/her place of work.
Children	The number of children the account holder had at time of application
Cards	The primary type of card used by the account holder as method of payment.
Credit history	The credit history status (received from the credit bureau) of the applicant at time of application.
Home ownership	The type of housing the account holder had at time of application.
Time at branch	The amount of time the account holder has been a customer at a specific branch at time of application.

3.3.6.2 Binning

A very important step in the development is the "Binning Phase". Binning is the process of transforming a numeric characteristic into a categorical one, setting the variable in a specific interval. It also includes the re-grouping of the characteristics within a specific variable. The binning process uses a frequency distribution calculation of a specific variable from the raw dataset (Newbold, Carlson & Thorne, 2009). Constructing these frequency tables for each variable will provide the scorecard developer with a dataset that contains variables that will be statistically more significant predictors of the target variable (a "Good" account). Each variable in the XYZ database is now analysed and grouped into categorical intervals (if required). The variables in the XYZ raw database are now transformed to represent their respective categories. Some variables required grouping data, others did not (e.g. Marital Status only compromises three possible entries – Married, Single or No Information). The categories of each variable are shown in Table 3.9 to Table 3.18.

Table 3.9: Age variable

AGE	
Category	**Frequency**
less than 25	144
from 25 to 30	265
from 30 to 40	604
from 40 to 50	737
from 50 to 55	356
more than 55	494

Table 3.10: Gender variable*

Gender	
Category	**Frequency**
male	1104
female	1496

Table 3.11: PTI variable

PTI	
Category	**Frequency**
less than 20%	823
from 20 to 25%	565
from 25 to 35%	589
from 35 to 40%	221
40% and greater	402
less than 20%	823

Table 3.12: Marital status variable*

Marital Status	
Category	**Frequency**
Married	1461
No Information	214
Single	925

*Required no grouping.

Table 3.13: Last work record variable

Last Work Record (in years)	
Category	**Frequency**
No Information	682
Less than 1	92
From 1 to 3	68
From 3 to 5	142
From 5 to 10	442
10 and greater	1174

Table 3.14 Children variable

Children	
Category	**Frequency**
No Information	458
None	1346
One	495
Two and more	301

Table 3.15: Cards variable*

Cards	
Category	**Frequency**
None	1512
Credit Card	779
Paycard	309

Table 3.16: Credit history variable*

Last Work Record (in years)	
Category	**Frequency**
None	2290
Average	48
Positive	249
with negative factors	13

Table 3.17: Home ownership variable*

Home Ownership	
Category	**Frequency**
No Information	87
Own	1493
Rent	1020

*Required no grouping.

Table 3.18: Time at branch variable

Time at Branch	
Category	**Frequency**
No Information	1557
Less than 1	225
From 1 to 3	88
From 3 to 5	156
From 5 to 10	218
10 and greater	356

After the binning has been completed for each variable, the raw data are converted to show the category and not the captured data-point (e.g. "more than 55" rather than "57" for the age variable). When the transformation of the raw data is complete, each entry (of the 2600 in the XYZ data) will take on this format:

Table 3.19: Transformed data

Age	Gender	PTI	Marital status	Last work record	Children	Cards	Credit history	Home ownership	Time at branch
30 – 40	M	25% – 35%	M	>= 10	>= 2	Pay card	None	Rent	-

3.3.6.3 Identifying good and bad accounts

To determine whether an account is "Good" or "Bad" the scorecard developer usually consults with the company. Most banks use the 90 day (or 3 month) in arrear status to identify accounts that have defaulted. Each company decides and captures their definition of default in their policies and the scorecard developer should consult closely with the company to determine the actual threshold (Siddiqi, 2006). For the purpose of this example the good and bad accounts have been assigned to the data. The good/bad distribution of the XYZ dataset is illustrated in Table 3.20 below.

Table 3.20: Good/bad distribution

XYZ Dataset	
Good	**Bad**
1800	800
2600 accounts	

3.3.6.4 Regression analysis

The next step in the development of a scorecard is to understand and provide a quantifiable measure of the significance that each variable in the dataset has in determining whether or not the account will turn out good or bad. Each category within each variable (e.g. "more than 55" for the AGE variable), needs to be statistically judged for contributing significance and assigned a score (number of points).

i) Understanding linear regression

Regression is a statistical methodology that the uses the relation between two or more variables to predict the outcome of a set target variable. This statistical concept is widely used throughout business, biological sciences, social and behavioural sciences as well as many other fields. Some examples of regression application are used to predict:

- sales of a product by using the relation between sales and advertising expenditure,
- the performance of an employee by using the relation between performance and aptitude test scores,
- the length of a hospital stay by using the relation between the length of stay and the severity of the operation and
- the vocabulary of a child by using the relation between the vocabulary, age and education level of the parents of the child (Kutner et al, 2005).

The regression model, using multiple predictor variables, is described by Equation 3.10:

$$\boldsymbol{Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_i X_i + \varepsilon_i} \qquad \textit{(3.10)}$$

where:

- Y denotes the response variable (The variable the other will predict)
- $X_1 \ldots X_i$ denotes the predictor variables (all other variables used in the dataset)

- β_0 denotes the y-intercept
- $\beta_1 \ldots \beta_i$ denotes the parameters of the model (coefficients to predict the response) (Kutner et al, 2005).

ii) Logistic regression

The primary distinction between linear regression and logistic regression is that the latter is more suited for the modelling of dichotomous categorical outcomes (e.g., default or no default). Both types of regression are based on many of the same assumptions and theory. While this makes modelling easier this still presents a minor problem with regard to predicting the actual outcome variable. Since the outcome has two possible values, the predictor variables (predicting change) have no relative meaning. Thus instead of modelling the value of the response or outcome variable, logistic regression instead calculates the odds of the outcome of the response variable. The natural logarithm of the odds (of the response variable) is linear across most of its range. This enables modelling using many of the same principals used in linear regression.

The logistic regression model, using multiple predictor variables, is described by Equation 3.11:

$$\boldsymbol{Ln\left[\frac{p}{1-p}\right] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_i X_i + \varepsilon_i} \qquad \boldsymbol{(3.11)}$$

where:

- p represents the probability of an event (default)
- $X_1 \ldots X_i$ denotes the predictor variables (all other variables used in the dataset)
- β_0 denotes the y-intercept
- $\beta_1 \ldots \beta_i$ denotes the parameters of the model (coefficients to predict the response) (Kutner et al, 2005).

3.3.6.5 Applying logistic regression in SAS

To apply logistic regression in SAS the scorecard developer will use the PROC LOGISTIC method. The XYZ database will use the following set (Figure 3.8) of programming code:

```
proc logistic data = master.Scorecard_database OutModel= ModelParam namelen=200 descend;
class
Age Gender PTI Marital_Status LastWorkRecord Children Cards Credit_History Home_Ownership
Time_at_Branch / param=glm;
model
GoodBad= Age| Gender| PTI| Marital_Status |LastWorkRecord| Children| Cards| Credit_History|
Home_Ownership|Time_at_Branch @1;
output out=toz_LOGISTIC_2 p=phat_new xbeta=xb;
ods output ParameterEstimates = coeff_est;
run;
```

Figure 3.8: Logistic regression code

The above coding performs three main functions. Firstly the model identifies all the categorical variables to be used in the dataset (under the "class" section in the above coding). Thus the model will use the categories in each variable, acquired through the binning process previously described, as predictor variables of the outcome variable.

Secondly in the "model" section of the code the logistic regression model is specified. Thus the outcome variable is specified as the "GoodBad" variable and all the others as predictor variables.

Lastly the logistic regression results are outputted into a separate dataset for viewing purposes. The output contains all the regression coefficients of the categories of each variable. The output of the coding is (only age variable, refer to Annexure A1 for full output):

Table 3.21: XYZ Bank logistic regression coefficients (age variable)

Variable	ClassVal0	DF	Estimate	StdErr	WaldChiSq	ProbChiSq
Intercept		1	1.58	0.87	3.09	0.08
Age	from 25 to 30	1	-0.37	0.19	3.70	0.05
Age	from 30 to 40	1	-0.30	0.16	3.52	0.06
Age	from 40 to 50	1	-0.32	0.15	4.46	0.04
Age	from 50 to 55	1	-0.34	0.18	3.50	0.06
Age	less than 25	1	-0.40	0.24	2.87	0.09
Age	more than 55	0	0.00	0.00	0.00	0.00

3.3.6.6 Normalising data

The next step in the development process is to normalise the estimates (regression coefficients) of the model output. The reason for doing this is to establish a quantifiable score for each category in each individual variable. The process of data normalisation (Figure 3.9) is to apply an individually calculated weight to each category within each variable and also weight the variables over all. For the XYZ Bank example the following code was applied to normalise the regression coefficients to reflect a score out of 1000:

```
proc sql ;
create table score_card as
select b.*,          sum(max_est1/counter) as sum_max,
                     case when est1=max_est1
                              then 1
                              else 0 end as max_cat,
                              round(1000*((est1)/calculated sum_max)) as score
                              from (select a.*, max(est1) as max_est1,
                                                        count(*) as counter
                              from (select *,     min(Estimate) as min_est,
                                        count(*) as counter,
                                                  case when calculated min_est=Estimate
                                                  then 0
                                                  else Estimate- calculated min_est end as est1
                              from coeff_est where variable ne 'Intercept'

                                        group by variable ) a
                                        group by variable ) b;
quit;
```

Figure 3.9: Data Normalisation Code

The SAS code above creates a table displaying the regression coefficients calculated in the previous step as well as the variables needed to normalise each variable category into a score out of one thousand. An estimate in each variable is normalised against the others in the same variable and then for the data as a whole. To get to each individual score the following calculations are done per variable:

- Min_est = category in the variable with smallest estimate (regression coefficient)
- Counter = counts the number of categories within the variable
- Est1 = difference between estimate and calculated minimum estimate
- Max_est1 = maximum of the calculated variable Est1.

The above calculations are repeated for each variable. The data calculated are then used to determine the next two variables:

- Sum_max = summation of Max_est1 of each variable
- Score = score for each category within each variable is calculated using Eq 3.12:

$$\mathbf{1000} \times \left(\frac{\boldsymbol{est1}}{\boldsymbol{sum_max}}\right) \qquad \textit{(3.12)}$$

The SAS data normalisation code creates a table that displays the calculated scores for each category in each variable as shown in the partial example below (refer to Annexure A2 for full output).

Table 3.22: Score table (age, cards and children variables)

Variable	ClassVal0	Estimate	min est	counter	est1	max est1	sum max	max cat	score
Age	less than 25	-0.40	-0.40	6	0	0.40	13.8	0	0
	from 50 to 55	-0.34	-0.40	6	0.06	0.40	13.8	0	4
	from 30 to 40	-0.30	-0.40	6	0.10	0.40	13.8	0	7
	from 25 to 30	-0.37	-0.40	6	0.03	0.40	13.8	0	2
	from 40 to 50	-0.32	-0.40	6	0.08	0.40	13.8	0	6
	more than 55	0.00	-0.40	6	0.40	0.40	13.8	1	29
Cards	credit card	-0.66	-2.07	3	1.41	2.07	13.8	0	102
	paycard	0.00	-2.07	3	2.07	2.07	13.8	1	150
	None	-2.07	-2.07	3	0.00	2.07	13.8	0	0
Children	no information	0.00	-0.42	4	0.42	1.35	13.8	0	30
	Two and more	0.93	-0.42	4	1.35	1.35	13.8	1	98
	One	0.41	-0.42	4	0.82	1.35	13.8	0	60

3.3.6.7 Scorecard results

The development of the scorecard has now been finalised and can be implemented in to the operations structure of XYZ Bank. For Mr. X to apply for a loan at XYZ Bank, he will have to provide the following matrix of information:

Table 3.23: Scorecard example (Mr X)

Variable	Information	Score
Age	from 30 to 40	7
Gender	Male	2
Marital Status	Married	37
Children	One	60
PTI	From 20 to 25%	83
Credit History	Average	153
Cards	Credit card	102
Home_Ownership	Rent	8
Last Work Record	From 5 to 10	21
Time_At_Branch	10 and greater	65
		538

By default, a scorecard cut off point will be exactly half of what the scoring range indicates, in this example 500. Mr. X achieved a rating of 538 and will thus be granted a loan by XYZ Bank. The most important part of post scorecard development is to analyse the effectiveness of the scorecard as well as to calibrate the acceptance criteria according to the risk appetite of the company.

Scorecard validity and effectiveness is usually checked through two statistical calculations:

i) Calculating the Gini coefficient[18] of the scorecard

Since the scorecard differentiates borrowers against the "Good" and "Bad" results of the credit deal, the quality of its performance can be evaluated from the point of view of classification (XYZ Bank example = 60%) (Scallan, 2007:1).

To calculate the Gini coefficient (in this example) the following are required:

- the mean of a distribution,
- the number of people (or percentiles) and
- the credit score of each person (or percentile).

The Gini coefficient is most prominently used to determine the inequality of but it can also be applied to any distribution to determine inequality. The standard Gini equation is as follows:

$$\boldsymbol{G} = \left(\frac{\boldsymbol{N+1}}{\boldsymbol{N-1}}\right) - \left(\frac{\boldsymbol{2}}{\boldsymbol{N(N-1)u}}\right)\left(\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{P_i X_i}\right) \qquad \textit{(3.13)}$$

where:

- u is the mean score of the population.
- P_i is the score rank P of person i, with score X (such that the person with the highest score receives a rank of 1 and the lowest score a rank of N. This effectively gives a higher weight to lower scores in the score distribution, which allows the Gini to meet the Transfer Principle).

[18] The Gini coefficient is a measure of the inequality of a distribution: 0 expresses total equality and 1, maximal inequality. Thus the scorecard's Gini value evaluates the distribution between "good" accounts and "bad" accounts.

ii) Applying the Kolmogorov-Smirnov[19] test

Evaluates the strength of distributions between the "goods" and "bads" within the scorecard %) (Scallan, 2007:9) and thus evaluates the quality of classification (XYZ Bank example = 50.01%)

The Kolmogorov-Smirnov test is defined by:

H_0 the data follow a specified distribution

H_a data *do not follow* the specified distribution

$$\boldsymbol{D = \max_{1 \le i \le N} \left(F(Y_i) - \frac{i-1}{N}, \frac{i}{N} - F(Y_i) \right)}$$

where:

- F denotes the theoretical cumulative distribution of the distribution being tested
- Y_i denotes the ith data point
- N denotes the number of data points.

The two tests above will be performed in the next chapter to test developed scorecards.

[19] The Kolmogorov-Smirnov test is a full non-parametric test used for comparing two distributions (in this case the "good" and "bad" accounts).

This concludes the walkthrough of basic scorecard development. The next chapter will focus on a data analysis of industry data acquired from Woolworths South Africa. The analysis will include the following:

- Construction of two basic scorecards on in-store card data (timeframe 2006, 2008) to investigate the effect the financial crisis has had on consumer credit in the retail sector.
- Credit appetite analysis to investigate the demand for consumer credit over the period of 2006 – 2009.
- An informed conclusion regarding the impacts and changes seen in consumer credit from before the financial crisis to during the financial crisis.

Chapter 4

Woolworths Credit Scorecards

4.1 Empirical study introduction

The previous two chapters have provided an in depth view of the global financial crisis, retail credit in South Africa as well as a step-by-step guide to basic credit scoring. The aim of this chapter's empirical study is to determine the effect the global financial crisis had on the retail credit market in South Africa. More specifically the study will be based on one of South Africa's largest retail groups, Woolworths.

The retail chain opened its first store in October 1931 in Cape Town. Max Sonnenburg entered the retail market in South Africa on the back of the great depression in the US. With an innovative store policy and customer approach the retailer soon opened a second store in Durban three years later. Building on the success and sound business strategy two more branches opened in Port Elizabeth and Johannesburg respectively (Woolworths, 2011).

Since their inception in the early 1930s, Woolworths' store count has grown to over 400 branches in South Africa, Africa and the Middle East. The retailer owes its popularity and growth mainly to its policy to always think of the customer first. Woolworths were the first retailer in South Africa to introduce sell-by dates on their food products, machine washable wool clothing as well as pre-washed lettuce and other fresh produce (Woolworths, 2011).

In order for the retail business to grow many financial decisions were made over the decades. One of the biggest steps for the retailer was to introduce financial services segment to their business. Woolworths introduced various financial products to the market over their long history.

These services today include:

- the in-store card: This product provides a credit line to a customer to buy any Woolworths and associated stores' products,
- the credit card: As any credit card the product extends the ability of the customer's credit line to any other retailer,
- personal loans: Woolworths also provides their customer base with short term personal loans and
- insurance: Woolworths provides a wide range of insurance products including home, car, travel and cell phone insurance.

In April 2008 Woolworths sold 50% plus one share to ABSA Group Ltd. The deal was completed at a cost of R875 million (Stewart, 2008). ABSA immediately set out to align the newly acquired segment to Basel II banking regulation. The ultimate goal was to determine the capital allocation required for the new business segment under the capital requirement rules of the Basel II accord. At the end of 2010 ABSA approached the IQ Business Group to help implement their PD[20], LGD[21] and EAD[22] methodology to calculate the risk weighted element of each financial product in the Woolworths financial services segment. These risk weighted elements (RWA[23]) are then used to determine the capital requirement for the Woolworths book.

The first step in the financial process in any bank or large retail company is to identify their target client base. This is especially true when providing clients with a credit product such as the in-store card. To reduce a company's credit risk (or possible default by a customer) the credit scoring approach is used, as discussed in the previous chapter.

[20] Probability of default (The chance that a customer will not pay the expected instalment).
[21] Loss given default (The loss made by Woolworths should a customer default on their account).
[22] Exposure at default (The amount owed by the customer at time of default).
[23] Risk weighted asset (The risk assigned to a customer's account as a function of PD, LGD and EAD).

Credit scorecards are developed and implemented for each product a company provides. In the case of Woolworths their most prolific product is the in-store card.

This empirical study will use the in-store card data obtained from Woolworths to achieve the following:

Two credit scorecards:

Two credit scorecards were constructed: one based on data from 2006 (prior to the global financial crisis) and the other on data from 2008 (at the height of the crisis). The two scorecards were constructed using exactly the same variables and the significance of these variables (their contribution to default) was compared to determine the possible effects of the global financial crisis.

Data analysis:

The primary theory behind this empirical part of the study is to show that the global financial crisis affected the retail credit market of South Africa in a tangible way. To prove this, the study will explore the total credit limit versus the uptake of credit by the in-store card customers. The in-store card data (over a 5 year period) will also be scrutinised to show any trends over the period before and during the global financial crisis.

4.2 Credit Scorecard Development (Pre-Crisis)

To determine the effects of the global financial crisis, two basic credit scorecards will be constructed, tested and evaluated. The data used for the two scorecards will be monthly snapshots from Woolworths' in-store card (ISC) book. The following months have been identified to act as sample windows:

- data from June 2006 (before the global financial crisis) and
- data from April 2009 (during the financial crisis).

The stages followed to build a credit scorecard (as outlined in Chapter 3) include steps for business case planning, staff involvement, production and

implementation. These stages are required if and when a credit scorecard is developed for a specific company and specific product. The goal of this empirical study is to use credit scoring as a tool to determine if the global financial crisis had a tangible effect on the retail industry in South Africa. Thus only the crucial steps used to create a basic credit scorecard will be applied. These steps are:

- data exploration,
- binning[24],
- identifying "good" and "bad" accounts,
- regression analysis and data normalisation and
- scorecard results.

4.2.1 Data exploration

4.2.1.1 Initial Investigation

The data provided by Woolworths (ISC) range from January 2006 until January 2011. For the purpose of creating two distinct credit scorecards, two months were chosen as samples for the credit scorecards. The variables available for inclusion in the calculation are depicted in Table 4.1 below.

Table 4.1: Variable dictionary of Woolworths ISC data

NO	FIELD (VARIABLE)	DEFINITION
1	**CUSTOMER_NO**	Unique number assigned to each customer. A customer may have one CUSTOMER_NO but many account numbers
2	**WFS_PRODUCT**	Product identifier. Identifies the type of product (Only in-store card numbers and their variants are included)
3	**STATEMENT_MONTH**	Month of statement file
4	**STATEMENT_YEAR**	Year of statement file
5	**WFS_ACCOUNT_NO**	Number of each account
6	**BUREAU_SCORE**	The credit bureau's score for the particular individual

[24] Binning: the transformation of numbers into categories (e.g. "23" to "20 – 25 years of age").

NO	FIELD (VARIABLE)	DEFINITION
7	**ACCOUNT_STATUS**	Current Status of account: A – Active, Z – Charged off, D – Dormant, I – Inactive, 9 – Waiting to be purged, 8 – waiting to be purged C – Closed, P – Purged
8	**SM_OPEN_BALANCE**	The account balance (all-inclusive) at the start of the billing cycle
9	**SM_END_BALANCE**	The account balance at the end of the billing cycle, all inclusive
10	**SM_CREDIT_LIMIT**	The credit limit set for the account.
11	**SM_RECOV_FEE**	Not applicable to ISC (values are set to zero).
12	**SM_COLL_FEE**	These are collection fees. Commission fees to agencies are an income statement item, and do not get charged to the customer's account
13	**SM_PURCHASES**	Purchases made by the customer during the billing cycle
14	**SM_PAYMENTS**	Payments made by the customer during the billing cycle
15	**SM_ARREAR**	The amount that is in arrears
16	**CURR_DUE**	This is the minimum amount the person has to pay before the next billing cycle
17	**SM_INTEREST_RAISE**	The interest increase for the month on the account. For Z, it is generally 0, except for where a credit limit still exists
18	**DELINQUENCY_CYCLE**	Number of cycles delinquent. Ranging from 0 to 8, S - special arrangement, W -Write off
19	**TOT_PAST_DUE**	
21	**SM_CBP_FEE**	Optional "Card Balance Protection" fee
22	**SM_LCP_FEE**	"Lost Card Protection" fee. Not applicable to PL
23	**BLOCK_CODE_1**	Codes reasons for the status of a given account
24	**OPEN_DATE**	The date the account was opened
25	**CLOSED_DATE**	The date the account was closed
26	**WRITE_OFF_DATE**	Date the account was written off
27	**WRITE_OFF_AMOUNT**	Amount that was written off given write off

After investigating the variables available for calculation, it is clear that the credit scorecard will be based on only a few variables. Many variables are not suitable for regression analysis purely because of their nature (e.g. customer number which does not contribute to the determination of whether or not an account will default). The following variables were included:

Table 4.2: Variable to be used in Scorecard

FIELD (VARIABLE)	DEFINITION
SM_OPEN_BALANCE	The account balance (all-inclusive) at the start of the billing cycle
SM_CREDIT_LIMIT	The credit limit set for the account
SM_PURCHASES	Purchases made by the customer during the billing cycle
SM_PAYMENTS	Payments made by the customer during the billing cycle
CURR_DUE	This is the minimum amount the person has to pay before the next billing cycle
SM_ARREAR	The amount in arrears
DELINQUENCY_CYCLE	Number of cycles delinquent. Ranging from 0 to 8, S - special arrangement, W - write off
BUREAU_SCORE	The credit bureau's score for the particular individual

The credit scorecards to be constructed will also follow a few assumptions:

- basic behavioural credit scorecards will be developed. Initial data exploration of the available variables indicates that the scorecard product will resemble a behavioural scorecard because of the nature of the variables. These variables do not contain data that can be captured at application (e.g. age, home ownership etc.),
- the "delinquency cycle" variable will be used to determine whether an account falls into a "good" or "bad" account status and
- an industry norm of 6 cycles (or 180 days in arrears) will used to determine whether an account is in default or not.

4.2.2 Cleansing the database

The initial investigation of the database and its variables has shown that each variable should be scrutinised individually to determine what types of data should be included. Before isolating the 8 chosen variables (indicated in Figure 4.2) entries should be removed from the database that should not be considered in the calculation. The scorecards will be built using ISC data only. Thus all entries will be removed that do not have a product code within the array: ('001', '002', '003', '004', '005', '006', '007', '009', '021').

This information was provided by Woolworths, with these provisions:

- The scorecards will only be developed on active accounts. Thus all entries that do not have an "A" account status will be removed.
- Investigation has shown that there are some missing data points in the "bureau score" variable. To keep the calculation as statistically sound as possible, these entries will be removed.
- Corporate accounts with "block_code" "P" have very large exposures and are not in scope for this study, they will also be excluded.

It is important to note that the coding from this point forward (until indicated) will be applied to the data for the first scorecard (June 2006).

To achieve the instructions indicated above, SAS code (see Figure 4.1) is used.

```
Libname Chp4 "C:\Users\jvdwalt\Desktop\Masters\WFS\DATA\Chapter 4 - Empirical Study";
/*Take out Product code which is not ISC */
DATA Chp4.Data2006;
SET Chp4.Data2006;
WHERE wfs_product in ('001', '002', '003', '004', '005', '006', '007', '009', '021');
RUN;
/*identify only Active accounts*/
DATA Chp4.Data2006;
SET Chp4.Data2006;
WHERE status_200606 in ('A');
RUN;
/*identify accounts that have populated bureau scores*/
DATA Chp4.Data2006;
SET Chp4.Data2006;
WHERE bureau_score_200606 > 0;
RUN;
/*Remove corporate accounts (block code "P", they have very large balances and are not in scope for this study)*/
DATA Chp4.Data2006;
SET Chp4.Data2006;
        WHERE block_code_1_200606 not in ('P');
RUN;
```

Figure 4.1: Data cleansing code

After the code in Figure 4.1 was applied, there was a reduction from 2 007 481 entries to 1 274 922 (732 599 entries removed). The next step in the cleansing process is to identify and remove outliers within the database. As discussed and explained in Chapter 2, the PROC UNIVARIATE function in SAS identifies the outliers within a dataset. The function also gives an output of the basic statistics of each variable included in the coding (as seen in the summary of Table 4.3 below).

Table 4.3: Basic Statistics (Scorecard Variables)

Field (Variable)	Number	Mean	Median	Std. Deviation
SM_OPEN_BALANCE	1 274 922	2 257	1 432	2825
SM_CREDIT_LIMIT		5 626	4 700	3961
SM_PURCHASES		341	0	667
SM_PAYMENTS		346	200	942
CURR_DUE		213	147	274
BUREAU_SCORE		682	683	102

The PROC UNIVARIATE procedure also provided the outliers for each variable. The outliers were grouped (as seen in Table 4.4 below) to determine the coding to be used to remove the variables. The entries will be removed on a minimum and maximum basis.

Table 4.4: Range for Outliers (Scorecard Variables)

Field (Variable)	Minimum	Maximum
SM_OPEN_BALANCE	-11 000	45 000
SM_CREDIT_LIMIT	0	50 000
SM_PURCHASES	0	50 000
SM_PAYMENTS	0	130 000
CURR_DUE	0	7 000

Initial inspection the maximum values seem far removed from the observed mean values of each variable. This phenomenon will be catered for when values are grouped into categories for each variable. The exclusion of the bureau score variable in Table 4.4 is due to the fact that no outlying data were found. After running the code indicated in Figure 4.2, the dataset size was reduced from 1 274 922 to 1 274 852 (70 entries removed).

```
/*Removing Outliers*/
DATA Chp4.Data2006Out;
                SET Chp4.Data2006;
                IF sm_open_bal_200606 < -11000 then delete;
                IF sm_open_bal_200606 > 45000 then delete;
                IF credit_limit_200606 => 50000 then delete;
                IF sm_purchases_val_200606 > 50000 then delete;
                IF sm_payments_val_200606 > 130000 then delete;
                IF curr_due_val_200606 > 7000 then delete;
        RUN;
```

Figure 4.2: Code to remove Outliers

4.2.3 Binning

The next step in the scorecard development is to bin the data. As explained in Chapter 2, this process is to put all the data entries into a specific category (or bin) within each variable. Each variable should be investigated and "binned" to resemble a distribution which most appropriately fits the variable (e.g. most would be a uniform distribution with a relatively smooth representation). In order to achieve this each variable was individually investigated and their bins determined through trial and error. Own discretion was used when the number of bins to be used were selected. The main aim of binning is to create categories that will be used in the logistic regression step.

4.2.3.1 Variable: Open Balance

The following bins were determined after segmentation was done in SAS (Table 4.5). The SAS code used to achieve the segmentation is displayed under Annexure B2.

Table 4.5: Variable: Open Balance (Bins)

Bin Number	Range	Number of Entries
1	$Bin\ 1 < 0$	72 795
2	$0 \leq Bin\ 2 \leq 300$	148 329
3	$300 < Bin\ 3 \leq 1000$	266 647
4	$1000 < Bin\ 4\ \leq 2500$	401 201
5	$2500 < Bin\ 5 \leq 5000$	235 727
6	$5000 < Bin\ 6 \leq 8000$	91 512
7	$Bin\ 7 > 8000$	58 641

The above numbers translate into the histogram depicted by Figure 4.3 below. It indicates a smooth uniform trend (applicable for this variable) and thus the "bins" can be considered for scorecard development.

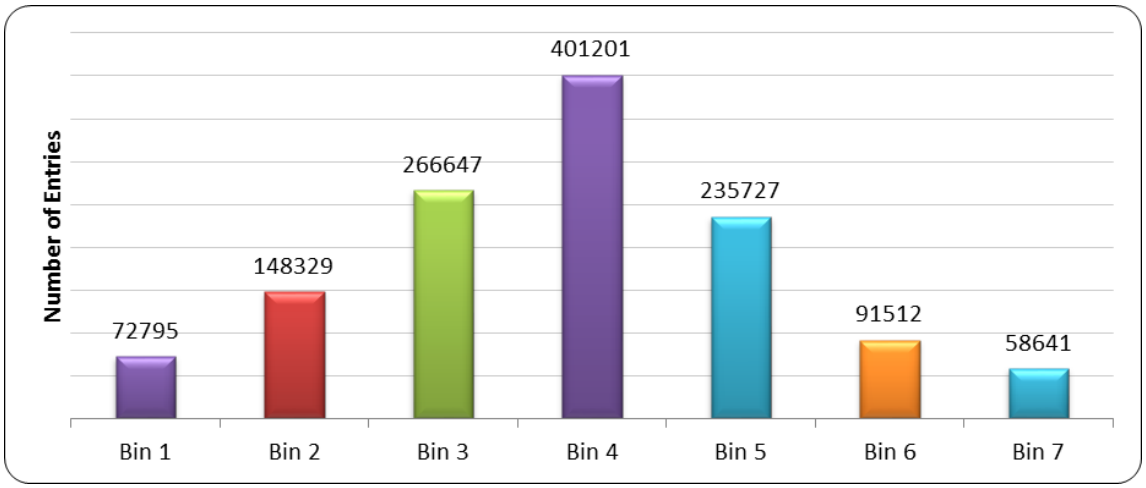


Figure 4.3: Histogram of Open Balance (Bins)

4.2.3.2 Variable: Credit Limit

The following bins were determined after segmentation was done in SAS (Table 4.6). An example of the SAS code used to achieve the segmentation is displayed under Annexure B2 (Only variable: Open_Balance is shown, but the other variables' categories were calculated with the same methodology)

Table 4.6: Variable: Credit Limit (Bins)

Bin Number	Range	Number of Entries
1	$Bin\ 1 \leq 1500$	148 476
2	$1500 < Bin\ 2 \leq 3000$	281 703
3	$3000 < Bin\ 3 \leq 6000$	360 698
4	$6000 < Bin\ 4\ \leq 10000$	304 484
5	$Bin\ 5 > 10000$	179 491

The above numbers translate into the histogram depicted by Figure 4.4 below. It indicates a smooth uniform trend (applicable for this variable).

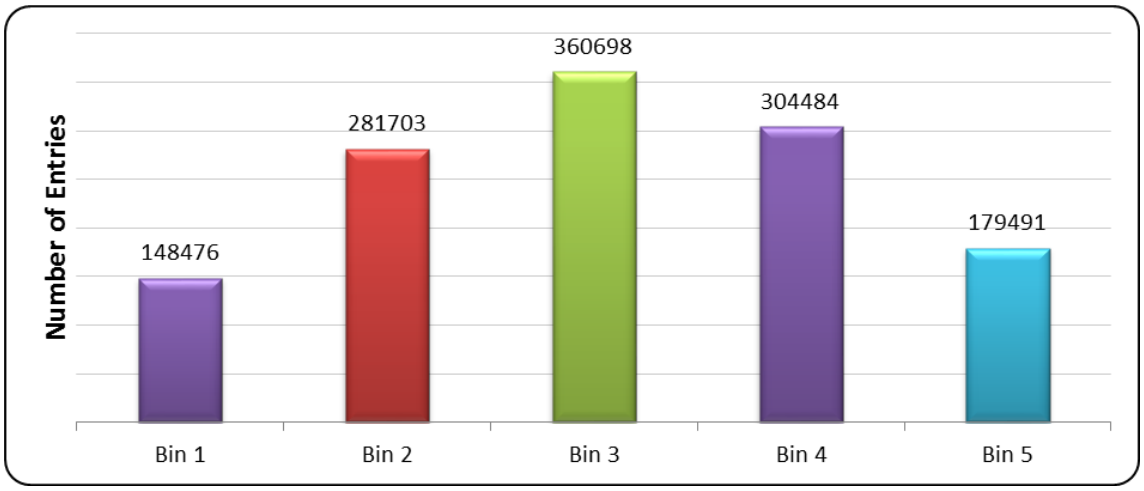


Figure 4.4: Histogram of Credit Limit (Bins)

4.2.3.3 Variable: Value of Purchases

The following bins were determined after segmentation was done in SAS (Table 4.7). An example of the SAS code used to achieve the segmentation is displayed under Annexure B2 Refer to first variable (Open Balance) for methodology.

Table 4.7: Variable: Value of Purchases (Bins)

Bin Number	Range	Number of Entries
1	$Bin\ 1 = 0$	640627
2	$0 < Bin\ 2 \leq 200$	146274
3	$200 < Bin\ 3 \leq 500$	194196
4	$500 < Bin\ 4 \leq 1000$	161377
5	$1000 < Bin\ 5 \leq 2500$	113807
6	$Bin\ 6 > 2500$	18571

The above numbers translate into the histogram depicted by Figure 4.5 below. Bin 1 refers to no purchases made during the billing cycle, this is expected consumer behaviour. If purchases are made a smooth uniform trend is depicted (applicable for this variable).

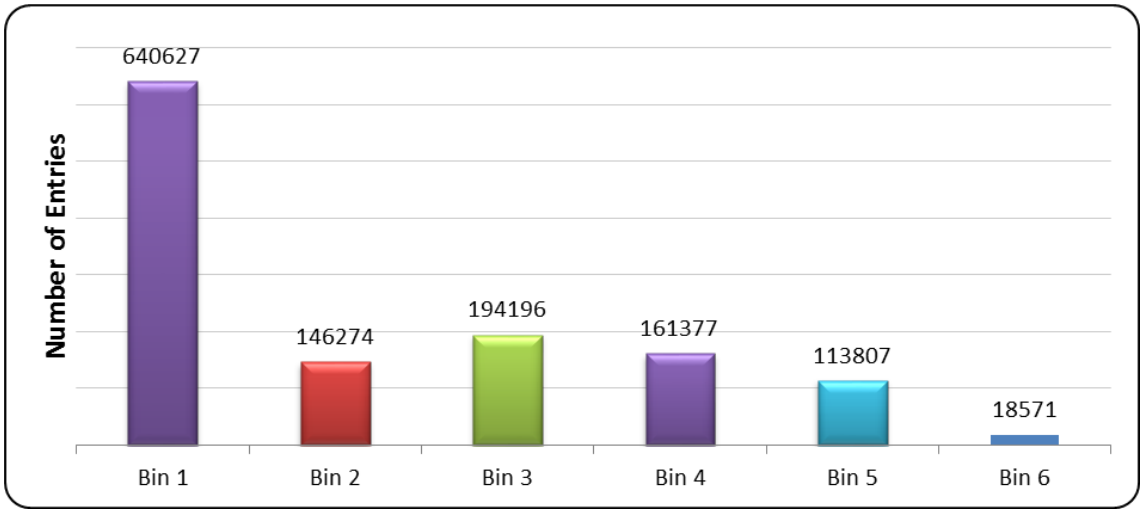


Figure 4.5: Histogram of Purchase Value (Bins)

The histogram in Figure 4.5 indicates a significant right-skewed distribution. It indicates that 50.25% of the entries have a zero value, which means that half of the active customer base did not utilise their in-store card during the billing cycle.

4.2.3.4 Variable: Value of Payments

The following bins were determined after segmentation was done in SAS (Table 4.8). An example of the SAS code used to achieve the segmentation is displayed under Annexure B2. Refer to first variable (Open Balance) for methodology.

Table 4.8: Variable: Value of Payments (Bins)

Bin Number	Range	Number of Entries
1	$Bin\ 1 = 0$	330 020
2	$0 < Bin\ 2 \leq 200$	390 441
3	$200 < Bin\ 3 \leq 400$	250 108
4	$400 < Bin\ 4\ \leq 600$	130 357
5	$600 < Bin\ 5 \leq 1000$	99 222
6	$Bin\ 6 > 1000$	74 704

The above numbers translate into the histogram depicted by Figure 4.6 below. A right skewed uniform distribution is depicted (applicable for this variable).

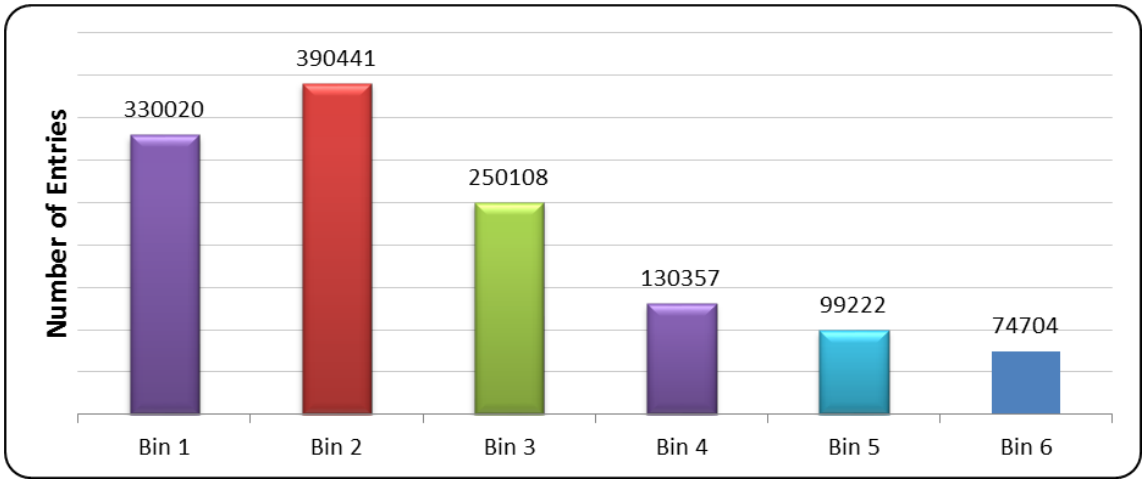


Figure 4.6: Histogram of Payments Value (Bins)

As with the "Purchase Value" variable, the "Payments Value" variable also has a right-skewed data distribution. It also indicates that most of the customer base did not provide a payment during the billing cycle (25.9%), but 30.6% of the customers did indeed make a payment of between R0 and R200. This most likely is due to a debit order set up by the customers to subtract the minimum amount due for the billing cycle.

4.2.3.5 Variable: Current Due Value

The following bins were determined after segmentation was done in SAS (Table 4.9). An example of the SAS code used to achieve the segmentation is displayed under Annexure B2. Refer to first variable (Open Balance) for methodology.

Table 4.9: Variable: Current Due Value (Bins)

Bin Number	Range	Number of Entries
1	$Bin\ 1 \leq 50$	289 393
2	$50 < Bin\ 2 \leq 150$	356 650
3	$150 < Bin\ 3 \leq 400$	410 747
4	$Bin\ 4 > 400$	218 062

The above numbers translate into the histogram depicted by Figure 4.7 below. A relatively uniform trend is illustrated (applicable for this variable).

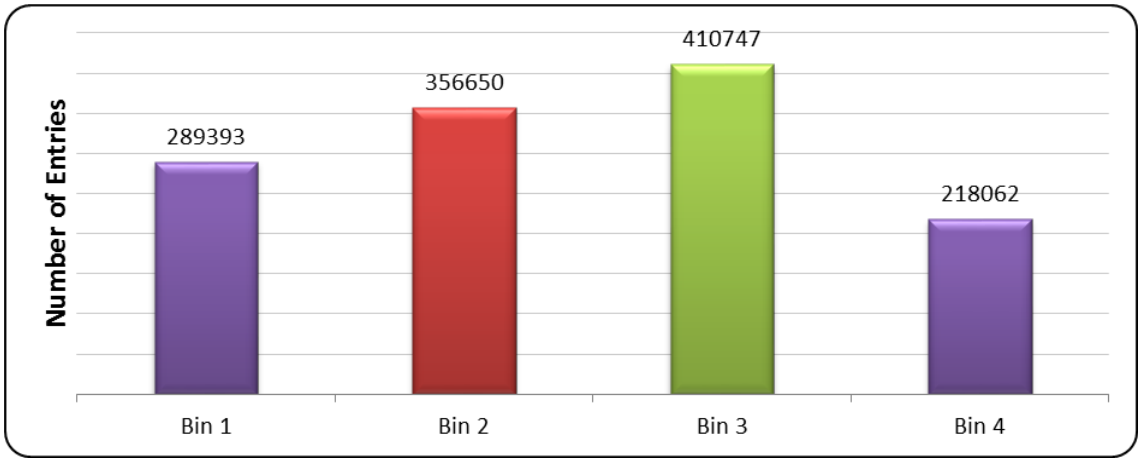


Figure 4.7: Histogram of Current Due Value (Bins)

4.2.3.6 Variable: Bureau Score

The following bins were determined after segmentation was done in SAS (Table 4.10). An example of the SAS code used to achieve the segmentation is displayed under Annexure B2. Refer to first variable (Open Balance) for methodology.

Table 4.10: Variable: Bureau Score (Bins)

Bin Number	Range	Number of Entries
1	$Bin\ 1 \leq 600$	324 910
2	$600 < Bin\ 2 \leq 700$	377 836
3	$700 < Bin\ 3 \leq 800$	379 355
4	$Bin\ 4 > 800$	192 751

The numbers translate into the histogram depicted by Figure 4.8 below. A relatively uniform trend is illustrated (applicable for this variable).

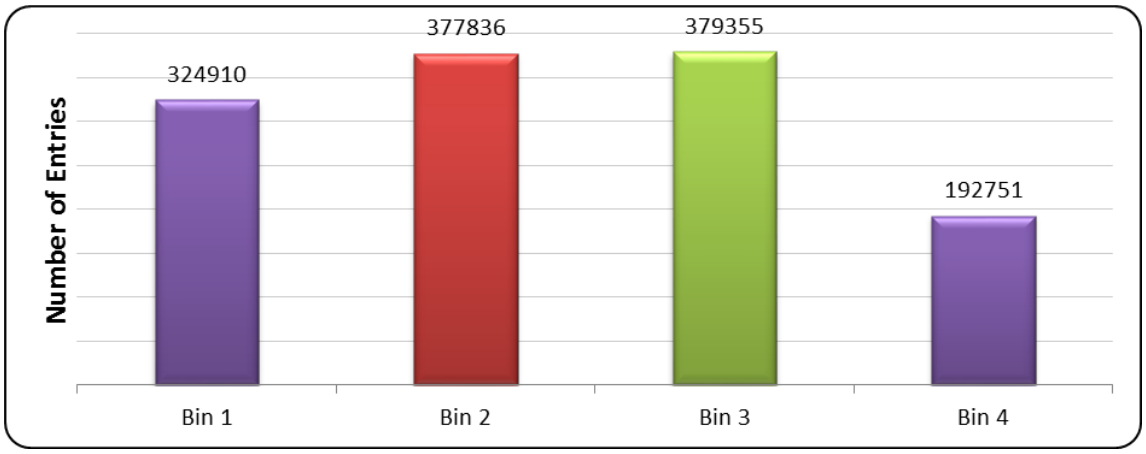


Figure 4.8: Histogram of Bureau Score (Bins)

The categories of each variable have now been identified. Along with creating a "Good "/ "Bad" variable in the next step, the data will be transformed to fit into the defined categories (bins) as defined above.

4.2.4 Identifying 'Good/Bad' Indicators

An additional variable will now be created that will act as the response variable in the regression analysis. This variable is called the Good/Bad indicator. The variable indicates whether an account is still viewed as "Good", implying the account is still performing, or as "Bad ", implying the account is in default. Confirmation from Woolworths indicated that the company applies the industry standard of 180 delinquent to identify defaulted accounts. The new variable will then be based on a six month delinquency cycle.

The next step is to take a sample of "Good" accounts as well as "Bad" accounts. As stated in Chapter 3, Siddiqi (2006) requests 2000 "Good" and 2000 "Bad" accounts to be used in the calculation sample. In this case, as sufficient data are available, 4000 of each will be included in the regression analysis. The code in Figure 4.9 performs the actions described above. This includes identifying

defaulted accounts, sampling 4000 "Good" and 4000 "Bad" accounts from the dataset and then combining the 2 datasets into the one to be used in the regression analysis. The SRS (Simple Random Sampling) method is used to determine the "Good" and "Bad" sample datasets.

```
/*indicate defaulted accounts with "GoodBad" variable*/
DATA Chp4.Data2006out;
        SET Chp4.Data2006out;
        IF delinquency_cycle_200606 < 6 then GoodBad = "Good";
        IF delinquency_cycle_200606 => 6 then GoodBad = "Bad";
run;/*Split bad accounts*/
DATA Chp4.Data2006_BAD;
        SET Chp4.Data2006out; WHERE GoodBad in ('Bad');
RUN;
/*Split good accounts*/
DATA Chp4.Data2006_Good;
                SET Chp4.Data2006out; WHERE GoodBad in ('Good');
RUN;
/*Take a sample of 4000 from bad accounts*/
PROC SURVEYSELECT DATA=Chp4.Data2006_BAD OUT=Chp4.Bad_Sample200606 METHOD=SRS
SAMPSIZE=4000 SEED=1234567;
RUN;
/*Take a sample of 4000 from good accounts*/
PROC SURVEYSELECT DATA=Chp4.Data2006_Good OUT=Chp4.Good_Sample200606
METHOD=SRS SAMPSIZE=4000 SEED=1234567;
RUN;
/*Combine 2 datasets*/
DATA Chp4.CombSample_200606;
        SET Chp4.Good_Sample200606 Chp4.Bad_Sample200606;
RUN;
```

Figure 4.9: Sample Dataset Procedure

After the sample has been constructed the entries within this sample needs to be transformed into the categories determined in the binning phase. This is accomplished by exporting the sample dataset to Microsoft Excel and applying the categorical formulas, which have been set apart in the binning segment. The format of the sample dataset then changes from Table 4.11 to Table 4.12.

Table 4.11: Sample Dataset Format (Extract)

Open Bal	Credit limit	Purchases	Payments	Curr due	Bur score	Good/ Bad
652.05	2300	0	250	41.21	596	Good
696.09	2200	0	100	60.73	769	
533.38	1400	203.9	100	64.86	638	

Table 4.12: Transformed Sample Dataset Format (Extract)

Open Bal	Credit limit	Purchases	Payments	Curr due	Bur score	Good/ Bad
300 $\rightarrow$ 1000	1500 $\rightarrow$ 3000	= 0	200 $\rightarrow$ 400	$\leq$ 50	$\leq$ 600	Good
300 $\rightarrow$ 1000	1500 $\rightarrow$ 3000	= 0	0 $\rightarrow$ 200	50 $\rightarrow$ 150	700 $\rightarrow$ 800	
300 $\rightarrow$ 1000	< 1500	200 $\rightarrow$ 500	0 $\rightarrow$ 200	50 $\rightarrow$ 150	600 $\rightarrow$ 700	

4.2.5 Regression Analysis and Data Normalisation

The next step is to perform logistic regression on the sample dataset set out above. After the logistic regression is completed the regression coefficients are then normalised to present a score out of 1 000. The same code structure as set out in Chapter 3 was used (see Figure 3.9 for logistic regression and Figure 3.10 for data normalisation). The output calculated is indicated in Table 4.13 below.

Table 4.13: Scorecard (2006/06 Data)

Variable	ClassVal0	DF	Estimate	min est	counter	est1	max est1	sum max	max cat	score
SM Open Bal 200606	> 8000	1	-7.85	-7.85	7	0.00	7.85	31.13	0	0
	5000 → 8000	1	-7.42			0.43			0	14
	1000 → 2500	1	-6.56			1.28			0	41
	2500 → 5000	1	-7.10			0.75			0	24
	0 → 300	1	-6.12			1.73			0	55
	300 → 1000	1	-5.80			2.04			0	66
	> 0	0	0.00			7.85			1	252
SM Payments Val 200606	600 → 1000	1	-2.20	-4.08	6	1.88	4.08		0	60
	200 → 400	1	-2.29			1.79			0	58
	> 1000	0	0.00			4.08			1	131
	400 → 600	1	-2.39			1.69			0	54
	0 → 200	1	-2.45			1.62			0	52
	= 0	1	-4.08			0.00			0	0
SM Purchases Val 200606	1000 → 2500	1	-8.52	-13.88	6	5.36	13.88		0	172
	= 0	1	-13.88			0.00			0	0
	0 → 200	1	-8.72			5.16			0	166
	> 2500	0	0.00			13.88			1	446
	500 → 1000	1	-8.85			5.03			0	162
	200 → 500	1	-8.15			5.73			0	184
bureau score 200606	600 → 700	1	-1.01	-3.27	4	2.26	3.27		0	73
	700 → 800	1	-0.17			3.10			0	100
	≤ 600	1	-3.27			0.00			0	0
	> 800	0	0.00			3.27			1	105
credit limit 200606	< 1500	0	0.00	0.00	5	0.00	1.18		0	0
	> 10000	1	1.18			1.18			1	38
	6000 → 10000	1	0.81			0.81			0	26
	3000 → 6000	1	0.63			0.63			0	20
	1500 → 3000	1	0.47			0.47			0	15
curr due val 200606	> 400	0	0.00	-0.62	4	0.62	0.88		0	20
	50 → 150	1	0.26			0.88			1	28
	150 → 400	1	0.08			0.69			0	22
	≤ 50	1	-0.62			0.00			0	0

The above scorecard (Table 4.13) gives a good indication what types of variables contribute the most to a performing account. It is clear that purchases (larger than 2500) contributes the most to a positive score (with 446). There exists an element of correlation in this statement though. A customer bordering on going into default would not spend much or at all, thus the act of purchasing (within the billing period) indicates that the account is utilised and in low danger of default. Other than the purchase variable, Table 4.14 below shows the most significant contributors:

Table 4.14: Significant Contributors (Variables and Categories – 2006/06 Data)

Variable	Category	Score
Open Balance	< 0	252
Payments	> 1000	131
Bureau Score	700 → 800	100
Bureau Score	> 800	105

4.2.5.1 Explanation of significant variables

Open Balance (> 0): An account that has a "less than 0" opening balance indicates that the account holder has a debit balance. This fact points to a client that pays his/her account regularly and often more than necessary. This contributes the most to a good behavioural score.

Payments (> 1000): An account holder which pays an amount larger than R1 000 during a billing cycle also indicates that the holder intends to stay away from default and up to date with payments.

Bureau Score (700 → 800): A bureau score from 700 to 800 indicates an individual that stays out of default problems by looking after all obligations. Thus the variable contributes 100 points toward a credit score.

Bureau Score (≥ 800): A bureau score above 800 is seen as an exceptional client, looking after all obligations.

4.2.5.2 Scorecard Report – Plug&Score (Lite)®

To review and evaluate the generated scorecard, the statistical software "Plug&Score" was used. Plug&Score is a software package specifically designed to evaluate and report on credit scorecards. The software has a simplistic design and easy to use interface.

Only the sample dataset and generated scorecard input is required and then the report is calculated. The scorecard report on the 2006/06 data reads as follows:

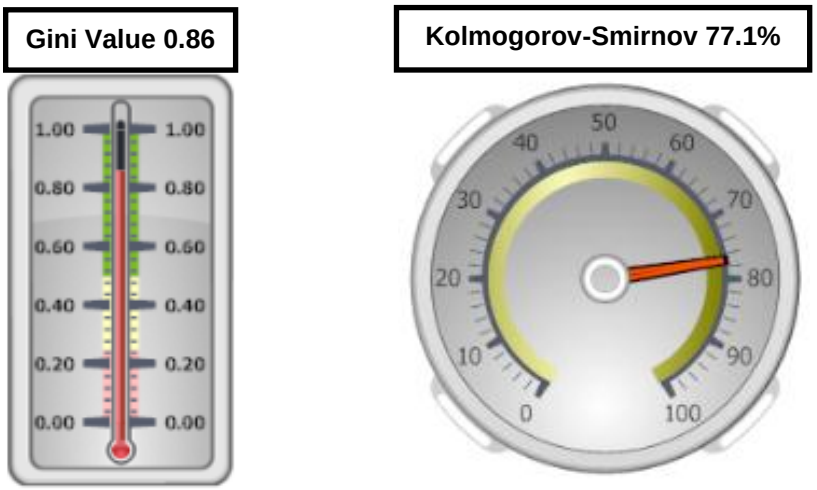


Figure 4.10: Gini coefficient and Kolmogorov-Smirnov Test (2006/06 scorecard)

The Gini coefficient (Chapter 3) indicates the quality of the classification of the sample data and ultimately the effectiveness of the scorecard. The evaluation of the quality of classification by the Gini coefficient is performed with the help of Table 4.15:

Table 4.15: Gini coefficient (classification quality)

Gini coefficient	Classification Quality
< 0.25	Low
0.25 $\rightarrow$ 0.5	Medium
> 0.5	High

Thus, the Gini value of 0.86 indicates that the scorecard (using 2006/06 data) *does* effectively determine "Good/Bad" classification.

The Kolmogorov-Smirnov value shows the maximal difference between the distributions of "Good" and "Bad". The higher this value, the better the scorecard will perform. The 77.1% Kolmogorov-Smirnov value in Figure 4.10 indicates that the distributions are more than sufficiently different in their distributions.

4.3 Credit Scorecard Development (Post-Crisis)

Section 4.2 followed the step-by-step guide set out in Chapter 3 to build a behavioural credit scorecard based on Woolworth's data from June 2006. This section will now show the scorecard results of Jul 2009 data, which is during the global financial crisis. Only the scorecard results will be shown, as exactly the same method was used to develop the 2009 scorecard. The scorecard results (Table 4.16) for the 2009/07 data are:

Table 4.16: Scorecard (2009/07 data)

Variable	ClassVal0	DF	Estimate	min est	counter	est1	max est1	sum max	max cat	score
SM Open Bal 200907	> 8000	1	-11.23	-11.23	7	0.00	11.23	38.13	0	0
	5000 → 8000	1	-9.77			1.46			0	38
	1000 → 2500	1	-6.54			4.69			0	123
	2500 → 5000	1	-7.97			3.26			0	86
	0 → 300	1	-5.28			5.95			0	156
	300 → 1000	1	-5.15			6.08			0	159
	< 0	0	0.00			11.23			1	295
SM Payments Val 200907	600 → 1000	1	-1.02	-5.50	6	4.48	5.50		0	117
	200 → 400	1	-2.96			2.53			0	66
	> 1000	0	0.00			5.50			1	144
	400 → 600	1	-1.93			3.57			0	94
	0 → 200	1	-4.09			1.40			0	37
	= 0	1	-5.50			0.00			0	0
SM Purchases Val 200907	1000 → 2500	1	0.50	-14.19	6	14.69	14.96		0	385
	= 0	1	-14.19			0.00			0	0
	0 → 200	1	0.77			14.96			1	392
	> 2500	0	0.00			14.19			0	372
	500 → 1000	1	0.57			14.76			0	387
	200 → 500	1	0.65			14.84			0	389
bureau score 200907	600 → 700	1	-1.02	-2.54	4	1.52	2.54		0	40
	700 → 800	1	-0.41			2.13			0	56
	≤ 600	1	-2.54			0.00			0	0
	> 800	0	0.00			2.54			1	67
credit limit 200907	< 1500	0	0.00	0.00	5	0.00	1.30		0	0
	> 10000	1	1.30			1.30			1	34
	6000 → 10000	1	1.25			1.25			0	33
	3000 → 6000	1	0.97			0.97			0	25
	1500 → 3000	1	0.46			0.46			0	12
curr due val 200907	> 400	0	0.00	-2.61	4	2.61	2.61		1	68
	50 → 150	1	-1.10			1.51			0	40
	150 → 400	1	-1.19			1.42			0	37
	≤ 50	1	-2.61			0.00			0	0

Once again, the purchase variable has the most significance in determining whether an account will default or not. Interestingly though, the scores for the purchase variable are even higher compared to those in the 2006 scorecard. This indicates that customers who were able to spend during the global financial crisis had an even lower risk of default (scores indicated in Table 4.17 below).

Table 4.17: Purchase Variable (2006 vs 2009 Data)

Variable	Category	2006 Score	2009 Score
SM_Purchases_Val	1000 → 2500	172	385
	= 0	0	0
	0 → 200	166	392
	> 2500	446	372
	500 → 1000	162	387
	200 → 500	184	389

4.3.1 Explanation of significant variables

Other than the purchase variable, the most significant contributors were the "Open_Balance" and "Payments" variables (as seen in Table 4.18 below). Although these two variables were also significant in the 2006 scorecard, they are even more important in the 2009 scorecard. The increased significance has decreased the emphasis placed on the bureau score. Thus the 2009 scorecard points to the lower risk of default for customers who were paying their accounts during the financial crisis.

Table 4.18: Significant Contributors (Variables and Categories – 2006/06)

Variable	Category	Score
SM_Open_Bal	1000 → 2500	123
	0 → 300	156
	300 → 1000	159
	< 0	295
SM_Payments_Val	600 → 1000	117
	> 1000	144

4.3.2 Scorecard Report – Plug&Score (Lite)®

The Plug&Score software produced the following scorecard report:

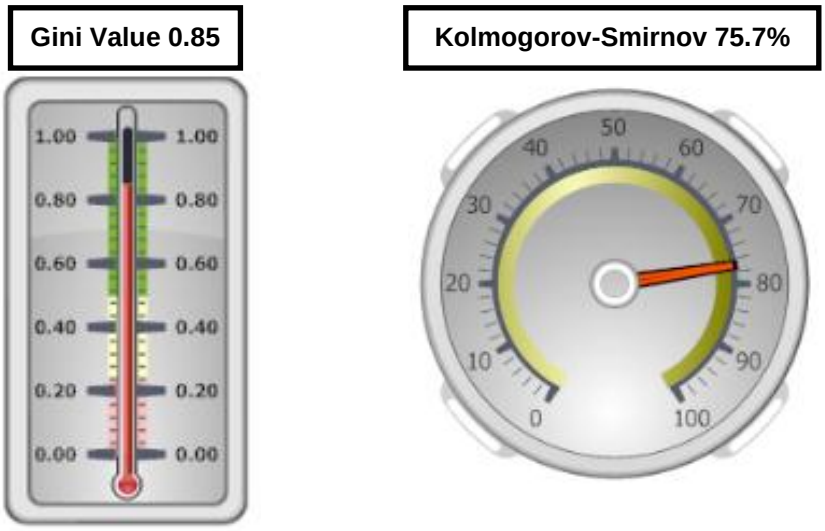


Figure 4.11: Gini coefficient and Kolmogorov-Smirnov test (2009/07 scorecard)

- The Gini value of 0.85 indicates that the scorecard (2009/07 data) effectively determines "Good/Bad" classification.
- The Kolmogorov-Smirnov value of 75.7% in Figure 4.11 indicates that the distributions are more than sufficiently different in their distributions.

4.4 Credit scorecard results and conclusion

The development and subsequent results of the two basic behavioural scorecards produced some interesting statistical points. Two of the most important findings are:

i) Before the global financial crisis (2006 data) Woolworths' retail customer risk was determined by their ability to purchase, utilising their account and having a good bureau score. During 2009 when the global financial crisis effects were felt in South Africa, the determination of risk shifted more towards the ability to purchase and to make payments on accounts and

ii) During 2009, customers had to achieve a higher credit score to be classified as a "good" account than in 2006. Moving the cut-off point[25] (score) to be approved for an in-store card from 150 to 200.

For the finding of such a cut-off point (acceptable score), the most attractive is the score range in which the share of the "bad" accounts decreases faster than that of the "good" accounts (Figure 4.12, 2006 data and Figure 4.13, 2009 data). The risk segmentation of the two data sets is illustrated in Annexure B4 and B5.

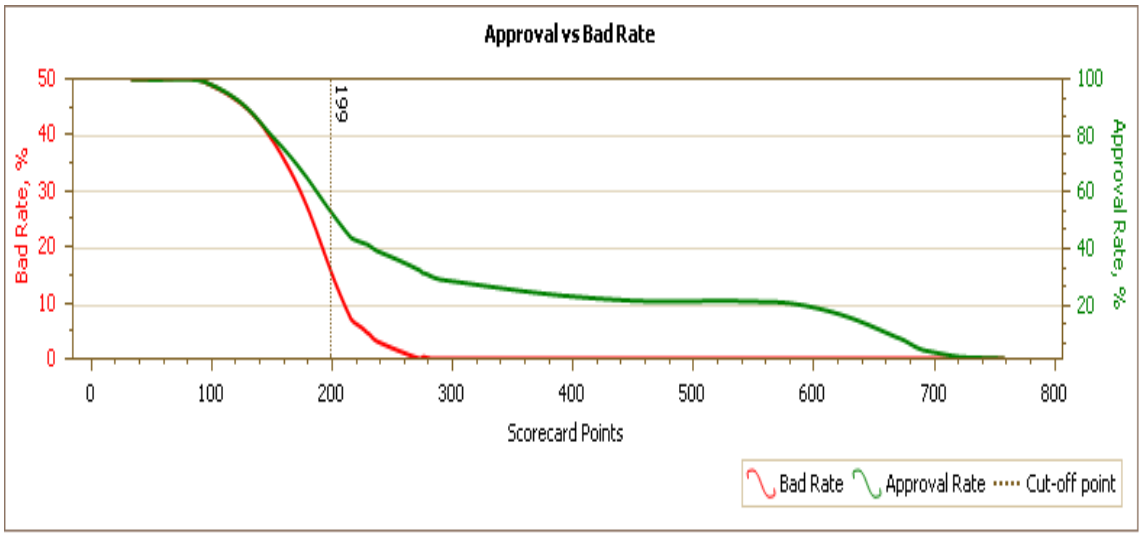


Figure 4.12: Approval vs bad rate (2006 data)

[25] Cut-off point: the decision barrier of a credit score. Above the cut-off point potential clients are deemed less risky, and below the cut-off point potential clients should be rejected due to a high risk factor.

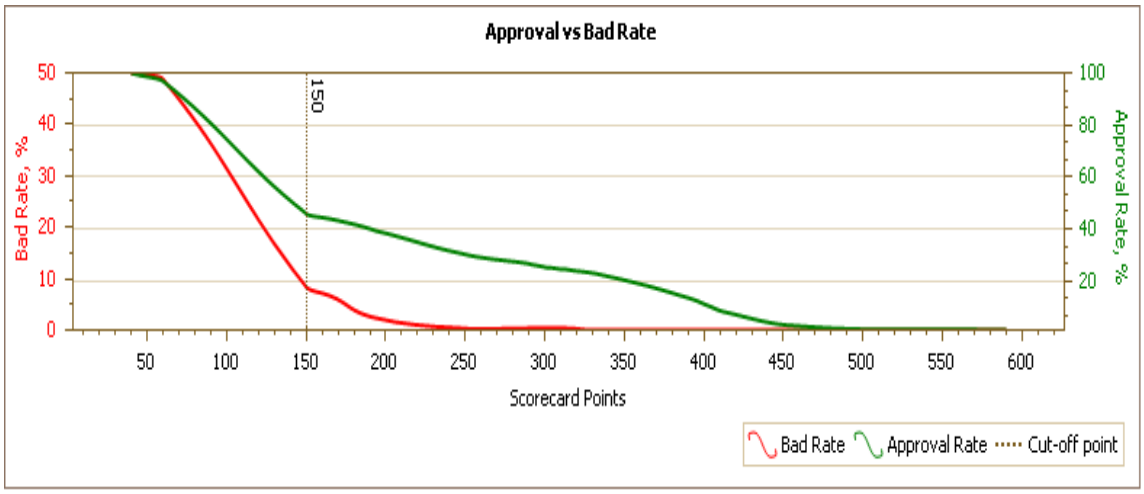


Figure 4.13: Approval vs bad rate (2009 data)

The retail credit market was definitely affected by the global financial crisis. It is evident from the results that Woolworth's clients who were able to make purchases, and – more importantly – make payments on their in-store accounts, had a lower risk of default compared to other customers. Section 4.5 provides a data analysis of all ISC data from January 2006 to December 2009 to assess changes in the retail credit market during the global financial crisis.

4.5 Data analysis (Woolworths ISC Data)

4.5.1 Customer base

The available Woolworths ISC data stretches from January 2006 to December 2009. The customer base grew over the period (Figure 4.14), but the actual customer growth rate declined. The customer growth rate dropped significantly during 2009, only achieving 0.52%. Whether through cost pressure or clients moving to cheaper retailers, the decline in customer growth over the initial period of the financial crisis indicates that the market became increasingly aware and cautious of taking on credit.

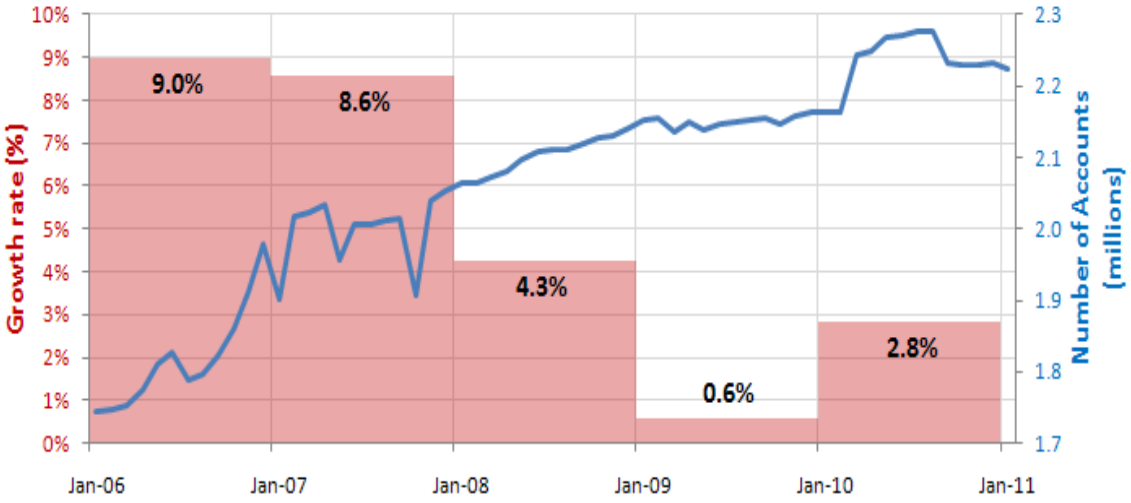


Figure 4.14: Woolworths ISC accounts

4.5.2 Defaults

Two methodologies were used to determine the number of defaulted accounts. The "Worst Ever" (WE) methodology assumes that if an account defaults once it would be treated as a defaulted account for the remainder of the accounts life time (up to December 2010) regardless what happens further on during the account data. The "Multiple Defaults" (MD) methodology assumes that when an account is not in default according to the default definition (180 days delinquent) the account has cured. Hence, an account can default, cure after a couple of months, and then default again.

Both the default methodologies were applied to the data to determine each time an account defaults. The number of defaults were counted and divided by the total number of "open" accounts. Accounts were defined as being "open" for the period between the first observation of the account and the last observation of the account.

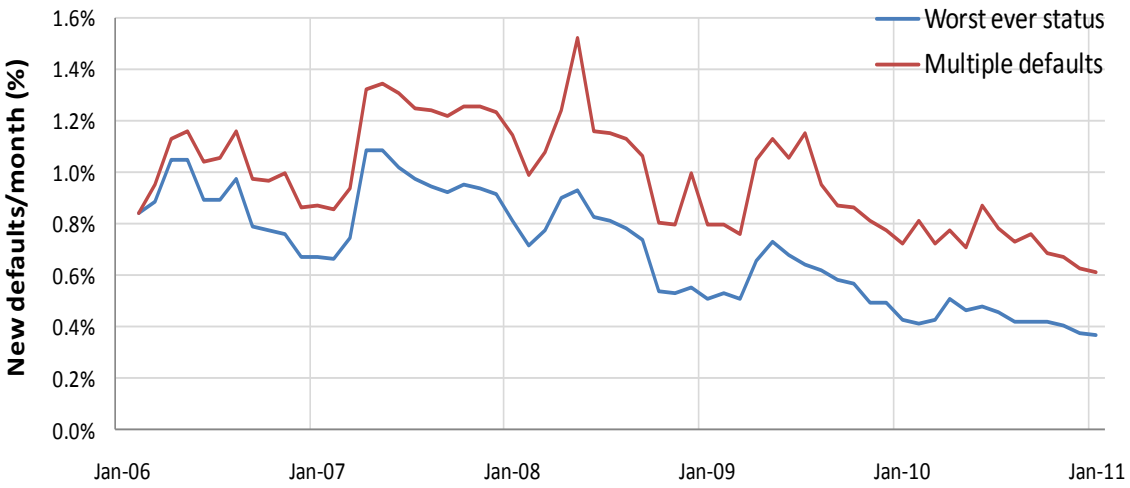


Figure 4.15: Percentage defaults each month

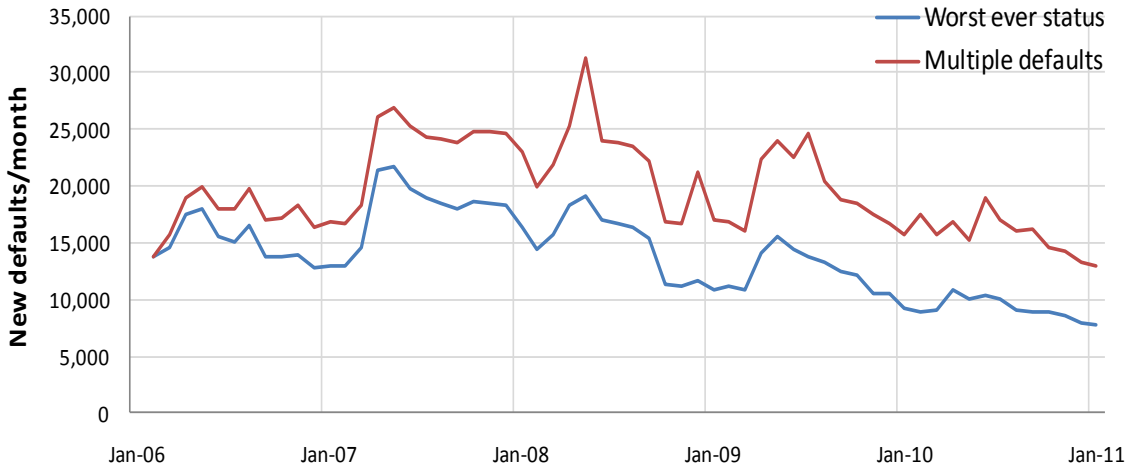


Figure 4.16: Number of new defaults each month

The overall downward trend in Woolworths' ISC defaults (Figure 4.15 and Figure 4.16) indicates a book that has performed quite well over the data window period. The two significant spikes in defaults did however occur during April 2008 and June 2009, which may be due to the increased pressures of the global financial crisis.

4.5.2.1 Default triggers

When examining Figure 4.17, a definite observable relationship between purchases made and monthly defaults can be seen. A spike in purchases is usually followed by a spike in number of defaults. This indicates that over spending leads to more defaults in the following billing cycles. For example, in January 2007 R630 million worth of purchases was made (up from R511 million in November 2006). In March of 2007 a spike in defaults occurred, increasing to 21346 defaults (up from 12955 in January 2007).

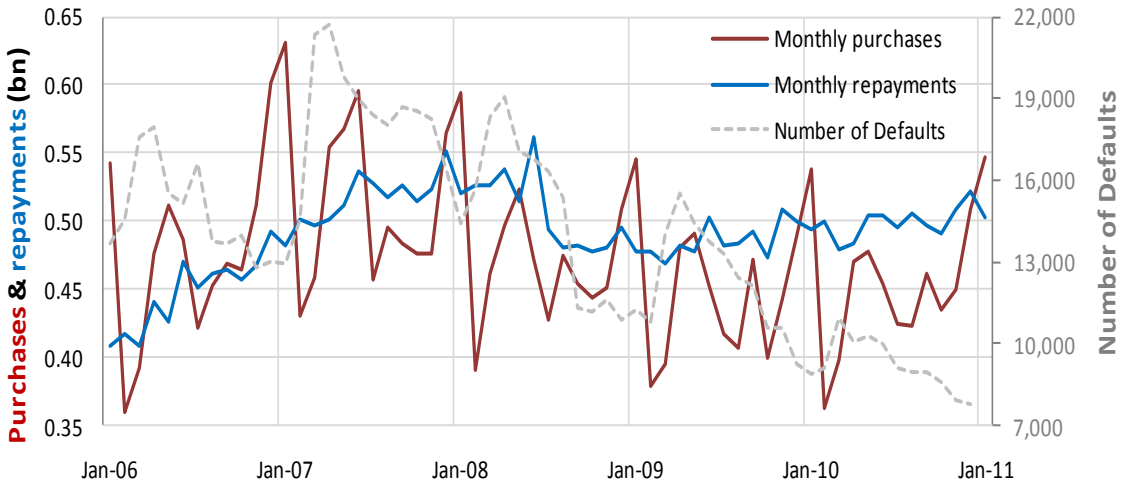


Figure 4.17: Purchases, payments and defaults

4.5.2.2 Credit limit versus credit uptake

Over the observed data period Woolworths has systematically decreased their total exposure towards their client base. Due to Woolworth's contractual obligation, they could not lower the credit limits already extended, but they could manage the new credit they extended. Thus over the course of 2007 to 2011, Woolworths actively decreased their total credit limit to try and control their exposure. At the same time, Woolworths' customers have been steadily taking on more credit (Figure 4.18) which in turn increased Woolworths' exposure towards their client base. This phenomenon explains the effects that the financial crisis has had on

both sides in the industry. Woolworths lowered their exposure to make sure the company did not take on unnecessary risk. The customers increased their demand for credit due to the financial pressure of the global financial crisis.

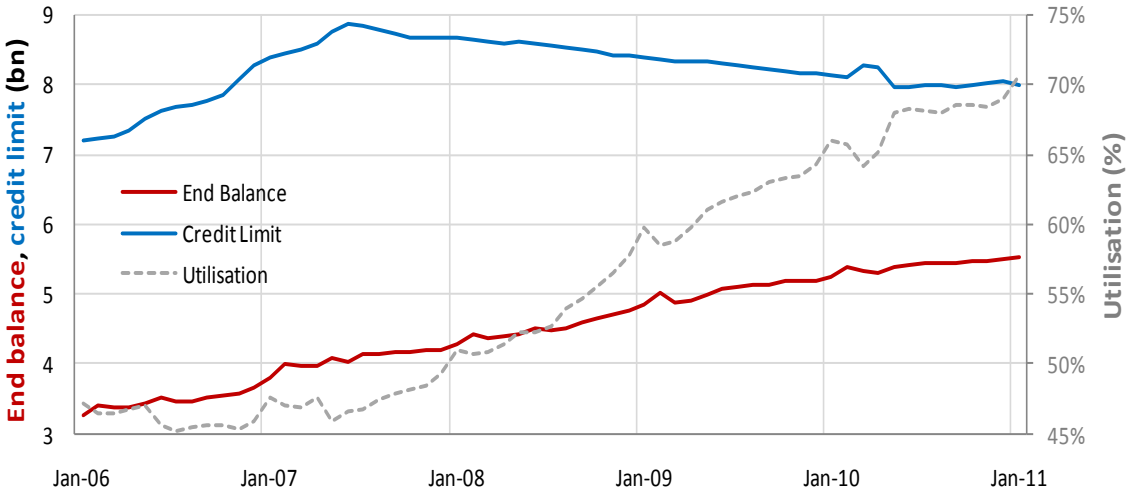


Figure 4.18: End balance versus credit limit (exposure)

The increase in credit uptake (End Balance) and decrease in exposure (Credit Limit) takes the utilisation figure up from 47% in January 2006 to 69% in December 2010.

As was the case with the scorecard results in Section 4.4, these analyses show that the global financial crisis definitely had an impact on the retail credit market in South Africa. Chapter 5 will provide a conclusion to the study by bringing the literature research and empirical study together. Chapter 5 will also explore insights into possible future study areas.

Chapter 5

Conclusion

5.1 The credit milieu

In 2011, the global financial milieu is undergoing a variety of different stresses, among them political drivers for democracy and fears of sovereign defaults. Many these pressures have been amplified by the global financial crisis as governments, economists and banks scrutinise their economic positions. Although the global economy is slowly recovering after the financial crisis, pressure has been placed on commodity prices and the disruptions to oil supply globally (IMF, 2011). The IMF has observed that the global fears of a double dip recession were (up until April 2011) 'unfounded'. With the emergence of the US debt crisis coupled with the already severe European debt crisis, these double dip recessionary fears could yet be proven correct.

South Africa is, at this time (2011), undergoing a contraction of credit in most industries. The value of new mortgages granted by banks for the quarter ended March 2011 decreased by 7.85% quarter-on-quarter from R26.9 billion to R24.86 billion (NCR, 2011). Vehicle finance indicated a decrease from R28.1 billion (December 2010) to R27.5 billion (March 2011) (NCR, 2011). Unsecured credit decreased by 0.84% from R16.8 billion to R16.7 billion. The only credit industry to increase in the quarter under review was retail credit facilities such as in-store cards, credit cards and bank overdrafts. These facilities saw an increase of 1.81% from R10.3 billion to R10.4 billion (NCR, 2011). Thus banks and companies that provide large credit facilities are lending less, but short term credit is on the increase in the country.

The empirical study undertaken in Chapter 4 confirms the statement made above, namely that credit facilities such as in-store cards have been increasing since the global financial crisis. Woolworths has made sure not to overextend credit to their

customers by reducing the total credit limit exposure. Their customers have however been utilising their credit facilities more since the global financial crisis. According to the NCR, the South African credit market has contracted since the introduction of the National Credit Act in 2007. The introduction of the act came at an opportune time to help protect consumers, as well as companies, from reckless borrowing and lending.

Consumers who have credit facilities such as credit cards and in-store cards have been able to pay their obligated amounts on a more on-going basis since the global financial crisis. Figure 5.1 below indicates an upward trend in current[26] accounts since 2008. Although marginal, it is important to remember that the number of accounts has also decreased since the crisis.

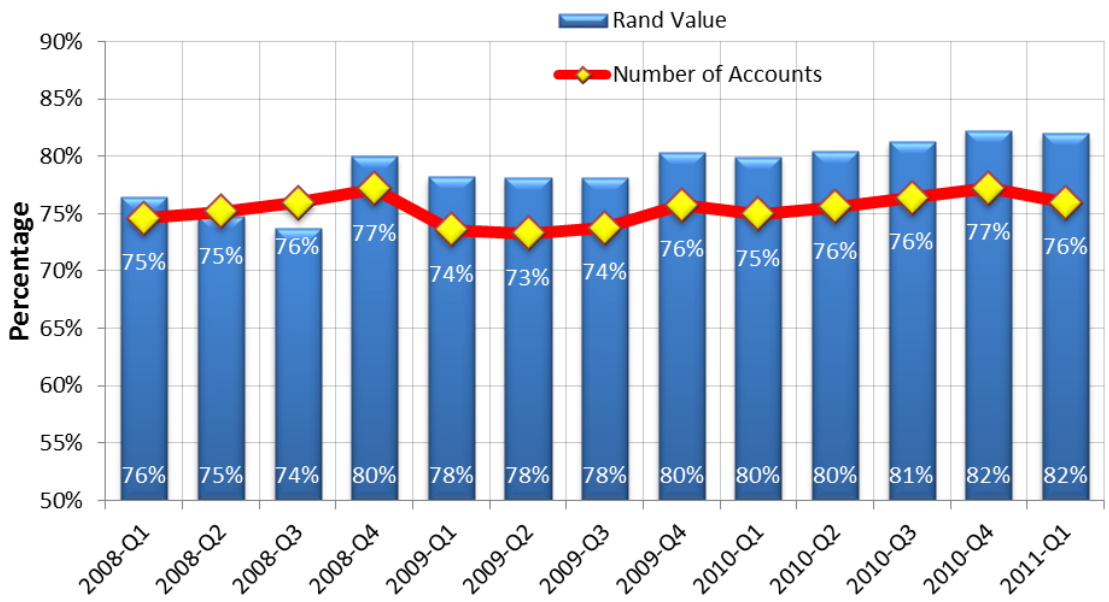


Figure 5.1: Credit facilities at a "current" position

Source: NCR, 2011

As with most other financial industries, the global financial crisis had a significant impact on the South African retail credit market. Credit cards and in-store cards

[26] Current accounts: accounts that are being serviced on their agreed timelines. Thus, customers are paying when payment is due.

will always be a popularly utilised credit facility. It provides a hassle free way of purchasing goods and services whilst not being immediately liable. Once the consumer is placed under financial pressure from any type of economic downturn, the easy accessible and revolving products such as these credit facilities are the first to be consulted. Consumers would automatically focus on their more expensive and important credit lines such as their mortgage or vehicle finance. As illustrated by the NCR 2011 data, large mortgage or vehicle type credit extensions are the first to decline whereas credit facilities such as credit cards and in-store cards are utilised more readily.

Credit scoring has had, and will have, an important role within the credit industry. Every company (whether a bank or other credit institution) needs to determine the risk of each potential client before considering the extension of credit. Keeping databases of these credit scores allows these companies to later rebuild their credit scorecards to reflect the actual nature of their specific business. The data will also allow a bank or business to evaluate the level of risk they are currently holding on their books. Knowing these details will enable executives and board members to compare it to the companies set risk appetite. Should the business decide to grow their books, they will need to take on more risk. If the businesses profits are under pressure due to a large amount of defaults, the business can re-evaluate their credit scoring acceptance criteria to reduce the risk on their books.

The biggest challenge for these companies is not the actual development of credit scorecards but the data that back these scorecards. It is essential for companies to set up their data capturing framework as well as data warehousing facilities before attempting to introduce credit scoring. An accurate client database is crucial for the development of an accurate credit scorecard. The science and mathematics driving credit scoring is an ever evolving subject, but an effective scorecard does not depend on the latest methodologies but the accuracy and effort put in by the company developing it. As stated in Chapter 3, the development of credit scoring is a company or enterprise wide exercise that should include many departments and individuals.

Credit scoring remains just one tool used by the financial world to control and mitigate risk. The ability to quantify these risks is found in the ever advancing field of quantitative banking. The problem though is that the financial world is continuously seeking new ways of making money or increasing return. This continuous pursuit of wealth often disregards these risk tools and frameworks and in return new risk mitigation methods are designed on the back of failings and economic crises.

As more and more problems emerge in the global economy and political revolutions engulf the Middle East and North Africa, every economic move of regulators and countries come under intense scrutiny. Leading nations such as the US, China and the UK sway the economic world with every decision or statistic published. These changes usually have their knock-on effects to the rest of the world, whether it is to withdraw from investing in emerging countries or buying gold and Swiss francs to 'ensure' safety. Economic researchers surely have endless research possibilities to look forward to as the global economic industry evolves in the 21st century.

5.2 Future research possibilities

Although this study only focussed on the effects of the financial crisis on the retail credit market in South Africa, various different avenues can be explored to determine the any impact of the crisis on South Africa. Table 5.1 lists and explains different avenues of possible future research topics on the effects of the global financial crisis on South Africa.

Table 5.1: Research Possibilities

TOPIC	DESCRIPTION
The role of the NCA in South Africa during the global financial crisis	Economists in South Africa suggest different elements may have limited the effects of the financial crisis in South Africa. One of these elements was the introduction of the National Credit Act in 2007, preceding the financial crisis. Researching the topic could reveal the actual role the act played during the crisis in South Africa.
The effect credit scoring can have on the profitability and risk control of a small credit provider.	Although all large credit institutions already make use of complex credit scoring, small credit providers usually have no set scoring method or at most a decision tree setup. The research can focus on how simple to complex credit scoring may assist smaller credit lenders to actively grow their book and have proper control over their risk levels and profitability.
The key drivers and trends of the housing market in South Africa during the financial crisis.	The housing bubble in the US played a big part in the origin of the financial crisis. The South African housing market has also undergone a slump since the effects of the crisis reached South Africa. A study can be undertaken to determine the effects, drivers and trends of the South African housing market during the financial crisis.
The effects of the financial crisis and other internal factors on consumer prices in South Africa.	The financial crisis, coupled with high petrol prices, taxes and the immanent high-way toll fees are all having an impact on consumer prices and inflation in South Africa. A study can be undertaken to determine the main drivers behind this trend and a possible forecast of where it may go.

5.3 Final assertion

The research conducted in this dissertation was primarily done to ascertain the effect the global financial crisis had on the retail credit industry, specifically retail credit facilities such as in-store cards by using credit scoring. Chapter 4 empirically shows that Woolworths (a major retailer in South Africa) was indeed affected by the global financial crisis as their customers took out more credit and they tried to restrict their exposure. This is of importance to the retail credit industry as the same trends can possibly be seen within their own business and will be of benefit when constructing credit score cards, new products or engaging new markets.

References

ACEA. 2009. FAQ on the economic crisis. European Automobile Manufacturer's Association. http://www.acea.be/index. Date of Access: 25 August 2011.

ALTMAN, E.I., SAUNDERS, A. 1997. Credit Risk Measurement: Developments over the last 20 Years. Journal of Banking & Finance, 21(11-12):1721-1742.

ANDERSON, R. 2007. The Credit Scoring Toolkit. Oxford, UK: Oxford University press. 296 p.

ANON. 2008. The Global Financial Crisis and Emerging Economies: Role Model South Africa. http://www.boell.de/worldwide/africa/africa-5426.html. Date of access: 10 May 2008.

ARTICLESBASE. 2010. Rebuilding Your Finances After an Economic Crisis. http://www.articlesbase.com. Date of Access: 25 August 2011.

ASHCRAFT, A.B. & SCHUERMANN, T. 2008. Understanding the Securitization of Subprime Mortgage Credit. New York: Federal Reserve Bank of New York (Staff Reports no. 318). 76 p.

BAILY, M.N., LITAN, R.E., & JOHNSON, M.S. 2008. The Origins

of the Financial Crisis. http://www.brookings.edu. Date of Access: 16 April 2010.

BAKER, D. 2005. The Housing Bubble Fact Sheet. Washington, DC: Centre for economic and Policy Research. 5 p.

BAKER, R. 2011. The Cause and Effects of the Lehman Brothers Crash. http://prospectsforcapitalism.wikispaces.com/search/view/lehman+brothers. Date of Access: 25 November 2011.

BALCHIN, N. 2009. The Impact of the Global Financial Crisis in Africa. AfriGrowth Agenda. Cape Town: Mthente Research and Consulting Services.

BAXTER, R. 2009. The global economic crisis and its impact on South Africa and the country's mining industry. http://www.resbank.co.za. Date of access: 25 May 2010.

BBC. 2008. Why is Northern Rock in trouble? http://news.bbc.co.uk/2/hi/business/7101523.stm. Date of access: 10 August 2009.

BERNANKE, B.S. 2005. The Economic Outlook. http://www.whitehouse.gov/cea/econ-outlook20051020.html. Date of Access: 15 August 2011

BIS. 2010. Basel III: A global regulatory framework for more resilient banks and banking systems. Basel: Bank of International Settlements. 69 p.

BOORMAN, J. 2009. The Impact of the Financial Crisis on Emerging Market Economies: The Transmission Mechanism, Policy Response and Lessons. http://www.emergingmarketsforum.org. Date of Access: 15 May 2010.

BUANEWS. 2009. Retail trade figures drop: BauNews. http://www.bizcommunity.com/Article/196/160/37026.html. Date of access: 15 June 2010.

BURKE, S. 2009. Missing Values, Outliers, Robust Statistics & Non-parametric Methods. High Wycombe, Buckinghamshire, UK: RHM Technology Ltd. 6 p.

CAGAN, C. 2007. Mortgage Payment Reset. http://www.facorelogic.com/landingpages/caganform.jsp?id=8037. Date of Access: 16 April 2010.

CAIRE, D., BARTON, S., DE ZUBIRIA, A., ALEXIEV, Z., DYER, J., BUNDRED, F. & BRISLIN, N. 2006. A HANDBOOK FOR DEVELOPING CREDIT SCORING SYSTEMS IN A MICROFINANCE CONTEXT. Washington, D.C.:United States Agency of International Development. 79 p.

CAIRE, D. & KOSSMANN, R. 2003. Credit Scoring: Is It Right for Your Bank? Bannock Consulting. London. 12 p.

CALI, M., MASSA, I. & VELDE, W.D. 2008. The Global Financial Crisis: Financial Flows to Developing Countries Set to Fall by One Quarter. London, UK: Overseas Development Institute.

CAMERON, B. 2009. Financial crisis and credit clampdown 'are forcing us to save. http://www.persfin.co.za/index.php?fSectionId=593&fArticleId=4910631. Date of access: 5 May 2009.

CHAMBERS, M., GARRIGA, C. & SCHLAGENHAUF, D. 2005. Accounting for Changes in the Home Ownership Rate. St Louis, FL: Florida State University.

CHOMSISENPHET, S. &PENNINGTON-CROSS, A. 2006. The Evolution of the Subprime Mortgage Market. *Federal Reserve Bank of St. Louis Review*, 88(1): 31-56, January/February 2006.

CIA. 2011. CIA World Factbook. https://www.cia.gov/library/publications/the-world-factbook/wfbExt/region_eur.html. Date of Access: 15 August 2011.

COHEN, T. 2011. What can South Africa learn from Greece's meltdown? http://www.dailymaverick.co.za. Date of Access: 25 August 2011.

COLLYNS, C. 2008. The Crisis through the Lens of History. *Finance and Development*, 45(4):20, December.

COLTON, K.W. 2003. Housing in the Twenty-First Century: Achieving Common Ground. Cambridge, MA.: Harvard University Press. 528 p.

DATAMOLOGY, 2010. Data Exploration. http://www.datamology.com/exploration.shtml Date of access: 5 May 2011.

DE GRAUWE, P. 2010. The Greek crisis and the future of the Eurozone. http://www.econ.kuleuven.be. Date of access: 20 April 2010.

DETRIXHE, J. 2011. U.S. Loses AAA Credit Rating as S&P Slams Debt Levels, Political Process. http://www.bloomberg.com. Date of Access: 15 August 2011.

DONOHOE, M. 2008. What the credit crunch means for South Africa. http://www.tradeinvestsa.co.za/feature_articles/843629.htm. Date of access: 10 May 2009.

DURAND, D. 1941. Risk elements in Consumer Installment financing.New York: National Bureau of Economic Research. 128 p.

EMBER, S. 2009. How a Housing Bubble Burst Into a Global Financial Crisis. Voice of America radio broadcast (3 Aug 2009). http://www.voanews.com/learningenglish/home/a-23-2009-08-03-voa2-83140657.html. Date of Access: 25 April 2010.

ESTES, D.A. 2006. Understanding the Business Credit Scoring Process. http://www.allbusiness.com/finance/business-loans-business-credit/4103226-1.html. Date of Access: 20 May 2011.

FEASIBILITY. 2009. The National Credit Act (NCA) - one year later consumers see benefits. www.feasibility.co.za. Date of access: 10 June 2010.

FERGUSON, N. 2008. The Ascent of Money: Episode 1. 35mm. British Broadcasting Corporation (BBC). UK.

FINMARK TRUST. 2011. The Financial Vulnerability of South African Consumers during the Second Quarter of 2011. http://www.finmarktrust.org.za/pages/Research-and-Publications/Consumer-Financial-Vulnerability-Index.aspx?randomID=d4d5f722-3256-40a6-a2a5-cf6419754c26&linkPath=7&lID=7_4. Date of Access: 10 October 2011.

FISHER, R.A. 1936. The use of multiple measurements in taxonomic problems. Ann. Eugenics, 7, 179-188.

FREDDIE MAC. 2011. Primary Mortgage Market Survey Archives. http://www.freddiemac.com. Date of Access: 15 August 2011.

GARRIGA, C., GAVIN, W.T. & SCHLAGENHAUF, D. 2006. Recent Trends in Homeownership, *Federal Reserve Bank of St. Louis Review*, 88(5): 397-411, September/October 2006.

GREENSPAN, A. 2007. The Age of Turbulence: Adventures in a New World. New York: Penguin. 231 p.

GRUENSTEIN, D. & ZHAI, R. 2005. Borrowers in higher minority areas more likely to receive prepayment penalties on subprime loans. http://www.responsiblelending.org/mortgage-lending/research-analysis/borrowers-in-higher-minority-areas-more-likely-to-receive-prepayment-penalties-on-subprime-loans.html. Date of Access: 20 April 2010.

HARRISON, E. 2011. The Sovereign Debt Crisis and Currency Sovereignty. http://www.creditwritedowns.com/2011/07/sovereign-debt-crisis.html. Date of Access: 20 August 2011.

HIGH, R. 2000. Dealing with 'Outliers': How to Maintain Your Data's Integrity. http://cc.uoregon.edu/cnews/spring2000/outliers.html. Date of access: 5 May 2011.

HUNT, J.P. 2008. Credit Rating Agencies and the 'Worldwide Credit Crisis': The Limits of Reputation, the Insufficiency of Reform, and a Proposal for Improvement. http://www.law.berkeley.edu/files/RAShortWhitePaperJH090908KTrev-1.pdf. Date of access: 3 June 2009.

IMF. 2009. What the Global Financial Crisis Means for Sub-Saharan Africa. Speech by Takatoshi Kato, Deputy Managing Director, IMF, 12th AU Summit, Addis Ababa, Ethiopia, 3 February 2009.

IMF. 2009. World Economic Outlook: Crisis and Recovery. Washington, DC: IMF Multimedia Services Division. 228 p.

IMF. 2010. Direction of Trade Statistics (DoTS). http://www.imf.org. Date of Access: 15 August 2011.

IMF. 2010. World Economic Outlook: Rebalancing Growth. Washington, D.C.: IMF Multimedia Services Division. 195 p.

IMF. 2011. IMF eLibrary Data - International Financial Statistics Public Data. http://elibrary-data.imf.org/DataReport.aspx?c=1449326&d=33061&e=169393. Date of Access: 15 August 2011.

IMF. 2011. World Economic Outlook. Washington, DC.: IMF Multimedia Services Division. 230 p.

Inside Mortgage Finance. 2007. The 2007 Mortgage Market Statistical Annual. http://www.insidemortgagefinance.com/catalog/newsletters/.

INVESTOPEDIA. 2011. Mortgage-Backed Security (MBS). http://www.investopedia.com. Date of Access: 20 August 2011.

JARA, A., MORENO, R. & TOVAR, C.E. 2009. The global crisis and Latin America: Financial impact and policy responses. BIS Quarterly Review (June 2009). Basel, Switzerland: Bank of International Settlements. 16 p.

JELAŠIĆ, R. 2009. Effects of the financial crisis on Central and Eastern Europe. (Speech by Mr Radovan Jelašić, Governor of the National Bank of Serbia, at the "Global Crisis in Europe and Central Asia" panel discussion of the IMF and WB Spring Meeting, Washington, 24 April 2009) http://www.bis.org/review/r090428d.pdf. Date of access: 10 April 2010.

KAZADI, M.L. 2011 Public perceptions of the impact of the global financial crisis on the South African economy. CPUT Theses & Dissertations. Paper 380. 286p.

KUTNER, M.H., NACHTSHEIM, C.J., NETER, J. & LI, W. 2005. Applied Linear Statistical Models. 5th ed. Columbus, OH: McGraw-Hill Irwin. 1396 p.

LARRIAN, G., REISEN, H. & VON MALTZAN, J. 1997. Emerging market risk And Sovereign Credit Ratings. http://fiordiliji.sourceoecd.org/vl=17153162/cl=19/nw=1/rpsv/cgi-bin/wppdf?file=5lgsjhvj7ctc.pdf. Date of Access: 3 June 2009.

LEWIS, E.M. 1992. An introduction to credit scoring. Athena Press: San Rafael, Calif.

LIBERTA. 2011. Prime Interest Rate in South Africa. http://liberta.co.za/blog/prime-interest-rate-in-south-africa-current-and-historical/. Date of Access: 1 November 2011.

LIN, J. 2008. The Impact of the Financial Crisis on Developing Countries, World Bank Draft, Washington D.C.:World Bank.

LONGLEY, R. 2009. The Economic Stimulus Package - How it Could Affect you. http://usgovinfo.about.com/od/moneymatters/a/ecstimhub.htm. Date of access: 4 September 2009.

MARKETWATCH, 2011. Senate approves debt-ceiling legislation. http://www.marketwatch.com. Date of Access: 20 August 2011.

MARQUEZ, J. 2008. An Introduction to Credit Scoring For Small and Medium Size Enterprises. Washington, DC: World Bank. 47 p.

MAYS, E. 2001. Handbook of Credit Scoring. Middelsex, UK: Global Professional Publishing. 382 p.

MESTER, L.J. 1997. What's the Point of Credit Scoring?. https://www.phil.frb.org/research-and-data/publications/business-review/1997/september-october/brso97lm.pdf. Date of access: 3 June 2009.

MILLS, S. 2008. How the credit crisis affects your finances. http://money.ninemsn.com.au/article.aspx?id=447851. Date of access: 15 March 2009.

MISHKEN, F.S. 1991. Anatomy of a Financial Crisis. Cambridge, MA.: National Bureau of Economic Research. (Working Paper 3934) 24 p.

MONEYWEB. 2011. Food, fuel costs push inflation higher. http://moneyweb.co.za/mw/view/mw/en/page295043?oid=530323&sn=2009+Detail&pid=294690. Date of Access: 1 March 2011.

MUCHA, T. 2009. A World of Trouble: Is the nightmare over? Online article from the global post. http://www.globalpost.com/dispatch/commerce/090908/is-the-global-recession-over. Date of Access: 20 June 2009.

NCR. 2006. National Credit Act Implementation Schedule. http://www.ncr.org.za/index.php?option=com_content&view=article&id=44.Date of Access: 10 November 2011.

NCR. 2009. Impact of Financial Crisis & NCR initiatives. http://www.ncr.org.za. Date of access: 10 June 2010.

NCR. 2011. Consumer Credit Market Report: First Quarter (March 2011). Midrand, South Africa: National Credit Regulator. 27 p.

NEDBANK. 2007. About the NCA. http://www.nedbank.co.za/website/content/nca/. Date of Access: 25 May 2010.

NEWBOLD, P., CARLSON, W. & THORNE, B. 2009. Statistics for Business and Economics. 7th ed. Canada: Pearson Education. 963 p.

NEW YORK TIMES. 2010. Economic Stimulus (Jobs Bills). http://www.nytimes.com.

Date of Access: 15 August 2011.

OANDA. 2011. Historical Exchange Rates. http://www.oanda.com/currency/historical-rates/. Date of Access: 18 August 2011.

PERRY, J. & WALKER, M. 2009. Central Banks Cut Rates: ECB, Bank of England Near Zero; U.K. to pump US$105.5 billion into Economy. Belgium, Brussels: Wall Street Journal (Europe).

RAPPEPORT, A. 2009. US economy shrinks at rate of 6.1%. http://www.ft.com/cms/s/0/06299c8e-3437-11de-9eea-00144feabdc0.html#axzz1UbW6i0Ii. Date of access: 25 August 2009.

ROONEY, B. 2011. EU unveils Greek rescue and expands bailout fund. http://money.cnn.com. Date of Access: 20 August 2011.

SAINFO. 2009. South Africa can tough out the crisis. http://www.southafrica.info/business/economy/policies/economy-301090.htm. Date of access: 25 August 2009.

SARB **see** SOUTH AFRICAN RESERVE BANK.

SAS. 2003. Data mining using SAS Enterprise Miner: a case study approach. Cary, N.C.: SAS Institute Inc. 49p.

SEAGER, A. 2009. UK budget deficit hits record £90bn. http://www.guardian.co.uk

Date of Access: 15 April 2010.

SCALLAN, G. 2007. Gini Coefficient Technical Review. Score Plus. http://www.scoreplus.com. Date of Access: 10 November 2011. 15p.

SCHMIDT, D. 2009. The Financial Crisis and Its Impact on China. China analysis 67. Germany: Trier University.

SCOTT, M. (2009), Europe's Banking Blues, Business Week Online. http://www.businessweek.com/globalbiz/content/jan2009/gb20090121_513375.htm. Date of Access: 25 July 2010.

SHELLHAMMER, G. 2009. The Financial Crisis Today, A “Recovery”? Far from it! http://www.globalcrisisnews.com/comments/the-financial-crisis-today-a-recovery-far-from-it/id=1030/. Date of access: 10 September 2009.

SHILLER, R.J. 2008. The Subprime Solution: How Today’s Global Financial Crisis Happened, and What to Do about It. Princetion, NJ: Princeton University Press. 136 p.

SIDDIQI, N. 2006. Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring. Hoboken, N.J.: John Wiley & Sons. 196 p.

SOUTH AFRICAN RESERVE BANK. Bank Supervision Department. 2008. Annual Report. Pretoria. 108 p.

STEWART, H. 2008. We are in the worst financial crisis since Depression, says IMF. http://www.guardian.co.uk/business/2008/apr/10/useconomy.subprimecrisis. Date of Access: 15 August 2011.

STEWART, R. M. 2008. ABSA to buy control of Woolworths financial services. www.marketwatch.com. Date of Access: 20 July 2011

SUZMAN, H. 2011. South Africa's Debt Trap – Indebtedness. http://www.hsf.org.za/past-events/quarterly-roundtable-series/quarterly-roundtable-series-south-africas-debt-trap. Date of Access: 15 August 2011.

TAM, Y.M. 2009. What Does Foreclosure Entail? Economic Information Newsletter. Federal Reserve Bank of St. Louis. http://liber8.stlouisfed.org/newsletter/2009/200904.pdf. Date of Access: 25 April 2010.

TEMKIN, K., JOHNSON, J.E.H. & LEVY, D. 2002. Subprime Markets, the Role of GSEs, and Risk-Based Pricing. Washington, DC: U.S. Department of Housing and Urban Development.

TOPÇU, N. 2009. The Global Financial Crisis and Its Effect to Eastern Europe Countries. Department of Economics. Eskişehir: Anadolu University. 20 p.

TRADING ECONOMICS. 2011. China Stock Market Index. http://www.tradingeconomics.com/china/stock-market. Date of Access: 18 August 2011.

TURNHAM, J., HERBERT, C., NOLDEN, S., FEINS, J. & BONJOUNI, J. 2004. Study of Homebuyer Activity through the HOME Investment Partnerships Program. http://www.huduser.org/publications/pdf/Homebuy.pdf. Date of Access: 20 April 2010.

UBS. 2007a. Mortgage Strategist. (26 June 2007 issue). Basel, Switzerland: UBS Investment Research . http://www.ubs.com. Date of Access 20 April 2010.

UBS. 2007b. Mortgage Strategist. (23 October 2007 issue). Basel, Switzerland: UBS Investment Research . http://www.ubs.com. Date of Access 20 April 2010.

UNCTAD. 2009. As financial and economic crisis begins to bite, developing countries ponder ways to support each other. Press Release (04/02/2009). http://www.unctad.org/templates/webflyer.asp?docid=10995&intItemID=1528&lang=1. Date of Access: 16 April 2010.

UNENF. 2010. World Financial Crisis. http://www.unenf.org. Date of Access: 16 April 2010.

UNITED NATIONS. 2009. The global financial crisis: impact, responses and way forward. Cairo, Egypt: United Nations. 15 p.

US CENSUS. 2010. 2010 Census Data. http://2010.census.gov/2010census/data/. Date of Access: 15 August 2011.

WATTS, A. 2010. Europe Agrees 500bn Euro Rescue Package. http://news.sky.com/skynews/Home/World-News/Europe-Agrees-500bn-Euro-Aid-Package-To-Stabilise-Greece-Crisis/Article/201005215628704?lpos=World_News_Third_Home_Page_Article_Teaser_Region__2&lid=ARTICLE_15628704_Europe_Agrees_500bn_Euro_Aid_Package_To_Stabilis.Date of access: 11 May 2010.

WORLD BANK, 2009. Impact of Global Financial Crisis on South Asia. http://search.worldbank.org. Date of Access: 4 April 2010.

WORLD BANK. 2011. World Bank, World Development Indicators. http://www.worldbank.org. Date of Access: 15 August 2011.

WHITLOCK, C. 2009. Germany's Financial Turnabout. http://www.washingtonpost.com/wp-yn/content/article/2009/02/25/AR2009022503354.html. Date of Access: 20 April 2010.

WIZARD. 2007. Everything you need to know about the NCA. http://www.wizardmidrand.com/home-loans/nca/nca-faq.html. Date of Access: 10 May 2009.

WOOLWORTHS. Our History. http://www.woolworths.co.za. Date of Access: 20 July 2011.

WTO. 2010. World Trade Organisation: International Trade Statistics 2010. Geneva, Switzerland: WTO Publications. 250 p.

ZINI, M. 2008. The Impact of the Financial Crisis on South Africa. http://africacan.worldbank.org/the-impact-of-the-financial-crisis-on-south-africa. Date of access: 5 May 2009.

ZUCKERMAN, M. 2007. The Credit Crisis Grows. http://www.usnews.com/articles/opinion/mzuckerman/2007/12/13/the-credit-crisis-grows.html. Date of access: 3 March 2009.

ZUMBRUN, J. 2010. Greenspan Says Crisis 'By Far' Worst, Recovery Uneven. http://www.bloomberg.com/apps/news?pid=newsarchive&sid=a1aLQ51QXlDA&pos=3. Date of Access: 1 November 2011.

Annexes

A1: Regression coefficients for XYZ Bank example

Variable	ClassVal0	DF	Estimate	StdErr	WaldChiSq	ProbChiSq
Intercept		1	1.53	0.87	3.09	0.08
Age	from 25 to 30	1	-0.37	0.19	3.70	0.05
	from 30 to 40	1	-0.30	0.16	3.52	0.06
	from 40 to 50	1	-0.32	0.15	4.46	0.03
	from 50 to 55	1	-0.34	0.18	3.50	0.06
	less than 25	1	-0.40	0.24	2.87	0.09
	more than 55	0	0.00			
Gender	female	1	-0.03	0.10	0.09	0.76
	male	0	0.00			
PTI	40% and greater	1	-1.52	0.17	84.80	0.00
	from 20 to 25%	1	-0.38	0.16	5.80	0.02
	from 25 to 35%	1	-0.98	0.14	46.25	0.00
	from 35 to 40%	1	-0.85	0.19	20.63	0.00
	less than 20%	0	0.00			
Marital_Status	Married	1	0.51	0.20	6.61	0.01
	Single	1	0.53	0.20	6.66	0.01
	no information	0	0.00			
LastWorkRecord	10 and greater	1	-0.66	0.16	17.76	0.00
	from 1 to 3	1	-1.28	0.32	16.08	0.00
	from 3 to 5	1	-0.81	0.24	11.03	0.00
	from 5 to 10	1	-0.99	0.18	29.54	0.00
	less than 1	1	-1.12	0.28	15.84	0.00
	no information	0	0.00			
Children	None	1	-0.42	0.14	8.31	0.00
	One	1	0.41	0.19	4.73	0.03
	Two and more	1	0.93	0.25	13.71	0.00
	no information	0	0.00			
Cards	None	1	-2.07	0.21	96.19	0.00
	credit card	1	-0.66	0.23	8.12	0.00
	paycard	0	0.00			
Credit_History	None	1	2.48	0.79	9.92	0.00
	average	1	2.11	0.86	6.01	0.01
	positive	1	4.88	0.86	32.16	0.00
	with negative factors	0	0.00			

Variable	ClassVal0	DF	Estimate	StdErr	WaldChiSq	ProbChiSq
Home_Ownership	Own	1	-0.83	0.31	7.20	0.01
	Rent	1	-0.72	0.32	5.01	0.03
	no information	0	0.00			
Time_At_Branch	10 and greater	1	0.89	0.18	24.27	0.00
	from 1 to 3	1	0.06	0.27	0.04	0.83
	from 3 to 5	1	0.32	0.22	2.10	0.15
	from 5 to 10	1	0.53	0.21	6.42	0.01
	less than 1	1	0.11	0.18	0.41	0.52
	no information	0	0.00			

A2: Data normalisation for XYZ Bank example

Variable	ClassVal0	Estimate	min_est	counter	est1	max_est1	sum_max	max_cat	score
Age	less than 25	-0.40	-0.40	6	0.00	0.40	13.8	0	0
	from 50 to 55	-0.34	-0.40	6	0.06	0.40	13.8	0	4
	from 30 to 40	-0.30	-0.40	6	0.10	0.40	13.8	0	7
	from 25 to 30	-0.37	-0.40	6	0.02	0.40	13.8	0	2
	from 40 to 50	-0.32	-0.40	6	0.08	0.40	13.8	0	6
	more than 55	0.00	-0.40	6	0.40	0.40	13.8	1	29
Cards	credit card	-0.66	-2.07	3	1.41	2.07	13.8	0	102
	paycard	0.00	-2.07	3	2.07	2.07	13.8	1	150
	None	-2.07	-2.07	3	0.00	2.07	13.8	0	0
Children	no information	0.00	-0.42	4	0.42	1.35	13.8	0	30
	Two and more	0.93	-0.42	4	1.35	1.35	13.8	1	98
	One	0.41	-0.42	4	0.82	1.35	13.8	0	60
	None	-0.42	-0.42	4	0.00	1.35	13.8	0	0
Credit_History	None	2.48	0.00	4	2.48	4.88	13.8	0	180
	with negative factors	0.00	0.00	4	0.00	4.88	13.8	0	0
	average	2.11	0.00	4	2.11	4.88	13.8	0	153
	positive	4.88	0.00	4	4.88	4.88	13.8	1	354
Home_Ownership	Rent	-0.72	-0.83	3	0.11	0.83	13.8	0	8
	Own	-0.83	-0.83	3	0.00	0.83	13.8	0	0
	no information	0.00	-0.83	3	0.83	0.83	13.8	1	60
LastWorkRecord	no information	0.00	-1.28	6	1.28	1.28	13.8	1	93
	from 5 to 10	-0.99	-1.28	6	0.29	1.28	13.8	0	21
	from 3 to 5	-0.81	-1.28	6	0.47	1.28	13.8	0	34
	from 1 to 3	-1.28	-1.28	6	0.00	1.28	13.8	0	0
	10 and greater	-0.66	-1.28	6	0.62	1.28	13.8	0	45
	less than 1	-1.12	-1.28	6	0.16	1.28	13.8	0	12
Marital_Status	Single	0.53	0.00	3	0.53	0.53	13.8	1	38
	Married	0.51	0.00	3	0.51	0.53	13.8	0	37
	no information	0.00	0.00	3	0.00	0.53	13.8	0	0
PTI	less than 20%	0.00	-1.52	5	1.52	1.52	13.8	1	110
	from 35 to 40%	-0.85	-1.52	5	0.67	1.52	13.8	0	49
	from 25 to 35%	-0.98	-1.52	5	0.55	1.52	13.8	0	40
	from 20 to 25%	-0.38	-1.52	5	1.14	1.52	13.8	0	83
	> = 40%	-1.52	-1.52	5	0.00	1.52	13.8	0	0
Gender	female	-0.03	-0.03	2	0.00	0.03	13.8	0	0
	male	0.00	-0.03	2	0.03	0.03	13.8	1	2

Variable	ClassVal0	Estimate	min_est	counter	est1	max_est1	sum_max	max_cat	score
Time_At_Branch	no information	0.00	0.00	6	0.00	0.89	13.8	0	0
	less than 1	0.11	0.00	6	0.11	0.89	13.8	0	8
	from 5 to 10	0.53	0.00	6	0.53	0.89	13.8	0	38
	from 3 to 5	0.32	0.00	6	0.32	0.89	13.8	0	23
	from 1 to 3	0.06	0.00	6	0.06	0.89	13.8	0	4
	10 and greater	0.89	0.00	6	0.89	0.89	13.8	1	65

B1: Statistics and outliers (June 2006 data)

Variable: SM_Open_Bal_200606 (sm_open_bal)

Moments			
N	1274922	**Sum Weights**	1274922
Mean	2257.73275	**Sum Observations**	2878433152
Std Deviation	2825.13934	**Variance**	7981412.28
Skewness	-2.5160079	**Kurtosis**	5835.06602
Uncorrected SS	1.66744E13	**Corrected SS**	1.01757E13
Coeff Variation	125.131699	**Std Error Mean**	2.50206199

Basic Statistical Measures				Tests for Location: μ_0=0				
Location		**Variability**		**Test**	**Statistic**		**p Value**	
Mean	2257.7	Std dev	2825	Student s t	t	902.4	Pr > |t|	<.0001
Median	1432.0	Variance	7981412	Sign	M	558426.5	Pr >= |M|	<.0001
Mode	0.0	Range	1302441	Signed rank	S	3.923E11	Pr >= |S|	<.0001
		Interquartile Range	2391					

Quantiles (Definition 5)				Extreme Observations			
Quantile	**Estimate**	**Quantile**	**Estimate**	**Lowest**		**Highest**	
100% Max	603663.33	25% Q1	560.94	**Value**	**Obs**	**Value**	**Obs**
99%	13047.43	10%	24.91	-698777	1.1E6	46748.7	1.06E6
95%	7712.06	5%	-0.12	-442941	905449	49993.7	786891
90%	5485.35	1%	-104.73	-228570	892009	61036.3	168531
75% Q3	2952.26	0% Min	-698777.35	-202915	1.13E6	490817.4	1.19E6
50% Median	1431.99			-113643	1.03E6	603663.3	461993

Variable: credit_limit_200606 (sm_credit_limit)

Moments			
N	1274922	**Sum Weights**	1274922
Mean	5626.53577	**Sum Observations**	7173394236
Std Deviation	3961.45946	**Variance**	15693161
Skewness	0.92476011	**Kurtosis**	0.34668411
Uncorrected SS	6.03689E13	**Corrected SS**	2.00075E13
Coeff Variation	70.4067231	**Std Error Mean**	3.50843478

Basic Statistical Measures				Tests for Location: $\mu_0=0$				
Location		**Variability**		**Test**	**Statistic**		**p Value**	
Mean	5626	Std dev	3961	Students t	t	1603	Pr > |t|	<.0001
Median	4700	Variance	15693161	Sign	M	632816	Pr >= |M|	<.0001
Mode	15000	Range	80000	Signed rank	S	4.01E+11	Pr >= |S|	<.0001
		Interquartile Range	5600					

Quantiles (Definition 5)				Extreme Observations			
Quantile	**Estimate**	**Quantile**	**Estimate**	**Lowest**		**Highest**	
100% Max	80000	**25% Q1**	2400	**Value**	**Obs**	**Value**	**Obs**
99%	15000	10%	1500	0	1.27E+06	50000	1.22E+06
95%	15000	5%	1000	0	1.27E+06	50000	1.22E+06
90%	11850	1%	500	0	1.27E+06	60000	1.18E+06
75% Q3	8000	0% Min	0	0	1.27E+06	70000	1.08E+06
50% Median	4700			0	1.27E+06	80000	1.26E+06

Variable: SM_Purchases_Val_200606 (sm_purchases_val)

Moments			
N	1274922	**Sum Weights**	1274922
Mean	341.744569	**Sum Observations**	435697670
Std Deviation	667.44719	**Variance**	445485.751
Skewness	12.4314516	**Kurtosis**	909.377708
Uncorrected SS	7.16856E11	**Corrected SS**	5.67959E11
Coeff Variation	195.305866	**Std Error Mean**	0.59111925

Basic Statistical Measures				Tests for Location: μ_0=0				
Location		**Variability**		**Test**	**Statistic**		**p Value**	
Mean	341.7	Std Deviation	667.4	Students t	t	578.1314	Pr > |t|	<.0001
Median	0	Variance	445486	Sign	M	317135.5	Pr >= |M|	<.0001
Mode	0	Range	92739	Signed Rank	S	1.01E+11	Pr >= |S|	<.0001
		Interquartile Range	451.6					

Quantiles (Definition 5)				Extreme Observations			
Quantile	**Estimate**	**Quantile**	**Estimate**	**Lowest**		**Highest**	
100% Max	92739.26	25% Q1	0	Value	Obs	Value	Obs
99%	2843.34	10%	0	0	1.27E+06	60042	174082
95%	1506.56	5%	0	0	1.27E+06	61962	1.17E+06
90%	1025.43	1%	0	0	1.27E+06	65313	1.03E+06
75% Q3	451.58	0% Min	0	0	1.27E+06	85366	1.25E+06
50% Median	0			0	1.27E+06	92739	16561

Variable: SM_Payments_Val_200606 (sm_payments_val)

Moments			
N	1274922	**Sum Weights**	1274922
Mean	346.75774	**Sum Observations**	442089071
Std Deviation	942.220889	**Variance**	887780.204
Skewness	127.594133	**Kurtosis**	47676.2143
Uncorrected SS	1.28515E12	**Corrected SS**	1.13185E12
Coeff Variation	271.723103	**Std Error Mean**	0.83447037

Basic Statistical Measures				Tests for Location: □$_0$=0				
Location		**Variability**		**Test**	**Statistic**		**p Value**	
Mean	346.8	Std Deviation	942.2	Students t	t	415.5	Pr > |t|	<.0001
Median	200	Variance	887780	Sign	M	472432.5	Pr >= |M|	<.0001
Mode	0	Range	453259	Signed Rank	S	2.23E+11	Pr >= |S|	<.0001
		Interquartile Range	400					

Quantiles (Definition 5)				Extreme Observations			
Quantile	**Estimate**	**Quantile**	**Estimate**	**Lowest**		**Highest**	
100% Max	453259	25% Q1	0	Value	Obs	Value	Obs
99%	2764.72	10%	0	0	1.27E+06	133665	1.20E+06
95%	1155.38	5%	0	0	1.27E+06	140206	481070
90%	800	1%	0	0	1.27E+06	140597	1.18E+06
75% Q3	400	0% Min	0	0	1.27E+06	233130	935209
50% Median	200			0	1.27E+06	453259	1.19E+06

Variable: curr_due_val_200606 (curr_due_val)

Moments			
N	1274922	**Sum Weights**	1274922
Mean	231.76255	**Sum Observations**	295479173
Std Deviation	274.745026	**Variance**	75484.8295
Skewness	19.4734503	**Kurtosis**	4264.24975
Uncorrected SS	1.64718E11	**Corrected SS**	9.62372E10
Coeff Variation	118.545911	**Std Error Mean**	0.24332573

Basic Statistical Measures				Tests for Location: $\mu_0=0$				
Location		**Variability**		**Test**	**Statistic**		**p Value**	
Mean	231.8	Std Deviation	274.7	Students t	t	952.5	Pr > |t|	<.0001
Median	147.2	Variance	75485	Sign	M	589632.5	Pr >= |M|	<.0001
Mode	0	Range	72787	Signed Rank	S	3.48E+11	Pr >= |S|	<.0001
		Interquartile Range	242.9					

Quantiles (Definition 5)				Extreme Observations			
Quantile	**Estimate**	**Quantile**	**Estimate**	**Lowest**		**Highest**	
100% Max	72768.4	25% Q1	58.03	Value	Obs	Value	Obs
99%	1321.84	10%	14.3	-18.26	246758	6992.8	40910
95%	782.56	5%	0	0	1.27E+06	12896	944773
90%	558.17	1%	0	0	1.27E+06	14867	168531
75% Q3	300.89	0% Min	-18.26	0	1.27E+06	42473	1.19E+06
50% Median	147.21			0	1.27E+06	72768	461993

Variable: bureau_score_200606 (bureau_score)

Moments			
N	1274922	**Sum Weights**	1274922
Mean	682.372427	**Sum Observations**	869971619
Std Deviation	102.201488	**Variance**	10445.1441
Skewness	-0.0627114	**Kurtosis**	-0.9618842
Uncorrected SS	6.06961E11	**Corrected SS**	1.33167E10
Coeff Variation	14.9773765	**Std Error Mean**	0.09051393

Basic Statistical Measures				Tests for Location: $\mu_0=0$				
Location		**Variability**		**Test**	**Statistic**		**p Value**	
Mean	682.4	Std Deviation	102.2	Students t	t	7538.9	Pr > |t|	<.0001
Median	683	Variance	10445	Sign	M	637461	Pr >= |M|	<.0001
Mode	826	Range	534	Signed Rank	S	4.06E+11	Pr >= |S|	<.0001
		Interquartile Range	169					

Quantiles (Definition 5)				Extreme Observations			
Quantile	**Estimate**	**Quantile**	**Estimate**	**Lowest**		**Highest**	
100% Max	883	25% Q1	599	Value	Obs	Value	Obs
99%	869	10%	546	349	747352	883	1.26E+06
95%	842	5%	520	365	1.05E+06	883	1.26E+06
90%	820	1%	478	365	803035	883	1.26E+06
75% Q3	768	0% Min	349	365	704323	883	1.27E+06
50% Median	683			365	331072	883	1.27E+06

B2: SAS Binning Code

```
/*Open Balance smaller than 0 *
DATA Chp4.Data2006_OpenBal_bin1;
SET Chp4.Data2006out;
IF sm_open_bal_200606 < 0 then output;
RUN;
/*Open Balance between 0 - 300 */
DATA Chp4.Data2006_OpenBal_bin2;
SET Chp4.Data2006out;
IF sm_open_bal_200606 => 0 and sm_open_bal_200606 <= 300 then output;
RUN;
/*Open Balance between 300 - 1000 */
DATA Chp4.Data2006_OpenBal_bin3;
SET Chp4.Data2006out;
IF sm_open_bal_200606 > 300 and sm_open_bal_200606 <= 1000 then output; RUN;
/*Open Balance between 1000 - 2500 */
DATA Chp4.Data2006_OpenBal_bin4;
SET Chp4.Data2006out;
IF sm_open_bal_200606 > 1000 and sm_open_bal_200606 <= 2500 then output; RUN;
/*Open Balance between 2500 - 5000 */
DATA Chp4.Data2006_OpenBal_bin5;
SET Chp4.Data2006out;
IF sm_open_bal_200606 > 2500 and sm_open_bal_200606 <= 5000 then output;RUN;
/*Open Balance between 5000 - 8000 */
DATA Chp4.Data2006_OpenBal_bin6;
SET Chp4.Data2006out;
IF sm_open_bal_200606 > 5000 and sm_open_bal_200606 <= 8000 then output; RUN;
/*Open Balance larger than 8000 */
DATA Chp4.Data2006_OpenBal_bin7;
SET Chp4.Data2006out;
IF sm_open_bal_200606 > 8000 then output;
RUN;
```

B3: Basic Statistics and Outliers (July 2009 Data)

Variable: SM_Open_Bal_200907 (sm_open_bal)

Moments			
N	1483179	**Sum Weights**	1483179
Mean	2646.20878	**Sum Observations**	3924801286
Std Deviation	3122.13186	**Variance**	9747707.38
Skewness	-4.9713388	**Kurtosis**	1768.20462
Uncorrected SS	2.48434E13	**Corrected SS**	1.44576E13
Coeff Variation	117.985092	**Std Error Mean**	2.56362476

Basic Statistical Measures				Tests for Location: $\mu_0=0$				
Location		**Variability**		**Test**	**Statistic**		**p Value**	
Mean	2646.2	Std Deviation	3122	Students t	t	1032.2	Pr > |t|	<.0001
Median	1660.8	Variance	9747707	Sign	M	629867	Pr >= |M|	<.0001
Mode	0	Range	1025965	Signed Rank	S	5.31E+11	Pr >= |S|	<.0001
		Interquartile Range	2722					

Quantiles (Definition 5)				Extreme Observations			
Quantile	**Estimate**	**Quantile**	**Estimate**	**Lowest**		**Highest**	
100% Max	332616.8	25% Q1	701.06	Value	Obs	Value	Obs
99%	14478.92	10%	17.5	-693348.6	1.30E+06	54755	654139
95%	9380.83	5%	-0.83	-202866.2	1.33E+06	83287	1.27E+06
90%	6558.42	1%	-143.15	-79982.5	1.10E+06	95004	5339
75% Q3	3423.19	0% Min	-693348.6	-79572.3	179062	104902	477396
50% Median	1660.8			-47978.1	1.28E+06	332617	1.39E+06

Variable: credit_limit_200907 (sm_credit_limit)

Moments			
N	1483179	**Sum Weights**	1483179
Mean	5322.99219	**Sum Observations**	7894950235
Std Deviation	4124.00124	**Variance**	17007386.3
Skewness	1.05325617	**Kurtosis**	0.49557334
Uncorrected SS	6.72497E13	**Corrected SS**	2.5225E13
Coeff Variation	77.4752451	**Std Error Mean**	3.38627328

Basic Statistical Measures				Tests for Location: $\mu_0=0$				
Location		**Variability**		**Test**	**Statistic**		**p Value**	
Mean	5323.0	Std Deviation	4124	Students t	t	1571.9	Pr > |t|	<.0001
Median	4000	Variance	17007386	Sign	M	734396	Pr >= |M|	<.0001
Mode	1000	Range	83968	Signed Rank	S	5.39E+11	Pr >= |S|	<.0001
		Interquartile Range	5700					

Quantiles (Definition 5)				Extreme Observations			
Quantile	**Estimate**	**Quantile**	**Estimate**	**Lowest**		**Highest**	
100% Max	83968	25% Q1	2000	Value	Obs	Value	Obs
99%	15000	10%	1200	0	1.48E+06	50000	1.42E+06
95%	15000	5%	1000	0	1.48E+06	55000	1502
90%	12000	1%	500	0	1.48E+06	70000	1.28E+06
75% Q3	7700	0% Min	0	0	1.48E+06	80000	83855
50% Median	4000			0	1.48E+06	83968	1.27E+06

Variable: SM_Purchases_Val_200907 (sm_purchases_val)

Moments			
N	1483179	**Sum Weights**	1483179
Mean	266.469545	**Sum Observations**	395222033
Std Deviation	605.850361	**Variance**	367054.66
Skewness	13.3499774	**Kurtosis**	1011.03118
Uncorrected SS	6.49722E11	**Corrected SS**	5.44407E11
Coeff Variation	227.361953	**Std Error Mean**	0.49747194

Basic Statistical Measures				Tests for Location: μ_0=0				
Location		**Variability**		**Test**	**Statistic**		**p Value**	
Mean	266.5	Std Deviation	605.9	Students t	t	535.6474	Pr > |t|	<.0001
Median	0	Variance	367055	Sign	M	319009	Pr >= |M|	<.0001
Mode	0	Range	93305	Signed Rank	S	1.02E+11	Pr >= |S|	<.0001
		Interquartile Range	316					

Quantiles (Definition 5)				Extreme Observations			
Quantile	**Estimate**	**Quantile**	**Estimate**	**Lowest**		**Highest**	
100% Max	93304.85	25% Q1	0	Value	Obs	Value	Obs
99%	2588.34	10%	0	0	1.48E+06	58085	1.19E+06
95%	1242.3	5%	0	0	1.48E+06	60224	777395
90%	810.98	1%	0	0	1.48E+06	74120	47280
75% Q3	316	0% Min	0	0	1.48E+06	80238	1.46E+06
50% Median	0			0	1.48E+06	93305	527252

Variable: SM_Payments_Val_200907 (sm_payments_val)

Moments			
N	1483179	**Sum Weights**	1483179
Mean	309.149945	**Sum Observations**	458524707
Std Deviation	602.732047	**Variance**	363285.92
Skewness	10.1021212	**Kurtosis**	214.803987
Uncorrected SS	6.80571E11	**Corrected SS**	5.38818E11
Coeff Variation	194.964306	**Std Error Mean**	0.49491145

Basic Statistical Measures				Tests for Location: □₀=0				
Location		**Variability**		**Test**	**Statistic**		**p Value**	
Mean	309.1	Std Deviation	602.7	Students t	t	624.7	Pr > |t|	<.0001
Median	180	Variance	363286	Sign	M	528913.5	Pr >= |M|	<.0001
Mode	0	Range	44621	Signed Rank	S	2.80E+11	Pr >= |S|	<.0001
		Interquartile Range	377					

Quantiles (Definition 5)				Extreme Observations			
Quantile	**Estimate**	**Quantile**	**Estimate**	**Lowest**		**Highest**	
100% Max	44621.3	25% Q1	0	Value	Obs	Value	Obs
99%	2400	10%	0	0	1.48E+06	30765	6397
95%	1000	5%	0	0	1.48E+06	34171	1.44E+06
90%	700	1%	0	0	1.48E+06	39068	564947
75% Q3	377	0% Min	0	0	1.48E+06	41618	30393
50% Median	180			0	1.48E+06	44621	5339

Variable: curr_due_val_200907 (curr_due_val)

Moments			
N	1483179	**Sum Weights**	1483179
Mean	204.832715	**Sum Observations**	303803581
Std Deviation	228.996737	**Variance**	52439.5056
Skewness	2.15670563	**Kurtosis**	10.0699747
Uncorrected SS	1.40006E11	**Corrected SS**	7.77771E10
Coeff Variation	111.796955	**Std Error Mean**	0.18803232

Basic Statistical Measures				Tests for Location: $\mu_0=0$				
Location		**Variability**		**Test**	**Statistic**		**p Value**	
Mean	204.8	Std Deviation	229.0	Students t	t	1089.3	Pr > |t|	<.0001
Median	130.1	Variance	52440	Sign	M	673986.5	Pr >= |M|	<.0001
Mode	0	Range	11672	Signed Rank	S	4.54E+11	Pr >= |S|	<.0001
		Interquartile Range	207.58					

Quantiles (Definition 5)				Extreme Observations			
Quantile	**Estimate**	**Quantile**	**Estimate**	**Lowest**		**Highest**	
100% Max	11654.1	25% Q1	55.83	Value	Obs	Value	Obs
99%	1093.95	10%	7.69	-18.26	61920	3758.4	1.06E+06
95%	710.62	5%	0	0	1.48E+06	4177.8	654139
90%	498.44	1%	0	0	1.48E+06	6732.6	5339
75% Q3	263.41	0% Min	-18.26	0	1.48E+06	7867.6	477396
50% Median	130.11			0	1.48E+06	11654	1.39E+06

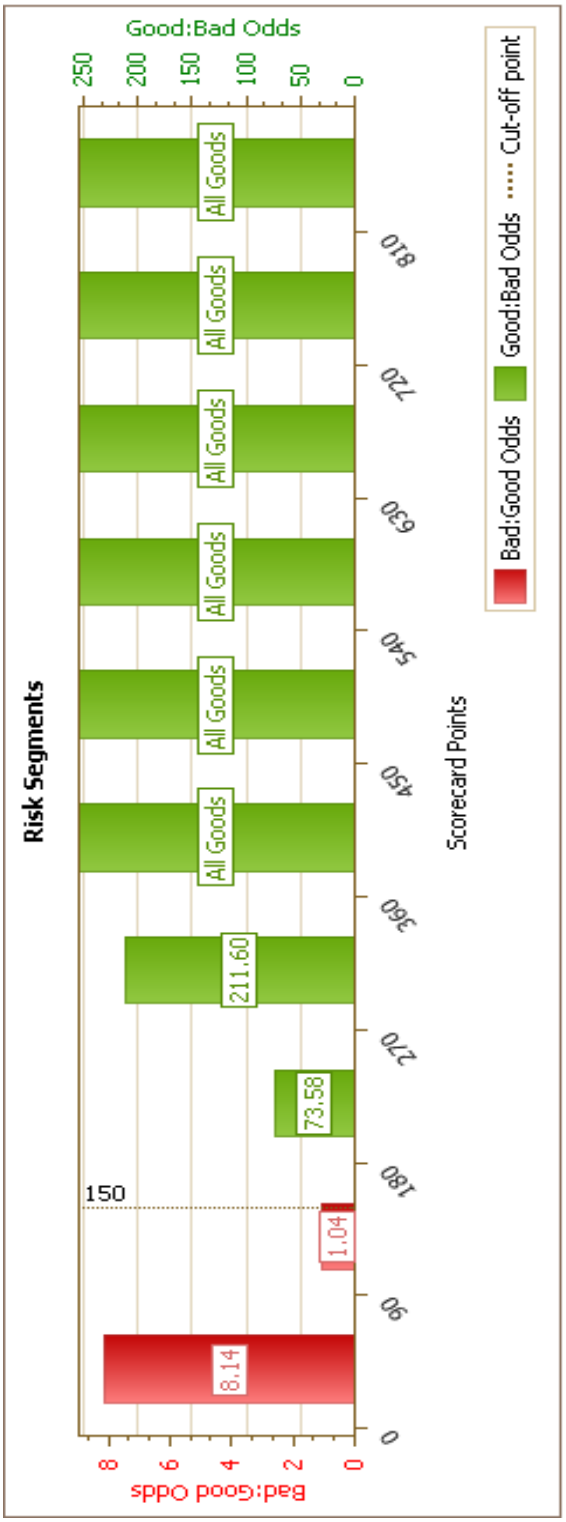
Variable: bureau_score_200907 (bureau_score)

Moments			
N	1483179	**Sum Weights**	1483179
Mean	657.768324	**Sum Observations**	975588165
Std Deviation	109.398657	**Variance**	11968.0661
Skewness	0.07976026	**Kurtosis**	-1.0551727
Uncorrected SS	6.59462E11	**Corrected SS**	1.77508E10
Coeff Variation	16.6317916	**Std Error Mean**	0.08982872

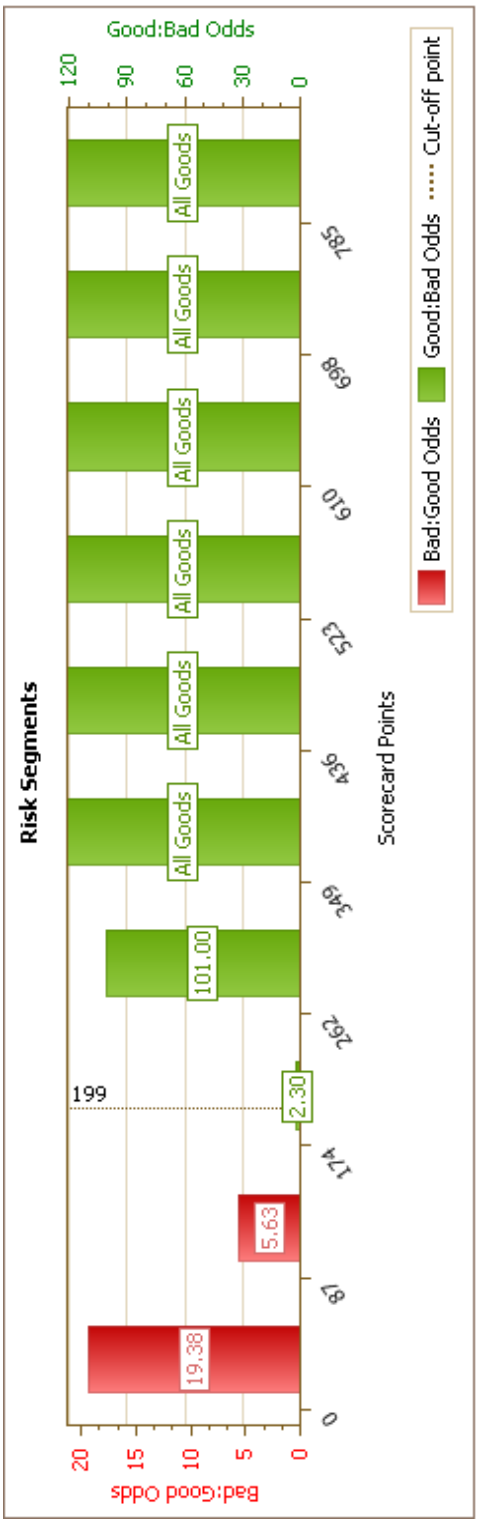
Basic Statistical Measures				Tests for Location: μ_0=0				
Location		**Variability**		**Test**	**Statistic**		**p Value**	
Mean	657.8	Std Deviation	109.4	Students t	t	7322.5	Pr > |t|	<.0001
Median	655	Variance	11968	Sign	M	741589.5	Pr >= |M|	<.0001
Mode	728	Range	542	Signed Rank	S	5.50E+11	Pr >= |S|	<.0001
		Interquartile Range	185					

Quantiles (Definition 5)				Extreme Observations			
Quantile	**Estimate**	**Quantile**	**Estimate**	**Lowest**		**Highest**	
100% Max	883	25% Q1	564	Value	Obs	Value	Obs
99%	869	10%	516	341	799162	883	1.46E+06
95%	834	5%	493	346	1.48E+06	883	1.46E+06
90%	809	1%	457	346	1.39E+06	883	1.47E+06
75% Q3	749	0% Min	341	346	1.16E+06	883	1.47E+06
50% Median	655			346	854270	883	1.47E+06

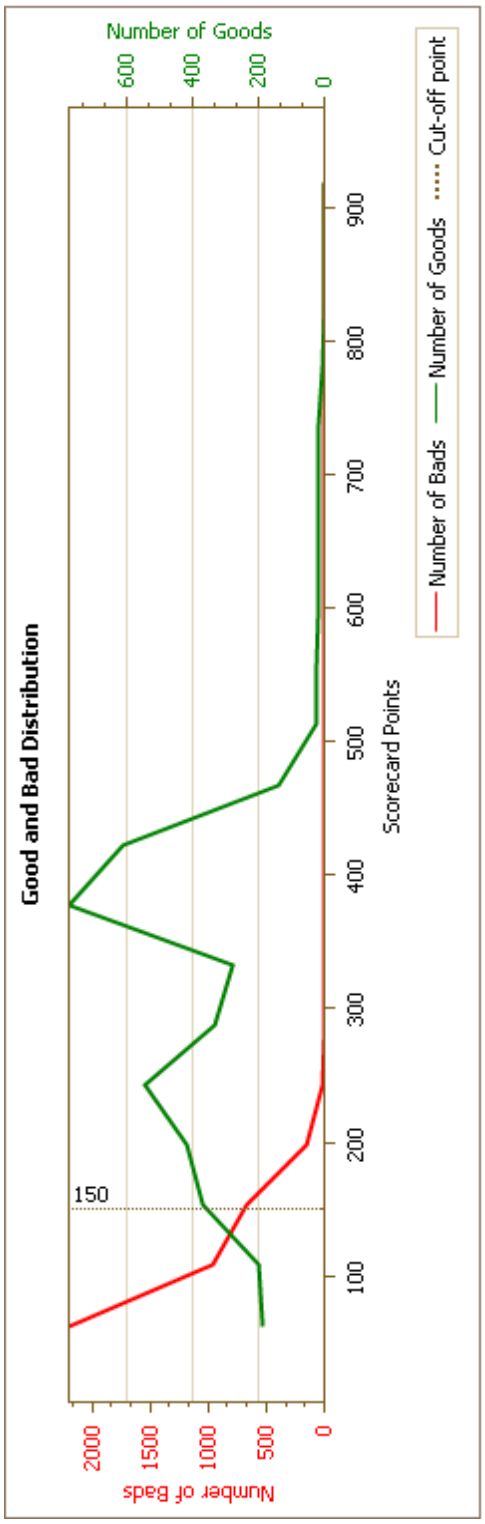
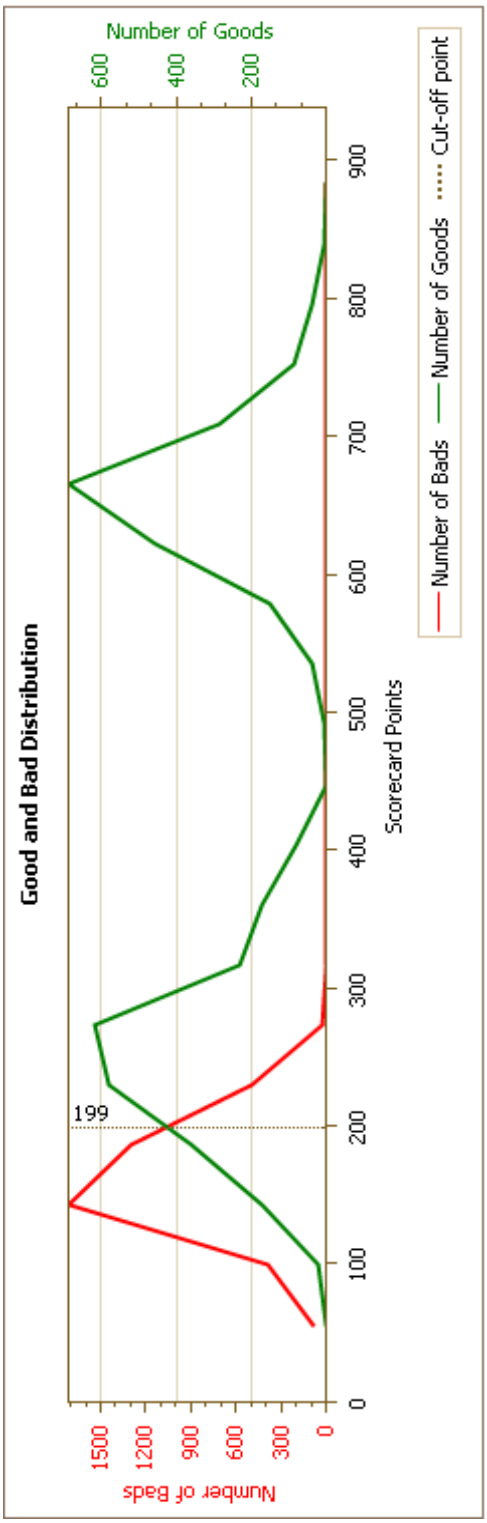
B4: Risk segmentation (2006 data)



B5: Risk segmentation (2009 data)



B6: Good vs. bad distribution (2006 data)



B7: Good vs. bad distribution (2009 data)