AOSIS

An optimised credit scorecard to enhance cut-off score determination

CrossMark
click for updates

Authors:
Nico Kritzinger[1] (iD)
Gary W. van Vuuren[1] (iD)

Affiliations:
[1]Department of Business Mathematics and Informatics (BMI), Faculty of Natural Sciences, North-West University, South Africa

Corresponding author:
Nico Kritzinger,
nkritzinger@mfc.co.za

Background: Credit scoring is a statistical tool allowing banks to distinguish between good and bad clients. However, literature in the world of credit scoring is limited. In this article parametric and non-parametric statistical techniques that are used in credit scoring are reviewed.

Aim: To build an optimal credit scoring matrix model to predict which clients will go bad in the future. This article also illustrates the use of the credit scoring matrix model to determine an appropriate cut-off score on a more granular level.

Setting: Data used in this article are based on a bank in South Africa and are Retail Banking specific.

Methods: The methods used in this article were regression, statistical analysis, matrix and comparative study.

Results: The matrix provides uplift in the Gini-coefficient when compared to a one-dimensional model and provides greater granularity when setting the appropriate cut-off.

Conclusion: The article provides steps to construct a credit scoring matrix model to optimise separation between good and bad clients. An added contribution of the article is the manner in which the credit scoring matrix model provides a greater granularity option for establishing the cut-off score for accepting clients, more appropriately than a one-dimensional scorecard.

Introduction

One of the most important elements driving a bank's existence and continuance is the ability to grant credit to the appropriate client who is less likely to go bad. Two general techniques exist to assess possible future bads; firstly this can be done subjectively by a credit manager or loan officer, and secondly it can be accomplished objectively by means of credit scoring (Allen, De Long & Saunders 2004:734–736). Credit scoring (which surfaced in the mid-1900s) is a statistical tool allowing banks to distinguish between good and bad clients. In banking especially credit scoring has grown in the last 25 years (2015) mainly due to a wider range of banking products and large number of credit applications (Abdou & Pointon 2011:65). The general idea behind credit scoring is to use client characteristics from the past to predict whether the client to be financed will either be good or bad in the future. Credit scoring has proved its worth in credit evaluation; however, the literature has been limited (Abdou & Pointon 2011:60). Various modelling techniques exist to model credit scoring with most of them being statistical; for example, linear regression, discriminant analysis, probit analysis, logistic regression, decision trees, expert systems, neural networks and genetic programming (Abdou & Pointon 2011:66–68). Given this range of statistical techniques, however, no optimal technique for scorecard construction exists (Abdou & Pointon 2011:68).

Various benefits exist for using scorecards for credit assessment; for example, less information is required to make a decision. Credit scoring removes the *bias* by not only looking at accepted applicants, but by considering the entire application population, credit scoring considers characteristics of both good and bad applicants. Credit scorecards are based on large data samples, include legally acceptable characteristics and demonstrate correlation between variables and repayment or bad behaviour. They also include a large number of characteristics, enjoy efficient processing time, minimise process time and cost and produce fewer errors. Credit scorecards are based on real historical data and the interrelation between characteristics is considered (Abdou & Pointon 2011:61–62). Given these benefits, however, various criticisms also exist for using scorecards for credit assessment. For example, no economic factors are considered, there are

misclassification problems, any characteristic is up for consideration, indirect discrimination is possible, no standardisation exists across the market, training analysts is expensive, the final model is statistically 'incomplete' as not all characteristics are part of the final model, the data are historical, characteristics are assumed to be constant over time, and credit scoring imposes a dichotomous outcome (Abdou & Pointon 2011:62–63).

Credit scoring – which allows banks to determine a level of risk of an applicant or borrower – is based on statistical numerical scores to determine whether they would either be a good or bad client in the future (Siddiqi 2006:5). In the credit scoring milieu, two types of credit scoring models have been employed, namely: application scoring and behavioural scoring (Lim & Sohn 2007:427). Application scoring (AS) is used where the scorecard is built for a specific credit organisation by utilising the credit organisation's historical data. It is a credit assessment performed at the application stage. Behavioural scoring (BS) indicates the way a borrower's characteristics of payment behaviour change after the loan is made, and is based on time-dependent characteristics of borrowers (Lim & Sohn 2007:427).

Another form of scoring that exists is the credit bureau score (CBS) which can be described as an economic barometer of the way the borrower performs with other organisations. The CBS is a representation of how borrowers manage their existing credit obligations and the CBS is an external score, with *external* meaning that the scorecard was built by the bureau using data from various credit organisations. As with all scorecards, the credit bureau scoring model is constructed using history, but it does consider the probability of change in behaviour when the score is updated. The model is a dynamic one in this sense, but becomes historic in actual use. One of the key strengths of the CBS is the relatively complete nature of economic information on which it is based; that is, it consists of external risk-related data from diverse businesses contributing information on their own debtors to the credit bureau. In this manner a borrower is assessed on the actual credit performance with all his or her accounts with different institutions such as banks, finance houses, service providers, credit card companies, retailers, public authorities, tax registers, etc. (Anderson 2007:285; TransUnion 2015). Anderson (2007) also indicates the importance of credit bureaus: the facilitation of information gathering from public sources, especially court orders, and the sharing of borrower performance information (Anderson 2007:10). The private nature of this kind of information and the credit bureaus' ability, gives them a monetary incentive to collect, to record and to exchange reliable, valuable and up to date information on all credit performance (TransUnion 2015).

In South Africa there are four major credit bureaus: TransUnion, Experian, Compuscan and Xpert Decision Systems (XDS) which capture, update and store the credit histories of the credit active consumers in South Africa.

The credit bureaus obtain and build their scorecards on data regularly supplied by the credit lenders to the credit bureaus (TransUnion 2015). TransUnion emphasise that to reduce exposure to risk a predictive scoring system is needed (TransUnion & Fair Isaac Corporation [FICO] 2009:1). The bureau score of TransUnion is referred to as the Empirica score which aims to assess risk at the origination stage, evaluate client risk in conjunction with expected performance, assess clients with no historical credit history, manage existing clients from a credit limit and collections perspective, rank clients according to credit risk and identify cross-sell opportunities and to grow the client base. The Empirica score consists of five categories to assess the credit risk of a client: demographics, judgements taken against the client, defaults experiences of the client, payment profiles and enquiries of the client (TransUnion & FICO 2009:2).

As the objective of credit scoring is to distinguish between good (usually up to date or one payment in arrears) and bad (usually three or more payments in arrears, in legal or written off) clients the problems that arise from credit scoring would be related to classification (Abdou & Pointon 2011:66). Problems that exist in credit scoring include the lack of a theoretical reason as to why characteristics are chosen in the final scorecard. Siddiqi (2006) gave some guidelines regarding number of variables (characteristics) to be used within scorecards. Scorecards should comprise between 8 and 15 characteristics to ensure a stable scorecard as the predictive power will remain strong even if the profile of one or two variables changes. Scorecards containing too few variables are more susceptible to minor changes from the applicant's profile making the scorecard unable to withstand the test of time (Siddiqi 2006:88). Anderson (2007) mentions that although a huge number of variables can potentially be used for credit scorecard developments usually only between 6 and 15 variables best explain consumer behaviour (Anderson 2007:393). However, it is generally believed that there is no optimal number of variables to be used in scorecards (Abdou & Pointon 2011:67).

In general, two scenarios which will be explained in the next section are considered when implementing a credit scorecard, namely maintaining the same reject rate or approval rate or keeping the same bad rate. However there is disagreement on the appropriate cut-off score when assessing credit at implementation and the determination of the sample size when building a scorecard (Abdou & Pointon 2011:67–68). In addition there is no optimal technique when building scorecards (Abdou & Pointon 2011:68).

Credit scorecard implementation strategies

Once an organisation has its credit scorecard built it is mostly used to set minimum levels of score which represents a threshold of risk called the cut-off score. This cut-off score could represent a level of accepting clients, profit level or any other level based on the objectives of the organisation
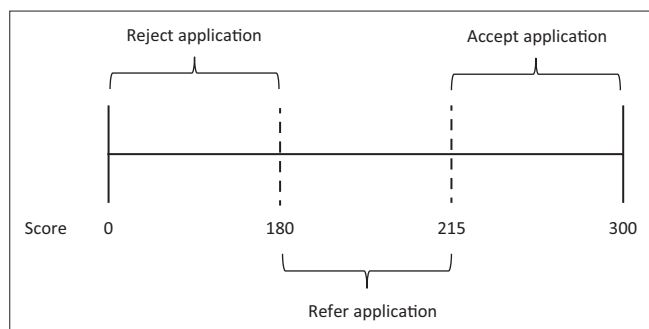
FIGURE 1: Simple cut-off score strategy.

(Siddiqi 2006:146). A simple illustration of implementing a credit scorecard with a cut-off score strategy is presented in Figure 1.

From Figure 1 the implementation strategy indicates that every client scoring below 180 will be rejected, in other words not being considered, a score above 215 will be accepted and clients scoring between 180 and 215 will be referred to credit managers for further analysis to determine whether the client will be rejected or accepted.

In general when building a new credit scorecard it should outperform the previous credit scorecard, which should lead to an outcome of a higher approval rate when the same bad rate is maintained or lead to a lower bad rate when the same current approval rate is maintained (Siddiqi 2006:148). This describes the two general scenarios when implementing a credit scorecard: firstly the lender may wish to implement its credit scorecard such that the lender keeps at the same reject rate or approval rate of clients to reduce bad rates in the future. This first scenario could be chosen when the company wishes to be conservative, where the aim is better risk management in a tough economic environment. Secondly the lender may wish to keep at the same bad rate to gain market share to expand the business; this scenario could be chosen by the company in a competitive environment. According to Anderson (2007), one of the most obvious approaches when setting a cut-off score is where any account is accepted that provides a profit. Since the introduction of credit scoring, lenders have gained experience to apply scientific approaches to enhance the business, use credit scoring for forecasting and portfolio valuation, take potential profitability into consideration, use credit scoring for risk-based pricing, account management and incorporate other aspects of borrower behaviour such as response, revenue and retention. In the event of choosing the cut-off score strategy it should be noted that several assumptions were made during the credit scorecard development and a credit scorecard is a tool that must fit in with a company's general strategies (Anderson 2007:67, 240). Credit scorecards play an important role when distinguishing good clients from bad clients; however, there has been disagreement on the most appropriate cut-off score when strategically implementing a credit scorecard. It is commonly known that there is no optimal cut-off score decision which is differently derived based on the environment and country (Abdou & Pointon 2011:67).

Problem statement and objective

The aim of this article is to build a Credit Scoring Matrix Model (CSMM) to address the issues raised, especially obtaining an optimal model (most favourable) and determining a more appropriate cut-off score when assessing credit of borrowers at the application stage of a loan. A more appropriate cut-off score, in this sense, refers to the cut-off made from the CSMM. The primary objectives of this study are to improve credit risk measurement and management in the world of credit scoring and accurately determine the appropriate cut-off score. The focus includes credit risk, credit scoring and credit risk management. Historical credit scoring modelling techniques are evaluated and the effectiveness of proposed primary objectives on credit risk management are assessed with the aim of implementing these in the retail banking environment.

Literature review

Statistical techniques used in credit scoring can be divided into two categories namely non-parametric techniques and parametric techniques. Non-parametric statistical techniques do not require many assumptions about the underlying data, if any, whereas parametric statistical techniques require several (Anderson 2007:172). This section gives an overview of the different types of non-parametric and parametric statistical techniques used in the credit scoring world.

Non-parametric statistical techniques

Expert systems

Expert systems involve the use of expert judgement or human expert knowledge to solve problems and explain the outcomes as to why certain credit applicants are rejected. Abdou and Pointon (2011) presented the three components of an expert system: relying on facts and rules, an interface communicating the expert's conclusion, and updating the expert's decisions and recommendations (Abdou & Pointon 2011:72). Disadvantages of expert systems include subjectivity, inconsistency and individual expert preferences. Advantages of expert systems include the use of qualitative characteristics within its judgemental evaluation and the vast experience of the expert from the past (Abdou & Pointon 2011:61).

Decision trees

Thomas, Edelman and Crook (2002) mentioned one of the earliest decision trees, which was a type of expert system with a rule set (Anderson 2007:172–173). A decision tree or recursive partitioning analysis (RPA) is defined as a classification technique where a dependent variable is analysed as a function of continuous explanatory variables (Abdou & Pointon 2011:71). A decision tree comprises three nodes with the root node the start (at the top of the tree) followed by subsequent levels called the child nodes and the bottom of the decision tree referred to as the terminal nodes. The aim of a decision tree is to indicate

a possible turn of events or consequences, with each event indicating the outcome. Figure 2 presents a simple example of a decision tree.

Advantages of decision trees include the ability to identify patterns, finding and exploiting interactions, results that are transparent and easy to implement, computationally simple, and quick and easy identification of extremely high and low risk categories. Disadvantages include decision trees becoming too busy, which could result in over-fitting and unreliable results. Where interactions are not an issue regression techniques provide better results and decision trees prove to be relatively inflexible (Anderson 2007:172–174).

Neural networks

Neural networks (NNs) are structures which allow training processes to take place in which the linear and non-linear variables help to distinguish variables to obtain better decision-making results. Andou and Pointon (2011) alluded to the use of NNs which could be successful in credit card fraud, bankruptcy prediction, bank failure prediction, option pricing and mortgage application (Abdou & Pointon 2011:73). Disadvantages of NNs include computational intensity, multiple iterations are required before a final model is obtained, they are expensive to implement and maintain, they are opaque and there is always a significant chance of over-fitting. In credit scoring NNs are seldom used, but have advantages where there are fewer data. NNs are also used for fraud scoring (Anderson 2007:175). Compared to more generally used techniques, such as discriminant analysis and logistic regressions, NNs have the highest average correct classification rate and statistical measures indicate that they represent the data better than logistic regression (Abdou & Pointon 2011:73).

Genetic algorithms

Genetic algorithms – first proposed in the 1960s – are the use of genetic operations to transform data according to fitness value (Abdou & Pointon 2011:74; Anderson 2007:176).
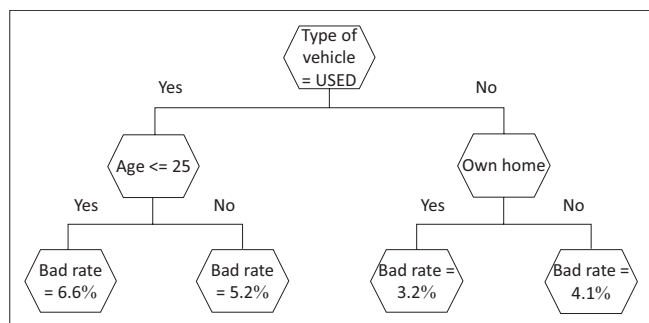

FIGURE 2: Decision tree.

Genetic algorithms have been used in financial services, computer sciences and engineering. Disadvantages of genetic algorithms include computational problems and slower run times. An advantage of genetic algorithms is that alternate solutions may be obtained when they are not readily apparent. The primary uses for genetic algorithms are: providing an exhaustive search if many solutions are possible, the aim is optimisation and not necessarily the best model, identifying good solutions which are not easy to find and when there are multiple targets. Genetic algorithms work best in a rapidly changing environment where new solutions are to be found (Anderson 2007:176). Table 1 summarises the assumptions for the non-parametric techniques.

Parametric statistical techniques

Discriminant analysis

Discriminant analysis is another statistical technique where the aim is to determine group membership where there are two or more known groups (Anderson 2007:169). Andou and Pointon (2011) indicate the earliest proposal of discriminant analysis, where multiple discriminant analysis was applied while examining car loan applications (Abdou & Pointon 2011:69). Discriminant analysis uses a classification tool to minimise the distance between cases in a group and maximise the differences between cases from different groups. The problem with discriminant analysis is that it suffers from all the assumptions associated with the statistical technique used. Linear discriminant analysis is the most common, which suffers from high misclassification errors when predicting rare groups (Anderson 2007:169–170). Problems such as using linear functions instead of quadratic functions, group definitions, inappropriateness of prior probabilities and classification error prediction also surface when using discriminant analysis and multivariate normal distributions and equal variances are assumed (Abdou & Pointon 2011:70).

Probit analysis

Probit analysis aims to transform a linear combination of independent variables into its cumulative probability value from a normal distribution. Under probit analysis normal distributions of the threshold values are assumed and the coefficient estimates can be tested individually for significance using a likelihood ratio test, which is not possible within discriminant analysis. A problem with probit analysis is that multicollinearity can cause incorrect signs for the coefficients which are not an issue within discriminant analysis applications (Abdou & Pointon 2011:70).

TABLE 1: Non-parametric techniques assumptions.

Assumption number	Expert systems	Decision trees	Neural networks	Genetic algorithms
1	Non-statistical approach that makes no assumptions	Non-statistical approach that makes no assumptions	Non-statistical approach that makes no assumptions	Existence of a fitness function which is to be maximised
2	Primitive form of a decision tree	–	–	A mapping from bit-strings to potential solutions

Linear regression

Galton (1889, quoted in Anderson 2007) introduced linear regression, a statistical technique to explain linear relationships described by:

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \cdots + \beta_j x_{jt} \qquad \text{[Eqn 1]}$$

In Equation 1, y_t is the dependent variable or endogenous variable value at time t, x_{jt} is the independent variable or exogenous variable j at time t, and β_j is the coefficient or change in predicted value y per unit of change in x_j at time t.

In Equation 1 the dependent variable y is predicted by using the values of independent variables x_j. The prediction of y is determined by calculating the coefficients β by minimising the sum of the squared error terms. The problem with linear regression is that it makes numerous assumptions such as linearity, homoscedasticity, a normally distributed error term, independent error terms, additivity, uncorrelated predictors and the use of relevant variables (Anderson 2007:166–167).

Logistic regression

Logistic regression is one of the widely used statistical techniques used in credit scoring. The difference between linear regression and logistic regression is that with the latter the outcome variable is binary, that is, 1 or 0, good or bad, etc. (Abdou & Pointon 2011:71). Anderson (2007) indicated that investigations on human populations were where logistic regression originated (Anderson 2007:171). Hand and Henley (1997) indicated that a comparison between discriminant analysis and logistic regression in the world of credit scoring concluded that logistic regression gave superior classification results (Hand & Henley 1997:533).

A logistic regression function can be presented as:

$$\ln\left(\frac{p(G)}{1 - p(G)}\right) = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \cdots + \beta_j x_{jt} \qquad \text{[Eqn 2]}$$

In Equation 2 $p(G)$ is the probability of being good. The left-hand side represents the natural logarithm of the odds of being a good client. For logistic regression the regression coefficients (i.e. the β coefficients) in Equation 2 are obtained by using the maximum likelihood (ML) method (Harrell 2015:220). ML is a general statistical technique to estimate parameters and make statistical conclusions in various situations (Harrell 2015:181). Logistic regression follows assumptions which include categorical target variable, log odds linear relationship function, independent error terms, uncorrelated error terms and the use of relevant variables. A disadvantage of logistic regression is its intense computations; however, logistic regression has been the primary choice for building credit scoring models because it predicts a binary outcome, the final probability cannot fall outside of the boundaries 0 to 1 and logistic regression provides robust estimates of the actual probability (Anderson 2007:170–171). Table 2 summarises the assumptions for the parametric techniques.

Data and methodology

Data

Data that were used in the research study are based on a bank in South Africa and are retail banking specific. External credit bureau data were also used where applicable and were based on South Africa specific credit bureau data. Table 3 presents the frequency and source data used in more detail.

Methodology

The credit scoring matrix model (CSMM) may be split up into two components: firstly an internal application scorecard which is built from internal information and secondly, a credit bureau score obtained from credit bureaus. The methodology that is required to construct the internal application scorecard is presented in Figure 3, which is further explained in the sections that follow.

Bad definition

A client's historical performance can be divided into three categories namely 'good', 'bad' and 'indeterminate' (Anderson 2007:138). A client who is seen as a 'bad' can be based on several considerations, such as organisational objectives, product, being risk-averse, being risk-seeking, an easily interpretable definition, accounting policies on write-offs, consistent definitions or regulatory requirements (Siddiqi 2006:38–40). Classifying a client as a 'good' client requires the same considerations as mentioned as when classifying a client as a 'bad' client. A client is classified as 'indeterminate' if the client does not conclusively fall into either the 'bad' or the 'good' category. As a rule of thumb, 'indeterminates' should not exceed 10% to 15% of the development sample used to build the scorecard (Siddiqi 2006:43–44).

TABLE 2: Parametric techniques assumptions.

Assumption number	Discriminant analysis (DA)	Probit analysis	Linear regression	Logistic regression
1	Suffer from any and all assumptions associated with the statistical technique used	Linearity with log odds function	Linearity	Linearity with log odds function
2	Linear DA was the original; however, logistic regression DA is now preferred due to fewer assumption violations	-	Homoscedasticity	–
3	-	Normally distributed error terms	Normally distributed error terms	Logistic distributed error terms
4	-	Independent error terms	Independent error terms	Independent error terms
5	-	Additivity	Additivity	Additivity
6	-	Uncorrelated predictors	Uncorrelated predictors	Uncorrelated predictors
7	-	Use of relevant variables	Use of relevant variables	Use of relevant variables

TABLE 3: Data used for research study.

Topic	Data requirement	Frequency	Source
Scorecard development and CSMM	Historical client data with performance	Monthly data	Retail bank in South Africa
Bureau data	Historical client data with performance	Monthly data	South African credit bureau

Note: Data availability for Table 3 includes data from January 2002 up to and including December 2014.
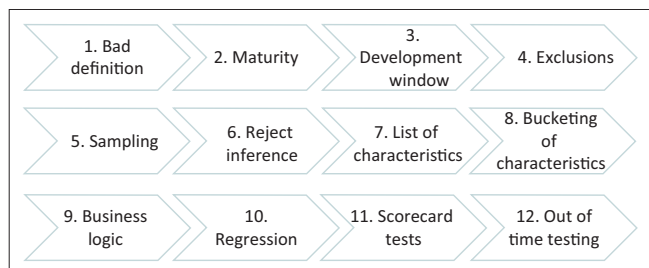
FIGURE 3: Scorecard methodology.

Maturity

Various methods exist to obtain reassurance that the correct 'bad' definition is determined. Firstly there are the analytical methods such as roll rate analysis (comparing delinquency buckets from the past with delinquency buckets observed currently and then calculating which clients remained in the same delinquency bucket, which moved to a better bucket and which rolled to a worse delinquency bucket) (Siddiqi 2006:41) and current versus worst delinquency comparison (comparing the worst ever delinquency status with the most current delinquency status) (Siddiqi 2006:42). Secondly, there is the consensus method where various stakeholders from the credit risk department, operational departments or other areas come together to obtain consensus on the definition of a 'bad' client (Siddiqi 2006:40–43). One can also use the maturity approach where an outcome is determined which is sufficient to capture the 'bad' clients, for example of all clients booked at a specific vintage date which of them were 'bad' clients in the next 'X' number of months (Siddiqi 2006:77). Although one would wish to use as many outcomes as possible to capture most of the 'bad' clients, the current portfolio of the bank could be vastly different from the portfolio in earlier years.

Development window

Building a scorecard requires the identification of two windows: firstly the development or observation window which is the period where the data are observed and secondly, the outcome or performance window during which we measure the level of clients being 'bad' after a certain time period has elapsed for clients from the observation window (Anderson 2007:77, 344). Siddiqi (2006:35) mentions that the development window should be chosen where the level of bad rate is deemed to be stable; however, Anderson (2007:344) indicates that one could come to a point where applications are not a representation of today's applications the longer the outcome period, as also mentioned in the 'Maturity' section.

Exclusions

Data used to build a scorecard should contain only information on which one intends to use them. Clients such as frauds, staff, deceased, out of country, preapproved or underage should be excluded when building a scorecard because these clients are given finance based on non-score-dependent criteria. Any client that is not a normal client or is not going to be scored should be excluded from the scorecard development (Anderson 2007:338; Siddiqi 2006:31).

Sampling

To ensure that a group of representative clients is used during the scorecard development, sampling is used (Anderson 2007:61). It is the general belief that a minimum of 1500 bads, 1500 goods and 1000 rejects are used when developing a credit scorecard. These minimum numbers were derived in the 1960s as the work required to obtain data was demanding and sufficient computer power was lacking (Anderson 2007:350). Two steps are required during the sampling phase when developing a scorecard: firstly, data are sampled which represent the client population (Siddiqi 2006:63). Generally, the total client population is sampled into a smaller data set, especially when working with a large amount of data which can take longer to process. Secondly, the representative total sample is split up into a development sample on which the scorecard is built, and a validation sample on which the scorecard is tested (Siddiqi 2006:63). It is important to note that random samples can pose difficulties as important subgroups of the total sample must also be adequately represented. To achieve this a technique such as over-sampling can be used. Over-sampling in essence adjusts the subgroups' distribution against the total sample (Anderson 2007:351).

Reject inference

Application scorecards are built based on the through-the-door population known as the 'All Good Bad (AGB)' scorecard, that is, all applications, not just accepted applications (Siddiqi 2006:101). If application scorecards are built using only the accepted clients (known as the 'Known Good Bad (KGB) scorecard') there will be bias towards accepted clients from the past when the new application scorecard is applied as the rejected clients from the past were ignored during development (Anderson 2007:65). Reject inference accounts for influence from past decision-making during the AGB scorecard development process. As an example, reject inference accounts for 'cherry-picked' applicants; for example suppose 100 out of 1000 applications have a very high delinquency, credit managers decline 90 of the 100 and accept 10, with subsequent performance indicating that the 10 accepted applicants perform well and are classified as good accounts. Building only a KGB scorecard would see these high delinquency accounts as good applicants. Reject inference accounts for these 'cherry-picked' accounts. From a decision-making view reject inference provides expected performance based on all applications, that is, the through-the-door (TTD) population. Suppose a bank rejects all applications below a score of 180; however, the bank feels it has been too conservative and now rejects all applications below a score of 165. If the bank has never accepted these applicants in the past, how will the bank account for this

additional risk being taken when moving the cut-off? Reject inference accounts for this through the estimated bad applicants for the rejected applications. It is important to realise that with the reject inference process there will always be a level of uncertainty; however, it can be minimised using better reject inference techniques. Reject inference leads to better decision-making; however, it is not 100% accurate (Siddiqi 2006:99–101). Various rejection inference techniques exist, such as random supplementation, augmentation, extrapolation, cohort performance and bivariate two-step (Anderson 2007:79).

List of characteristics

A scorecard consists of a group of characteristics to separate good clients from bad clients (Siddiqi 2006:5). Characteristics to be included in the development sample are an important step when building a scorecard and business input is recommended when characteristics are included. Characteristics should be selected based on: expected predictive power, reliability and robustness, ease in collection, interpretability, human intervention, legal issues surrounding the usage of certain types of information, creation of ratios based on business reasoning, future availability and changes in competitive environment (Siddiqi 2006:60–62).

Bucketing of characteristics

Bucketing of characteristics before the regression step has advantages such as: it is easier to deal with outliers, relationships are easier to understand, non-linear dependencies can be modelled with linear models, unprecedented control in the development process of the scorecard is allowed, increased knowledge of the portfolio, and the user may develop insights into the behaviour of risk predictors (Siddiqi 2006:78).

The predictive power of each characteristic attribute (a group of attributes form a characteristic) after bucketing can be determined by weights of evidence (WoE) presented by Equation 3 for each attribute i:

$$\left[\ln\left(\frac{\text{Distribution}_{Good_i}}{\text{Distribution}_{Bad_i}} \right) \right] \times 100 \qquad \text{[Eqn 3]}$$

The predictive power of each characteristic is measured by the information value (IV) presented by Equation 4:

$$IV_{Characteristic} = \sum_{i=1}^{n} \left(\%Distribution_{Good_i} - \%Distribution_{Bad_i} \right)$$
$$\times \ln\left(\frac{\%Distribution_{Good_i}}{\%Distribution_{Bad_i}} \right) \qquad \text{[Eqn 4]}$$

In Equation 4, $\%Distribution_{Good_i}$ is the distribution of the good clients for attribute i and $\%Distribution_{Bad_i}$ is the distribution of the bad clients for attribute i. The IV of each characteristic is interpreted as follows (Anderson 2007: 192–193; Siddiqi 2006:78–79):

- IV < 0.02   = Unpredictive
- 0.02 ≤ IV < 0.1   = Weak
- 0.1 ≤ IV < 0.3   = Medium
- IV ≥ 0.3   = Strong

Business logic

WoE and IV are used as statistical measures when bucketing characteristics as discussed in the 'Bucketing of characteristics' section; however, business logic also needs to be considered. Suppose a company under its policy rules refers clients to the credit managers that have a debt service ratio greater than 50%. Then the debt service ratio, if a characteristic in the scorecard, should be bucketed with a break at 50% to minimise the distortion of the policy rule on the scorecard, as the clients affected by the policy rule are now isolated (Siddiqi 2006:87).

Regression

The pioneers of credit scoring are Fair and Isaac who started their consultancy, Fair Isaac, in 1956 (Anderson 2007:40). In recent times Fair Isaac has become known as FICO. In the 'Literature review' section various statistical techniques were described that may be used in the world of credit scoring. FICO in addition have their own methods, which is part of their modelling software called FICO Model Builder (FICO MB7) and one of these methods is called 'Scorecard – Divergence'. The divergence statistic measures the distance between the scores of the bad clients' distribution and the scores of the good clients' distribution and is presented by Equation 5 (Anderson 2007:189–190):

$$\text{Divergence} = \frac{\left(Mean_{Goods} - Mean_{Bads} \right)^2}{\left[\frac{\left(Variance_{Goods} + Variance_{Bads} \right)}{2} \right]} \qquad \text{[Eqn 5]}$$

The 'Scorecard-Divergence' method in FICO MB7 is a generalised additive model of bucketed predictors that provides pattern constraints and a penalty term to reduce over-fitting and smooth weight patterns. The fitting objective function optimises the model weights to maximise the divergence between binary outcome classes, subject to user-defined constraints (FICO 2014).

Scorecard tests

Various scorecard measurements can be carried out to determine the predictiveness of the scorecard, such as Akaike's Information Criterion (AIC), Schwarz's Bayesian Criterion (SBC) and the Kolmogorov-Smirnov (KS) statistic. The most common measurement used is called the Gini-coefficient which provides a single value representing the predictive power of the scorecard over the entire range of possible scores (Anderson 2007:205). Statistics such as correlation between characteristics scores, the population stability index (PSI)– which measures the stability of the scorecard – are also analysed when building a scorecard (Anderson 2007:194–200). A PSI less than 0.1 indicates no significant change from development, a PSI between 0.1 and 0.25 indicates a small change in distribution from development that needs investigation and a PSI greater than 0.25 indicates a significant shift from the development population (Siddiqi 2006:137).

Out-of-time testing

The out-of-time window is a validation period which falls outside the development window of the scorecard

(Anderson 2007:78). The out-of-time window is a validation period where the scorecard tests as mentioned in the 'Scorecard tests' section can be carried out to test the scorecard on a period that did not form part of the scorecard building development window.

Results

The CSMM build consisted of two components, namely the internal application scorecard and the credit bureau score. The first part of this section will present the results obtained when building the internal application scorecard, after which the credit bureau score (Empirica score) will be added as a second dimension to illustrate the optimal model. The section concludes with the illustration of improved cut-off score determination from the CSMM.

Internal application scorecard

The methodology followed when building the internal application scorecard was described in the 'Methodology' section. Firstly the 'bad', 'indeterminate' and 'good' definition is presented in Table 4 and maturity analysis in Figure 4.

Figure 4 presents the bad rate percentage of booked clients for each vintage date, that is, for booked clients at a specific date the percentage that went bad after 3 months, 6 months, 9 months, 12 months, 18 months, 24 months, 30 months and

TABLE 4: Internal application scorecard 'Bad' definition.

Classification	Definition
Bad	An account is seen to be bad if it is ever more than three payments in arrears and the balance outstanding is greater than ZAR100, OR if the account is in Legal OR if the account is written off
Indeterminate	An account is seen to be indeterminate if it is ever two payments in arrears and the balance outstanding is greater than ZAR100
Good	An account is seen to be good If it is not 'bad' or 'indeterminate'

36 months. The 24-month outcome was used when building the internal application scorecard for two reasons: firstly, capturing as many bad clients as possible is desired and secondly, too much history is undesirable as applications at development could not represent the applications of today. Based on the 'Bad' outcome decision and data availability up to and including December 2014 (as mentioned in the 'Data' section) the development window and out-of-time window are presented in Table 5.

Exclusions used in the development include company clients, as the scorecard is intended for individual consumers, staff clients, unemployed clients, clients with missing identification numbers (ID), applicants aged-less than 18, non-South-African citizens, loan amounts larger than R500 000 (as this is not normal business for the company), frauds, deceased clients, application disputes and pending decisions. Exclusions were approximately 11% for the development window. Sampling was performed using a simple random sampling technique in the statistical analysis software (SAS) and Figure 5 presents the bad rate percentage for accepted clients between the total population after exclusions and the sample after exclusions.

Reject inference (fuzzy augmentation technique) was carried out by using the bureau score at outcome from the credit bureau TransUnion to infer the rejected applications. The Gini-coefficient for the bureau score at outcome was determined as approximately 81%, indicating this as a good proxy to infer the rejected applications. The bad rate after inference for the development window was calculated as 6.3%. Figure 6 presents the rejects inference results.

Figure 6 illustrates the higher bad rate for rejected applications after reject inference (as expected). Characteristics considered in the development of the internal application scorecard included 32 characteristics consisting of both borrower and
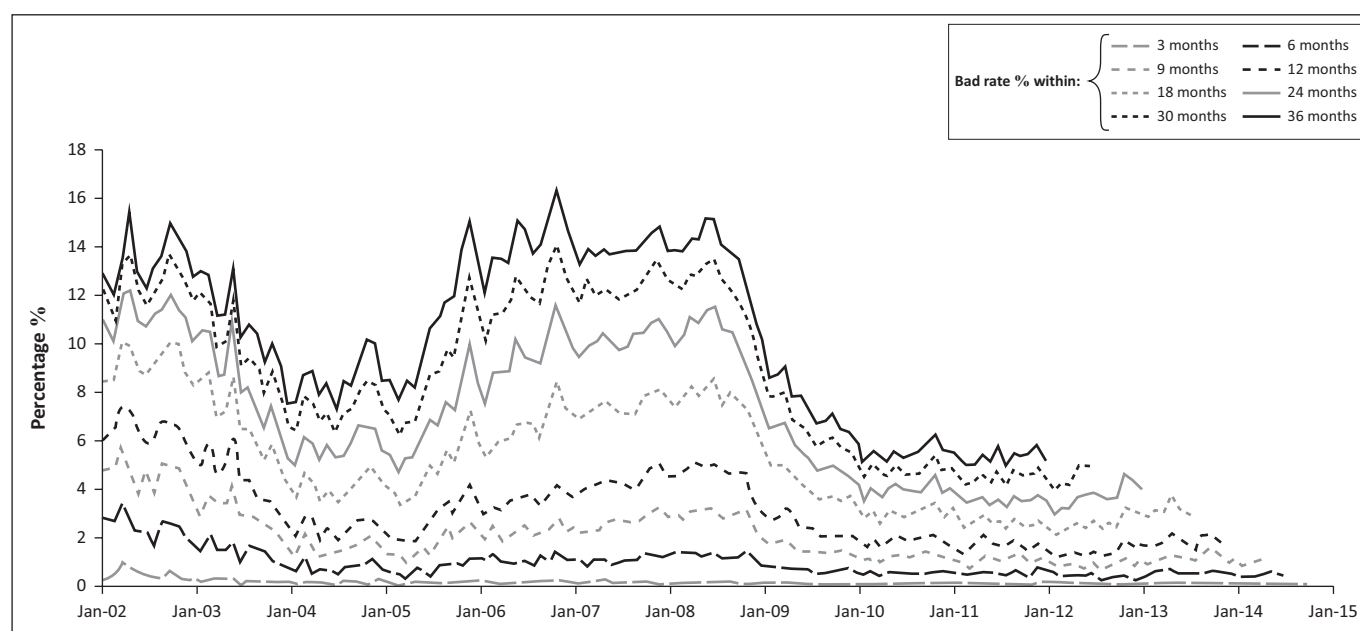

FIGURE 4: Maturity analysis.

transaction type characteristics. The bucketing of the characteristics was done in the FICO MB7 software as follows:

- Step 1: Firstly auto-binning ('Coarse Fine Supervised') in FICO Model Builder was carried out
- Step 2: Secondly after auto-binning, bucketing was carried out such that:

TABLE 5: Internal application scorecard build windows.

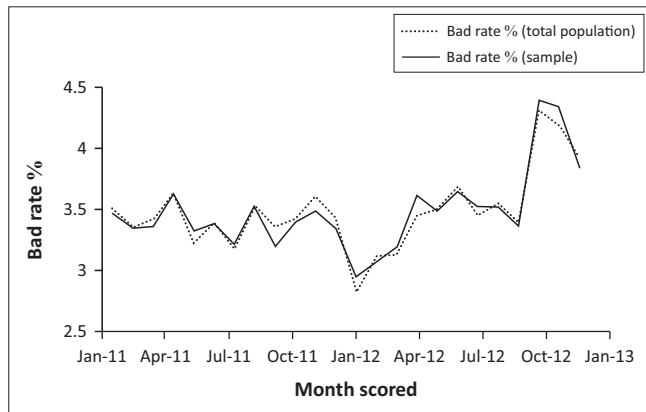| Window	Definition
Development	January 2011 up to and including December 2011 (80% development, 20% hold-out)
Out-of-time	January 2012 up to and including December 2012

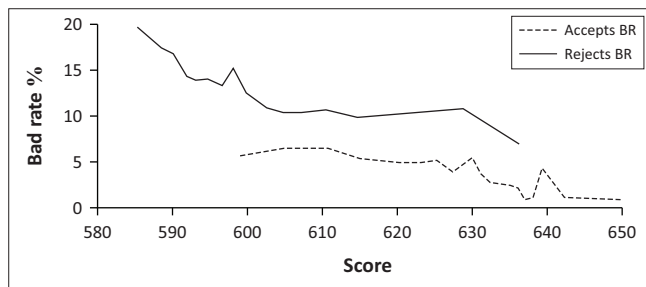**FIGURE 5:** Accepted clients' bad rate (%).

FIGURE 6: Bad rate for accepts and rejects (at observation).

TABLE 6: Internal application scorecard performance measures results.

Measurement	Development window	Hold-out
Divergence	0.316	0.253
Gini-coefficient	0.305	0.274
KS-statistic	21.838	19.843

- There should be a logical relationship between the bad rate and the characteristic
- Buckets must contain at least 2% of total and not less than five bads per bucket
- A tail-end bucket can contain around or less than 2% of total population of the characteristic if the characteristic has at least three buckets
- Step 3: Apply business rationale

After the bucketing process regression was carried out and based on the 'Scorecard-Divergence' method from FICO. Table 6 presents the predictiveness statistics of the internal application scorecard from the FICO MB7 software.

Table 6 presents predictive results for both the development sample as well as the hold-out or validation sample from FICO Model Builder. The divergence, Gini-coefficient and KS-statistic show deterioration from the development window to hold-out sample. Table 6 indicates a Gini-coefficient of 30.5%, which excludes bureau information and is based only on internal company information. No correlation (correlation above 50% is deemed significant) between characteristics was observed as indicated in Table 7.

Figure 7 illustrates the score trend of the internal application scorecard. The score trends in Figure 7 illustrate the high bad rates for low scores and the higher the score the lower the bad rate for both the development and hold-out sample. The correlation analysis and score trend for the out-of-time window (OOT) are presented in Table 8 and Figure 8 respectively.

No correlation (correlation above 50% is deemed significant) between characteristics was observed as indicated in Table 8. The score trends in Figure 8 illustrate the high bad rates for low scores and the higher the score the lower the bad rate. The PSI which measures the change in distribution against the development sample is presented in Figure 9.

Figure 9 presents the population distribution for the OOT window. A PSI less than 0.1 is considered to be stable (as evident from Figure 9).

TABLE 7: Correlation (development).

Window	Characteristic 1	Characteristic 2	Characteristic 3	Characteristic 4	Characteristic 5
Development					
Characteristic 1	100%	−4%	8%	3%	4%
Characteristic 2	−4%	100%	20%	13%	10%
Characteristic 3	8%	20%	100%	25%	6%
Characteristic 4	3%	13%	25%	100%	3%
Characteristic 5	4%	10%	6%	3%	100%
Hold-out					
Characteristic 1	100%	−2%	7%	1%	6%
Characteristic 2	−2%	100%	19%	12%	11%
Characteristic 3	7%	19%	100%	26%	7%
Characteristic 4	1%	12%	26%	100%	4%
Characteristic 5	6%	11%	7%	4%	100%

Note: The score trend is a summary of grouped scores equally distributed as far as possible and is presented in Figure 7.

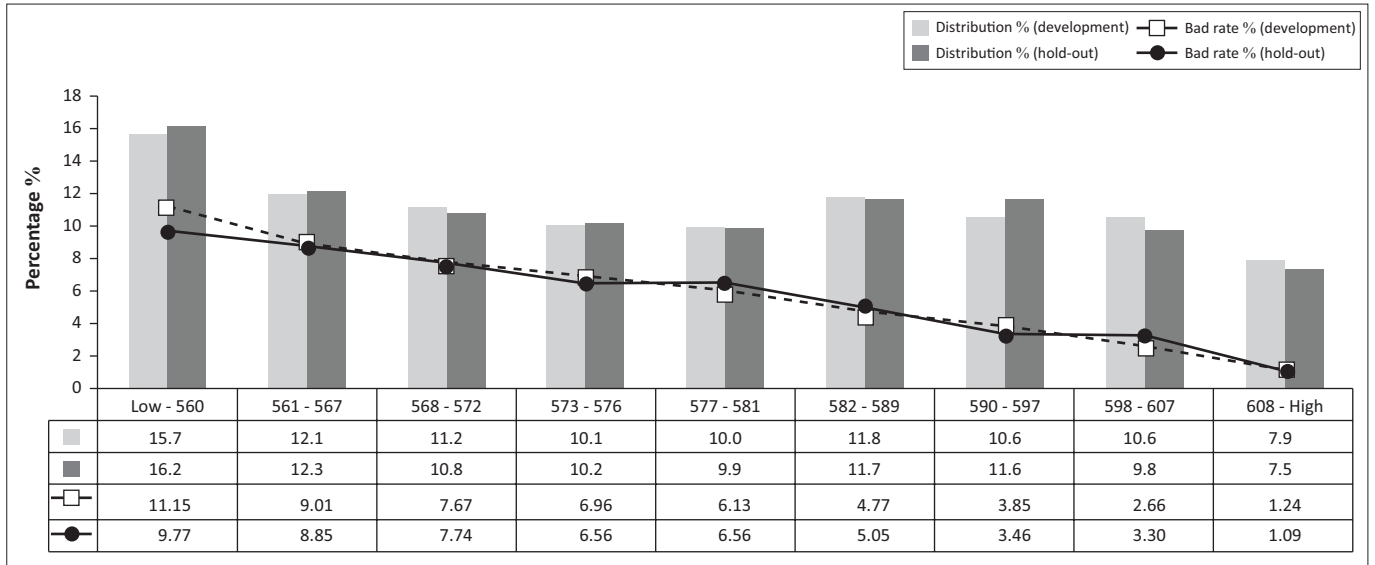|	Low - 560	561 - 567	568 - 572	573 - 576	577 - 581	582 - 589	590 - 597	598 - 607	608 - High
	15.7	12.1	11.2	10.1	10.0	11.8	10.6	10.6	7.9
	16.2	12.3	10.8	10.2	9.9	11.7	11.6	9.8	7.5
	11.15	9.01	7.67	6.96	6.13	4.77	3.85	2.66	1.24
	9.77	8.85	7.74	6.56	6.56	5.05	3.46	3.30	1.09

FIGURE 7: Score trend (internal application scorecard).

TABLE 8: Correlation (out-of-time).

Variable	Characteristic 1	Characteristic 2	Characteristic 3	Characteristic 4	Characteristic 5
Characteristic 1	100%	−5%	6%	3%	2%
Characteristic 2	−5%	100%	18%	13%	9%
Characteristic 3	6%	18%	100%	24%	5%
Characteristic 4	3%	13%	24%	100%	2%
Characteristic 5	2%	9%	5%	2%	100%

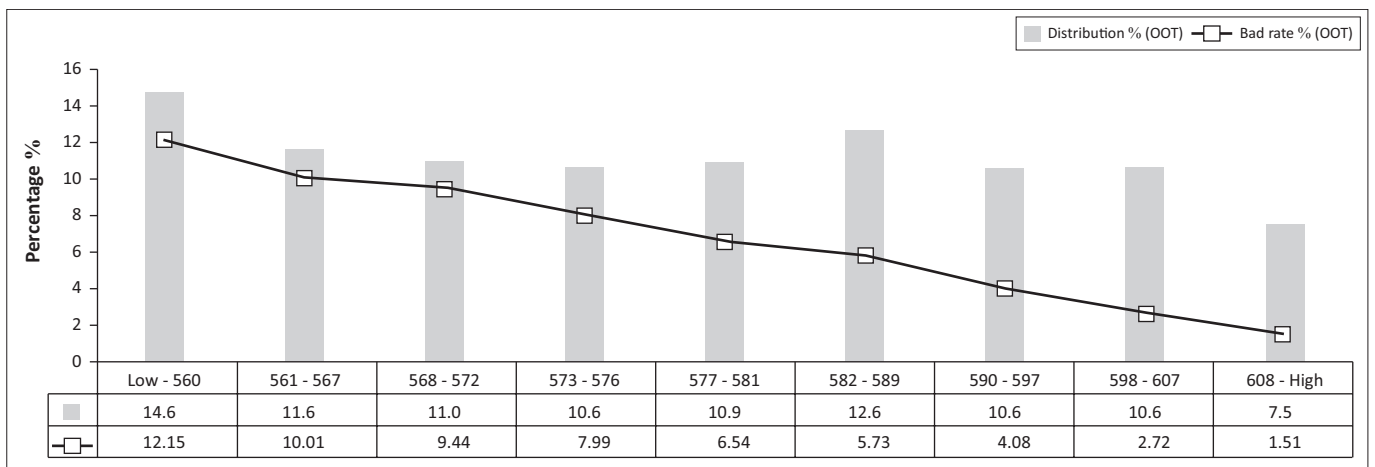|	Low - 560	561 - 567	568 - 572	573 - 576	577 - 581	582 - 589	590 - 597	598 - 607	608 - High
	14.6	11.6	11.0	10.6	10.9	12.6	10.6	10.6	7.5
	12.15	10.01	9.44	7.99	6.54	5.73	4.08	2.72	1.51

FIGURE 8: Score trend (internal application scorecard) for out-of-time window.

Bureau scorecard

The credit bureau TransUnion in South Africa provides on-request bureau scores (Empirica scores) to the lending institutions. The Gini-coefficient of the Empirica score was calculated as 42.0% for the development window, as shown in Figure 10.

The score trend of the Empirica score is presented in Figure 11.

Figure 11 illustrates again how high the bad rates are for lower scores and the higher the score the lower the bad rate (as expected).

Credit scoring matrix model

Siddiqi (2006) presented three approaches to implementing a multi-scorecard solution: the sequential approach, the matrix approach and the matrix-sequential hybrid approach. In the sequential approach an applicant is scored sequentially on each scorecard with each scorecard having its own cut-off. The sequential approach is best used when 'hurdles' are being used, that is, an applicant must have a minimum bureau score before sequentially moving on to the next scorecard. In the matrix approach multiple scorecards are used concurrently with the decision-making based on a combination of the cut-offs from each scorecard (Siddiqi 2006:144). This matrix approach is used when a balanced choice needs to be made
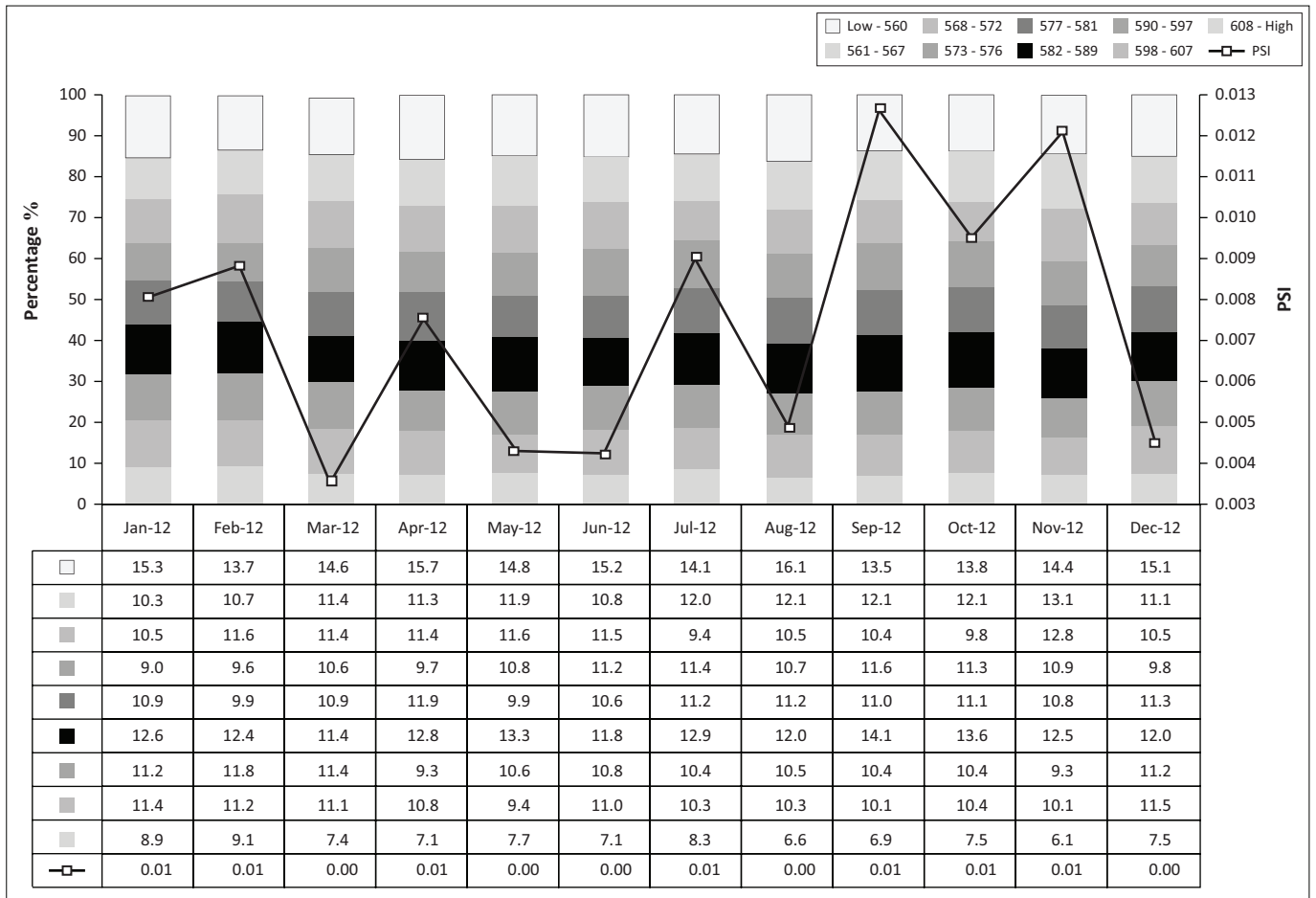
	Jan-12	Feb-12	Mar-12	Apr-12	May-12	Jun-12	Jul-12	Aug-12	Sep-12	Oct-12	Nov-12	Dec-12
	15.3	13.7	14.6	15.7	14.8	15.2	14.1	16.1	13.5	13.8	14.4	15.1
	10.3	10.7	11.4	11.3	11.9	10.8	12.0	12.1	12.1	12.1	13.1	11.1
	10.5	11.6	11.4	11.4	11.6	11.5	9.4	10.5	10.4	9.8	12.8	10.5
	9.0	9.6	10.6	9.7	10.8	11.2	11.4	10.7	11.6	11.3	10.9	9.8
	10.9	9.9	10.9	11.9	9.9	10.6	11.2	11.2	11.0	11.1	10.8	11.3
	12.6	12.4	11.4	12.8	13.3	11.8	12.9	12.0	14.1	13.6	12.5	12.0
	11.2	11.8	11.4	9.3	10.6	10.8	10.4	10.5	10.4	10.4	9.3	11.2
	11.4	11.2	11.1	10.8	9.4	11.0	10.3	10.3	10.1	10.4	10.1	11.5
	8.9	9.1	7.4	7.1	7.7	7.1	8.3	6.6	6.9	7.5	6.1	7.5
	0.01	0.01	0.00	0.01	0.00	0.00	0.01	0.00	0.01	0.01	0.01	0.00

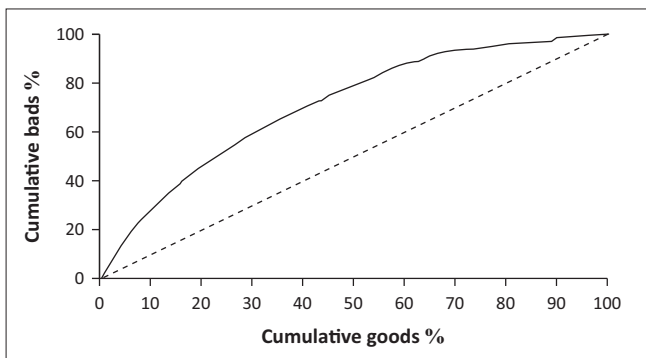**FIGURE 9:** Population stability index.

FIGURE 10: Empirica score Gini-coefficient (development window).

from different types of ideally independent information (Siddiqi 2006:144–145). The matrix-sequential hybrid approach is used when an applicant is firstly put through the sequential approach, after which the applicant goes through the matrix approach. The matrix-sequential hybrid approach is more versatile than the sequential approach and simpler than the matrix approach. This approach is best used when three independent scorecards are used, balancing several competing interests and used in conjunction with policy rules for applicants that are prequalified (Siddiqi 2006:146).

The CSMM is a matrix approach except that the decision-making is based on a cut-off determined after the internal application scorecard and bureau scorecard are combined (discussed in the 'Cut-off score determination' section). The CSMM is represented as follows:

$$
\begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_i \end{bmatrix} \times \begin{bmatrix} B_1 & B_2 & \dots & B_j \end{bmatrix} = \begin{bmatrix} C_{11} & \cdots & C_{1j} \\ \vdots & \ddots & \vdots \\ C_{i1} & \cdots & C_{ij} \end{bmatrix} \quad \text{[Eqn 6]}
$$

In Equation 6 A_i represents internal application scorecard score i, B_j represents Empirica Scorecard score j, C_{ij} represents the matrix's score ij; $i = 1 \dots x$, $j = 1 \dots y$, x = score range of internal scorecard and y = Empirica Scorecard.

In Equation 6 C_{11} represents the matrix score where there is a low internal application score and low Empirica score, C_{ij} represents a high internal application score and high Empirica score, scores to the bottom left represent high internal application scores and low Empirica scores and scores to the top right represent low internal application scores and high Empirica scores. Using the scores from the internal application scorecard and combining with the Empirica scores, the bad rates for the CSMM are represented in Table 9.

Table 9 indicates high bad rates in the top left quadrant with low internal application scores and low Empirica scores. The lowest bad rates are at the bottom right quadrant as expected
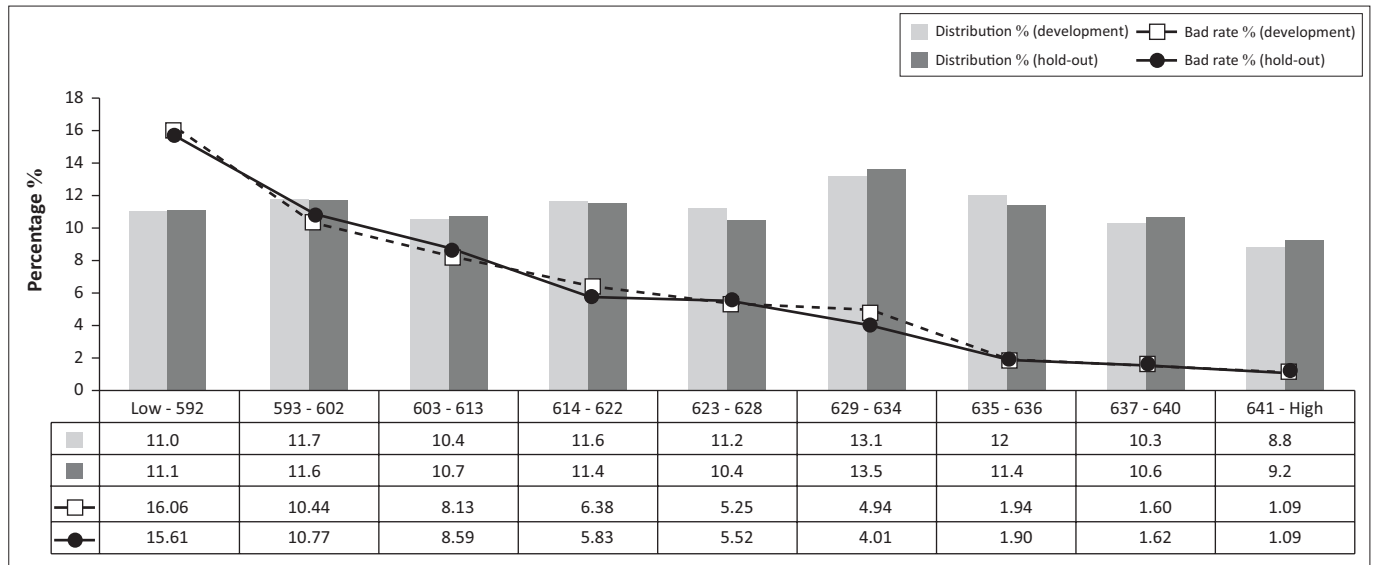
FIGURE 11: Score trend (Empirica score).

TABLE 9: Credit Scoring Matrix Model (bad rate % per matrix cell).

Bad rate %		Bureau score (Bad rate %)										Bad rate (%)
		A	B	C	D	E	F	G	H	I	J	
Internal Scorecard (Bad Rate)	1	20.6	16.7	15.2	13.3	11.9	10.7	9.7	9.9	9.8	4.9	12.1
	2	18.8	15.6	12.5	11.6	10.4	8.3	8.4	8.3	6.4	4.8	10.0
	3	18.7	15.5	12.8	10.2	9.3	6.7	6.9	7.3	6.8	4.0	9.1
	4	18.8	14.8	12.7	9.7	8.3	7.9	5.6	4.5	6.6	3.4	8.2
	5	17.3	12.5	12.1	10.5	10.5	6.0	6.1	5.6	5.8	2.9	7.7
	6	17.6	12.8	12.6	9.5	9.4	5.2	5.8	5.2	4.6	2.2	6.8
	7	18.5	12.1	10.7	10.3	7.7	6.0	6.4	4.7	4.9	2.1	6.2
	8	18.1	13.7	11.0	8.8	6.5	5.4	4.0	2.9	3.5	1.6	4.9
	9	16.7	10.6	10.5	6.4	6.0	4.7	4.7	3.5	2.8	1.1	3.8
	10	12.4	12.9	6.3	5.0	4.4	3.0	3.2	2.4	1.4	0.8	2.1
Bad rate	–	18.4	14.1	11.6	9.2	8.1	6.2	5.9	4.8	4.8	1.9	6.3

(high internal application scores and high Empirica scores). Top right and bottom left quadrant are marginal client bad rates with either a low internal application score or high Empirica score (top right quadrant) or high internal application score and low Empirica score (bottom left quadrant).

To illustrate the benefit of using the CSMM from a predictive point of view the Gini-coefficient was calculated. Firstly, the Gini-coefficient for the internal application scorecard was determined by grouping each score into the 10 score bands used in Table 9 and was calculated as 29%. Secondly, the Gini-coefficient for the bureau scorecard (Empirica score) was determined by grouping each score into the 10 score bands used in Table 9 and was calculated as 41%. Thirdly, the Gini-coefficient for the CSMM was calculated with a value of 46% by ranking the bad rates from the matrix from worst to best. The Gini-coefficients are summarised in Figure 12.

In Figure 12 the uplift in Gini-coefficient is $\frac{46-41}{41} = 12\%$.

The theoretically correct way to compute the uplift from the CSMM is to compare it to a one-dimensional scorecard containing both internal application information and external

bureau information; it was calculated as 43%, indicating that the CSMM still gives uplift. However it is not recommended to implement a one-dimensional scorecard containing both internal application information and external bureau information for various reasons: external bureau information is much stronger than internal application information, hence when building a one-dimensional scorecard containing both internal application information and external bureau information, the bureau information will completely overpower the internal application information. Internal application information is more susceptible to application information manipulation, especially to a certain secured retail product. Given that the application information can be manipulated, the second axis of the CSMM acts as a hedge. Converting the internal application information to one axis of a CSMM opens the door to marginal clients, for example you can have applicants with high internal application scores, but low bureau scores and you can have applicants with low internal application scores, but high bureau scores; this opens up opportunities. In addition clients with no bureau history can now also be considered, as they would have internal application information. Although the Gini-coefficient of 30.5% of the internal application scorecard (excluding bureau information) might seem too low, the necessity of this information within the overall CSMM
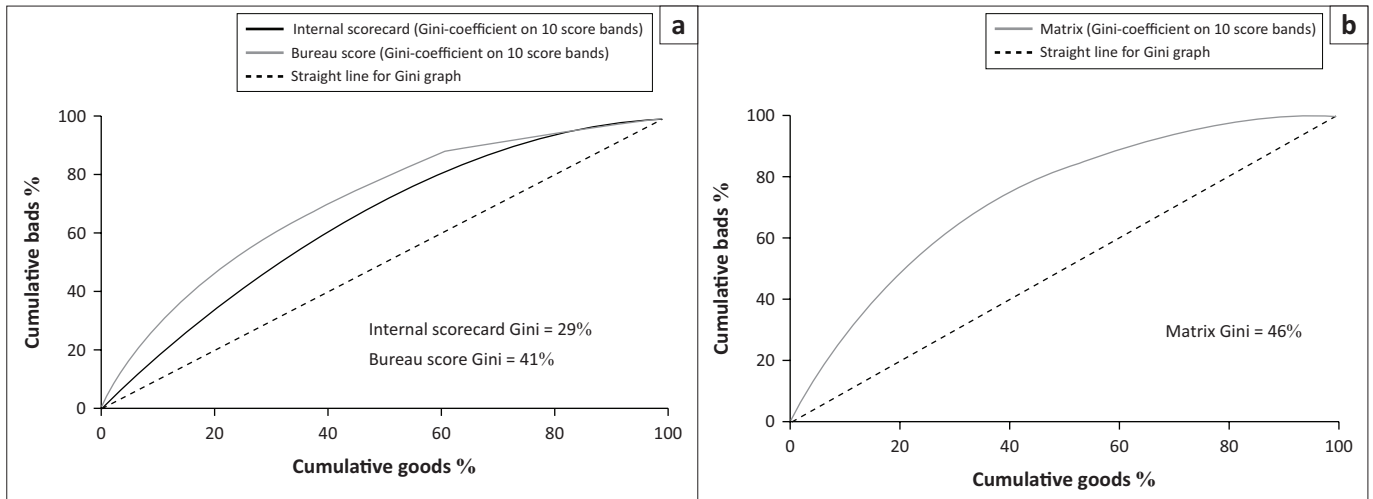
FIGURE 12: Gini-coefficients: (a) Internal scorecard & bureau score and (b) Matrix.
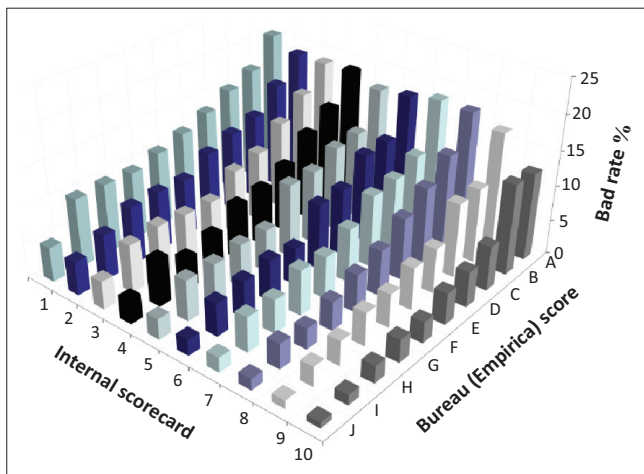

FIGURE 13: Credit scoring matrix model.

TABLE 10: One-dimensional score (internal application score).

Internal application score	Bad rate (%)	Cut-off score
Low–559	12.13	-
560–561	9.95	-
562–565	9.12	-
566–568	8.24	Cut-off score
569–572	7.70	-
573–576	6.84	-
577–581	6.18	-
582–588	4.92	-
589–597	3.79	-
598–High	2.07	-

is paramount, for example, in preventing the booking of high loan to value (LTV) applicants, which could have devastating effects.

Whenever there is a need to change to a latest bureau score it would be easier to change one axis of the CSMM instead of doing an entire new scorecard rebuild, which will be the case if the bureau score is embedded within a one-dimensional scorecard containing both internal application information and external bureau information. The CSMM in addition does not fully put reliance on bureau information or solely on internal processes; thus, if credit bureaus are unable to provide information or if internal processes fail one can fall back on either the internal application scorecard or bureau scorecard respectively in such extreme cases. Given the uplift in Gini-coefficient from the CSMM against the bureau scorecard and the numerous reasons given above the CSMM can be regarded as the optimal model (most favourable) from a credit scoring perspective in retail banking.

The CSMM is presented in Figure 13.

Cut-off score determination

To illustrate the effectiveness of the CSMM, consider a one-dimensional approach and assume only the internal application scorecard is available to score clients presented in Table 10. The bad rate data in Table 10 correspond to the bad rate data in Table 9 for the internal application scorecard dimension.

Using Table 10, suppose that the decision was made that a bad rate of 8.24 or higher for applicants would not be considered, which will set the cut-off score at 568 for the internal application score. Table 10 is a straight cut-off from the one-dimensional scorecard. Secondly, assume the internal application scorecard and bureau (Empirica) score in matrix format are at hand to assess an applicant. Using Table 9, suppose that the decision was made that a bad rate of 8.24 or higher for applicants will not be considered, which will have a diagonal cut-off presented in Table 11.

Table 11 presents information that is not a straight cut-off as for the one-dimensional scorecard, but a more granular two-dimensional cut-off. Having such a strategy cut-off has some advantages such as: clients with a low bureau (Empirica) score, but high application score can be considered, clients with a low application score, but high bureau score (Empirica) can be considered, negative effects of application form manipulation for internal scorecard are reduced because the bureau (Empirica) hedges this manipulation and a greater dimension of information is taken into account to determine whether an applicant will be bad in future.

TABLE 11: Two-dimensional Credit Scoring Matrix Model.

Bad rate (%)		Bureau score (bad rate %)										Bad rate (%)
		A	B	C	D	E	F	G	H	I	J	
Internal Scorecard (Bad Rate)	1	20.6	16.7	15.2	13.3	11.9	10.7	9.7	9.9	9.8	4.9	12.13
	2	18.8	15.6	12.5	11.6	10.4	8.3	8.4	8.3	6.4	4.8	9.95
	3	18.7	15.5	12.8	10.2	9.3	6.7	6.9	7.3	6.8	4.0	9.12
	4	18.8	14.8	12.7	9.7	8.3	7.9	5.6	4.5	6.6	3.4	8.24
	5	17.3	12.5	12.1	10.5	10.5	6.0	6.1	5.6	5.8	2.9	7.70
	6	17.6	12.8	12.6	9.5	9.4	5.2	5.8	5.2	4.6	2.2	6.84
	7	18.5	12.1	10.7	10.3	7.7	6.0	6.4	4.7	4.9	2.1	6.18
	8	18.1	13.7	11.0	8.8	6.5	5.4	4.0	2.9	3.5	1.6	4.92
	9	16.7	10.6	10.5	6.4	6.0	4.7	4.7	3.5	2.8	1.1	3.79
	10	12.4	12.9	6.3	5.0	4.4	3.0	3.2	2.4	1.4	0.8	2.07
Bad rate	–	18.44	14.07	11.57	9.15	8.10	6.15	5.85	4.83	4.80	1.90	6.27

Note: Values in bold represent the cut-off score.

Conclusion

In the last 60 years (to 2016), credit scoring played (and still plays) an important role in the credit industry, which mitigates and controls future bad clients. However, literature is limited, which encourages further research on credit scoring. In the last 25 years (to 2015) credit scoring has grown, particularly in the banking industry, making the separation between good and bad clients more critical to have effective credit risk management and to reduce future bad debts, which emphasises the significance of this research.

This article presented a literature background on statistical techniques used in the world of credit scoring illustrating both weaknesses and advantages of each technique. The methodology then led to an application scorecard which includes the decision on the initial bad definition, outcome decision or maturity, development window, exclusions to be used in the development, sampling, reject inference, characteristics considered to predict bad clients, bucketing process, business intervention, regression technique, relevant scorecard tests and out-of-time testing.

The contribution from this article is the construction of a scoring model to optimise the separation between good and bad clients in the form of a CSMM. The CSMM consists of two components, namely, the internal application scorecard (built from information specific to the organisation) and the credit bureau or Empirica score (based on external data). These provide insights on client's credit performance with all other credit organisations. The separation of internal organisation client credit performance information and external client credit performance information by the CSMM is important. Internal information, contained in the internal application score, indicates a direction of client credit performance; however, the external information contained in the Empirica score acts as a hedge should the internal application score not capture all client-related credit performance.

During the construction phase of the internal application scorecard it was decided to use a 24-month outcome period to capture appropriate number of bad clients and to work on a population, which is a reflection of the current portfolio. Reject inference results indicated the higher bad rate for rejected clients (as expected). The internal application scorecard gave a Gini-coefficient of 30.5% for the development window with the relevant scorecard tests indicating no correlation between characteristics and stability. Combining the internal application scorecard with the Empirica score a CSMM was constructed that distinguishes between good and bad clients on a more granular level that in addition enables the setting of a more appropriate cut-off score, which was highlighted as a problem in past literature. It was illustrated that having a CSMM gives a relative percentage uplift in the Gini-coefficient of 12.0% to distinguish between good and bad clients more effectively. This will lead to fewer clients being initially selected, which would have resulted in future bad clients creating effective credit risk management and reducing future bad debts. How a CSMM establishes the cut-off score more appropriately to accept clients than a one-dimensional scorecard by providing a greater granularity option, adds to the contribution of this work.

Possible future research includes the combination of an internal application scorecard with an internal bureau scorecard which can be built on specific bureau information that works best for the organisation in question. In addition, possibilities exist to investigate the optimal construction when building a matrix.

Acknowledgements

Competing interests

The authors declare that they have no financial or personal relationships that may have inappropriately influenced them in writing this article.

Authors' contributions

N.K. was fully responsible for the writing of this article, while G.W.v.V. was the appointed supervisor from the university and gave valuable guidance in writing this article.

References

Abdou, H. & Pointon, J., 2011, 'Credit scoring, statistical techniques and evaluation criteria: A review of the literature', *Intelligent Systems in Accounting, Finance & Management* 18(2–3), 59–88. https://doi.org/10.1002/isaf.325

Allen, L., De Long, G. & Saunders, A., 2004, 'Issues in the credit risk modelling of retail markets', *Journal of Banking and Finance* 28, 727–752. https://doi.org/10.1016/S0378-4266(03)00197-3

Anderson, R., 2007, *The credit scoring toolkit, theory and practice for retail credit risk. Management and decision automation*, Oxford University Press, Oxford.

Hand, D.J. & Henley, W.E., 1997, 'Statistical classification methods in consumer credit scoring: A review', *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 160(3), 523–541. https://doi.org/10.1111/j.1467-985X.1997.00078.x

Harrell, F.E., 2015, *Regression modeling strategies*, 2nd edn., Springer International Publishing, Cham, Switzerland.

Lim, M.K. & Sohn, S.Y., 2007, 'Cluster-based dynamic scoring model', *Expert Systems with Applications* 32, 427–431. https://doi.org/10.1016/j.eswa.2005.12.006

Siddiqi, N., 2006, *Credit risk scorecards, developing and implementing intelligent credit scoring*, John Wiley & Sons, Inc. Hoboken, NJ.

Thomas, L.C., Edelman, D.B. & Crook, J.N., 2002, *Credit scoring and its applications*, SIAM, Philadelphia, PA.

TransUnion, 2015, *Credit bureau, credit reporting companies. Learn about your credit*, viewed 02 April 2015, from https://www/mytransunion.co.za

TransUnion & Fair Isaac Corporation (FICO), 2009, *Empirica. Minimise your credit risk. Increase your profitability*, TransUnion, Johannesburg.