Knowledge-Based Systems 103 (2016) 89–103

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Predicting creditworthiness in retail banking with limited scoring data

Hussein A. Abdou[a,b,*], Marc D. Dongmo Tsafack[c], Collins G. Ntim[d], Rose D. Baker[e]

[a] Huddersfield Business School, University of Huddersfield, Huddersfield, West Yorkshire HD1 3DH, UK
[b] Management Department, Faculty of Commerce, University of Mansoura, Mansoura, Dakahlia, Egypt
[c] Salford Business School, University of Salford, Salford, Greater Manchester, M5 4WT, UK
[d] Huddersfield Business School, University of Huddersfield, Huddersfield, West Yorkshire, HD1 3DH, UK
[e] Salford Business School, University of Salford, Salford, Greater Manchester, M5 4WT, UK

ARTICLE INFO

Article history:
Received 11 April 2015
Revised 21 March 2016
Accepted 25 March 2016
Available online 12 April 2016

JEL classification:
E50
G21
C45

Keywords:
Predicting creditworthiness
Credit scoring
Cascade correlation neural networks
CART
Limited data

ABSTRACT

The preoccupation with modelling credit scoring systems including their relevance to predicting and decision making in the financial sector has been with developed countries, whilst developing countries have been largely neglected. The focus of our investigation is on the Cameroonian banking sector with implications for fellow members of the Banque des Etats de L'Afrique Centrale (BEAC) family which apply the same system. We apply logistic regression (LR), Classification and Regression Tree (CART) and Cascade Correlation Neural Network (CCNN) in building our knowledge-based scoring models. To compare various models' performances, we use ROC curves and Gini coefficients as evaluation criteria and the Kolmogorov-Smirnov curve as a robustness test. The results demonstrate that an improvement in terms of predicting power from 15.69% default cases under the current system, to 7.68% based on the best scoring model, namely CCNN can be achieved. The predictive capabilities of all models are rated as at least very good using the Gini coefficient; and rated excellent using the ROC curve for CCNN. Our robustness test confirmed these results. It should be emphasised that in terms of prediction rate, CCNN is superior to the other techniques investigated in this paper. Also, a sensitivity analysis of the variables identifies previous occupation, borrower's account functioning, guarantees, other loans and monthly expenses as key variables in the forecasting and decision making processes which are at the heart of overall credit policy.

1. Introduction

The capability of statistical credit scoring systems to improve decision-making and time efficiencies in the financial sector has widely attracted researchers and practitioners particularly in recent years (see for example, [4,37,43–45,49,51,53,54]). Credit scoring systems are now regarded as virtually indispensible in developed countries. In developing countries statistical scoring models are needed not least to support judgemental techniques subject to each bank's individual policies. In building a scoring system a number of particular client's characteristics are used to assign a score. These scores can provide a firm basis for the lending and re-lending decision [9,17,23,48,49,52,53].

Background of the Cameroonian banking sector: Credit scoring is not popular in Africa at present. It appears neither to have been applied nor considered in the case of the Cameroonian banking sector[1] and across the BEAC family. Cameroon is one of the developing countries in west and central Africa and is estimated to have a population just over 19 million people. The labour force was estimated in 2009 to be 7.3 million. Employment derives mainly from three sectors. Firstly, from industry: petroleum production and refining, aluminium production, food processing, light consumer goods, textiles, lumber, ship repair; secondly, from ser-

[1] The Bank of Issue for Cameroon is the "Bank of the Central African States" (Banque des Etats de L'Afrique Centrale, BEAC) which was created on November 22nd 1972. It was introduced to replace the "Central Bank of the State of Equatorial Africa and Cameroon" (Banque des Etats de l'Afrique Equatoriale et du Cameroun, BCEAC) which had been operating since April 14th 1959. BEAC is the central bank for the following six countries, in no particular order of priority: Cameroon, Central African Republic, Chad, Republic of the Congo, Equatorial Guinea and Gabon. Together these six countries also form the "Economic and Monetary Community of Central Africa" (Communauté Economique et Monétaire de l'Afrique Centrale, CEMAC). BEAC's headquarters are located in Yaounde, the capital of Cameroon. The issued currency is the "CFA Franc", which stands for "Financial Cooperation in Central Africa" (Coopération Financiere en Afrique Centrale) and is pegged to the Euro at a rate of €1 = CFA665.957 [8].

* Correspondence author. Tel.: +44 1484473872; fax: +44 1484473148.
E-mail address: h.abdou@hud.ac.uk (H.A. Abdou).

http://dx.doi.org/10.1016/j.knosys.2016.03.023

vices; and finally, from the main sector which is agriculture, predominantly coffee, cocoa, cotton, rubber, bananas, oilseed, grains and root starches. The Gross Domestic Product (GDP) in 2008 was US$20.65 billion. Total domestic lending was US$1.3 billion which represented approximately 6.3% of its GDP. By contrast, in an advanced economy such as the Netherlands with a population only 2 million fewer than the Cameroon, domestic lending represented an estimated 219% of their GDP (CIA, 2009). Thus, there is at least a case for investigating the scope for the growth of the credit industry in the Cameroonian market (for details see Appendix A) including the selection of appropriate scoring techniques.

In Cameroon and across BEAC, a judgemental and traditional system called Tontines[2] remains very popular. Cameroonian banks are reluctant to take risks so most people rely on Tontines to overcome loss of income and, in the case of small entrepreneurs, to raise funds to finance their operations. Members' behaviour is to some extent guaranteed by the wish not to be excluded from help and solidarity which is important in the context of a background of great social and economic uncertainty. Tontines have some drawbacks as credit tools. They can only be used for the short-term as the debt will have to be repaid at the end of the Tontine's cycle; the interest on Tontine credit is relatively high (between 5–10% per month); a huge sum of money cannot be easily obtained to fund a large investment [31,35].

The aims of this paper are: firstly, to identify and investigate the currently used approaches to assessing consumer credit in the Cameroonian banking sector; secondly, to build appropriate and powerfully predictive scoring models to predict creditworthiness then to compare their performances with the currently used traditional system; and finally and freshly to discern which of the variables used in building the scoring models are most important to the decision making process.

Our practical contribution emerges from the foregoing. It would clearly be in the interests of both borrowers and banks to have decision making models which make credit available on terms which reflect the needs of borrowers and their ability to repay. Provision of such a service requires a sensitive and efficient credit scoring system. This is essential to establishing and monitoring the creditworthiness of borrowers in the joint interests of themselves and their lenders. The credit scoring system of choice needs to be tailored to the particular society and credit granter. The range of available models has to be compared and the preferred scoring systems should include direction of credit grantors' attention to the crucially relevant variables. However, in so far as Tontines are in use across six BEAC countries, a scoring system which potentially improves on these is likely to respond to the needs of more than one of the countries. Investors within and beyond the six stand to benefit from a more stable banking system which adopts a powerful scoring system to predict the soundness and profitability of banks and their borrowers. The rest of our paper is organised as follows: section two reviews related studies; section three deals with the research methodology, section four explains the results and section five comprises the conclusion with policy recommendations and suggestions for future research.

2. Related studies

The purpose of credit scoring is to provide a concise and objective measure of a borrower's creditworthiness. Historically, Fisher [28] is the first to have used discriminant analysis to differentiate between two groups. Possibly the earliest application of applying multiple discriminant analysis is by Durand [24] who investigated car loans. Altman [62] introduced a corporate bankruptcy prediction scoring model based on five financial ratios.

Advances in information processing have fuelled progress in credit scoring techniques and applications. Conventional statistical techniques including logistic regression have been widely used and compared with non-parametric techniques such as classification and regression tree (CART) in building scoring models (e.g. [7,9,12,13,16,30,39,51,55,58,61]). Logistic regression deals with a dichotomous dependent variable which distinguishes it from a linear regression model, and makes the assumption that the probability of the dependent variable belonging to any of two different classes relies on the weight of the characteristics attached to it [1,4,5,37,41,48]. LR varies from other conventional techniques such as discriminant analysis in that it does not require the assumptions necessary for the discriminant problem [4,22]. Classification and regression tree is a tree-like decision model which is also used for classification of an object within two or more classes [18]. CART can be used to analyse either quantitative or categorical data and is widely used in building scoring models (e.g. [10,13,16,32,39,59,60]).

Advanced statistical techniques such as neural networks have been widely used in building scoring models ([1,4–6,9,18,29,38,42,55,56]. Also, by way of comparison between neural networks and other non-parametric techniques such as CART, Davis et al. [21] compared CART with Multilayer Perceptron Neural Network for credit card applications, and found comparable results for decision accuracy. Zurada and Kunene [63] found in their investigation of loan granting decisions comparable results for neural networks and decision trees across five different data-sets. A neural network is a system made of highly interconnected and interacting processing units that are based on neurobiological models mimicking the way the nervous system works. It usually consists of a three layered system comprising input, hidden, and output layers [1,4,5,33]. A Cascade Correlation Neural Network (CCNN) is a special type of neural network used for classification purposes. CCNN can avoid Multilayer Perceptrons Neural Network's drawbacks, such as the design and specification of the number of hidden layers and the number of units in these layers [19,27]. Various scoring models' evaluation criteria including receiver operating characteristic (ROC) curves and Gini coefficients are widely used and serve to assess the predictive capabilities of scoring models [2,4,11,18,20,46].

World-wide evolution of thought and practice in credit scoring can be substantially attributed to increasingly rigorous models of personal and corporate finance, increasingly powerful and discriminating statistical techniques and enormously more potent and economic processing capacity. This progress has been matched by a huge increase in the global demand for credit, not least in Africa including the BEAC family. All countries stand to benefit from wisely supervised credit's contribution to a healthy economy. Credit scoring already plays a key role in developed countries but our early investigation revealed that this is not the case for Cameroon and across BEAC, where judgemental approaches with

[2] A Tontine is a scheme in which members of a group combine resources to create a kitty [35]. Under a complex Tontine scheme the kitty is divided into lots and then auctioned. A small auction is held whereby a pre-set nominal fee is deducted from the kitty for every bid and the winner is the person ready to accept the least funds [31]. The difference between the original fund raised and the amount the member receives after the auction is a fee which is paid to the recipient of that lot at that session. The money usually has to be repaid within one or two months [35]. The fee paid by the 'beneficiary' at a particular session can be seen as interest paid on that money over the length of time before the loan is repaid. It also acts as an investment yielding a dividend for the other members since the sum of fees collected during the lending activities are then divided and distributed to the members of the Tontine at the end of each round of meetings. Despite relying solely on a tacit judgemental technique to select its members who do not even need to provide collaterals, Tontines are estimated to handle about 90 per cent of individuals' credit needs in Cameroon, and across BEAC, whereas the commercial and savings and loan banks realize a volume of about 10 per cent of all national loan business [35]. Tontines experience very high repayment rates relying on trust among members and most of all on their fear of being cast out of the Tontine.