Expert Systems with Applications 36 (2009) 11402–11417

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Genetic programming for credit scoring: The case of Egyptian public sector banks

Hussein A. Abdou *

Salford Business School, University of Salford, Salford, Greater Manchester, M5 4WT, UK

ARTICLE INFO

Keywords:
Genetic programming
Credit scoring
Weight of evidence
Egyptian public sector banks

ABSTRACT

Credit scoring has been widely investigated in the area of finance, in general, and banking sectors, in particular. Recently, *genetic programming* (GP) has attracted attention in both academic and empirical fields, especially for credit problems. The primary aim of this paper is to investigate the ability of GP, which was proposed as an extension of genetic algorithms and was inspired by the Darwinian evolution theory, in the analysis of credit scoring models in Egyptian public sector banks. The secondary aim is to compare GP with *probit analysis* (PA), a successful alternative to logistic regression, and *weight of evidence* (WOE) measure, the later a neglected technique in published research. Two evaluation criteria are used in this paper, namely, *average correct classification* (ACC) rate criterion and *estimated misclassification cost* (EMC) criterion with different *misclassification cost* (MC) ratios, in order to evaluate the capabilities of the credit scoring models. Results so far revealed that GP has the highest ACC rate and the lowest EMC. However, surprisingly, there is a clear rule for the WOE measure under EMC with higher MC ratios. In addition, an analysis of the dataset using Kohonen maps is undertaken to provide additional visual insights into cluster groupings.

1. Introduction

Credit scoring models are widely used by financial institutions, especially banks, to assign credit to good applicants and to differentiate between good and bad credit. Using credit scoring can reduce the cost of the credit process and the expected risk of being a bad loan, enhancing the credit decision, and saving time and effort (Lee, Chiu, Lu, & Chen, 2002; Ong, Huang, & Tzeng, 2005). Particularly, with the fast growth in the credit industry and the huge loan portfolio management, credit scoring is regarded as a one the most important techniques in banks and has become a very critical tool during recent decades.

It is believed that the Egyptian banking sector has been "tough" since 1999, and is "expected to remain so" for the performance of the banking sector in Egypt has shown an "ongoing profitability weakness due to revenue pressure" and a high incidence of problem loans (Central Bank of Egypt, CBE, 2006/2007; Oldham & Young, 2004). The Egyptian banking sector is being reformed to deal with this problem, which was approved in September 2004 by the President of the Arab Republic of Egypt. The main objective of this reform plan was to develop a more effective financial instrument, to strengthen the system's infrastructure, and to enhance competitiveness through increased private participation within the overall development strategy. The main pillars of the reforming plan are: firstly, banking sector consolidation and privatization through reducing the number of operating banks; secondly, financial and managerial restructuring; thirdly, solution of the bad loans problem, and finally, updating the supervision sector at the banking sector. This reforming plan also included the privatization of one of the public sector banks (CBE, 2006/2007; Oldham & Benaddi, 2005).

Egyptian banks' lending activities remarkably expanded during the last two decades. Banks' credit activities witnesses an increase, compared with the previous period, of LE28 billion (7.90%) against LE19.90 billion (6.10%) constituting LE381.80 billion or 37.40% of banks total assets and 54.50% of total deposits at the end of December, 2007. Also the pickup in foreign currency loans witnessed an increase by LE17.80 billion (16.90%) constituting LE123 billion at the end of December 2007, as well. Loans and advances exceeding one year, excluding discounts, also expanded; they went up by LE27.60 billion or 7.80%, to LE380 billion at the end on December 2007 (CBE, 2007/2008).

In view of the size of lending activities and to make efficient decisions in the granting of credit for consumer loans, it is posited that different statistical scoring techniques can be beneficially introduced to supplement the judgemental techniques, which currently are based on single numerical evaluation systems and the CBE's own perspective of creditworthiness. Indeed discussions with key banking personnel have suggested that all public sector banks in Egypt are using judgemental techniques in their evaluation process. Therefore, the role that scoring techniques can play is critical in helping to reduce the current and/or the expected risk they face; because of an inadequate risk-reduction through efficient diversification, and to support the banking sector reforming plan as currently applied.

* Tel.: +44 1612 953001; fax: +44 1612 955022.
E-mail address: h.abdou@salford.ac.uk

0957-4174/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2009.01.076

The categorisation of good and bad credit is of fundamental importance, and is indeed the objective of a credit scoring model (Lee et al., 2002; Lim & Sohn, 2007). The need of an appropriate classification technique is thus evident. But what determines the categorisation of a new applicant? Characteristics, such as marital status, income and age, have been recommended (Chen & Huang, 2003). The classification techniques themselves can also be categorised into conventional methods and advanced statistical techniques. The former include weight of evidence, multiple linear regression, discriminant analysis, probit analysis and logistic regression. The latter comprise various approaches and methods, such as, fuzzy algorithms, genetic algorithms, expert systems, and neural nets (Hand & Henley, 1997).

A few number of studies have investigated the use of WOE measure in this field, also results were comparable with those from other techniques (Abdou, 2009; Bailey, 2001; Banasik, Crook, & Thomas, 2003; Siddiqi, 2006). Furthermore, PA has been investigated, as well, and compared with other statistical scoring models (Banasik et al., 2003; Greene, 1998; Guillen & Artis, 1992); also classification results were very close to other techniques, such as logistic regression and better than discriminant analysis (Banasik et al., 2003). GP models were proposed by Koza (1992) based on Darwin's evolution theory. The use of GP applications is a rapidly growing area (Chen & Huang, 2003), and number of applications has increased during the last couple of decades, such as bankruptcy prediction (Etemadi, Rostamy, & Dehkordi, 2009; McKee & Lensberg, 2002), scoring applications (Huang, Chen, & Wang, 2007; Huang, Tzeng, & Ong, 2006), classification problems (Lensberg, Eilifsen, & McKee, 2006; Ong et al., 2005; Zhang & Bhattacharyya, 2004) and financial returns (Xia, Liu, Wang, & Lai, 2000).

However, unlike other published works, which used other statistical techniques, such as neural networks (Abdou, 2009; Lee & Chen, 2005; Malhotra & Malhotra, 2003; Oldham & Benaddi, 2005; Tsai & Wu, 2008; West, 2000), discriminant analysis and logistic regression (Elliott & Filinkov, 2008; Lee et al., 2002; Desai, Crook, & Overstreet, 1996), the focus of chosen methodologies in this paper is on two types of GP models, namely, the program model and the team model, as well as conventional techniques, such as WOE and PA. WOE has been mainly neglected in published work, yet may have much to offer, whilst PA can be a successful alternative to logistic regression (see below).

Here, the focus of the chosen environment for credit scoring investigation is upon the Egyptian public sector banks. As stated by Oldham and Young (2004), the main problems with the Egyptian banking sector exist in the large public sector banks, whose assets represent more than 50% of the whole system. The author was not aware of any other studies having investigated the use of statistical scoring models in evaluating consumer loans in whole Egyptian public sector banks. Since statistical techniques have not been used in the Egyptian public sector banks, the sample selection bias problem should be less serious compared with other studies, and this highlights the importance of the present study.

This paper is organized as follows: part two covers methodology, including data collection and sampling method; part three explains the empirical results for both the whole sample scoring models and the validated scoring models; part four compares the classification and misclassification results for different techniques; and finally, part five concludes the study results and suggests areas for future research.

2. Research methodology

In this paper, three different credit scoring modelling techniques are used in building the scoring models. The first model is the WOE measure, which is one of the earliest techniques used in credit scoring, which has a few applications in the field (Abdou, 2009; Bailey, 2001; Banasik et al., 2003). The second model is the PA model, which is also usually used with other statistical techniques for comparative purposes (Guillen & Artis, 1992; Pindyck & Rubinfeld, 1997). Finally, GP models are applied as proposed by Koza (1992) as an extension to genetic algorithms, and inspired by the Darwin's evolution theory (Koza, 1994). Here two types of GP models are used, a program model/evolved program, *which is a single program*, and a team model, *which is a combination of single programs*. The advantage of applying the team model is that the currently selected software creates this model in order to produce better results than any of the single program models can achieve.

The proposed models are discussed in Section 2.1 and the evaluation criteria in Section 2.2. The data collection, sampling method and variables' identification are discussed in Section 2.2.2. Data cases in the validated scoring sample are divided judgementally into two samples: a training dataset (67%), and a testing dataset (33%).

2.1. Proposed scoring models

2.1.1. Weight of evidence measure

The WOE measure has a long history in credit scoring models. It focuses on the odds ratio of good scores to bad scores. The information odds (IO) ratio is used to analyse the difference between two distributions without affecting the overall population. Thus, the IO equation is as follows:

$$IO = (\text{Good scores sub} - \text{classification percent})/(\text{Bad scores sub} - \text{classification percent})$$

i.e. the number of good scores within a given category as a percentage of the number of good scores for all categories, divided by the number of bad scores for a given category as a percentage of the number of bad scores for all categories.

WOE can be calculated from the IO using the logarithmic function, which can be considered as raw scores, as follows:

$$WOE = Ln(IO).$$

The information value (IV), *or total strength of the characteristics*, is used to identify the strength of different variables, as an alternative to other statistical tests, such as chi-square. IV can be calculated, from as follows:

$$IV = \Sigma[(G\% - B\%) \times WOE].$$

The importance of IV as a measure can be seen by its provision of the maximum contribution to the attributes that generate the maximum impact on the score. IV is sometimes adjusted by a discretionary multiplicative factor, for example, by multiplying by 100, or calling it Power, and multiplying by 1000 (Bailey, 2001; Siddiqi, 2006).

Bailey (2001) recommends the following values as a guideline:

Less than 0.03	:	poor prediction
From 0.03 to less than 0.10	:	weak prediction
From 0.10 to less than 0.30	:	average prediction
From 0.30 to less than 0.50	:	strong prediction
Over 0.50	:	very strong prediction

Of course, there is a subjective element attached to the categorised definitions. Nevertheless, IV is widely used in industry, because of its predictive capability. Point Score for the WOE, is determined as follows:

$$Point\ Score = \Sigma\{[P/Ln(2) \times R_W] \times [WOE + c]\}$$

where P is the score at which the odds are doubled; R_W is the correlation coefficient (from a multiple regression) between the