European Journal of Operational Research 202 (2010) 528–537

Contents lists available at ScienceDirect

European Journal of Operational Research

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

Innovative Applications of O.R.

Credit scoring for profitability objectives

Steven Finlay *

Department of Management Science, Lancaster University, Lancaster, UK

ARTICLE INFO

Article history:
Received 11 August 2008
Accepted 18 May 2009
Available online 24 May 2009

Keywords:
OR in banking
Credit scoring
Genetic algorithms
Profitability

ABSTRACT

In consumer credit markets lending decisions are usually represented as a set of classification problems. The objective is to predict the likelihood of customers ending up in one of a finite number of states, such as good/bad payer, responder/non-responder and transactor/non-transactor. Decision rules are then applied on the basis of the resulting model estimates. However, this represents a misspecification of the true objectives of commercial lenders, which are better described in terms of continuous financial measures such as bad debt, revenue and profit contribution. In this paper, an empirical study is undertaken to compare predictive models of continuous financial behaviour with binary models of customer default. The results show models of continuous financial behaviour to outperform classification approaches. They also demonstrate that scoring functions developed to specifically optimize profit contribution, using genetic algorithms, outperform scoring functions derived from optimizing more general functions such as sum of squared error.

1. Introduction

The first research into credit scoring (the application of quantitative methods to consumer risk assessment) was undertaken by Durand, who used quadratic discriminant analysis to classify credit applications as good or bad payers (Durand, 1941). Since then, the most popular approaches to consumer risk assessment have continued to treat lending decisions as binary classification problems (Hand and Henley, 1997; Thomas et al., 2002; Finlay, 2008a). Data about individuals is collected from the time lending decisions are made and their behaviour observed over a period of a few months or years. On the basis of their behaviour individuals are classified as good or bad payers. Classification or regression methods are then applied to create predictive models that are applied to new credit applications in the future. Those that the model predicts have a high likelihood of being a good payer are accepted, while those with a low likelihood are rejected. In more recent times lenders have moved on to apply classification approaches to other aspects of customer behaviour, constructing models to predict behaviours such as the likelihood to respond to a mailing, the propensity to revolve a balance on a credit card or likelihood of attrition to a rival product (Thomas et al., 2002). It is therefore, becoming increasingly common to create and use a number of different models of customer behaviour in combination to make decisions about who to lend to and on what terms.

There are a number of arguments that can be raised against applying traditional classification approaches to credit scoring problems, and in this paper two of them are explored. First, the loss function of interest to the user is often different from the loss function used during model development. This has long been recognized as an issue for forecasting problems in general (Fildes and Makridakis, 1988) and has been discussed in a number of more recent papers in relation to credit scoring (Finlay, 2005, 2009b; Hand, 2005; Hand et al., 2008; Andreeva et al., 2007). Consider logistic regression, the most popular technique used for constructing credit scoring models (Thomas et al., 2001a; Crook et al., 2007). The dependent variable, y, can take values of 1 or 0 and a model is derived that maximizes the likelihood over the set of n observed cases:

$$\prod_{i=1}^{n}\left(P_i^{y_i}(1-P_i)^{(1-y_i)}\right), \tag{1}$$

where P_i is the posterior probability that $y_i = 1$, calculated as a function of the independent variables, x_i. Yet, for many practitioners likelihood is of little interest, and the accuracy of point estimates for individual observations are not important. Instead, practitioners are primarily interested in the properties of the distribution of model scores and the accurate ranking of scores within this distribution (Thomas et al., 2001a).

The second argument against using classification approaches is more strategic in nature. The construction of binary classification models of behaviour is accepted practice within the credit scoring community, but often it is not a true representation of a commercial lender's main objectives. A model that has been constructed to minimize the misclassification of good and bad payers, based on default behaviour, is at best a crude approximation to the primary objective of identifying those customers that will generate a

* Tel.: +44 07964 316465.
E-mail addresses: steve.finlay@btinternet.com, s.finlay@lancaster.ac.uk

0377-2217/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.ejor.2009.05.025

positive contribution to profit. At worst, it might be a misspecification of the problem, if those classified as good or bad actually generate a loss or profit respectively, despite their eventual repayment classification. For example, consider a credit card account that was highly utilized and maintained a revolving balance, but defaulted for a small sum at the end of the observation period. The account would be classified as bad under most definitions of good/bad payer, but may have generated a positive contribution to profit over the period of observation.

To summarise, this paper looks at two issues concerning the appropriateness of applying binary classification methodologies to the development of credit scoring model(s):

(1) The objective function does not represent customer behaviour in terms of meaningful financial measures. Instead, crude measures such as good/bad, response/non response etc. are used as substitutes.
(2) The modelling process does not take into account prior information about the lender's decision making criteria. Instead, the objective function used to generate most credit scoring models maximizes/minimizes some generalised measure such as likelihood or the sum of squared errors.

The remainder of this paper is in three parts. First, the treatment of continuous financial measures for consumer risk assessment is discussed. Second, an empirical study is undertaken comparing the properties of continuous models of financial behaviour with a traditional good/bad payer model constructed using logistic regression. Third, a genetic algorithm (GA) is used to construct models of financial behaviour that takes into account prior information about the decision rules (the cut-off strategy) to be employed. The GA derived models are then compared with the first set of models developed using more traditional methods. It is worth pointing out that GAs have been applied to credit scoring problems before (Fogarty and Ireson, 1993; Desai et al., 1997; Yobas et al., 2000; Finlay, 2005). However, a key element that differentiates this study from previous ones is that it is the first to consider the use of a GA to optimize continuous financial measures.

1.1. The treatment of continuous financial objectives in credit scoring

This is by no means the first paper to discuss the need to consider financial measures of customer behaviour when making credit granting decisions, and a number of approaches have been proposed. The traditional approach is to weight the results produced from a binary model of behaviour with prior information about the portfolio. For example, with linear discriminant analysis (Lachenbruch, 1982) the discriminant score function, S, is defined as:

$$S_i = [x_i - 0.5(u_1 + u_2)]^T \sum^{-1}(u_1 - u_2), \tag{2}$$

where x_i is the vector for observation i with k independent variables. u_1 and u_2 are the k means vectors for goods and bads respectively. $\sum$ is the common covariance matrix.

If the objective is to use the score to accept only those likely to generate a positive contribution, the rule to assign observations to each class (the decision to accept or reject them) is augmented with average profit and loss information (Thomas et al., 2002):

Classify applicant i as good (and accept the application) if:

$$S_i > \ln\frac{LP_2}{RP_1}.$$

Otherwise classify as bad (and reject the application) where P_i is the prior probability of an observation being in class i, $i = 1$ for good payers, $i = 2$ for bad payers. L is the average loss for a bad payer, $L \geqslant 0$, R is the average return from a good payer, $R > 0$.

In general, for any method that generates probability estimates of class membership, the general form of the decision rule to optimize the lending decision will be:

$$S_i = R * P(G/x_i) - L(1 - P(G/x_i)). \tag{3}$$

Accept applicant i if $S_i > 0$: where $P(G/x_i)$ is the posterior probability of observation i being a good payer. However, this strategy is simplistic because R and L are assumed to take the same values for all model scores, and in situations where the ranking of accounts is of primary importance, there is no change to the ranking of accounts to adjust for those that generate greater/lesser contribution; i.e. the underlying model remains the same regardless of the values of R and L chosen. In practice, R and L may vary in which case a more appropriate formulation of the score function is:

$$S_i = E(R/x_i, G)P(G/x_i) - E(L/x_i, B)(1 - P(G/x_i)), \tag{4}$$

where $E(R/x, G)$ and $E(L/x, B)$ are the expected revenue/loss given x and the good/bad status respectively. Therefore, a practical implementation of (4) would involve building models to estimate $P(G/x)$, $E(R/x, G)$ and $E(L/x, B)$. However, R and L are not always easy to determine, and this may be why (4) has received little attention to date, in terms of empirical study relating to customer decision making.

Eq. (4) is an improvement on Eq. (3), but it still represents an incorrect specification of the problem. Normally, the good/bad definition is based on delinquency states. Goods and bads are considered to be mutually exclusive so that $P(G/x) + P(B/x) = 1$ (see Li and Hand (2002) and Finlay (2008b) for notable exceptions to this). Correspondingly, it is assumed that the return, R, only applies to goods and losses, L, applies only to bads. The possibility of bad payers generating a return prior to default is ignored, as is the possibility of loss from a good. This feature can be accommodated within (4) by allowing R and L to take values in the range $\pm\infty$, or by extending (4) so that the score function becomes:

$$S_i = P(G/x_i)[E(R_G/x_i, G) - E(L_G/x_i, G)] + [1 - P(G/x_i)] \times [E(R_B/x_i, B) - E(L_B/x_i, B)], \tag{5}$$

where the subscripts G and B for R and L represent the return/loss on good and bad payers respectively. However, if in the final analysis one is only interested in the return or loss generated from an account, the eventual delinquency status becomes something of a moot point. $P(G/x)$ becomes redundant, leading to a much simpler expression of the score function:

$$S_i = E(R/x_i) - E(L/x_i) = E(C/x_i), \tag{6}$$

where C is the net contribution $(R - L)$.

A number of other approaches to dealing with financial outcomes have also been proposed. Cyert et al. (1962) adopted a Markov chain approach to examine the movement between different delinquency states over time, in order to estimate the bad debt component of account contribution. The idea of using Markov processes was developed further by Thomas et al. (2001b) who proposed a profit maximization model based on a Markov processes/dynamic programming approach to determine optimal credit limits to assign to accounts. Oliver and Wells (2001) and Beling et al. (2005) suggested the definition of efficient frontier strategies based on profitability objectives, and an approach described by Thomas et al. (2002) is to develop binary models of different aspects of customer behaviour such as default, usage, acceptance and attrition. These are then used in combination to segment a population, with decisions made on the basis of the financial properties of each segment. While offering the potential to improve the decision making process, none of these approaches model continuous financial measures of behaviour directly. All of them are