Available online at www.sciencedirect.com
ScienceDirect
European Journal of Operational Research 183 (2007) 1447–1465
EUROPEAN JOURNAL OF OPERATIONAL RESEARCH
www.elsevier.com/locate/ejor

Recent developments in consumer credit risk assessment

Jonathan N. Crook [a,*], David B. Edelman [b], Lyn C. Thomas [c]

[a] *Credit Research Centre, Management School and Economics, 50 George Square, University of Edinburgh, Edinburgh EH8 9JY, United Kingdom*
[b] *Caledonia Credit Consultancy, 42 Milverton Road, Glasgow G46 7LP, United Kingdom*
[c] *School of Management, Highfield, Southampton SO17 1BJ, United Kingdom*

Received 1 July 2006; accepted 1 September 2006
Available online 15 February 2007

Abstract

Consumer credit risk assessment involves the use of risk assessment tools to manage a borrower's account from the time of pre-screening a potential application through to the management of the account during its life and possible write-off. The riskiness of lending to a credit applicant is usually estimated using a logistic regression model though researchers have considered many other types of classifier and whilst preliminary evidence suggest support vector machines seem to be the most accurate, data quality issues may prevent these laboratory based results from being achieved in practice. The training of a classifier on a sample of accepted applicants rather than on a sample representative of the applicant population seems not to result in bias though it does result in difficulties in setting the cut off. Profit scoring is a promising line of research and the Basel 2 accord has had profound implications for the way in which credit applicants are assessed and bank policies adopted.

Keywords: Finance; OR in banking; Risk analysis

1. Introduction

Between 1970 and 2005 the volume of consumer credit outstanding in the US increased by 231% and the volume of bank loans secured on real estate increased by 705%.[1] Of the $3617.0 billion of outstanding commercial bank loans secured on real estate and of consumer loans in the personal sector in December 2005, the former made up 80%.[2] The growth in debt outstanding in the UK has also been dramatic. Between 1987 and 2005 the volume of consumer credit (that is excluding mortgage debt) outstanding increased by 182% and the growth in credit card debt was 416%.[3] Mortgage debt

* Corresponding author. Tel.: +44 131 6503820; fax: +44 131 6683053.
E-mail addresses: j.crook@ed.ac.uk (J.N. Crook), davide24@ntlworld.com (D.B. Edelman), l.thomas@soton.ac.uk (L.C. Thomas).

[1] Data from the Federal Reserve Board (series bcablci_ba.m and bcablcr_ba.m deflated), H8, *Assets and Liabilities of Commercial Banks in the United States.*

[2] Data from the Federal Reserve Board (series bcablci_ba.m and bcablcr_ba.m deflated), H8, *Assets and Liabilities of Commercial Banks in the United States.*

[3] Data from ONS Online.

0377-2217/$ - see front matter
doi:10.1016/j.ejor.2006.09.100

outstanding increased by 125%.[4] Within Europe growth rates have varied. For example between 2001 and 2004 nominal loans to households and the non-profit sector increased by 36% in Italy, but in Germany they increased by only 2.3% with the Netherlands and France at 28% and 22% respectively.[5]

These generally large increases in lending and associated applications for loans have been underpinned by one of the most successful applications of statistics and operations research: credit scoring. Credit scoring is the assessment of the risk associated with lending to an organization or an individual. Every month almost every adult in the US and the UK is scored several times to enable a lender to decide whether to mail information about new loan products, to evaluate whether a credit card company should increase one's credit limit, and so on. Whilst the extension of credit goes back to Babylonian times (Lewis, 1992) the history of credit scoring begins in 1941 with the publication by Durand (1941) of a study that distinguished between good and bad loans made by 37 firms. Since then the already established techniques of statistical discrimination have been developed and an enormous number of new classificatory algorithms have been researched and tested. Virtually all major banks use credit scoring with specialised consultancies providing credit scoring services and offering powerful software to score applicants, monitor their performance and manage their accounts. In this review we firstly explain the basic ideas of credit scoring and then discuss a selection of very exciting current research topics. A number of previous surveys have been published albeit with different emphasises. Rosenberg and Gleit (1994) review different types of classifiers, (Hand and Henley, 1997) cover earlier results concerning the performance of classifiers whilst (Thomas, 2000) additionally reviews behavioural and profit scoring. Thomas et al. (2002) discuss research questions in credit scoring especially those concerning Basel 2.

2. Basic ideas of credit scoring

Consumer credit risk assessment involves the use of risk assessment tools to manage a borrower's account from the time of direct mailing of marketing material about a consumer loan through to the management of the borrower's account during its lifetime. The same techniques are used in building these tools even though they involve different information and are applied to different decisions. Application scoring helps a lender discriminate between those applicants whom the lender is confident will repay a loan or card or manage their current account properly and those applicants about whom the lender is insufficiently confident. The lender uses a rule to distinguish between these two subgroups which make up the population of applicants. Usually the lender has a sample of borrowers who applied, were made an offer of a loan, who accepted the offer and whose subsequent repayment performance has been observed. Information is available on many sociodemographic characteristics (such as income and years at address) of each borrower at the time of application from his/her application form and typically on the repayment performance of each borrower on other loans and of individuals who live in the same neighbourhood. We will denote each characteristic as x_i which may take on one of several values (called "attributes") x_{ij} for case i.

The most common method of deriving a classification rule is logistic regression where, using maximum likelihood, the analyst estimates the parameters in the equation:

$$\mathrm{Log}\left(\frac{p_{gi}}{1-p_{gi}}\right) = \beta_0 + \boldsymbol{\beta}^{\mathbf{T}}\mathbf{x_i}, \tag{1}$$

where p_{gi} is the probability that case i is a good.

This implies that the probability of case i being a good is

$$p_{gi} = \frac{e^{\boldsymbol{\beta}^{\mathbf{T}}\mathbf{x_i}}}{1+e^{\boldsymbol{\beta}^{\mathbf{T}}\mathbf{x_i}}}.$$

Traditionally the values of the characteristics for each individual were used to predict a value for p_{gi} which is compared with a critical or cut off value and a decision made. Since a binary decision is required the model is required to provide no more than a ranking. p_{gi} may be used in other ways. For example in the case of a card product, p_{gi} will also determine a multiple of salary for the credit limit. For a current account, p_{gi} will determine the number of cheques in a cheque book and the type of card that is issued. For a loan product p_{gi} will often determine the interest rate to be charged.

In practice the definition of a good varies enormously but is typically taken as a borrower who

[4] Data from the Council of Mortgage Lenders.

[5] Source: OECDStatistics, Financial Balance Sheets – consolidated dataset 710.