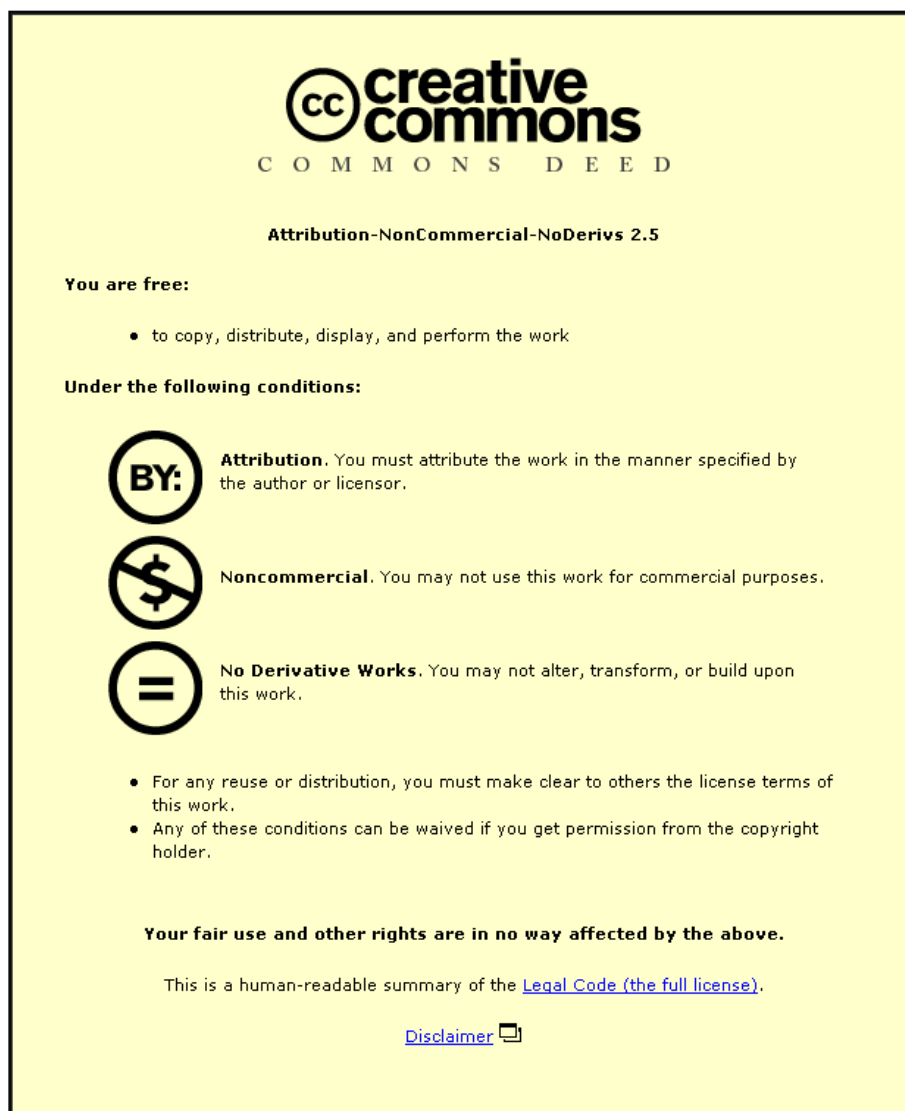


This item was submitted to Loughborough University as a PhD thesis by the author and is made available in the Institutional Repository (<https://dspace.lboro.ac.uk/>) under the following Creative Commons Licence conditions.



For the full text of this licence, please go to:
<http://creativecommons.org/licenses/by-nc-nd/2.5/>

University Library

Author/Filing Title HAMILTON, R.

.....

Class Mark T

Please note that fines are charged on ALL
overdue items.

FOR REFERENCE ONLY

0403270731



**[CREDIT] SCORING: PREDICTING, UNDERSTANDING AND EXPLAINING CONSUMER
BEHAVIOUR**

by

ROBERT HAMILTON

A Doctoral Thesis

**Submitted in partial fulfilment of the requirements
for the award of
Doctor of Philosophy
of Loughborough University**

**Loughborough University
Business School**

August 2005

© by Robert Hamilton 2005

 Loughborough University Pilkington Library
Date SEPT 2006
Class T
Acc No. 0403270731

[CREDIT] SCORING: PREDICTING, UNDERSTANDING AND EXPLAINING CONSUMER BEHAVIOUR

**By
ROBERT HAMILTON**

ABSTRACT

This thesis stems from my research into the broad area of (credit) scoring and the predicting, understanding and explaining of consumer behaviour. This research started at the University of Edinburgh on an ESRC funded project in 1988.

This work, which is being submitted as the partial fulfilment of the requirements for the award of Doctor of Philosophy of Loughborough University, consists of an introductory chapter and a selection of papers published 1991 – 2001 (inclusive). The papers address some of the key issues and areas of interest and concern arising from the rapidly evolving and expanding credit (card) market and the highly competitive nature of the credit industry. These features were particularly evident during the late 1980's and throughout the 90's

Chapter One provides a general background to the research and outlines some of the key (practical) issues involved in building a (credit) scorecard. Additionally, it provides a brief summary of each of the research papers appearing in full in Chapters 2 – 9 (inclusive) and ends with some general limitations and conclusions. The research papers appearing in Chapters 2 – 9 (inclusive) are all concerned with predicting, understanding and explaining different types of consumer behaviour in relation to the use of credit cards. For example discriminating between 'GOOD' and 'BAD' repayers of credit card debt on the basis of different definitions of good and bad, the identification of 'slow payers' using different statistical methods; examining the characteristics of credit card users and non-users, and identifying the characteristics of credit card holders most likely to return their credit card.

Keywords: Credit scoring; Behavioural scoring; Discriminant analysis, Credit cards; Scorecard

ACKNOWLEDGEMENTS

This research has taken place over a number of years and to the many people that have helped, contributed and supported my research I say a very grateful and heartfelt thank you.

In no particular order I would especially like to mention. Professor Jonathan Crook; Professor Lyn Thomas, Mr. David Edelman; Professor Barry Howcroft; Professor Ian Morison; Dr. David Coates; the ESRC, the various financial institutions who provided the data and in some cases funding for the research; the various journal editors and the anonymous referees; the secretaries who helped prepare the various articles; colleagues at the University of Edinburgh and Loughborough University Business School; the many practitioners I have met over the years and Kay Harris for carefully, patiently and diligently putting all the material together for this thesis

I would also like to thank those present at my Oral Examination: Professor Ian Davidson, Director, Loughborough University Business School for his continued support, Professor Christine Ennew and Professor Gary Akehurst for their encouragement and constructive comments

DEDICATION

To my parents and especially my mother, Elizabeth McKean Hamilton (nee Thaw), who always believed and trusted in me and was always there to support me.

To Ruth Elizabeth, my daughter, for making each and every day rewarding

To Irene for her support, encouragement and belief

GLOSSARY OF TERMS

Attribute: A set or range of values that a characteristic (variable) can attain.

Behavioural scoring: A scoring system for assessing the performance of an existing account (cardholder).

Bespoke credit scorecard: A scorecard whose development is based on the credit grantor's own experience of the product for which their use is intended. Normally this involves using the credit grantor's own data collected from the credit grantor's own accounts.

Categorical variable (characteristic): A variable that has a discrete set of possible answers

Characteristic: Any variable that could appear in a scorecard. Characteristics are made up of Attributes.

Continuous variable (characteristic): A variable whose range of possible values is numeric and very large (infinite)

Credit scoring: The term for using a linear predictive model for assessing and ranking customers or applicants for credit. Typically used more generally to include all types of predictive credit models used for decision making in the accept/reject situation.

Generic scorecard: A scorecard that has been generated when there is insufficient data to build a bespoke scorecard. These scorecards can be based upon the experience of other credit grantors and/or of another credit product.

Linear Discriminant Analysis: A statistical technique that involves deriving the linear combination of two or more independent variables (characteristics) that will discriminate best between the *a priori* defined groups (e.g. goods and bads).

Logistic Regression A logistic form of regression analysis in which the dependent variable takes one of two values, typically 0 or 1.

Revolvers. Typically credit card users that pay less than the previous months outstanding balance by the due date

Robust scorecard: A scorecard that will perform as expected for a reasonable length of time

Scorecard A table listing the characteristics that provide predictive information in the scoring system, the attributes of each characteristic and the score points (weights) associated with each attribute.

SOURCE: Various

CONTENTS

	PAGE
ABSTRACT	ii
ACKNOWLEDGEMENTS	iii
DEDICATION	iv
CERTIFICATE OF ORIGINALITY	v
GLOSSARY OF TERMS	vi–vii
CONTENTS	viii–xi
LIST OF TABLES	xii–xv
LIST OF FIGURES	xvi
 CHAPTER	
1 Introduction, Structure, Methodology and Conclusions	1
Introduction	1
Structure of the Thesis	3
General Methodology of (Credit) Scoring	4
Summary of the Research Papers	17
Conclusions	31
References	35
 2 A Comparison of Discriminators under Alternative Definitions of Credit Default	 39
Introduction	40
Data and Variables	47
Results	53
Conclusions	65
References	69

3	Methods for Credit Scoring Applied to Slow Payers	77
	Introduction	78
	Methodologies for Credit Scoring	80
	Results	87
	Conclusions	92
	References	93
4	The Degradation of the Scorecard over the Business Cycle	96
	Introduction	98
	Methodology	99
	Changes in Discriminating Functions	102
	Effects of Changes in Cut-off Scores	105
	Discussion	113
	Conclusion	114
	References	118
5	A Comparison of a Credit Scoring Model with a Credit Performance Model	121
	Introduction	122
	Variables and Methodology	124
	Results	128
	Conclusion	141
	References	147

6	Credit Card Holders: Characteristics of Users and Non-Users	150
	Introduction	151
	Data, Variables and Methodology	153
	Results	156
	Conclusions	161
	References	165
7	Credit Cards: Haves, Have-Nots and Cannot-Haves	170
	Introduction	171
	Methodology	173
	Ownership of Credit Cards	176
	Granting of Credit Cards	179
	Conclusions	182
	References	184
8	Customer Retention: A Behavioural Model	186
	Introduction	187
	Methodology	189
	Results	192
	Conclusion	197
	References	199
9	Revolving Credit Card Holders: Who are They and How Can They be Identified?	203
	Introduction	205
	Sample Period, Variable Selection and Methodology	207
	Results	212
	Conclusions and Further Research	215
	References	218

APPENDIX

A	Credit Scoring Using Discriminant Analysis: A Teacher’s Guide	221
	Introduction	222
	Building a Bespoke Credit Scorecard	223
	Using Discriminant Analysis	226
	Validation	230
	References	235
B	A Practical Approach to Maximising Customer Retention in the Credit Card Industry	237
	Introduction	239
	The Basic Elements of a Retention Strategy	240
	Methodology	243
	Results	246
	Conclusion	252
	References	259

LIST OF TABLES

		PAGE
Chapter 1		
Table 1	Typical Application Form Questions/Characteristics	8-9
Chapter 2		
Table 1	The Original 24 Predictor Variables	41
Table 2	Significance of the Estimated Function	54
Table 3	Classification Matrices	55
Table 4a	Standardised Coefficients and Structure Matrices Lax Definition	57
Table 4b	Standardised Coefficients and Structure Matrices: Stringent Definition	58
Appendix 1		
Table A1	Spouse's Income (Lax)	73
Table A2	Spouse's Income (Stringent)	73
Appendix 2		
Results for Functions without Postcode		
Table 2a	Significance of the Estimated Functions	74
Table 2b	Classification Matrices	74-75
Table 2c	Standardised Coefficients and Structure Matrices	75-76
Chapter 3		
Table 1	Results of Applying the System to the Hold-out Sample of 200	91
Table 2	Results of Applying the System to the 801 Clients used to Build the System	92

Chapter 4

Table 1	Values of the Coincident Indicator for the UK Economy	100
Table 2	The Samples	101
Table 3	Classification Matrices	103
Table 4	Standardised Canonical Discriminant Function Coefficients	104
Table 5	Total Effect Cross-Tabulation	106
Table 6	Modal Groups: Total Effects	107
Table 7	Two Cross-Tabulations	109
Table 8	Modal Groups	111-112

Appendix

Table A1	The Sociodemographic Variables	117
----------	--------------------------------	-----

Chapter 5

Table 1	Significance of the Estimated Functions	128
Table 2	Classification Matrices	129
Table 3	Standardised Coefficients	131-132
Table 4	Structure Matrices (Pooled Within Groups Correlation Coefficients)	133-134

Appendix 1	The Original 23 Predictor Variables	143
------------	-------------------------------------	-----

Appendix 2	Spouse's Income (Bads)	144
	Spouse's Income (Defaulters)	145
	Spouse's Income (Slow Payers)	146

Chapter 6

Table 1	The Original 24 Predictor Variables	153
Table 2	Variables in the Analysis	155
Table 3	Significance of the Estimated Function	156
Table 4	Classification Matrix	157
Table 5	Standardised Coefficients and Structure Matrix	159

Appendix Result for Function Excluding Postcode

Table 1(a)	Significance of the Estimated Function	168
Table 1(b)	Classification Matrix	168
Table 1(c)	Standardised Coefficients and Structure Matrix	169

Chapter 8

Table 1	Residual Discrimination and Test of Significance	192
Table 2	Classification of Results	193
Table 3	Within-Groups Structure Coefficients	194

Appendix	Twenty-Two Original Variables	202
----------	-------------------------------	-----

Chapter 9

Table 1	Description of Selected Independent Variables	212
Table 2	Ranking of Selected Variables	213
Table 3	Classification Table	214

Appendix A

Table 1	Application Form Information	225
Table 2	Residential Status	228
Table 3	Classification of Results	232

Appendix B

Table 1	Residual Discrimination and Test of Significance	247
Table 2	Classification of Results	247
Table 3	Within Groups Structure Coefficients	249
Appendix 1	Twenty-Two Original Variables	257

LIST OF FIGURES

	PAGE
Chapter 1	
Figure 1 Number of Children	11
Chapter 2	
Figure 1 Number of Children	49
Chapter 3	
Figure 1 Relationship of Credit Risk with Age	84
Figure 2 K-S Distance Between the Two Distributions	86
Figure 3 Classification Tree	89
Chapter 5	
Figure 1 Borrower Sets	125
Chapter 9	
Figure 1 Credit Card Issuer's Portfolio	206
Figure 2 Sample Period	207

CHAPTER 1

INTRODUCTION, STRUCTURE, METHODOLOGY AND CONCLUSIONS

“Credit scoring uses statistical techniques to measure the likelihood that an applicant will be a good credit risk ” (Credit Industry, 1993)

Introduction

Credit scoring¹ is the use of decision models that aid (financial) lenders in the granting of consumer credit (Thomas *et al* , 2002) and as stated above statistical techniques are used to measure the likelihood that an applicant will be a good credit risk.

The underlying assumption is that “... it is possible, using statistical techniques, to predict the future performance of groups with particular characteristics from the past performance of other groups with the same characteristics” (Credit Industry, 1993, Guide to Credit Scoring, p4). Consequently, credit scoring uses application form data relating to a large sample of existing customers each of whom, based on their own credit history will be classified as either ‘goods’ or ‘bads’ depending on the organisations pre-specified definition of ‘good’ and ‘bad’. The statistical technique used will then calculate a ‘weight’ or score for each attribute and the sum of the scores will provide an overall score for each consumer, which will then determine whether or not the consumer is predicted to be a ‘good’ or ‘bad’ risk. That is, credit scoring is predicting the future performance of consumers (i e applicants) based on the past performance of existing customers with the same characteristics

My introduction to credit scoring started in 1988 when I was a Research Associate, Department of Business Studies, University of Edinburgh working with Professor Lyn C.

¹ Credit scoring refers to the techniques that aid lenders to make a decision to accept or reject a new application for credit and will use the information from the application form, which is typically the only information they have about a new applicant

Thomas and Professor Jonathan N Crook on an Economic and Social Research Council (ESRC) funded project on Credit Scoring and Credit Control. At this time the academic literature tended to focus more on the statistical techniques used to build a scorecard (e.g. Eisenbeis, 1978; Frank, Massy and Morrison, 1965; Reichart, Cho and Wagner, 1983) rather than on the practice of credit scoring or the practical issues relating to building a scorecard.

Consequently, some of the key issues and areas of interest (most of which had not been raised or addressed in previous academic literature) covered in the research by Crook, Hamilton and Thomas included

- (i) using different definitions of 'goods' and 'bads'
- (ii) the relative importance of the various discriminating/predictor variables;
- (iii) given the nature of the data, how to satisfy the assumptions of the statistical techniques;
- (iv) the effects of total sample size and different numbers of 'goods' and 'bads';
- (v) the strengths and weaknesses of different statistical techniques;
- (vi) the 'shelf-life' of a credit scorecard;
- (vii) building a generic scorecard.

Although credit scoring as a lending tool was first discussed in the 1950's it was not until (i) the 1960's and the significant increase in the number of applications for credit from mail order firms and (ii) the 1970's and the growth in credit card applications, that credit scoring was more generally adopted as a means of speeding up the decision process (Lewis, 1994). However, the ever-growing use of credit scoring did not in itself lead to an overwhelming acceptance of the techniques. Rather, the event that ensured the acceptance of credit scoring (Thomas *et al* , 2002) was the Equal Opportunities Acts and its amendments in the U.S. in 1975 and 1976, which outlawed discrimination in the granting of credit unless it was "empirically derived and statistically valid". Another 'seal of approval' can be found in the second Guide to Credit Scoring, 1993 which states "credit scoring calculates the level of risk and reduces the element of subjectivity in lending decisions" and "is one of the most consistent, accurate and fair forms of credit assessment available".

The increasing level of acceptance of the use of statistical and modelling techniques to aid the lending decision making process has encouraged the use of scoring in other decision making areas including:

- Behavioural scoring²
- Account profitability
- Customer retention
- Collection possibilities/strategies for charged-off accounts
- Credit card fraud detection

There are a number of other factors that have also helped the growth in the use of modelling techniques and scoring to help understand, explain and predict the behaviour of potential and existing customers. These factors include the proliferation of available (cardholder) data and the falling cost of computer processing power and storage capacity (Frank, 1996b).

Given the above developments and the support of several major UK banks my research interests in this area continued at Loughborough University Business School³ when I was researching the behaviour of customers in the areas of customer retention and revolving/non-revolving credit cardholders.

Structure of the Thesis

As my research learning, interests and opportunities closely followed the developments in the credit card industry the structure of this thesis does likewise. The aims of this chapter include:

² For example, once a customer has been issued with a credit card the lender then has to decide on the customer's credit limit and this can change over time depending on how the card is being used. The techniques that aid this decision making process are called behavioural scoring.

³ Between 1990-2003 I was also an Associate Member of Loughborough University Banking Centre.

- (i) to provide a background to (credit) scoring and to my research;
- (ii) to place the research in the context of firstly my own learning experiences and secondly the developments that have taken place in the credit card industry (in relation to (credit) scoring predicting, understanding and explaining consumer behaviour) since my research started.

Therefore, the remainder of Chapter 1 includes an outline of the methodology behind building a scorecard, a summary of each of the research papers appearing in later chapters and a conclusions section that includes some general limitations of the research. Chapters 2 - 9 (inclusive) are the research papers as they appeared in the various refereed academic journals each with their own references and notes. The appendices contain other published work in this area involving R. Hamilton.

General Methodology of (Credit) Scoring⁴

Most of the research papers summarised in the next section and presented in full in Chapters 2 – 9 (inclusive) involved the building of a scorecard. Therefore, this section provides a general outline of the methodology behind the building of a scorecard especially when using one of the more commonly used statistical techniques, linear discriminant analysis⁵. This outline presents the general methodology as a process involving six stages or steps:

- | | |
|--------|--------------------|
| Step 1 | the data |
| Step 2 | weight of evidence |
| Step 3 | variable selection |
| Step 4 | multicollinearity |
| Step 5 | validation |
| Step 6 | interpretation |

⁴ Each individual paper presented in the later chapters has its own methodology section

⁵ Also see appendix A

STEP 1 *The Data*

As stated earlier, credit scoring is predicting the future performance of consumers based on the past performance of existing customers with the same characteristics in the accept/reject situation. Similarly, behavioural scoring can involve (i) predicting the future performance of existing customers based on the past performance of other existing customers with the same characteristics (e.g. predicting attrition; predicting revolving card holders) or (ii) predicting the future performance of consumers based on the past performance of existing customers with the same characteristics (e.g. target mailing/direct marketing).

In many respects credit scoring is data driven in that typically the bulk of the information that the lender has about the applicant is the information (data) requested on the application form. However, support for using socio-economic and demographic variables to predict, explain and understand consumer behaviour is grounded in microeconomic theory and the marketing literature.

The main determinants of how much a consumer will purchase, according to basic microeconomic theory (Sloman, 2003), are the own price, the number and prices of related goods, the consumer's income and tastes. Consequently, when estimating or forecasting demand organisations will, typically using a statistical technique like regression analysis, try to identify and explain the relationship between the dependent variable (e.g. sales) and the independent variables (e.g. price, advertising expenditure, age, income) using relevant socio-economic and demographic data

Similarly the use of socio-economic and demographic variables (characteristics) as proxy measures of beliefs, attitudes and intentions is to be found in the various prediction models used to predict and understand consumer behaviour in the marketing literature. The Theory of Reasoned Action (TRA), developed in 1967 was revised and expanded by Ajzen and Fishbein (1975, 1980) in the 1970s and is a well-developed and tested behavioural prediction

model to predict consumer behaviour (Karjaluoto, 2002). TRA⁶, in trying to predict a specific behaviour (see Karjaluoto, 2002) uses:

- (i) Environmental influences – physical environment, social environment and marketing environment and,
- (ii) Personal variables – values, goals, desired ends, other knowledge, beliefs and attitudes; personality traits; lifestyle patterns; demographic characteristics and; psychological characteristics.

Empirical studies of consumer decision making in relation to financial services have also made extensive use of socio-economic variables as predictors of financial behaviour (see for example, Eisenbeis, 1997, Lundy, 1992, Davis *et al* , 1992).

Therefore, fundamental to building a scorecard is the collection and use of historical (socio-economic and demographic) data and a number of key issues must be addressed in the early stages of development:

- *Defining good and bad.* generally, 'good' can be defined as behaviour that is acceptable to the lending organisation and 'bad' is behaviour that (i) the lending organisation would like to alter after accepting the customer or (ii) leads the lender to wish they had rejected the customer⁷. Therefore, as discriminant analysis involves deriving the linear combination of two (or more) independent variables that will discriminate between the *a priori* groups (Hair *et al* , 1987) the data must include one variable that allows each case to be a known member of one of the mutually exclusive and exhaustive groups (e.g. 'good' or 'bad').

⁶ Later Ajzen (1991) added a third element, the concept of perceived behavioural control, to the original theory and this addition resulted in the newer theory known as the Theory of Planned Behaviour (TPB)

⁷ Whether a case is good or bad is determined only by its performance once accepted

• *The sample* Lewis (1994) points out that while there is no magic number the result from a scorecard built on 1500 'good' and 1500 'bad' cases⁸ will be effective and robust⁹. However, when selecting a random sample of the population, several key questions need to be addressed

- (i) The population: as stated earlier the underlying assumption is that people with the same characteristics will behave in the same way. Therefore in credit scoring the sample (from existing customers) should be representative of people who might apply for credit in the future. Whereas, with behavioural scoring the sample (from existing customers) should be representative of the behaviour of existing customers.
- (ii) Time period: the objective of having a representative sample brings with it different time dimensional issues depending on the type of scoring. With credit scoring, the time period between accepting the customers (and therefore obtaining the application form details) and the scoring of new applicants should not be so long that the sample is no longer representative of new applicants. Crook *et al* (1992) considered the idea of a credit scorecard having a 'shelf-life' and this is covered in Chapter 4. With behavioural scoring, one would normally wish to differentiate between the behavioural period (i.e. during this period the variables selected will reflect how the customer has used the product in question) and the outcome period (e.g. when a certain outcome may or may not have occurred). The importance of identifying the different time periods is shown in Chapter 9 (Hamilton and Khan, 2001);
- (iii) The number of 'goods' and the number of 'bads'. as the aim is to select a sample representative of the population, theoretically the sample should have the same 'goods': 'bads' odds as the population. In most instances however, because one group will be significantly larger than the smaller group this is not desirable because

⁸ In practice much larger samples are used (Thomas, 2002)

⁹ Robust in this context means that the scorecard will perform as expected for a reasonable length of time (Lewis, 1994)

(a) it might result in too few cases being in the smaller group to build a robust model, and (b) the prior probabilities are used to obtain a rule for classifying the cases into one of the groups. Morrison, 1969 argues that the effective sample size is really governed by the smaller group. However, in practice (Thomas *et al.*, 2002) the sample tends to be either 50 50 or between 50.50 and the true population

- *Available characteristics (variables)* the characteristics or variables used to build the scorecards presented in later chapters come from a combination of (a) the customer's application form, and (b) information relating to how the customer has used the product (i.e. transaction history) For credit scoring¹⁰, most of the discriminating variables (i.e. right-hand side or independent variables) will be derived from the questions asked on the application form as this will typically be the only information the organisation will have about a new applicant¹¹, see Table 1. On the other hand, the dependent variable (or the left-hand side variable) which is the variable that determines group membership, will relate to how the existing customer has used the product (e.g. repayment history).

Table 1: Typical application form questions/characteristics

Postcode
Age
Time at present address
Residential status
Occupation
Number of children

¹⁰ Credit scoring involves building a model, based on the behaviour (to determine group membership) and the characteristics (to discriminate between group membership) of existing customers to predict the behaviour of future applicants

¹¹ In practice the lending organisation will also use information obtained from a credit reference agency or credit bureau

Number of other dependants
Home telephone
Applicant's net monthly income
Household net monthly income
Household monthly outgoings
Applicant's employment status
Years at present employment
Cards held
Bank accounts held

For behavioural scoring, in addition to considering which characteristics from the application form to use, most of the information will relate to how the customer has used the product in question (i.e. transactional characteristics). For example: number of missed payments; number of times over credit limit; payment as a percentage of balance outstanding; maximum and/or minimum balance over the time period.

- *Grouping or classing the attributes (responses)*: before the variables can be used to build a scorecard the attributes, for each characteristic, need to be grouped or coarse classified (Thomas *et al* , 2002) to form fewer classes or groups with all attributes in the same group getting the same value (e.g. weight of evidence). This is necessary because without grouping the attributes

- (a) there could be many more attributes than could be used to build a robust scorecard (Lewis, 1994);
- (b) some characteristics could have many attributes with very few cases: too few cases to allow conclusions to be drawn.

Additionally, (i) grouping or classing could help the organisation to better understand the behaviour of their customers, especially if it is performed manually¹² and (ii) for continuous characteristics, grouping or classing will render more meaningful results when adjacent attributes (values) are grouped together (e.g. age, income).

In Chapters 2 – 9 (inclusive) for both the categorical and continuous variables the grouping was performed on the basis of similarity of $g_i / (g_i + b_i)$ where g_i is the number of 'goods' with attribute i and b_i is the number of 'bads' with attribute i .

STEP 2 Weight of Evidence

One of the basic assumptions of linear discriminant analysis is that all discriminating variables are measured at the interval or ratio level of measurement (Klecka, 1980). Therefore, having already grouped the variables (characteristics) on the basis of $g_i / (g_i + b_i)$ each group, to satisfy this assumption, was then given a value based on the weight of evidence, W_{ij} (Banasik *et al* , 1995):

$$W_{ij} = \ln (g_i / b_i) + \ln (B_T / G_T)$$

$$W_{ij} = \ln (g_i - B_T / b_i - G_T)$$

$$W_{ij} = \ln (g_i / G_T / b_i / B_T)$$

- Where W_{ij} = the weight of evidence for group i for variable j
- g_i = the number of 'goods' for group i
 - b_i = the number of 'bads' for group i
 - G_T = the total number of 'good' cases in the sample

¹² While this might be viewed as the art part of credit scoring some statistics (i.e. science) can be used for guidance, see Thomas *et al*, 2004

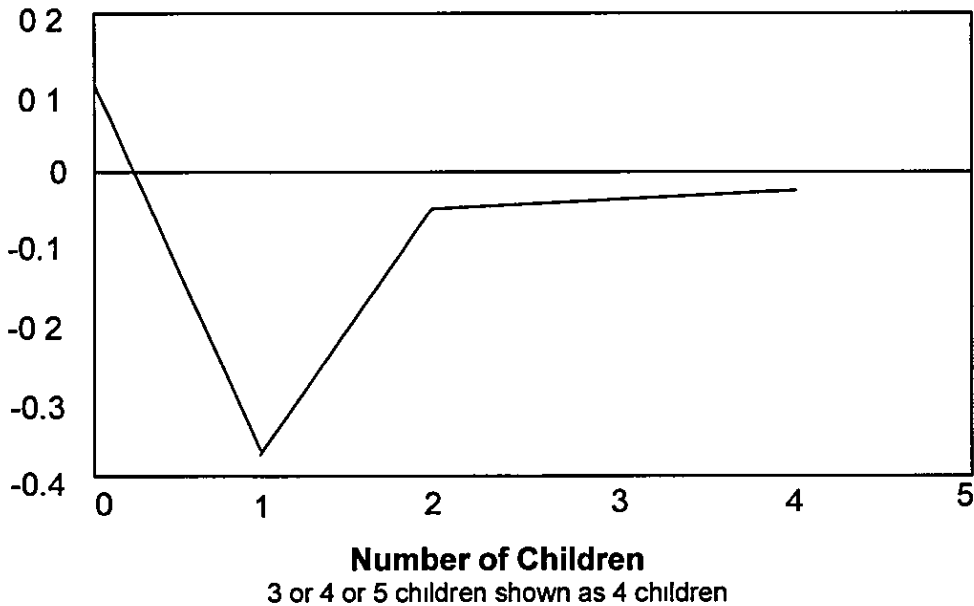
B_T = the total number of 'bad' cases in the sample

This method was selected over alternative methods (see Crook *et al* , 1991; Boyle *et al*., 1991) as it does not result in creating even more variables. For example, if one introduces binary (dummy) variables then one is creating, for each characteristic (N-1) dummy variables where N = the total number of groups

Using the weight of evidence (a measure of risk) as the value for each group, rather than the original values, also allows the relationship between risk and the characteristic to be non-monotone (i.e. need not always move in the same direction). Normal regression involving a continuous variable requires the risk will be monotone (and linear) in that variable (Thomas *et al* , 2002).

Figure 1

$\ln (g_i/b_i) + \ln (B_T/G_T)$



Source Crook *et al* , 1991

Figure 1 however, shows that in reality this is not necessarily the case: initially the measure of risk is high, then falls but rises again as the number of children increases. So using 'Number of Children' as a continuous predictor variable will be unhelpful because the number of children does not monotonically reflect risk

However, by giving each group a value based on the weight of evidence one is rearranging the groups so that they are monotone in risk but not necessarily in their original values. This rearrangement allows one to better understand, predict and explain the behaviour of consumers where the relationship between risk and the characteristic could be non-monotone

Not using the original values to derive the scorecard also has important implications for Step 6: Interpretation (see page 16).

STEP 3 Variable Selection

In the research papers where it involved building a scorecard, one of the objectives (see Hamilton and Khan, 2001) was to maximise the predictive power of the model while minimising the number of predictor variables (or characteristics). Thomas *et al* (2002) pointed out that if one aims to construct a scorecard that is both understandable and acceptable to managers it should not have much more than 20 characteristics in it. This problem of having too many variables is not so great with credit scoring where the number of potential discriminators is limited to the application form information (and any additional information obtained from a credit reference agency). However, with behavioural scoring one could start with as many as 200/300 characteristics resulting in more than 1000 attributes. Moreover, this problem will be compounded if, after grouping (coarse classifying the attributes), one has created dummy variables for each characteristic. Therefore, variable selection could involve, depending on the initial number of characteristics/attributes, as many as three stages:

- (I) initially calculating descriptive statistics (i.e. frequencies; cross-tabulations) to identify *inter alia* too many missing cases, correlation between variables, characteristics that might not be available for through-the-door consumers;
- (II) although not actually testing the discriminatory power (Thomas *et al* , 2002) using the χ^2 -statistic to aid grouping the attributes (see footnote 11) and also to help identify poor predictor characteristics;
- (III) using the stepwise method of variable selection to ensure that only the most important discriminating variables remained in the final algorithm to construct the scorecard. Stepwise selection (Norusis, 1990) combines the features of both forward entry and backward elimination in that the variable with the greatest discriminatory power is entered first, given the other variables in the equation (at the first step there are no other variables). Subsequent variables are then considered on the same basis while variables already in the model are also considered for elimination. The entry and removal criteria were set relatively high to (i) help eliminate variables (characteristics) too dependent on each other and (ii) ensure that only those variables that contributed significantly to the distance between the two groups remained in the final algorithm.

STEP 4 *Multicollinearity*

In addition to predicting risk, a common objective of the research papers was to understand and explain the behaviour of the consumers and to compare the discriminatory power of the characteristics that best discriminate between the 'goods' and 'bads'. However, when using any multivariate technique, such analysis is both difficult and potentially suspect when the independent or predictor variables are highly correlated. This problem of multicollinearity (i.e. highly correlated independent variables) can lead to estimated coefficients that are both

unstable and hard to interpret because the variables that are highly correlated¹³ are measuring almost the same thing (Morrison, 1969). For example, the estimated coefficients could have the wrong sign and/or be artificially low.

In credit scoring most of the variables (characteristics), see Table 1, relate to income and expenditure and one should therefore expect several variables to be, to a greater or lesser extent, related to one variable – income (e.g. Household income, Applicant's income). Similarly, one would imagine the relationship between the variable Age and several other variables to be significantly strong (e.g. Number of Children, Number of Other Dependants). Therefore to identify variables that were too dependent on other variables, in addition to using the stepwise method of variable selection, each independent variable was linearly regressed against the other independent variables and a measure of the degree of linear association was obtained. The measure used was $(1 - R^2_i)$ where R^2_i is the squared multiple correlation coefficient when the i th independent variable is considered the dependent variable and is regressed against all the other independent variables (Norusis, 1990).

Having identified the existence of multicollinearity other statistics (i.e. correlation coefficients, regression analysis) were used to identify which pairs or groups of variables were highly correlated and all such variables, apart from one, were removed from the equation¹⁴. This process continued until all the independent variables left in the final equation had a $(1 - R^2_i)$ value greater than 0.79. Consequently, the number of variables has been further reduced and for the remaining variables only 20% (or less) of their variation could be explained by changes in the value of other variables remaining in the model (i.e. a relatively low level of dependency).

STEP 5 *Validation*

¹³ Given the nature of the data there could be many variables that are highly correlated

¹⁴ If performed carefully removing such variables will not affect the discriminatory power of the model

To answer the question, “How well do the variables discriminate?” or to assess the predictive performance of the model, normally one uses (i) the classification matrix and (ii) a suitable Chance Criterion

However, one common source of misinterpretation (Morrison, 1969) comes if testing how predictive the model is and one is using the same sample of cases to test the model as was used to develop the model. Deriving a classification matrix on this basis can lead to an upward bias and the results obtained will be much better than if the model was tested on a completely independent sample (Thomas *et al.*, 2002) To avoid such bias the usual procedure involves using a holdout sample. Now the model is developed using a random selection of, say 80 per cent of the original sample¹⁵ (the analysis sample) and the remaining 20 per cent of the original sample (the holdout sample) are used to test the model Both samples should (a) be representative of the true population and (b) have the same proportion of ‘goods’ and ‘bads’ as the original sample.

The classification matrix is a 2 x 2 table that compares actual group membership for each case (e.g. ‘good’ or ‘bad’) with the predicted group membership for each case. In particular, the diagonal elements in this table provide the percentage of cases correctly classified by the model, which can then be compared with the percentage of cases that would be correctly classified by chance.

The appropriate chance model (Hair *et al.*, 1987) given that we are using unequal sized groups and wish to correctly classifying cases into both groups (rather than simply trying to maximise the number of cases correctly classified by allocating all cases to the largest group) is the proportional chance criterion.

$$C_{prop} = p^2 + (1 - p)^2$$

¹⁵ There are no hard and fast rules for dividing the sample (Hair *et al.*, 1987) but if dividing the sample in this way the original sample must be sufficiently large

Where p = the proportion of cases in one of the groups.

STEP 6 *Interpretation*

Having analysed the percentage correctly predicted, an aim common to all the research papers is to understand and explain the behaviour of the consumers. In this respect the output from the computer package provides certain useful statistics (Klecka, 1980):

- (i) Standardised coefficients¹⁶: these values can be used to determine which variables contribute most to determining the scores on the discriminant function;
- (ii) Pooled within groups correlations: these values also provide information with respect to the relative importance of the variables however unlike the standardised coefficients they are not affected by relationships with other variables (i.e. multicollinearity);
- (iii) Partial F (to remove) statistics: throughout the variable selection procedure variables can enter and then be removed from the function given (a) the variable's absolute contribution (i.e. it must be greater than the criterion set) and (b) its relative contribution (i.e. the other variables in the function). However, at the final step this statistic can be used to obtain the rank order of the unique discriminating power of each selected variable.

Therefore, the ranking of variables on two of the three statistics can be affected by relationships with other variables. Consistency, however, in terms of ranking across all three measures would suggest that multicollinearity is not a significant problem with the model and one could be more confident about their understanding and explanation of consumer behaviour.

¹⁶ The unstandardised coefficients are used to compute the discriminant scores for each case.

When analysing the relative importance of each characteristic it must also be remembered that the values used to discriminate between the (two) groups was the weight of evidence. As discussed earlier, this value was rarely monotonically related to the original value for each attribute (Crook *et al* , 1991). Therefore to understand and explain the behaviour of a consumer in terms of a specific characteristic (e.g. age) one must examine the weight of evidence (W_{ij}) for each individual attribute (e.g. each age group) and not the original value.

Summary of the Research Papers

Chapters 2 – 9 (inclusive) contain each of the research papers as they appear in the various refereed academic journals. This section provides a brief summary outline of the papers, identifying some key issues, aims and results. The summaries appear in the same order as they appear in Chapters 2 – 9.

- (i) Crook, J.N. , Hamilton, R. and Thomas, L.C., "A Comparison of Discriminators Under Alternative Definitions of Credit Default";
- (ii) Boyle, M , Crook, J N , Hamilton, R. and Thomas, L.C , "Methods for Credit Scoring Applied to Slow Payers";
- (iii) Crook, J.N., Thomas, L.C. and Hamilton, R , "The Degradation of the Scorecard over the Business Cycle";
- (iv) Crook, J.N., Hamilton, R. and Thomas, L C , "A Comparison of a Credit Scoring Model with a Credit Performance Model",
- (v) Crook, J.N., Hamilton, R. and Thomas, L.C., "Credit Card Holders: Characteristics of Users and Non-Users",

- (vi) Crook, J.N., Thomas, L. C. and Hamilton, R , "Credit Cards. Haves, Have-Nots and Cannot-Haves";
- (vii) Hamilton, R., Howcroft, J. B. and Saunders, J , "Customer Retention: A Behavioural Model";
- (viii) Hamilton, R. and Khan, M , "Revolving Credit Card Holders: Who Are They and How Can They Be Identified"?

(i) Crook, J. N. , Hamilton, R. and Thomas, L. C., "**A Comparison of Discriminators Under Alternative Definitions of Credit Default**", in *Credit Scoring and Credit Control*, Thomas, L.C., Crook, J.N. and Edelman, D. (eds.), Oxford University Press, December 1991, pp217-246, ISBN 0 19 853651 8.

(R. Hamilton's contribution 33%)

This paper was first presented at the conference on Credit Scoring and Credit Control, organized by the Institute of Mathematics and its Applications, University of Edinburgh, August 1989.

Earlier research in the broad area of credit scoring tended to focus on (i) the different aspects of credit granting policy and (ii) the relative attributes of different mathematical or statistical techniques for predicting consumer behaviour in relation to financial products. However, the aim of this paper was to compare the ranking of the predictor variables and the model's predictive ability when default is defined according to two different time periods (i.e. a 'stringent' criterion and a 'lax' criterion). This issue had not been previously addressed in any published work. Additionally it reinforces the importance of clearly defining the definition of 'bads' given the purpose of the scorecard.

The sample consisted of 1001 individuals who held a bank credit card (and who had used it) and the data, supplied by a financial institution, comprised of 24 variables most of which stemmed from the information obtained from the customers application form.

In order to achieve the stated aim several important issues/questions in relation to the methodology had to be addressed:

- (i) The alternative definitions of 'good' and 'bad' customers;
- (ii) The units of measurement for the predictor variables;
- (iii) Creating meaningful categories within each variable,
- (iv) Identifying the presence of multicollinearity;
- (v) How to assess the predictive performance of the model;
- (vi) The total number of 'goods' and the total number of 'bads'.

In this respect little guidance could be found in the published literature given the competitive nature of the credit card industry and the proprietary nature of credit scoring models.

The article showed that using application form data it is possible to discriminate between 'goods' and 'bads' and for both definitions of default the models correctly predicted a greater proportion of cases than would be expected by chance. Additionally, using discriminant analysis it was possible to identify the relative importance of each of the predictor variables.

(ii) Boyle, M., Crook, J. N., Hamilton, R and Thomas, L C , **"Methods for Credit Scoring Applied to Slow Payers"**, in *Credit Scoring and Credit Control*, Thomas, L C , Crook, J.N and Edelman, D. (eds), Oxford University Press, December 1991, pp75-90, ISBN 0 19 853651 8.

(R. Hamilton's contribution 25%)

This paper was also presented at the conference on Credit Scoring and Credit Control, organized by the Institute of Mathematics and its Applications, University of Edinburgh, August 1989.

Again the data used for this research came from a credit card provider and consisted of the application form information for 1001 accepted credit card holders. However, in this paper the definition of 'bad' was a credit card holder whose account was at least one month delinquent at the end of the period under consideration (i.e. a 'slow' payer). A strength of using this definition of 'bad' was that it provided a larger number of 'bads' in the sample than if the definition of 'bad' had been, for example 'ever been 3 or more months delinquent'. The aim of this paper was to identify the strengths and weaknesses of two different techniques used in credit scoring: linear discriminant analysis and recursive partitioning. One of the strengths of recursive partitioning is that it can deal with non-linear relationships between variables, linear discriminant analysis cannot. Additionally, the paper considers the benefits, in terms of the percentage correctly classified, of combining important predictor variables rather than simply using them independently. For example, using recursive partitioning, postcode and employment category were identified as two very important predictor variables which were then combined to create a new variable (instead of the two original variables) that was then, using discriminant analysis, used to build a new scorecard card.

The results of this research suggested that:

- (i) it is possible to build a model to identify 'slow' payers,
- (ii) both techniques have their own strengths,
- (iii) creating compound variables can improve the percentage correctly classified when using discriminant analysis;
- (iv) systems can be built that benefit from the strengths of both techniques

(iii) Crook, J N., Thomas, L.C. and Hamilton, R., **"The Degradation of the Scorecard over the Business Cycle"**, *IMA Journal of Mathematics Applied in Business and Industry*, 4(1), 1992, pp111-123, ISSN 09530061.

(R Hamilton's contribution 33%)

This paper was first presented at the conference on Credit Scoring and Credit Control (II), organized by the Institute of Mathematics and its Applications, University of Edinburgh, September 1991.

Typically credit scorecards are built using data relating to two or three consecutive years of usage for applications over three to five years previous. Therefore, continuing trying to understand the principles, methodologies and approaches associated with credit scoring this paper, using the same statistical technique (i.e. discriminant analysis), is looking to examine the 'shelf life' of a scorecard especially when there is a change in the state of the national economy. This involved:

- (i) building a credit scoring model for each of the two different years selected;
- (ii) comparing the default rate for each of the two years;
- (iii) examining the effects of changing the cut-off score/decision rule in terms of the proportion of applicants that would be accepted (rejected) by one model but rejected (accepted) by the other model,
- (iv) examining the characteristics of applicants that would be accepted (rejected) by one model but rejected (accepted) by the other model.

The sample used for this research contained many more cases and therefore provided significant numbers in each category for each variable. This should, in theory, make any results (more) statistically robust. Additionally for this research the data consisted of (i) credit card holders, split into non-defaulters ('goods') and defaulters ('bads') with defaulters being individuals who have missed three consecutive payments and (ii) rejected applicants. The

variables (characteristics) again came from the applicants application form and where available, information about how the credit card has been used.

The results of this research showed that:

- (i) the lending organisation would make different accept/reject decisions if different scorecards were developed using data for one year rather than another, even if the years are adjacent to each other. This stems from the having different default rates (and hence different prior probabilities) between the two years;
- (ii) even maintaining the same reject rate across different scorecards would not result in the same applicants being accepted (rejected);
- (iii) when deciding between different data the lending organisation should examine the costs associated with the two types of error (i.e. the loss in revenue of rejecting a 'good' customer and the losses associated with accepting a 'bad' customer) across the alternative scorecards.

(iv) Crook, J N , Hamilton, R. and Thomas, L.C., "**A Comparison of a Credit Scoring Model with a Credit Performance Model**", *The Service Industries Journal*, 12(4), October 1992, pp558-579

(R Hamilton's contribution 33%)

This paper recognises that within a credit card issuer's portfolio of card users (and within a given time period) one can, with respect to repayment history, identify different groups of card user. For example¹⁷:

- (i) those who have never missed a payment;
- (ii) those who have missed at least one payment;
- (iii) those who have missed three consecutive payments;

¹⁷ At this level these groups are not mutually exclusive

- (iv) those who have missed 1 or 2 payments but not 3 consecutive payments.

Therefore the aim of this paper was to investigate whether or not the characteristics of card user differed across the different groups. Specifically, three discriminant functions, (i.e. two credit scoring models and one credit performance model), were estimated using the following definitions:

- (i) 'GOODS': an individual who has never missed even one payment;
- (ii) 'DEFAULTERS' an individual who has missed three consecutive payments;
- (iii) 'SLOWS': an individual who has missed 1 or 2 or 3 consecutive payments but not necessarily three;
- (iv) 'BADs'. an individual who has missed 1 or 2 consecutive payments but never 3 consecutive payments

And the groupings for the three models were: (I) 'GOODS' and 'SLOWS'; (II) 'GOODS' and 'DEFAULTERS' and (III) 'BADs' and 'DEFAULTERS'. The first two models may assist the credit-granting organisation to decide whether or not to issue credit. Model three may be used to identify, in advance, existing customers most likely, at some time to move to becoming (three payments) delinquent having only ever missed one or two payments. The rationale for this research from the card issuers' point of view might be that when building a traditional scorecard for the accept/reject decision the definition of *bad* is normally an individual who has missed three consecutive payments. Whereas possibly the most profitable cardholder would be an individual who misses one or two consecutive payments (and therefore pays interest on the outstanding debt) but never three consecutive payments as some credit providers may pass the debt to a collection agency at that stage. The results of this work showed that the relative importance of the different variables (characteristics) in terms of their discriminating power, varied across the different models.

- (v) Crook, J N., Hamilton, R. and Thomas, L C , "**Credit Card Holders: Characteristics of Users and Non-Users**", *The Service Industries Journal*, 12(2), April 1992, pp251-262.

(R Hamilton's contribution 33%)

At the time of working on this paper certain developments and proposals were being discussed in relation to credit card services in the United Kingdom. For example

- (i) the introduction of annual fees, by some card providers, to all card holders;
- (ii) differential pricing by retailers on the basis of payment methods. For example, consumers paying by credit card might be charged a higher price than consumers using cash or cheque;
- (iii) in the period 1984-1989 the total number of credit cards in circulation was rising by an average of 10% per annum starting from 16.9 million in 1984 (The Monopolies and Merger Commission, 1989).

Within the portfolio of any credit card issuing organisation a number of distinct subsets can be identified: card holders who default, card holders who do not default and card holders who do not use the credit card issued. Therefore in light of the issues already identified, the aim of this paper was to predict those who are most likely to use, as opposed to those who would not use their credit card. Segmentation of this type might help credit providers to target their products more closely to the needs and behaviour of consumers. Additionally, card holders who do not use their card(s) could actually be costing the card issuer money in the form of issuing and administration costs.

Recognising that within the industry credit scoring techniques were (and still are) being applied to other decision-making situations this paper used the methodology outlined and discussed earlier. The definition of 'bads' in this case being a cardholder who does not use their card. Again the data used was application form information and subsequent behaviour details supplied by a UK credit card issuer.

The results show that with the aid of discriminant analysis it is possible to discriminate between the two groups of card holder (i.e. users and non-users) and that the most powerful

discriminating variables (characteristics) are: Postcode; Age of Card Holder, Applicant's Income; Years as an Account Holder; Years at Present Address; Residential Status.

For the card-issuing organisation the results suggest *inter alia* that:

- (i) using traditional credit scoring techniques it is possible to segment the market,
- (ii) they could use different promotional material for the different groups of consumer;
- (iii) it might be profitable to introduce different pricing strategies¹⁸ for the different customer segments.

(vi) Crook, J.N., Thomas, L.C. and Hamilton, R , **"Credit Cards: Haves, Have-Nots and Cannot-Haves"**, *The Service Industries Journal*, 14(2), April 1994, pp204-15, ISSN 0264 2069.

(R. Hamilton's contribution 33%)

Again trying to understand and explain the behaviour and attitudes of consumers in relation to credit cards this paper aims to investigate who has credit cards and, for those who do not have a credit card, whether or not they would be given a credit card if they applied for a credit card. The key developments in the credit card market at the time of writing were still (i) the introduction of annual charges by some card issuers and (ii) the number of credit cards in circulation increasing year on year (MMC, 1989)

To achieve the above aims two data sets were used:

Application Form Data

¹⁸ It was recently announced that a credit card issuer was to introduce a charge of £15 per annum to cardholders who fail to use their card with the definition of a *non-user* being one who fails to spend at least £250 on credit every six months, (The Sunday Express, 09/05/04)

A credit card issuer provided application form data and subsequent performance history for over 1000 credit card holders;

The Family Expenditure Survey (1986)

This is a government-backed survey of the income and expenditure pattern of UK households that for the first time in 1986 provided data relating to credit card ownership¹⁹ From the 7,178 households included in the survey 13,549 people were identified who could legally own a credit card as they were aged 18 or over. Additionally, using the income and expenditure information collected at the individual level nine variables were identified that were common to both data sets. The nine variables were residential status; length of residence at present address; outgoings; 'phone ownership; age, occupational status; current account ownership; income; and spouse's income

Therefore, using the application form data supplied by the card issuer a scorecard was constructed, based on the methodologies²⁰ outlined earlier, using the nine common variables This *generic scorecard* was then used to split the Family Expenditure Survey sample into four categories:

- (i) those who own a credit card and would get a credit card using the *generic scorecard*,
- (ii) those who do not own a credit card but would get a credit card using the *generic scorecard*;
- (iii) those who own a credit card but would be rejected using the *generic scorecard*,
- (iv) those who do not own a credit card and would be rejected using the *generic scorecard*.

¹⁹ The relevant question in the Family Expenditure Survey did not differentiate between credit card and charge card ownership. However, as they are used in similar ways, apart from repayment terms, we have treated them all as credit cards for this research.

²⁰ In constructing the scorecard the definition of 'bad' was missing three consecutive payments during the performance period.

Having constructed the scorecard the accept/reject decision depends on the cut-off score chosen and those with scores greater than the cut-off would be accepted, those below, rejected. In this research we used two different cut-off scores; one which gives a 3% rejection rate (this rate minimised the misclassification errors) and one which gives a 13% rejection rate. The latter is nearer cut-off levels used by the industry.

The results of this research suggest that:

- (i) although credit card ownership is increasing it is not uniform across all characteristics. Occupation, income and age show marked differences, in terms of card ownership between the various categories;
- (ii) the vast majority of individuals that do not have a credit card do not because they do not want one (i.e. using the *generic scorecard* and a high rejection rate, around 83% of the sample without a card would be given a card);
- (iii) the most important discriminators, when looking at who could and who could not get a credit card are phone ownership, current account ownership and income of spouse;
- (iv) the largest group who do not have credit cards because they do not want them consists of people of retirement age.

(vii) Hamilton, R., Howcroft, J. B. and Saunders, J., "**Customer Retention: A Behavioural Model**", *IMA Journal of Mathematics Applied in Business & Industry*, 6(4), (1995), pp333-342, ISSN 0953 0061.

(R. Hamilton's contribution 60%)

This paper was first presented at the conference on Credit Scoring and Credit Control (III), organized by the Institute of Mathematics and its Applications, University of Edinburgh, September 1993.

At the time, the credit card industry had been experiencing:

- (i) a fall in the number of applications being received each month,
- (ii) a constant decline in the number of credit cards held by consumers²¹,
- (iii) an increasing number of card issuers.

This consumer behaviour could at least in part be explained with reference to the introduction of annual fees, which meant that many cardholders were becoming less willing to hold more than one or two credit cards. Consequently, card issuing organisations were being more aggressive with respect to their marketing campaigns and were particularly keen to encourage not only *their* customers to retain *their* card but also for customers of other card issuers to transfer their balances.

Using data provided by a major credit card issuer in the UK the aim of this paper was to construct a behavioural scorecard to identify the characteristics and/or behaviour of customers most likely to close²² their credit card account (i.e. 'segmentation for customer retention'). The data related to the characteristics and the behaviour of a sample of 27,099 cardholders over a 15-month period and consisted of 70 variables. The methodology for this research closely followed the methodology presented earlier and ultimately resulted in 22 variables being considered for inclusion in the final model.

The results of this research showed that the scorecard performs better, as measured by the percentage correctly classified into both groups, than a chance model. Additionally, the most important predictor variables are related more to how customers use their credit card, (with respect to customer need; how the account is controlled and the relationship the card holder has with the card issuer)²³ than to their individual characteristics (or application form data).

The results, on a less positive note however, also suggested that an alternative segmentation model, where more than two groups could be identified, might be more useful. For example,

²¹ Card holders were using their card(s) more often and/or were using their card(s) for larger purchases as the value of turnover was still increasing during this period.

²² Closed in this respect refers only to customers who have made the decision to return their card without any involvement of the card issuer.

²³ The four most important discriminating variables related to (i) the customer's behavioural score (ii) interest paid in the previous year (iii) external status and (iv) circumstances of last credit limit change.

cluster analysis²⁴ would allow one to further segment cases on the basis of profitability (usage) into four groups. 'normal' high profit and low profit and 'closed' high profit and low profit

(viii) Hamilton, R and Khan, M , **"Revolving Credit Card Holders: Who Are They and How Can They Be Identified"?** *The Service Industries Journal*, Vol. 21, No.3, July 2001, pp37-48, ISSN 0264 2069.

(R. Hamilton's contribution 75%)

This paper was first presented at The Second International Stockholm Seminar on Risk Behaviour and Risk Management, Stockholm University School of Business, June 1997.

Building on previous research, this paper recognises that retaining cardholders, (see Hamilton, Howcroft and Saunders, 1995), is a necessary but not sufficient requirement to guarantee a portfolio of profitable cardholders. Arguably, cardholders should be segmented on the basis of whether or not they are likely to 'revolve' (i.e. pay interest on outstanding balances)

Database (or target) marketing, and the use of modelling techniques, had recently been introduced to play a key role in the marketing strategies of credit card issuers for several reasons, (see Frank, 1996), including

- (i) increased competition,
- (ii) the increasing availability of cardholder data;
- (iii) rising industry comfort level with scoring;
- (iv) falling data processing and storage costs.

²⁴ Discriminant analysis can be used to form more than two groups but unlike discriminant analysis, cluster analysis does not require cases to be a member of a known group.

Using two quantitative techniques more commonly associated with credit scoring (i.e. linear discriminant analysis and logistic regression) the aim of this paper was to identify the characteristics of cardholders with the greatest propensity to revolve. The rationale being, such customers will be the most profitable as they are paying interest in addition to any annual fee and, given they seem comfortable with paying interest, could be targeted with other interest charging bank products. On the other hand, 'non-revolvers' might be targeted with alternative bank products that could be more profitable or less costly to issue and administer for the card issuer (e.g. a debit card, a gold card).

A major UK bank provided data relating to a random sample of 27,681 active credit cardholders, which contained 313 socio-demographic (application form data) and behavioural predictor variables. The methodology closely followed the methodology outlined earlier although certain key differences can be identified:

- (i) this research is concerned with likely consumer behaviour within a specific time period. Consequently, the behavioural variables (predictor variables) selected for consideration reflected the consumers behaviour in one time period and the outcome (or dependent variable) reflected the consumers behaviour in a later time period (i.e. if they had paid interest on their credit card balance at least once one, two, or three months later);
- (ii) unlike other published work in this area a shortage of data was not an issue. However, given the large number of original variables Chi-square tests were initially used on all 313 variables to test the association between the dependent variable and the independent variables. This resulted in only 55 variables being considered for inclusion in the final models,
- (iii) traditionally most organisations use discriminant analysis for credit scoring. However, with the increased variety of modelling techniques used for marketing strategies the credit scoring industry has also witnessed the increasing use of logistic regression for

model building. Consequently this research used both techniques and compares the results²⁵.

The main result of this research is again that the most important discriminating (or predictor) variables relate to how the cardholder has used his/her credit card (i.e. cash advances, minimum payment due, interest paid in previous periods) rather than application form data. This would imply that segmentation of this type couldn't be built into a scoring model used at the initial accept/reject stage

Conclusions

In this chapter I have provided a background to credit scoring, outlined a general methodology, considered some of the practical issues relating to credit scoring and provided a summary of some of the key issues stemming from the research papers appearing in full in chapters 2-9 (inclusive)

The main contributions of this research include:

- Identifying, analysing and addressing some of the practical issues relating to credit/behavioural scoring rather than focusing solely on the statistical techniques. For example, sample size, defining 'goods' and 'bads'; available and suitable data; classifying the attributes; and interpreting the research output in relation to predicting, understanding and explaining consumer behaviour;
- Studying the relative importance of the various card holder characteristics, both demographic and behavioural that help to predict, understand and explain consumer behaviour,

²⁵ Both techniques provided similar results which supports the findings of Banasik *et al*, 1995, Hand and Henley, 1997

- Examining issues not previously covered in the published literature. For example: the shelf-life of a scorecard; the characteristics of credit card users (non-users); the characteristics of consumers that have/do not have/cannot have a credit card, and identifying consumers most likely to revolve (their credit balance);
- Providing a background/introduction to credit scoring for non-practitioners,
- Disseminating the researchers' understanding of credit scoring to a wider audience. This was achieved via papers in refereed academic journals, conference presentations and articles in non-refereed (industry) publications

As already highlighted, the use of statistical techniques to assist in (i) the granting or refusal or the extension of consumer credit and (ii) the understanding of consumer behaviour has been and still is a very dynamic and evolving area to research. Consequently, the research I have presented here is not exhaustive in that it does not look at the use of similar approaches and techniques in relation to, for example, the provision of mortgages; small business scoring, fraud prevention, debt recovery and customer profitability. Additionally, given the confidential nature of the data used for the research and the highly competitive nature of the credit industry certain limitations²⁶, in relation to the data used in the research presented also need to be highlighted:

- Credit Bureau (Agency) Characteristics: in addition to using application form and behavioural characteristics normally credit bureau information is also used when building a scorecard. Details (characteristics) that might be available and of relevance could include (i) the status of a customer's past and present accounts and (ii) details of any county court judgements (CCJ's);

²⁶ This is in addition to any specific limitations highlighted in any of the articles

- Refused Applications: the Credit Industry in their Guide to Credit Scoring (1993) emphasised that when building a scorecard to make decisions about the granting of credit the sample should include, when appropriate, application form information from refused applicants²⁷. However, for the reasons stated above such information was not generally included in the scorecards presented in this research although rejected applicants were included in the sample used in the research paper presented in Chapter 4, (Crook, J.N., Thomas, L.C. and Hamilton, R., 1992).
- Costs (opportunity) of Misclassification: in the various research papers the models have been validated by comparing the percentage correctly classified by the model and the appropriate chance measure (see page 15) However the classification matrix has been derived without incorporating the opportunity costs associated with a misclassification error That is, the costs to the lending organisation of classifying an individual a GOOD (bad) when he/she is actually a BAD (good). Not surprisingly, given the confidential nature of the information, the true costs to the lending organisation of such errors were unavailable.

Although credit scoring has been in common use in the financial services industry in the Western world for some five decades there are still a number of areas/issues that lenders are seeking to improve and/or address. Some of these are old, some are new, some are technique based and some are practical For example (see Thomas *et al* , 2005)

- New approaches to the classification problem (i e. what is the 'best' classification technique or method)
- Changing the objective of the classification
- How to measure the performance of a scorecard.
- How to build a scorecard for a new product with little data.
- Incorporating information about refused applicants (i.e. reject inference)

²⁷ If such applicants were not included in the sample then the sample used to build the scorecard would not reflect the through-the-door population and this causes a "reject bias" (Thomas *et al*, 2002)

- How to price the product (e.g. credit card) according to risk.
- Develop profit-based scoring systems

Additionally, some UK banks recently announced that in an attempt to (i) tackle bad debts and/or (ii) identify people who are struggling to repay their debts/loans the banks are going to share, via the main credit reference agencies more 'positive' or 'white' data. This data, unlike 'negative' data (which show customers who have missed a payment or defaulted) will identify *inter alia* customers making minimum payments; how much is spent each month, how much cash has been withdrawn. This recent development, which is also an attempt by banks to counter the suggestion that banks encourage irresponsible borrowing, raises another question: How much, if any, of this new data should be incorporated into an accept/reject scorecard?

References

- Ajzen, I., 1991, "The theory of planned behaviour", *Organizational Behavior and Human Decision Processes*, 50, pp 179-211.
- Ajzen, I And Fishbein, M , 1980, *Understanding attitudes and predicting social behaviour*, London: Prentice-Hall, pp.249-259, ISBN 0-1393644-3-9
- Bailey, M , 2001, *Credit Scoring The Principles and Practicalities*, White Box Publishing, ISBN 0 – 9540053 – 0 – 9.
- Banasik, J , Crook, J N. and Thomas, L C , 1995, "Does Scoring a Subpopulation Make a Difference?", paper presented at the Credit Scoring and Credit Control Conference, University of Edinburgh, September.
- Boyle, M., Crook, J. N , Hamilton, R. and Thomas, L C , 1991, "Methods for Credit Scoring Applied to Slow Payers", in *Credit Scoring and Credit Control*, Thomas, L.C., Crook, J N. and Edelman, D (eds), Oxford University Press, December, pp 75-90, ISBN 0 19 853651 8.
- Brown, J., 2004, ' Capital One to charge for *not* using its cards', *Your Money*, The Sunday Express, 9th May, p.9.
- Credit Industry, "Guide to Credit Scoring, 1993", Credit Industry Working Party on Credit Scoring, c/o FLA, JBL Printers
- Crook, J.N , Hamilton, R and Thomas, L.C , 1991, "A Comparison of Discriminators Under Alternative Definitions of Credit Default", in *Credit Scoring and Credit Control*, Thomas, L C , Crook, J.N. and Edelman, D. (eds.), Oxford University Press, December, pp.217-246, ISBN 0 19 853651 8.

Crook, J N , Thomas, L C and Hamilton, R , 1992, "The Degradation of the Scorecard over the Business Cycle", *IMA Journal of Mathematics Applied in Business and Industry*, 4(1), pp.111-123, ISSN 09530061

Crook, J.N., Hamilton, R. and Thomas, L.C., 1992, "Credit Card Holders: Characteristics of Users and Non-Users", *Service Industries Journal*, 12(2), April, pp.251-262.

Crook, J.N., Hamilton, R and Thomas, L.C., 1992, "A Comparison of a Credit Scoring Model with a Credit Performance Model", *Service Industries Journal*, 12(4), October, pp 558-579

Crook, J.N., Thomas, L C. and Hamilton, R., 1994, "Credit Cards. Haves, Have-Nots and Cannot-Haves", *The Service Industries Journal*, 14(2), April, pp.204-15, ISSN 0264 2069.

Davis, R H., Edelman, D. B and Gammerman, A J., 1992, "Machine-learning algorithms for credit-card applications", *Journal of Mathematics Applied in Business & Industry*, 4, pp.43-51.

Eisenbeis, R A , 1977, "Pitfalls in the Application of Discriminant Analysis in Business, Finance and Economics", *Journal of Finance*, 32(3), pp 875-900.

Eisenbeis, R.A., 1978, "Problems in Applying Discriminant Analysis in Credit Scoring Models", *Journal of Banking and Finance*, Vol 2

Fishbein, M. and Ajzen, I., 1975, *Belief, attitude, intention, and behaviour: An introduction to theory and research*, Reading, MA: Addison-Wesley.

Frank, J.N., 1996a, 'Beyond Direct Mail', *Credit Card Management*, August, pp 54-58.

Frank, J.N , 1996b, 'Scoring Takes on a New Meaning', *Credit Card Management*, September, pp.155-159.

Frank, R E , Massy, W F and Morrison, D G , 1965, "Bias in Multiple Discriminant Analysis", *Journal of Marketing Research*, Vol. 2, pp 250-258

Hamilton, R and Khan, M., 2001, "Revolving Credit Card Holders: Who Are They and How Can They Be Identified"?, *The Service Industnes Journal*, Vol 21, No.3, July, pp.37-48, ISSN 0264 2069

Hamilton, R., Howcroft, J. B. and Saunders, J , 1995, "Customer Retention: A Behavioural Model", *IMA Journal of Mathematics Applied in Business & Industry*, 6(4), pp.333-342, ISSN 0953 0061

Hair, J F , Anderson, R E. and Tatham, R.L., 1987, *Multivariate Data Analysis*, New York Collier MacMillan

Hand, D J and Henley, W E., 1997, 'Statistical Classification Methods in Consumer Credit Scoring: A Review', *Journal of the Royal Statistical Society*, Vol 160, Part 3, pp.523-41. ISBN 0 19 853651 8

Karjaluoto, H., 2002, *Electronic Banking in Finland*, Jyväskylä Studies in Business and Economics, 18

Klecka, W.R., 1980, 'Discriminant Analysis', Sage University Paper No. 19, Series: Quantitative Applications in the Social Sciences.

Lewis, E.M., 1994, *An Introduction to Credit Scoring*, The Athena Press.

Lundy, M., 1992, "Cluster Analysis in Credit Scoring", in *Credit Scoring and Credit Control*, Oxford. Clarendon Press, ISBN 0-19-853651-8

Morrison, D G , 1969, On the Interpretation of Discriminant Analysis, *Journal of Marketing Research*, Vol VI, pp.156 – 63.

Norusis, M J., 1990, SPSS Advanced Statistics User's Guide, Chicago· SPSS Inc

Reichert, A K , Cho, C.C. and Wagner, G.M., 1983, "An Examination of the Conceptual Issues Involved in Developing Credit Scoring Models", *Journal of Business and Economic Statistics*, Vol 1, No 2, pp.101-104

Sloman, J., 2003, *Economics*, 5th Edition, FT/Prentice Hall, ISBN: 0-27-365574-4

The Monopolies and Merger Commission, 1989, Credit Card Services A Report on the Supply of Credit Card Services in the United Kingdom, London; H M S O

Thomas, L C., Edelman, D B and Crook, J N , 2002, Credit Scoring and Its Applications, SIAM, ISBN 0-89871-483-4.

Thomas, L. C., Oliver, R. W. and Hand, D. J., 2005, "A survey of the issues in consumer credit modelling research", *Journal of the Operational Research Society*, 56(9), September, pp.1006-1015.

CHAPTER 2

A COMPARISON OF DISCRIMINATORS UNDER ALTERNATIVE DEFINITIONS OF CREDIT DEFAULT

J N. Crook, R Hamilton and L C Thomas

(University of Edinburgh)

in *Credit Scoring and Credit Control*, Thomas, L. C., Crook, J. N. and
Edelman, D (eds), Oxford University Press, December 1991, pp. 217-246,
ISBN 0 19 853651 8

A COMPARISON OF DISCRIMINATORS UNDER ALTERNATIVE DEFINITIONS OF CREDIT DEFAULT

J.N Crook, R Hamilton and L.C. Thomas

(University of Edinburgh)

ABSTRACT

The aim of this paper is to compare the ranking of a selection of variables in terms of their ability to discriminate between "good" and "bad" repayers of bank credit card loans under a stringent definition and a lax definition of "good" and "bad". The sample consists of 1001 cardholders. It was possible to discriminate between "goods" and "bads" with a high degree of significance on both definitions of default and both definitions gave a better predictive performance than allocating the cardholders into each group by chance. The most important discriminators for the lax function were postcode, years at bank, applicant's employment status, years at present employment, whether or not a current account is held and spouse's income respectively. In the case of the stringent definition the most important discriminators were again postcode and employment status respectively followed by mortgage balance outstanding, years at bank, number of children and years at present employment.

1 INTRODUCTION

The use of credit scoring procedures increased dramatically during the 1960s in the US and the UK and by 1979 was used by over 30% of US credit grantors [6]. This increase was partly due to the rapid growth in applications for loans and credit cards in both countries and the relative speed with which such models predicted the credit worth of applicants. In the US this was also due to the stipulations of the Equal Credit Opportunities Act (1974) and subsequent amendments which outlawed the use of race, religion, nationality, sex, marital status and age as factors to be considered in the loan decision although lending

organisations could use credit scoring methods which were 'demonstrably statistically sound' and 'empirically derived'.

TABLE 1

THE ORIGINAL 24 PREDICTOR VARIABLES

- Postcode
- Age
- Number of children
- Number of other dependants
- Whether an applicant has a home phone
- Spouse's income
- Applicant's employment status
- Applicant's employment category
- Years at present employment
- Applicant's income
- Residential Status
- Years at present address
- Estimated value of home
- Mortgage balance outstanding
- Years at bank
- Whether a current account is held
- Whether a deposit account is held
- Whether a loan account is held
- Whether a cheque guarantee card is held
- Whether a major credit card is held
- Whether a charge card is held
- Whether a store card is held
- Whether a building society card is held
- Value of outgoings

The alternative definitions used for a defaulter were as follows. A case was defined as "bad" if.

- (a) "Stringent" definition - the person had ever been one or two or three cycles delinquent during the sample period.
- (b) "Lax" definition - the person had ever been 3 cycles delinquent during the sample period

Correspondingly the definitions of "good" corresponding to (a) and (b) were:

- (a¹) the person had never been one or 2 or 3 cycles delinquent
- (b¹) the person had never been 3 cycles delinquent.

Two separate discriminant analyses were therefore carried out between (1) a and a¹ and (2) b and b¹.

The literature on credit scoring can be divided into two groups First, those papers which consider different aspects of credit granting policy and second, those which consider the relative attributes of different techniques for predicting whether a specific credit applicant will or will not default on loans made to him. One of the first aspects of policy to be considered was the optimal number of contracts to be accepted. Hence Greer [19] argued that the optimal number, X^* , was that which maximised the present value of credit related profits ⁽¹⁾ which in turn consisted of the sum of the present values of (a) profit from credit sales in the current period, (b) profit from credit sales made in future time periods and (c) profit from cash sales beyond that which would have been made if credit had not been extended. Each is decomposed into revenues and costs as a function of the number of applicants, X , and simple differentiation gives the first order condition for maximum profits. Since the probability of default is assumed to be monotonically and positively related to the number of accepted applicants, the value of X^* indicates the maximum probability of default associated with any credit application which the firm should accept.

As Eisenbeis [15] remarks, Greer's model does not give an accept/reject rule for any individual credit applicant, but is an aggregative model relating to total revenues and costs from applicants as a whole. Alternatively Greer does incorporate the possibility that credit extended in one time period may lead to greater profits in later periods.

A second issue considered by the credit policy literature is that of how to decide whether or not to grant credit to an individual applicant. One of the earliest papers is by Mehta [25] who assumes that, given the amount of information available to the decision-maker, one of three decisions can be made: accept, reject, or gain more information. The expected cost of acceptance and of rejection are each linear functions of the number of product units, n , on which credit is sought. The strategy is chosen which minimises expected cost. Since expected cost is linearly related to n there are ranges of values of n for which the cost of extension exceeds that of rejection. The investigation cost is the expected cost in the light of the information which investigation would give. For example, the investigation may give information on those items which enter the acceptance or rejection cost calculations (probability of default, average credit period, average collection cost) stratified by the past experience the firm has had with this applicant, the credit agency rating, creditor reference and so on. Now consider the case where the investigation concerned say, past experience. For the relevant range of n the expected cost for all possible findings is calculated by weighting the cost of acceptance or rejection (whichever is appropriate, given n) for each possible finding by the expected proportion of occasions on which that finding has been made. By constructing a decision tree allowing for accept, reject, investigate decisions to be made following every possible finding at each round of investigation, and calculating the expected cost of investigation at the final stage and so working towards the top of the tree, the ranges of n for which the expected cost of each stage of investigation is less than that of acceptance or rejection can be found.

Bierman and Hausman [2] have proposed methods which allowed prior probabilities of default to be revised as information as to an applicant's payment history is obtained and an applicant returns for an equal amount of credit in each future time period. Since the outcome on each occasion is that either payment is made or it is not, with probability of repayment p ,

over a number of periods the cumulative outcomes follow a binomial process. On Bayesian assumptions p follows a Beta distribution with parameters r and n . After several time periods r and n are increased according to the number of repayments made and the number of periods which have elapsed. The expected monetary value is calculated and credit granted if it is positive. Dynamic programming is used to solve the problem over a finite number of time periods. Srinivasan and Kim [33] relax Bierman and Hausman's restrictive assumption that the firm collects debts and pays all of its variable costs on the same day.

Cyert *et al* [8] proposed that repayment behaviour could be modelled by the use of Markov Chains. A matrix of probabilities (transition matrix) is constructed where each element is the probability that a customer's debt will move from being a certain period old to being another period old e.g. one month old to 0 months old. Cyert *et al*. [9] considered different transition matrices for different risk classes of applicants. Dynamic programming techniques are then used to find the profit maximising (over n periods), credit limit for each state (age of debt). Adaptive Markov Chains, whereby the probability that an individual moves from one to another state is updated in the light of past payments have also been used [35] ⁽²⁾

A further aspect of credit to be considered is the question as to which is the optimal analysis method to use. Edmeister and Scharbaum [12] formulate the expected net present value of granting loans, given N applications and analysis method S , in terms of both expected profits and losses from repayers and defaulters respectively and administrative costs. The difference between this and the expected net present value without analysis is the value of the analysis, and is maximised by choice of S .

A different group of papers consider the relative advantages of different techniques which may be used to predict whether or not an individual applicant is likely to default. Many techniques have been proposed. The oldest technique is discriminant analysis [11], [32] although Mathematical Programming [17], Recursive Partitioning and a judgemental method based on Analytic Hierarchy Process methods have been proposed (see [34] and [3] for empirical comparisons).

The literature on the application of discriminant analysis to consumer credit scoring has considered a number of issues. Chandler and Coffman [6] have summarised the differences between empirical and judgemental credit evaluation. These are that empirical methods are based on actual and not perceived performance, that empirical methods produce more consistent evaluations than judgemental methods, that empirical methods involve validation whereas judgemental methods do not, empirical methods ascribe weights to an individual's many characteristics simultaneously whereas judgemental methods tend to concentrate on a small number of characteristics at any one time.

Other papers have compared the predictive accuracy of discriminant analysis with other methods of distinguishing between "good" and "bad" accounts. For example, Myers and Forgy [27] compared the predictive accuracy of discriminant analysis, stepwise regression, equal weights for all predicting variables, and finally, separate discriminant analyses estimated from subsamples ranked according to their scores on a discriminant analysis based on the entire sample. The sample consisted of 600 accepted loan contracts on mobile homes. Analysis was based on 300 cases with the remainder used as a hold-out sample to test the predictive accuracy of each model. Twenty-one out of forty-one predicting variables were found to be predictive of account payment at the 0.05 significance level or better. The equal weight model gave the greatest prediction accuracy using the correlation coefficient between actual and predicted score as the measure of predictive accuracy. However whilst the twenty-one included variables are described, their relative importance within the estimated functions is not disclosed. Moreover the sensitivity of results to alternative definitions of "good" or "bad" is not investigated; "good" being defined as those with 'no more than two or three late payments' in a given period and "bad" as 'made less than 18 payments' or repossessed.

Wiginton [36] compared the predictive performance of a logit ⁽³⁾ model with that of a linear discriminant analysis. Whilst the discriminant analysis model's predictive performance was no better than chance (allocating all cases to the largest group) the logit models correctly predicted 62% of cases in comparison with the proportion expected by chance of only 50%. Wiginton included only three variables in the empirical analysis 'years at present

employment', 'living status'⁽⁴⁾ and 'occupation type' but the relative importance of each is not given

Chandler and Coffman [7] applied discriminant analysis to a sample of 10,000 bank credit card accounts which were one month delinquent to distinguish between (a), those accounts which were never delinquent again in 6 months and (b) those accounts which became at least 3 months delinquent within the same 6 months. The aim was to construct a performance scoring model (as opposed to a new applicant scoring model) which could predict whether an individual who had been accepted would move from the first to the second category. The predicting variables are not divulged. As an indication of predictive accuracy the authors note that of a hold-out sample of 4,700 cases, 2,000 cases had scores less than a certain number and these 2,000 cases include 62% of those who actually became at least 3 months' delinquent and 56% of those who actually became one or 2 months' delinquent.

Overstreet and Kemp [30] compared the weights applied subjectively by loan offices with those derived from a credit scoring model. Unfortunately, the reported coefficients of the discriminant analysis which gave the scoring model would appear to be the unstandardised values, and therefore they do not indicate the relative importance of each. However, the 'significant' ⁽⁵⁾ discriminators were "loan type", "length of employment", "monthly income", "monthly fixed expenses", "amount currently owed to financial institutions", "existence of loan history" and "type of loan history". This model also does not consider alternative definitions of default. Overstreet and Kemp argue that by comparing the coefficients of a scoring model with those of a loan officer, the performance of the latter can be reviewed and improved.

However, an issue which has not been addressed in any published paper is to compare the ranking of the predictor variables and the model's predictive ability when default is defined according to a 'stringent' criterion with the ranking and predictive ability when the definition of default is relatively "lax". This is the aim of this paper. This paper consists of three further sections. Section 2 describes the data and variables used, Section 3 presents and discusses the results and Section 4 concludes.

2. DATA AND VARIABLES

2.1. *Introduction*

The sample consists of 1,001 individuals who held a bank credit card and who had used it in the sample period. Data was available on 24 sociodemographic and economic variables for which an a priori reason for their use as discriminators could be given. These variables are listed in Table 1 and it can be seen that most have been included in previously published discriminant analysis scoring models (see [4]).

2.2 *Use of Nominal Discriminators*

An immediate difficulty can be seen in that many of the variables are measured only at nominal level whilst use of discriminant analysis requires that all predictor variables be measured at least at interval level [22]. The literature suggests three alternative methods of using such data.

- a) For each of n such nominal values, $(n-1)$ dummy variables which take on values $(1, 0)$ are included as predictor variables. This method has two limitations. First that the required assumption of discriminant analysis that the predictor variables are multivariate normal is violated. Second, the practical problem exists that the degrees of freedom are considerably reduced when large numbers of such variables are included.
- b) Following Krzanowski [21] for every possible combination of nominal values a discriminant function is estimated using variables measured at interval level and above as predictor variables.
- c) To replace each such variable by one measured at interval or higher levels. Hence suppose a nominal variable takes on any of m possible values and let g_i and b_i be the

number of "goods" and "bads" respectively in the sample which take on the i^{th} nominal value ($i < m$) such that

$$G_T = \sum_{i=1}^m g_i \quad \text{and} \quad B_T = \sum_{i=1}^m b_i$$

i.e. G_T and B_T are the total number of "good" and "bad" cases respectively in the sample. Clearly each of G_T , B_T , g_i and b_i are measured at ratio level. Therefore we could replace the i^{th} value of a nominal variable by a combination of g_i , b_i , G_T and B_T and obtain a ratio level variable. Boyle *et al.* [3] describe several possible combinations which are related to the probability odds or log of the probability odds of the "goods" and "bads" taking on the i^{th} value of the nominal variable

Because of the outlined limitations of methods (a) and (b) and because, for reasons to be given later in this paper, we wished to apply the same procedure to variables measured at ratio level, method (c) was adopted. Of the possible combinations outlined by Boyle *et al.*, the specific form of the predictor variables chosen was:

$$x_j^1 = \ln x_j^1 = \ln \left(\frac{g_i}{b_i} \right) + \ln \left(\frac{B_T}{G_T} \right)$$

for case j

Furthermore, for many variables, e.g. postcode, there were so many different values (seventy for postcode) that the frequency distribution of cases left very few in certain categories - in some the number of "bads" was zero. We therefore aggregated the values of the nominal variables according to similarity of $g_i/(g_i+b_i)$ and nominal categories for which there were no "bads" were combined with those categories with the highest value of $g_i/(g_i+b_i)$.

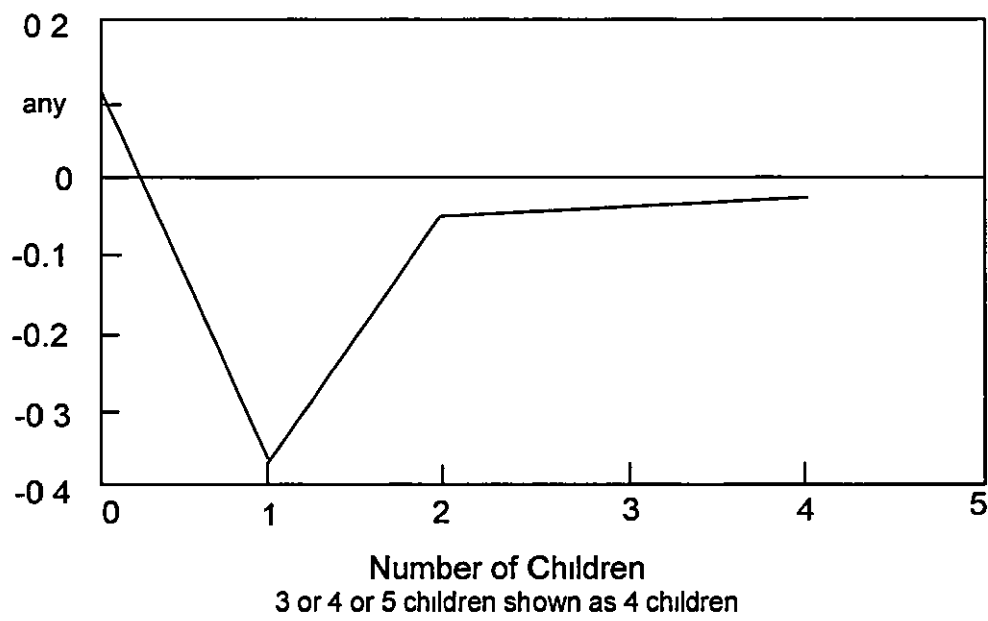
Turning to those variables which were measured at ratio level, it is sometimes the case that the proportion of "bads" is not monotone in these variables. Since the primary objective of the model is to gain maximum discrimination and prediction, not to describe, the aggregation

procedure was applied to these variables too. However, in these cases the original values of each variable were aggregated with adjacent values because on a priori grounds it seems unlikely that the probability of default would vary considerably between, say, very similar spouse's income values, and such differences in estimated probabilities $g/(g+b_i)$ were ascribed to large sampling errors due to relatively small sample sizes associated with each ratio value.

An implication of replacing the original values of ratio level variables by x^1_j values is that such variables take on values which are ranked by $\ln (g/b_i)+k$ (where k is a constant), which may not be monotonically related to the original values. For example, in the case of Number of Children under the "stringent" definition the relationship was as shown in Figure 1

FIGURE 1

$\ln (g/b_i) + \ln (B_T/G_T)$



2.3 Multicollinearity

Since the aim of the paper is to compare the ranking of variables in terms of their contribution to any discrimination between "goods" and "bads" for alternative definitions of "bads", it is particularly important to reduce the correlations between predictive variables to the extent that their coefficients become acceptably stable. If multicollinearity is high the matrix of standardised coefficients ⁽⁶⁾ is an unreliable guide to the relative contribution of each variable and the rankings of variables on this matrix will differ considerably from those on the matrix of pooled within-groups correlations between the discriminating variables and the discriminant scores (the structure matrix or 'discriminant loadings'). To reduce multicollinearity each predictor variable was, in turn, linearly regressed on the other 23 predictors and the Tolerance ($1-R^2$) was calculated in each case. Those predictors with a Tolerance of less than 0.8 (i.e. 20% or more of the variance in the variable was 'explained' by variation in the other predictors) were considered for deletion. Predictors in this group were deleted if they were not highly correlated with other predictors which were deleted. To decide which pairs of predictors were correlated we used the criterion as to whether the regression coefficient in the relevant regression equation was statistically different from zero at 5% (2 tail). We also considered the zero order bivariate correlation matrix and in this case values of at least 0.20 were taken as indicative of 'serious' collinearity. After such predictors were deleted we recalculated the Tolerances and deleted those which still had values of less than 0.8. In the case of the "stringent" definition of default (one or 2 or 3 cycles delinquent) the total list of variables selected for deletion when using the regression or regression and bivariate correlations was the same. In the case of the "lax" definition (3 cycles delinquent), use of the regression criterion implied that "current account" should be deleted whereas if the correlations are considered too then it is unclear if possession of a "cheque guarantee card" should be deleted instead. We have chosen to present the results which include "current account" rather than "cheque guarantee card" because it gives greater predictive accuracy. Hence the deleted variables were whether or not the applicant had a cheque guarantee card, applicant's employment category, years at present address, and age.

2.4 Variable Selection Criterion

To ensure that only those variables which contributed significantly to the discrimination were included in the final function, the predictors were selected by a step-wise procedure. The criterion for variable selection was the Mahalanobis Distance statistic (D^2)⁽⁷⁾. At each step the variable which results in the greatest D^2 when included, is added. Whether the change in D^2 which results from a variable's inclusion is statistically significant is tested by a partial-F test. Given the variables already in the equation the F on the change in D^2 following entry is calculated and compared with 1.0 (and the F on the change in D^2 if the variable is deleted is also compared with 1.0)⁽⁸⁾.

2.5 Assessment of Predictive Accuracy

To avoid bias in assessing the predictive performance of the model [16], the analysis was carried out on a random sample of 801 cases from the 1,001 cases and the predictive accuracy assessed from the hold-out sample. Of the remaining 200 cases, the choice of a 20% hold-out sample rather than a higher proportion was based on the desire to have the same proportion for both the "stringent" discriminant analysis and the "lax" discriminant analysis, and the fact that in the "lax" discriminant analysis, the total number of bads was only 44. If the hold-out sample had been, say, a randomly selected 50% of cases, the number of bads, on which the analysis was performed, could have been extremely low in comparison with the number of "goods". Of course the implication of a hold-out sample being a low proportion of the total sample is that the proportion of bads in the hold-out would be very low. However, we believe it was more desirable to complete the analysis on a more even split of "goods" to "bads" than the validation, although this is obviously open to question.

To assess the predictive performance of the model, the proportion of cases which is correctly classified by each function must be compared with the proportion which we would expect to be correctly classified by chance. However, two criteria for calculating the latter are available

(a) The Maximal Chance Criterion

$$C_{\max} = \text{Max } (p, 1-p)$$

where p is the proportion of cases in one of the groups e.g. "goods". That is, if over half of the cases were "good", the greatest proportion correctly classified by chance would be obtained by placing every one in the "good" category.

(b) The Proportional Chance Criterion:

$$C_{\text{prop}} = p^2 + (1-p)^2$$

The Maximal Chance Criterion is appropriate when the aim is to correctly classify the maximum proportion of cases regardless of whether they are, for example, "good" or "bad" ([20], [26]). If the function did not give a greater accuracy than this, we should allocate every case to the group with the greatest number of members. The Proportional Chance Criterion is appropriate when we wish to correctly classify cases into both groups. That is, if the membership of both groups is unequal, we wish the function to defy the odds by classifying cases correctly into the smaller group as well as the larger one.

In this paper we do not wish to maximise the proportion correctly classified regardless of whether they are "good" or "bad", but to correctly classify "bads" and "goods" and to use the chance criterion which specifically considers the proportion correctly classified by chance into both groups. Therefore we shall compare the proportion correctly classified by the model with C_{prop} .

2.6 Limitations

Certain limitations of our methodology must be acknowledged. First we did not include rejected applicants nor those who did not use their card and these omissions may possibly

lead to bias. Furthermore it is possible (given the very different sample sizes for the two groups in the "lax" case) that the covariance matrices for the two groups in each analysis may not be equal, contrary to the assumptions of linear discriminant analysis. However, in response to both criticisms, Reichert, Cho and Wagner [31] have argued that the predictive ability of linear discriminant analysis in the credit scoring context when covariance matrices differ between groups and when rejected applications are excluded from the sample is relatively robust. If the covariance matrices differ between the two groups it has been shown that the appropriate method is quadratic discriminant analysis, but this is more difficult to use because it is less robust to interactions between the variables and is less efficient as the number of predictors increases.

3 RESULTS

3.1 *Significance of the Function*

Table 2 shows the significance of each estimated function. A common test of the null hypothesis that the group means differ is to consider whether, prior to the estimation of a function, the variables would be able to discriminate between the two groups beyond the discrimination which has been achieved by earlier functions. The statistic used is Wilks' Lambda which is the ratio of the within groups sum of squares to the total sum of squares. Wilks' Lambda is inversely related to the degree of discrimination since a value close to zero indicates that the group centroids are very different relative to the within group variation. Wilks' Lambda can be converted into a χ^2 statistic ⁽⁹⁾. Table 2 shows that for both of the functions (which are not sequentially estimated) the group means are statistically different, that is that the mean score for defaulters is different from that for non-defaulters for both the "lax" and the "stringent" definitions.

TABLE 2
SIGNIFICANCE OF THE ESTIMATED FUNCTION

	Wilks' Lambda	χ^2	d f.	Significance
LAX (Ever been 3 cycles delinquent)	0.8820	99.54	12	0.000
STRINGENT (Ever been at least one cycle delinquent)	0.8367	141.2	14	0.000

3.2 Predictive Performance

Table 3 shows the predictive performance of both functions

In the case of the "stringent" definition of default the function correctly predicted 68.5% of the cases in the hold-out sample which is considerably in excess of the 52% expected by chance (and larger than the Cmax of 60%). However, the comparison with Cprop for the "lax" definition is more difficult because of the extremely dissimilar numbers of cases in the "good" and "bad" groups

Whilst the proportion correctly classified, at 98% is only percentage points above chance this is four out of a maximum possible six. In view of the grossly dissimilar membership sizes of the two groups corroborative evidence may be sought from the predictive performance of the function using the analysis sample, though we must be aware that this will bias upwards the model's performance. This supplementary evidence again suggests that the function correctly classifies four percentage points above chance, this time out of a possible nine.

TABLE 3

CLASSIFICATION MATRICES

		HOLD OUT SAMPLE			ANALYSIS SAMPLE		
		LAX DEFINITION					
		(Ever been 3 cycles delinquent)					
		Predicted Group			Predicted Group		
		Good	Bad	Total	Good	Bad	Total
Actual Group	Good	193	1	194	757	6	763
	Bad	3	3	6	32	6	38
Percentage correctly classified		98 00%			95 26%		
Cprop		94.18%			90 96%		

		HOLD OUT SAMPLE			ANALYSIS SAMPLE		
		STRINGENT DEFINITION					
		(Ever been at least 1 cycle delinquent)					
		Predicted Group			Predicted Group		
		Good	Bad	Total	Good	Bad	Total
Actual Group	Good	100	20	120	425	75	500
	Bad	43	37	80	172	129	301
Percentage correctly classified		68 50%			69 16%		
Cprop		52 00%			53 09%		

An alternative way of considering the predictive performances of the two functions might be to note that the "lax" function correctly classified 99% of the "goods" and 50% of the "bads" whereas the "stringent" function only 83% of the "goods" and 54% of the "bads", in both cases of the hold out samples.

3.3 *Rankings of the Variables*

Tables 4a and 4b show the rankings of the variables in terms of the standardised coefficients, the bivariate correlations between each predictor variable and the discriminant scores (structure coefficients), and the Partial-F statistic, for each function. Before we compare the rankings a cautionary note is in order. we are discussing the ability of values of $X^1_j = \ln (g_i/b_i) + \ln (B_T/G_T)$ (see 3 4) to distinguish between "goods" and "bads" and that for each ratio level variable the values of X^1_j are rarely monotonically related to the original values of the variable

TABLE 4a

STANDARDISED COEFFICIENTS AND STRUCTURE MATRICES

LAX DEFINITION						
(Ever been 3 cycles delinquent)						
Variable	Standardised Coefficients		Pooled Within Groups Correlations		Partial F (to remove)	
	Value	Rank	Value	Rank	Value	Rank
Postcode	0.56	1	0.499	1	29.94	1
Applicant's Employment Status	0.40	2	0.400	3	14.98	2
Years at Bank	0.37	3	0.440	2	12.02	3
Current Account	0.30	4	0.264	5	7.75	4
Spouse's Income	0.29	5	0.260	6	7.73	5
Residential Status	0.28	6	0.246	8	6.65	6
Phone	0.19	7	0.250	7	3.11	7
Years at Present Employment	0.18	8	0.295	4	2.93	8
Deposit Account	0.16	9	0.121	13	2.16	9
Estimated Value of Home	0.14	10	0.175	9	1.76	10
Outgoings	0.13	11	0.128	12	1.50	11
No of Children	0.12	12	0.095	14	1.24	12
Applicant's Income			0.164	10}	Not in function	
Mortgage Balance Outstanding			0.156	11}		
Charge Cards			0.061	15}		
Loan Account			0.053	16}		
Major Credit Cards			0.049	17}		
Store Cards			0.025	18}		
Building Society Cards			0.017	19}		
No of Other Dependents			0.008	20}		

TABLE 4b

STANDARDISED COEFFICIENTS AND STRUCTURE MATRICES

STRINGENT DEFINITION						
(Ever been at least 1 cycle delinquent)						
Variable	Standardised Coefficients		Pooled Within Groups Correlations		Partial F (to remove)	
	Value	Rank	Value	Rank	Value	Rank
Postcode	0.55	1	0.485	1	39.80	1
Applicant's Employment Status	0.44	2	0.472	2	24.30	2
No of Children	0.36	3	0.271	5	15.50	3
Residential Status	0.27	4	0.205	7	9.07	4
Mortgage Balance Outstanding	0.27	5	0.377	3	8.67	5
Years at Bank	0.24	6	0.329	4	6.59	6
Major Credit Cards	0.23	7	0.098	12	6.29	8
Outgoings	0.23	8	0.168	8	6.55	7
Years at Present Employment	0.21	9	0.256	6	5.31	9
Current Account	0.14	10	0.161	10	2.41	10
Estimated Value of Home	0.13	11	0.151	11	2.10	11
Spouse's Income	0.11	12	0.096	13	1.62	12
Charge Cards	0.11	13	0.163	9	1.30	13
Deposit Account	0.10	14	0.090	14	1.14	14
Building Society Cards			-0.068	15}	Not in function	
Store Cards			0.054	16}		
Phone			0.053	17}		
Loan Account			0.031	18}		
Applicant's Income			0.026	19}		
No of Other Dependents			-0.002	20}		

For each function separately, the ranks of the most important half dozen variables are very similar on all three criteria. Considering the "lax" definition first, the standardised coefficients place postcode as the variable with the greatest discriminating power, given the other variables in the function, followed in decreasing order of discriminatory power by applicant's employment status, years at bank, whether or not a current account is held, the level of spouse's income and residential status. The rankings on the basis of the partial-F statistics, which indicate the significance of the discrimination which that variable contributes over that contributed by the other variables in the function, are identical. But values of both of these criteria could be altered by intervariable correlation. This is not the case for the bivariate correlations between each variable and the discriminant scores. On this criterion the same variables are amongst the top six, but years at present employment is ranked fourth and not eighth as on the other two criteria, and residential status is ranked eighth. Interestingly, neither applicant's income nor the number of dependants was found to contribute significant discriminatory power beyond that contributed by variables already in the function. In terms of the correlations however, income was ranked tenth suggesting that it does discriminate between "goods" and "bads" but is slightly correlated with other variables which contribute greater discriminatory power (and so were included in the function).

Turning to the rankings for the "stringent" function, the rankings on the standardised coefficients and on the Partial-F statistics are identical. On these criteria the six variables with the greatest discriminatory power were postcode, applicant's employment status, number of children, residential status, mortgage balance outstanding and years at bank. The rankings were slightly different on the within group correlations, although the difference is mainly described by different rankings within the top six rather than including variables in this group which, on the other criteria, were outside it. The exceptions to this are years at present employment, ranked sixth on the correlation criterion rather than ninth, and residential status, ranked seventh on the correlation criterion rather than fourth.

When the rankings are compared between the two functions (and concentrating on the correlation rankings) postcode can be seen to be the most important variable in both cases

with the value of the bivariate correlations being similar. The ranking of applicant's employment status is similar and very high being second ("stringent") or third ("lax") as is years at bank (fourth and second respectively), although in this case the correlation coefficient is much higher under the "stringent" than for the "lax" definition of default. Likewise years at present employment is similarly ranked (sixth and fourth respectively) as is residential status (seventh and eighth, respectively).

However, there the similarity ends. Some variables have a markedly higher rank with greater correlations on the "stringent" criterion than on the "lax" one. Thus on the "stringent" definition, the outstanding balance on the applicant's mortgage is ranked third but is not even in the function on the "lax" definition, although it is ranked eleventh. The possession of a charge card, whilst ranked ninth on the "stringent" definition is also not included in the function on the "lax" definition. Similarly, on the "stringent" definition, number of children is ranked fifth but on the "lax" definition fourteenth, and the correlations between this variable and the discriminant scores are markedly different.

On the other hand, some variables are ranked much more highly on the "lax" definition than on the "stringent". The possession of a current account is ranked tenth on the "stringent" definition but fifth on the "lax", a similar ordering is true for spouse's income (thirteenth on the "stringent" definition and sixth on the "lax")

Interestingly, applicant's income was included in neither function because it did not contribute a significant amount of additional discriminating power beyond that contributed by the included variables. Since the degree of collinearity between the predictor variables was very low, we conclude that applicant's income has little discriminatory power in either case.

However, a limitation of these findings must be considered. This is that of the seventy postcodes for which data was available many had fewer than, say, five observations with

consequently high sampling variances for the values of g/b_i . Given that postcodes were aggregated only by similarity of $g/(g_i + b_i)$, (without regard to geographical proximity), the variance of the population values of $g/(g_i + b_i)$ between postcodes within an aggregated group is likely to be relatively high compared with that between groups. In short, postcodes may have been inappropriately aggregated and the number of "defaulters" in the holdout sample under the "lax" definition is possibly too small to assess the importance of this.

To consider this possibility further, the entire set of calculations were repeated with postcode excluded. The results are shown in Appendix 2. Briefly, the degree of discrimination is statistically significant under both definitions of default. Under the "lax" definition the proportion correctly classified at 97.50 exceeded the Cprop by 2.32 percentage points and the corresponding proportions under the "stringent" definition were identical to the function reported above in Table 4a which included postcode.

Tolerance tests under the "lax" definition led to the replacement of current account by cheque guarantee card in the group of predictors to be entered into the stepwise routine. Under the "stringent" definition the tolerance tests suggested that no replacement should be made. Turning to the rankings, under the "lax" definition the rankings of the most important seven variables were virtually identical to the results of Table 4a above. However, number of children, estimated value of home and deposit account were not included by the stepwise procedure whilst they were originally. Major credit card was included, but excluded originally. Under the "stringent" definition the rankings of the first twelve predictors were identical to the original results of Table 4b. Applicant's income replaced value of house as the least powerful discriminator included in the function.

In short, the ranking results are extremely robust with respect to the inclusion/exclusion of postcode. However, postcode is included in most commercial scoring systems and there is a valid a priori justification for its inclusion. Therefore further discussion of our results will refer to those which include this variable and are reported in Tables 4a and 4b.

3 4 *Interpretation of Variables*

Bearing in mind that the discrimination contributed by each variable has been based on the values of $X^1_j = \ln(g/b_j) + \ln(B_T/G_T)$ which it took on, we now try to interpret the above findings in terms of the untransformed values, X_j . To do this we must consider the relationships between the X^1_j values and the X_j values for each of the variables of interest.

In terms of postcode, the areas of the country which give the greatest X^1_j values are so heterogeneous that few conclusions can be drawn. In the case of employment status categories, on the "stringent" definition of default (those who missed at least one due payment) those categories which have the greatest X^1_j values are public sector employment and retired followed by government (non-military) and unemployed. The worst payers are the self-employed, and, slightly better, those who work in the private sector. On the "lax" definition of default (those who have ever been three cycles delinquent) public sector employees, the retired and government (non-military) employees are also the best payers followed by students. The worst payers, i.e. those who on average are most prone to default, are housewives, the military and the unemployed with private sector employees being only slightly better. In short, everything else equal, if it is desired to refuse credit to those who are ever likely to miss even one payment, the categories who are most likely to fall into this group are the self-employed, whilst if it is desired to refuse credit only to those who are likely to miss three consecutive payments, the categories most likely to fall into this group are housewives, the military and the unemployed.

Turning to the length of time for which an account was held at the bank, under both definitions of default the relationship between X^1_j and years is monotonic for one year and above. However, in both cases, those having an account for less than six months are less likely to default than are those with accounts for one or two years. In short, the longer the applicant has been with the bank, all else equal, the lower the chance that (s)he will either ever miss at least one payment or ever miss three in succession.

Years at present employment is also monotonically related to the proportion who ever miss a payment, (except marginally for those who have had the same job for the shortest time). Thus the chance that a payment is ever missed is negatively related to the length of time a person has been in the same job. In the case of those who miss three consecutive payments (but not less), the proportion who default is positively related to years up until 3 to 5 years and negatively related thereafter. The best payers are those who have been in the same job for at least ten years whilst the worst are those who have had the same job for 3 to 5 years.

Residential status is ranked seventh for those who have ever missed at least one payment and eighth for those who have ever missed three in succession. However, the ranking of the chance of default differs over the categories between the two definitions of default. For both definitions of default those who were most likely to miss three payments were those who were not tenants nor owners nor living with parents. However, in the case of those who missed at least one payment, this "other" category was followed by tenants in furnished accommodation. The least likely to miss at least one payment were tenants in unfurnished accommodation. On the other hand, those who were next most likely to miss three consecutive payments were tenants living in unfurnished housing, and the best payers were those living with parents. One interpretation is that those living in unfurnished accommodation rarely miss even one payment, but those who do are most likely to miss three consecutively than are those having alternative forms of accommodation.

We now consider the predictors where there is a marked difference in ranking between the two definitions of default. For both types of default, the higher the mortgage balance outstanding, the lower the proportion who avoid default. Since this predictor has the third highest discriminating power under the "stringent" definition, but has no significant incremental power on the "lax" definition, having a higher balance outstanding increases the chance that an applicant will miss at least one due payment but will not significantly increase the chance that (s)he will miss three in succession.

The number of children had a much greater correlation with the discriminant score when distinguishing between those who did and those who did not miss at least one payment than it had when distinguishing between those who did and those who did not miss three. The number of children is monotone in the proportion who miss at least one payment- the greater the number of children the greater the chance a payment is missed. But number of children is not monotonically related to X^1 , when considering the proportion of card holders who miss three cycles. This proportion is least for those without children, greatest for those with one child, and thereafter decreases as the number of children increases. So one may conclude that more children increases the chance that an applicant is likely to miss at least one payment but has much less effect on the chance that (s)he will miss three in succession, and if anything, reduces it.

Turning to spouse's income, there is no monotonic relationship between X^1 , and money income under either definition of default, as is shown in Appendix 1. However, one may note that in 72% of cases the spouse had no income and that in comparison to other income levels, for the "lax" definition, this group had a relatively high probability of repayment (except for spouses earning £15,000 plus), whilst on the "stringent" definition this group had a relatively low probability of repayment. We might therefore suggest that if the spouse earns nothing, or alternatively a relatively large amount, there is a lower chance that the applicant will miss three payments in a row than if the spouse earns an intermediate amount. But if the spouse earns nothing there is a greater chance that the applicant will miss at least one payment. We could also argue that if the spouse earns a relatively high amount, £15,000 or over, there is, on the whole, a relatively lower chance that an applicant will miss one or more consecutive payments and a relatively lower chance still that the applicant will become three cycle delinquent. Given the higher discriminating power of spouse's income in distinguishing between those who miss three consecutive payments and those who don't than in distinguishing between those who miss one or more payments and those who don't we might suggest that, whilst a high spouse's income can lead an applicant to avoid missing three consecutive payments, this is less important in leading one to avoid missing one or more payments. However, whilst the spouse earning

no income can have the same effect in terms of avoiding three cycle delinquency, this is not the case for avoiding missing at least one payment.

For both definitions of default not having a phone is associated with a higher probability of default. Therefore since having a phone is included in the "lax" function but not in the "stringent" one the results show that not having a phone is strongly associated with becoming three cycles delinquent but not with missing one or more payments.

Finally we consider credit cards held. Building Society or store cards has little effect on default probability on either definition. Having a charge card reduces the probability of an applicant missing at least one payment whilst it has no effect on the probability of missing three in succession. Alternatively, not having a major credit card increases the chances of missing at least one payment but is not associated with missing three consecutive payments.

4. CONCLUSION

We have shown that using discriminant analysis it is possible to significantly discriminate between those who miss one or more payments and those who do not, and between those who miss three consecutive payments and those who do not. In both cases our models correctly predict a greater proportion of cases correctly than would be expected by chance. Many predictors were identified, the most important being summarised as follows. Where a credit applicant lives strongly affects that chance that (s)he will miss one or more payments and that (s)he will miss three in succession. In addition the most likely to miss at least one payment ("stringent" definition of default) are the self employed, who have had an account with the bank for a year or less, who have had a job for only one year, who have at least three children and a low mortgage balance outstanding. Alternatively those most likely to miss three consecutive payments ("lax" definition of default) are (apart from living in certain

areas) housewives, military personnel and the unemployed, who have had an account with the bank for one or two years, who have been in the same job, if they have one, for three to five years, who do not have a current account and whose spouse earns £5,000 to £7,500. However we must temper these conclusions with caution in view of the limitations noted above of the method applied to these particular samples.

The support of the Economic and Social Research Council (ESRC) is gratefully acknowledged. The work was funded by ESRC under award number: R000 23 1152.

NOTES

1. Greer also formulated the model in terms of opportunity costs.
2. See Frydman *et al* [18] for evidence that a "mover - stayer" model is superior to stationary and non-stationary Markov chains
3. When using the logit model it is assumed that the cumulative density function relating the population probability of default, Π_i for case i to the values of the explanatory variables is

$$\Pi_i = 1/(1 + e^{-X_i' \beta}) \quad \dots (1)$$

where X_i and β are vectors of the explanatory variables and coefficients respectively
Using the sample values of Π_i , P_i , equation 1 implies

$$\ln P_i / (1 - P_i) = X_i' \beta + u_i$$

where u_i is a random error term. The β vector may be estimated using Generalised Least Squares.

4. "Living status" measures the same type of characteristic as our variable "residential status", although Wiginton used different nominal categories. He used "own", "rent", "live at home", and "abroad". In this study "residential status" was categorised as "owner", "with parents", "tenant furnished", "tenant unfurnished" and "other"
5. The criterion used to judge such significance is unclear.
6. The standardised coefficients, β^* , are those which result when the values of each predictor variable are divided by their standard deviation Since the units in which

two variables are measured differ by a factor of say, K, and therefore so does their standard deviations, calculating the ratio $X'_j = kX_j / k\sigma_j$ where σ_j is the standard deviation of X_j values, gives a variable X'_j which is independent of its original units. Hence the coefficient which maximises the ratio of between to within group variation when such data is used shows the relative contribution of each variable independent of its original units (see [26])

- 7 The Mahalanobis Distance is defined as

$$D_{g,b}^2 = (n - g) \sum_{i=1}^m \sum_{j=1}^m W_{ij}^* (X_{i,g} - \bar{X}_{i,g})(X_{j,g} - \bar{X}_{j,g})$$

where m = number of predictor variables in the model.

g, b = the groups of "good" and "bad" cases respectively

$\bar{X}_{i,g}$ = sample mean value of predictor i for group g

W_{ij}^* = an element from the inverse of the within group's covariance matrix.

8. An implication of a fixed value of F-to-enter and F-to-remove is that the significance of the F statistic varies as the degrees of freedom changes as the number of variables in the equation alters

9.
$$\chi^2 = \frac{(n - p + g - 1) \ln \Lambda_k}{2}$$

where p = number of predictor variables

g = number of groups

n = total number of cases

Λ_k = Wilks' Lambda after k functions have been estimated.

REFERENCES

- [1] Altman, E I , (1980), "Commercial Bank Lending Process, Credit Scoring, and Costs of Errors in Lending", *Journal of Financial and Quantitative Analysis*, Vol. 15(4), pp 813-832.
- [2] Bierman, H. and Hausman, W H , (1970), "The Credit Granting Decision", *Management Science*, Vol. 16(8), pp. B519-B532
- [3] Boyle, M., Crook, J.N., Hamilton, R and Thomas, L.C., (1989), "Methods for Credit Scoring Applied to Slow Payers", Department of Business Studies, University of Edinburgh, Working Paper.
- [4] Capon, N., (1982), "Credit Scoring Systems: A Critical Analysis", *Journal of Marketing*, Vol. 46, pp 82-91.
- [5] Carleton, W.T. and Lerner, E.M., (1969), "Statistical Credit Scoring of Municipal Bonds", *Journal of Money, Credit and Banking*, pp 750-764.
- [6] Chandler, G.C. and Coffman, J Y , (1979), "A Comparative Analysis of Empirical Vs Judgemental Credit Evaluation", *Journal of Retail Banking*, Vol. 1(2), pp. 15-25.
- [7] Chandler, G.C. and Coffman, J Y., (1983-4), "Applications of Performance Scoring to Accounts Receivable Management in Consumer Credit", *Journal of Retail Banking*, Vol 5(4), pp. 1-10.
- [8] Cyert, R.M., Davidson, H J. and Thompson, G.L., (1962), "Estimation of the Allowance for Doubtful Accounts by Markov Chains", *Management Science*, April.
- [9] Cyert, R M and Thompson, G.L., (1968), "Selecting a Portfolio of Credit Risks by Markov Chains", *Journal of Business*, pp 39-4

- [10] Dirickx, Y M I and Wakeman, L., (1975), "An Extension of the Bierman-Hausman Model for Credit Granting", *Management Science*, Vol 22(11), pp 1229-1237.
- [11] Durand, D., (1941), "Risk Elements in Consumer Instalment Financing, Financial Research Program", Study 8, Studies in Consumer Instalment Financing, NBER
- [12] Edminster, R O. and Schlarbaum G.G., (1974), "Credit Policy in Lending Institutions", *Journal of Financial and Quantitative Analysis*, Vol. 10, pp. 335-356
- [13] Eisenbeis, R.A., Gilbert, G G. and Avery, R.B., (1973), "Investigating the Relative Importance of Individual Variables and Variable Subsets in Discriminant Analysis", *Communications in Statistics*, Vol. 2(3), pp 205-219.
- [14] Eisenbeis, R.A., (1978), "Problems in Applying Discriminant Analysis in Credit Scoring Models", *Journal of Banking and Finance*, Vol 2, pp 205-219.
- [15] Eisenbeis, R.A., (1977), "Pitfalls in the Application of Discriminant Analysis in Business, Finance and Economics", *Journal of Finance*, Vol. 32, (3), pp 875-900.
- [16] Frank, R E., Massy, W F. and Morrison, D.G., (1965), "Bias in Multiple Discriminant Analysis", *Journal of Marketing Research*, Vol. 2, pp. 250-258
- [17] Freed, N and Glover, F , (1981), "Simple But Powerful Goal Programming Models for Discriminant Problems", *European Journal of Operations Research*, Vol. 7, pp. 44-60.
- [18] Frydman, H., Kallberg, J .G. and Kao, D.L., (1985), "Testing the Adequacy of Markov Chain and Mover-Stayer Models as Representations of Credit Behaviour", *Operations Research*, pp 1203-1214
- [19] Greer, C.C., (1967), "The Optimal Credit Acceptance Scheme", *Journal of*

Financial and Quantitative Analysis, Vol. 3, pp. 399-415.

- [20] Hair, J.F , Anderson, R.E. and Tatham, R L , (1987), *Multivariate Data Analysis*, New York: Collier Macmillan
- [21] Klecka, W R., (1980), *Discriminant Analysis*, Sage University Paper No 19, Series: Quantitative Applications in the Social Sciences.
- [22] Krzanowski, W.J., (1975), "Discrimination and Classification Using both Binary and Continuous Variables", *Journal of the American Statistical Association*, Vol. 70, pp.782-790.
- [23] Lachenbruch, P.A. and Goldstein, H., (1979), "Discriminant Analysis", *Biometrics*, Vol. 35, pp 69-85
- [24] Lieberman, L.R., (1972), "A Markov Decision Model for Selecting Optimal Credit Control Policies, *Management Science*, Vol. 18 (10), pp. B519-B525.
- [25] Mehta, D , (1968), "The Formulation of Credit Policy Models", *Management Science*, Vol. 15 (2), pp B30-B50.
- [26] Morrison, D G., (1969), "On the Interpretation of Discriminant Analysis", *Journal of Marketing Research*, Vol. 4, pp 156-163.
- [27] Myers, J R. and Forgy, E.W., (1963), "The Development of Numerical Credit Evaluation Systems", *Journal of the American Statistical Association*, Vol 58, pp. 799-806.
- [28] Norusis, M J /SPSS Inc (1988) *SPSS-X Advanced Statistics User's Guide*, Chicago: SPSS Inc.

- [29] Orgler, Y.E., (1970), "A Credit Scoring Model for Commercial Loans", *Journal of Money, Credit and Banking*, pp. 435-445
- [30] Overstreet, G A and Kemp, R.S., (1986), "Managerial Control in Credit Scoring Systems", *Journal of Retail Banking*, Vol. 8(182), pp 79-86.
- [31] Reichert, A K., Cho, C.C. and Wagner, G M , (1983), "An Examination of the Conceptual Issues involved in Developing Credit-Scoring Models", *Journal of Business & Economic Statistics*, Vol. 1(2,) pp 101-114
- [32] Smith, P.F., (1964), "Measuring Risk on Consumer Instalment Credit", *Management Science*, Vol 11(2), pp. 327-340.
- [33] Srinivasan, V. and Kim, Y.R., (1987), "The Bierman-Rausman Credit Granting Model A Note", *Management Science*, Vol. 33(10), pp. 1361-1362.
- [34] Srinivasan, V. and Kim, Y.R., (1987), "Credit Granting: A Comparative Analysis of Classification Procedures", *Journal of Finance*, Vol. 92(3), pp 665-683.
- [35] Thomas, L.C., "Behavioural Scoring", Department of Business Studies, University of Edinburgh, Working Paper no 88/32, 12 pp.
- [36] Wiginton, J.C., (1980), "A Note on the Comparison of Logit and Discriminant Models of Consumer Credit Behaviour", *Journal of Financial and Quantitative Analysis*, Vol. 15(3), pp. 757-770.

APPENDIX 1

Table A1 Spouse's Income (lax)

$$\ln(g/b_i) + \ln(B_T / G_T)$$

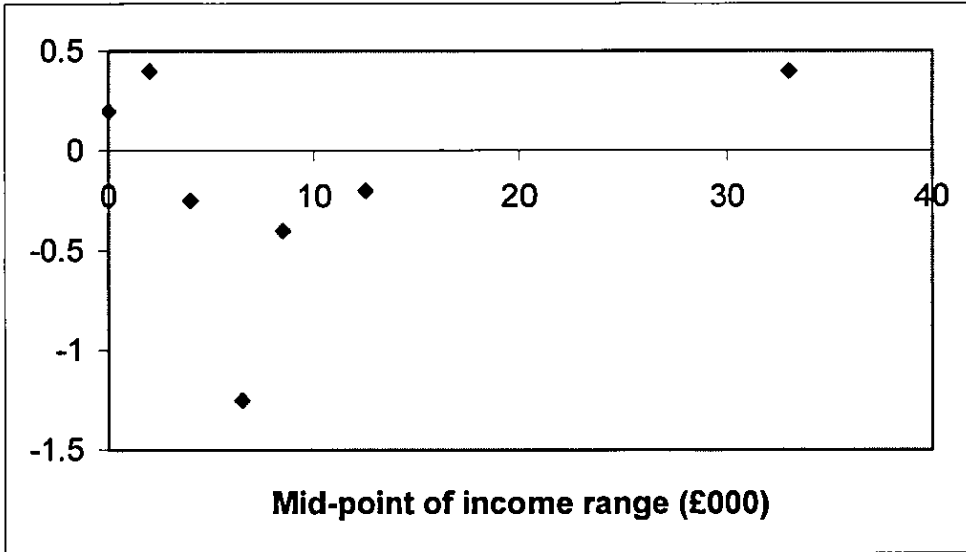
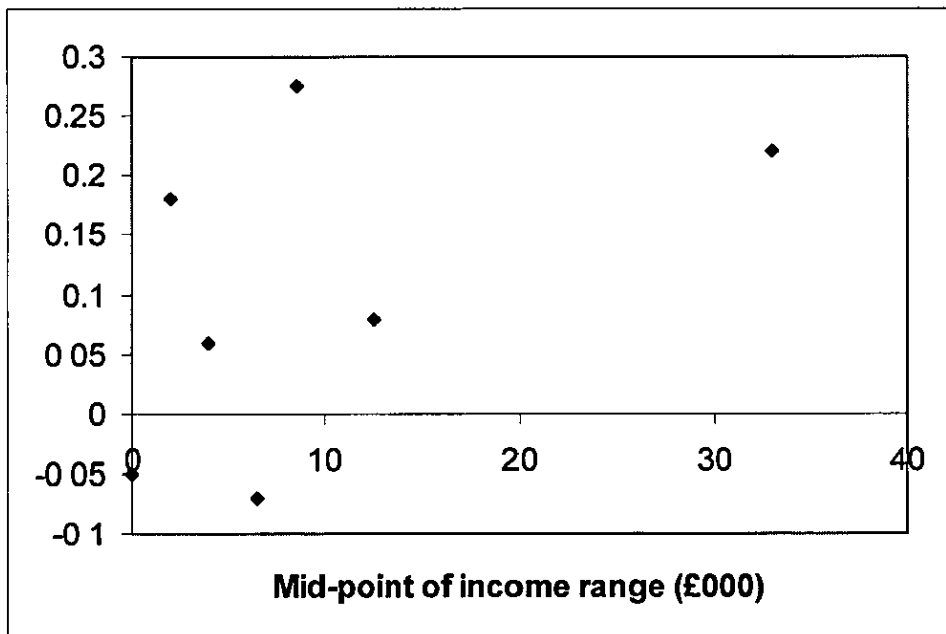


Table A2 Spouse's Income (stringent)

$$\ln(g/b_i) + \ln(B_T / G_T)$$



APPENDIX 2

RESULTS FOR FUNCTIONS WITHOUT POSTCODE

TABLE 2a

SIGNIFICANCE OF ESTIMATED FUNCTIONS

	Wilks' Lambda	χ^2	d f.	Significance
LAX (Ever been 3 cycles delinquent)	0.9197	66.55	9	0.000
STRINGENT (Ever been at least one cycle delinquent)	0.8788	102.4	13	0.000

TABLE 2b

CLASSIFICATION MATRICES

		HOLD OUT SAMPLE			ANALYSIS SAMPLE		
		LAX DEFINITION					
		(Ever been 3 cycles delinquent)					
		Predicted Group			Predicted Group		
		Good	Bad	Total	Good	Bad	Total
Actual Group	Good	194	0	194	761	2	763
	Bad	5	1	6	36	2	38
Percentage correctly classified		97.50%			95.26%		
Cprop		94.18%			90.96%		

HOLD OUT SAMPLE

ANALYSIS SAMPLE

STRINGENT DEFINITION

(Ever been at least 1 cycle delinquent)

		Predicted Group			Predicted Group		
		Good	Bad	Total	Good	Bad	Total
Actual Group	Good	102	18	120	425	75	500
	Bad	45	35	80	189	112	301
Percentage		68.50%			67.04%		
correctly classified							
Cprop		52.00%			53.09%		

Table 2c

STANDARDISED COEFFICIENTS AND STRUCTURE MATRICES

LAX DEFINITION

(Ever been 3 cycles delinquent)

	Standardised		Pooled Within		Partial F (to remove)	
	Coefficients		Group Correlations			
	Value	Rank	Value	Rank	Value	Rank
Applicant's Employment Status	0.48	1	0.495	2	14.91	1
Years at Bank	0.41	2	0.544	1	9.91	2
Spouse's Income	0.35	3	0.322	5	7.61	3
Residential Status	0.33	4	0.304	7	6.57	4
Cheque Card	0.28	5	0.405	3	4.71	5
Years at Present Employment	0.27	6	0.365	4	4.26	6
Phone	0.19	7	0.309	6	2.16	7
Outgoings	0.18	8	0.159	8	2.05	8
Major Credit Card	0.16	9	0.145	11	1.66	9
Mortgage Balance Outstanding			0.156	9}	Not in function	
Applicant's Income			0.147	10}		
Estimated Value of Home			0.102	12}		
Charge Card			0.066	13}		
Store Card			0.048	14}		
Deposit Account			0.039	15}		
Loan Account			0.037	16}		
Building Society Cards			0.036	17}		
No of Other Dependents			-0.035	18}		

STRINGENT DEFINITION

(Ever been at least one cycle delinquent)

Applicant's Employment Status	0 52	1	0 561	1	25 24	1
No of Children	0 39	2	0 322	4	13 47	2
Years at Bank	0 32	3	0 392	3	8 49	4
Mortgage Balance Outstanding	0 32	4	0 448	2	9 10	3
Residential Status	0 28	5	0 244	6	7.20	5
Major Credit Cards	0 26	6	0 116	10	5 87	7
Outgoings	0 25	7	0 200	7	5 96	6
Years at Present Employment	0 21	8	0 305	5	3 84	8
Current Account	0 18	9	0 192	9	2 82	9
Charge Card	0 17	10	0 194	8	2 65	10
Spouse's Income	0 13	11	0 114	11	1.57	11
Deposit Account	0 11	12	0 106	12	1.11	13
Applicant's Income	0 11	13	0 098	13	1 16	12
Estimated Value of Home			0 088	14}		
Phone			0 058	15}		
Store Cards			0 056	16}	Not in	
Building Society Cards			-0 053	17}	function	
No of Other Dependants			0 002	18}		
Loan Account			0 001	19}		
No of Children			0 001	19}		

CHAPTER 3

METHODS OF CREDIT SCORING APPLIED TO SLOW PAYERS

M Boyle, J N Crook, R Hamilton and L. C. Thomas

(University of Edinburgh)

in *Credit Scoring and Credit Control*, Thomas, L. C , Crook, J N. and
Edelman, D. (eds.), Oxford University Press, December 1991, pp. 75-90,
ISBN 0 19 853651 8.

METHODS FOR CREDIT SCORING APPLIED TO SLOW PAYERS

M. Boyle, J.N. Crook, R. Hamilton, and L C. Thomas

(University of Edinburgh)

ABSTRACT

The paper discusses various statistical methods used in credit scoring systems, including discriminant analysis, recursive partitioning analysis and hybrid methods which use both approaches. The methods are used to develop scoring systems to identify the slow payers in a population of credit card holders. This choice of slow as opposed to bad payers was made to lessen the effects of prior selection of the population by the credit card company. The paper points out the strengths and weaknesses of the various methods used.

1. INTRODUCTION

Credit scoring, the use of statistical techniques and mathematical models to aid the credit granting decision, has become of considerable importance in the last fifteen years. This is partly due to the rapid growth in the numbers seeking credit, especially consumer credit from credit-card companies, finance houses, mortgage companies, and partly to the legal restrictions placed on credit granters by, for example, the Equal Credit Opportunity Act of 1974 and 1976 in the United States, which made judgemental methods difficult to sustain. Hsia [9] gives a description of the Act and Chandler and Coffman [2] make a comparison of judgemental versus statistical approaches

Credit scoring techniques were first used to decide whether or not to grant credit to a new customer, but have spread to the subsequent decisions of whether to extend the credit allowed to existing customers. They are also used to decide which accounts to monitor carefully for delinquency, which methods of debt recovery to pursue, and to whom in the

client base to market a new product. The aim of this study is to compare some of the statistical techniques used in credit scoring and to point out how they can be combined to develop hybrid systems. The techniques will be compared by building scoring systems using application data and subsequent performance on 1001 applicants supplied by a credit card company.

Srinivasan and Kim [13] carried out a similar exercise at a more general level by comparing the results of five statistical and two non-statistical scoring systems using data on 215 commercial firms held by a supplier and they also concentrated on the statistical methods. Our exercise looks in more detail at the statistical techniques and since it uses consumer credit information has far more variables available and a larger sample.

The earliest statistically-based scoring system for consumer loans was a discriminant analysis system developed by Durand [4] in 1941. Myers and Forgy [11] outlined three versions of 'discriminant analysis' which are used in credit scoring. Eisenbeis, Gilbert and Avery [6] discuss methods of determining which of the variables in the application information should be part of a discriminant analysis scoring system. Eisenbeis [5] focuses on some of the problems in applying such scoring systems and what should be the objectives of a credit scoring system. As Eisenbeis points out most systems concentrate only on default rates, whereas profit maximisation might be a more appropriate criterion, though difficult to quantify. Eisenbeis also identifies other problems in applying discriminant analysis to credit scoring, namely the non-normality of the variable involved, the inequality in variance between the subgroups of acceptable and non-acceptable credit risks, difficulties in deciding which variables to remove from the analysis and the problem that the sample of credit histories used to develop the scoring system is usually censored in that not all previous applicants for credit were granted it. Reichert, Cho and Wagner [12] took an empirical approach to testing a discriminant analysis-based scoring system and the authors came to the conclusion that the system was fairly robust and relatively insensitive to a number of the assumptions which theoretically discriminant analysis requires but which are not usually satisfied in credit granting data.

Wiginton [14] performed a comparison of a discriminant analysis scoring system and a logit-based one, using oil company consumer credit data but concluded both systems were unsatisfactory. Logit models are akin to regression models in which the dependent variables are the log odds of the data belonging to one group as opposed to the other group. Grablowsky and Talley [8] compared a probit model with a linear discriminant model and concluded the former was superior

In practice, however, most credit scoring systems are based on discriminant analysis methodology or on a non-parametric binary tree classification suggested by Freidman [7] and outlined in [1] which following Srinivasan and Kim [13] we will call the recursive partitioning algorithm (RPA). In section two of this paper we describe how credit scoring systems can be built using these techniques and outline possible variations in scoring systems based on these techniques. We also describe hybrid systems which use both techniques to develop the final credit scoring system. Section three describes the performance of the various systems which were built using credit card company data, while the final section draws some conclusions about the strengths and weaknesses of these two techniques.

2. METHODOLOGIES FOR CREDIT SCORING

The initial credit granting decision is whether to extend credit to a new client on the basis of the application information the client has supplied together with possibly a reference to a credit agency, a bank opinion and an employer's reference. In order to make this decision the credit-grantor has available the credit histories and application forms of previous clients and possibly the application forms of those that were refused credit. Normally only a sample of the previous clients is used as the data set. This leads to a bias unless inference is made about the behaviour of rejected clients and they are also included in the sample. However as we will concentrate on slow payers, the population we are interested in is those who are accepted not those who apply. Thus we can ignore this difficulty in this paper

The credit grantor determines which ones of the credit histories are acceptable and which ones are unacceptable to him – i.e. he splits the data set into the “goods” and the “bads”.

We now consider the two main methodologies – discriminant analysis (DA) and recursive partitioning algorithm (RPA) which are used to assist in this problem.

Discriminant Analysis

Discriminant analysis considers the credit-granting problem as one of dividing the initial information set (in effect the observations) into two exclusive and exhaustive regions I_g and I_b so that if the information vector $\underline{x}$ of a client falls into I_g , credit is extended and if into I_b it is refused. Let the cost of misclassifying a client, who is really “good” as “bad” be L (L for lost profit) and that of classifying a client who is really “bad” as “good” be D (D for debt that will have to be written off). If *a priori* the probabilities of “goods” and “bads” in the population applying for credit are p_g and p_b , then the expected loss is.

$$p_g L \int_{I_b} f(\underline{x} | P_g) d\underline{x} + p_b D \int_{I_g} f(\underline{x} | P_b) d\underline{x} \quad (2.1)$$

where $f(\underline{x} | P_g)$ [$f(\underline{x} | P_b)$] is the density function over the initial information set for the population of “goods” (P_g) [“bads” (P_b)]. The objective is to determine I_b and I_g which minimise (2.1). Despite Eisenbeis’ [5] reservations it is often assumed that $L = D = 1$ so that (2.1) becomes the expected rate of misclassification. In that case the solution is to define

$$I_g = \{ \underline{x} | p_g f(\underline{x} | P_g) > p_b f(\underline{x} | P_b) \} \quad (2.2)$$

If the two populations have multivariate Normal information distributions so that $f(\underline{x} | P_g)$ is multivariate Normal with mean $\underline{\mu}_g$ and covariance matrix Σ , and $f(\underline{x} | P_b)$ is multivariate Normal with mean $\underline{\mu}_b$ and covariance matrix Σ , the rule (2.2) becomes the Fisher linear discriminant function, where one classifies $\underline{x}$ in I_g if:

$$\underline{x} \cdot \Sigma^{-1} (\underline{\mu}_g - \underline{\mu}_b) > \log (p_b/p_g) + \frac{1}{2} (\underline{\mu}_g + \underline{\mu}_b) \cdot \Sigma^{-1} (\underline{\mu}_g - \underline{\mu}_b) \quad (2.3)$$

This is a linear scoring rule in that one extends credit to a client if the weighted linear sum of the initial information responses – the LHS of 2.3 – exceeds some value – the RHS of (2.3)

In practice, the means and covariance are not known and so $\underline{\mu}_g$, $\underline{\mu}_b$ and Σ are replaced by the usual sample estimators $\underline{x}_g$, $\underline{x}_b$ and $\underline{S}$ of the means and covariance matrix. There is no assurance that this sample linear discriminant function will minimise the expected rate of misclassification, but it has proved satisfactory in practice when the populations have multivariate Normal information distributions. It has also proved fairly satisfactory in other situations – see the survey by Choi [3]. This is because Fisher actually developed this discriminant function in another way. If one looks at two univariate Normal populations with means μ_g and μ_b respectively and a common variance σ^2 , it is clear that an observation x would be classified in I_g if it is nearer to μ_g than μ_b . The risk of misclassifying then is clearly related to $(\mu_g - \mu_b) / \sigma$, since when this is large there is little overlap between the two populations. So Fisher felt that when dealing with two multivariate populations of information vectors, one should look for a linear combination of the information data so that for this linear combination the distance $(\mu_g - \mu_b) / \sigma$ is maximised. In other words he looked for a vector $\underline{a}$ of constants which maximises

$$(\text{Mean of } \underline{a} \cdot \underline{x} \text{ for } \underline{x} \text{ in population } P_1 - \text{Mean of } \underline{a} \cdot \underline{x} \text{ for } \underline{x} \text{ in population } P_2) / (\text{Standard deviation of } \underline{a} \cdot \underline{x}). \quad (2.4)$$

This turned out to be the LHS of (2.3) and so this discriminant function maximises the ratio of between group dispersions to that of within group dispersions. This property may well make the discriminant function more robust to changes in distributions

One of the major difficulties in applying this methodology to credit scoring systems is that many of the characteristics in the initial application form are qualitative not quantitative –

e.g. post-code, employment category, residential status – and so they correspond to discrete rather than continuous variables. There are several ways of dealing with this.

- i) Introduce binary variables, i.e. $\{0,1\}$ – variables for each possible outcome of each discrete variable. Thus if residential status is classified into N categories, one introduces $N-1$ binary variables where the first might be 1 if owner-occupier; 0 otherwise; and the second might be 1 if living with parents, 0 otherwise. These are then dealt with like the continuous variables in the discriminant analysis, but will lead to a large number of variables, which are clearly non-Normal.
- ii) A second approach is the location model (see Krzanowski [10]) which constructs a different linear discriminant function over the continuous variables for each possible combination of the values of the discrete variables. Thus for postcodes beginning EH and residential status, owner-occupier, there would be a linear discriminant function over age and income with a different one for other combinations of postcode and residential status.
- iii) Translate the qualitative variable into a quantitative one. If the qualitative variable has m values, let g_i be the number from the population of “goods” who take the i^{th} value and b_i be the number from the “bad” population who take the i^{th} value, where if

$$G = g_1 + g_2 + \dots + g_m \quad B = b_1 + b_2 + \dots + b_m$$

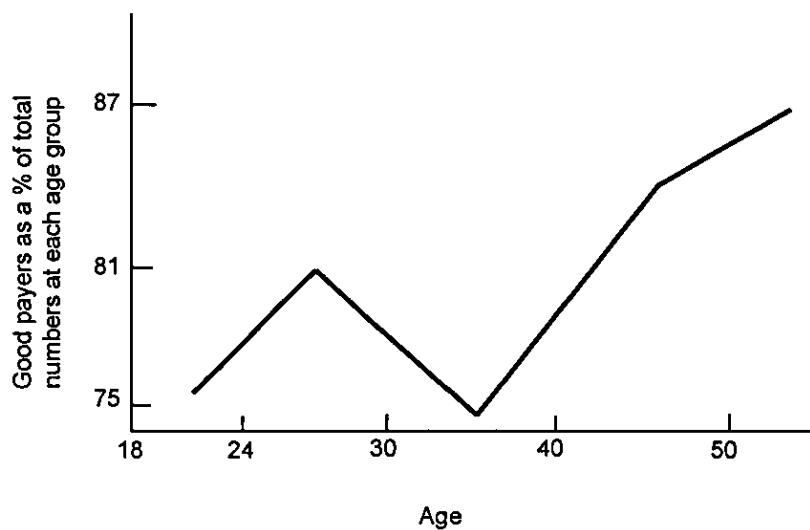
G is the total number of “goods” in the sample population and B is the total number of “bads”. Then one could translate the i^{th} value of the variable into a quantitative one depending on g_i , b_i , G and B . Possible choices would be $g_i/(b_i)$, $g_i/(g_i+b_i)$, $g_iB/(b_iG)$, $\log(g_iB/b_iG)$ or $\log g_i/(g_i+b_i)$ which are all related to estimates of probability odds or log probability odds of the “goods” and “bads” taking the i^{th} value of the variable.

Since for some variables, like postcode, there are a large number of values the variable can take, all methods would benefit from aggregating some values together, to ensure that the

aggregated values appear sufficiently often in the sample set to make the results statistically robust. Otherwise there will be too many variables in methods i) and ii) and in all three cases there would be a need for an enormous initial data set to ensure significant numbers in each value of a variable.

In this paper we have chosen to use the third method of dealing with qualitative data. The outcome values are grouped into blocks homogeneous in the proportion of “goods” and each block is ascribed the value of the ratio of “goods” to “bads” in that block. This procedure was chosen because the same methodology needed to be applied to the continuous variables such as income or age. It is often the case that credit risk appears not to be monotone in these variables. Figure 1 shows the age results when grouped in blocks of years.

Figure1: Relationship of credit risk with age



Since a credit scoring system is predictive rather than descriptive, it is acceptable to rearrange the age blocks in increasing order of credit risk by giving each block the value of $g_i/(g_i + b_i)$. Thus we will apply this procedure to all variables, discrete and continuous.

Returning to the ideas underlying Fisher's discriminant function, if the covariance matrices Σ_g and Σ_b are different for the "good" and "bad" groups, (2.2) leads to a quadratic discriminant function. In the case where the distribution is not known, the parameters μ_g , μ_b , Σ_b , Σ_g are replaced by their estimates $\underline{x}_g$, $\underline{x}_b$, S_b , S_g . In this case $\underline{x}$ is classified in I_g if

$$(\underline{x} - \underline{x}_b)S_b^{-1}(\underline{x} - \underline{x}_b) - (\underline{x} - \underline{x}_g)S_g^{-1}(\underline{x} - \underline{x}_g) + \log(|S_b| / |S_g|) > 2\log(p_b / p_g) \quad (2.5)$$

This involves many more coefficients in the scoring system – $(n^2 + n)$ – compared with n in the linear discriminant function and so is more difficult to implement. As it is less robust to interactions between the variables and is less efficient as the number of variables increases, most discriminant analysis scoring systems are built on linear discriminant functions.

Another problem in building a credit scoring system based on discriminant analysis is to determine which of the variables obtained from the initial information should be included in the discriminant function. Since high degrees of collinearity between the variables, where variables have a nearly linear relationship, lead to unstable coefficients, it is better to omit highly correlated variables. Similarly variables that add little or nothing to the discrimination of the scoring system can be dropped.

Recursive Partitioning Algorithm (RPA)

This nonparametric method forms a binary tree as an aid to classification by repeatedly splitting subsets of the information space, I , into two descendent subsets or nodes. The terminal nodes of the tree are designated as part of I_g or I_b depending on whether defining all the sample set in that node as "good" or "bad" minimises the error under the criterion considered. The formation of the tree thus depends on the splitting rule used and the rule to determine when a node is terminal and need not be split any more. The idea behind each split is that the two new sets are as homogeneous as possible and as different from each other as possible in terms of the concentration of "goods" and "bads" in the sets.

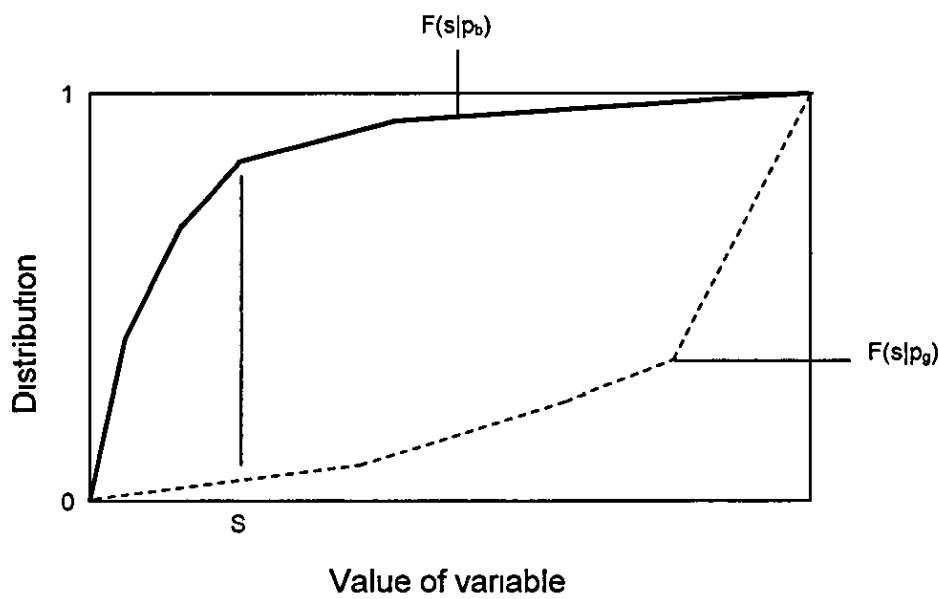
The algorithm starts with the whole information space I . Each variable which makes up the information space is considered in turn and the best splitting point for that variable is determined. To do this the values of the variables are reordered to be monotone in proportion of “goods” and a splitting rule is used. The myopic splitting rule suggested by Friedman [7] is one of the simplest. Let L , D , p_g and p_b be defined as in (2.1) and let $F(x|P_b)$ and $F(x|P_g)$ be the distribution functions of the values of this modified variable for the “bad” and “good” populations. The expected loss if this is the only split and accounts with values below the splitting point s are designated “bad” and accounts with values greater than s are designated “good”, is

$$p_g L F(s | P_b) + p_b D (1 - F(s | P_g)) \tag{2.6}$$

The myopic rule chooses the s that minimises (2.6).

If $p_g L = p_b D$, this rule becomes maximise the Kolmogorov-Smirnov (KS) distance $|F(x|P_b) - F(x|P_g)|$ which is the difference between the two cumulative distribution functions, see Figure 2.

Figure 2: K-S distance between the two distributions



More complicated splitting rules can be considered (see [1] for discussion) including ones that look ahead k -levels of splits before determining the best split. Having found the best split for each variable, the information set I is split into two groups using the best of these splits. The process is repeated on each of these subgroups to form further subgroups, though it may well be different variables that give the best splits on these subgroups. Subgroups are terminal nodes, and do not split further either if there are insufficient accounts in the subgroup to split or if the optimal split results in subgroups which are not sufficiently distinguishable. If $p_g L = p_b D$, a terminal node is defined to be in I_g if the majority of the sample set in that node are "good".

The process is continued until all nodes have been split on or are terminal. The tree thus constructed is really over fitted and the next step is to prune it back to a less complex tree. This is usually done by repeating the process but instead of using the whole of the data set, subsets of the data set are used and the resultant tree is tested on the data not used in its construction. In this way, one can construct a more robust if less complex tree. Other ways involve minimising a cost function which is a combination of the number of terminal nodes and the classification error, see Breiman [1] for details.

3. RESULTS

The credit scoring systems were constructed and tested on data supplied by a bank's credit card organisation. The initial application data and subsequent credit history over two years of 1001 card holders recruited over a twelve month period were made available. Since these applicants had passed the bank's credit granting system, their default rate was likely to be very low. It was therefore determined to build a credit scoring system to try and identify the 'slow' payers as opposed to the defaulters, where the identification of slow was taken to be that the account was at least one month delinquent at the end of the period under consideration. This criterion was chosen both because it gave a reasonable number of unsatisfactory accounts and also to test whether it is possible to identify at the outset accounts which though acceptable should be more carefully monitored.

The 1001 accounts were split into a set of 801 accounts on which to build the system and a holdout sample of 200 accounts (152 good, 48 slow) for testing

The application form gave rise to 24 information variables including postcode, age of applicant, applicant and spouse's income, employment category, residential status, etc. This information was used to construct six different scoring systems.

3.1 *Linear Discriminant Analysis (LDA) using all 24 variables*

The methodology outlined in section two was employed on all 24 variables. For each variable the good-bad ratio for each value was calculated, values with similar ratios were aggregated together, and a modified variable taken whose values are the good-bad ratios. Discriminant analysis was applied using these modified variables

The results were similar whether the discriminant function was built on all 24 in one go, or whether variables were introduced stepwise one at a time to the discriminant function. 'Postcode' and 'years at Bank' were the most important variables both on their effect on the discriminant function using standardised coefficients and on the correlation between their value and that of the discriminant function value. Thereafter the ranking of the variables was different under standardised coefficients from that under correlation with a discriminant function. This is because of the dependency between the variables.

3.2 *LDA using 11 variables*

Analysis of the correlation matrices of the 24 variables shows some significant dependency between the variables. Using this, the standardised coefficients of the variables in the discriminant function and the correlation of the variables with the discriminant value suggested that four variables could be removed because they had little impact on the discriminant function and another nine were highly correlated with more significant variables. The discriminant methodology was then applied to the remaining 11 variables – postcode, age of applicant, number of children, employment category, income, residential

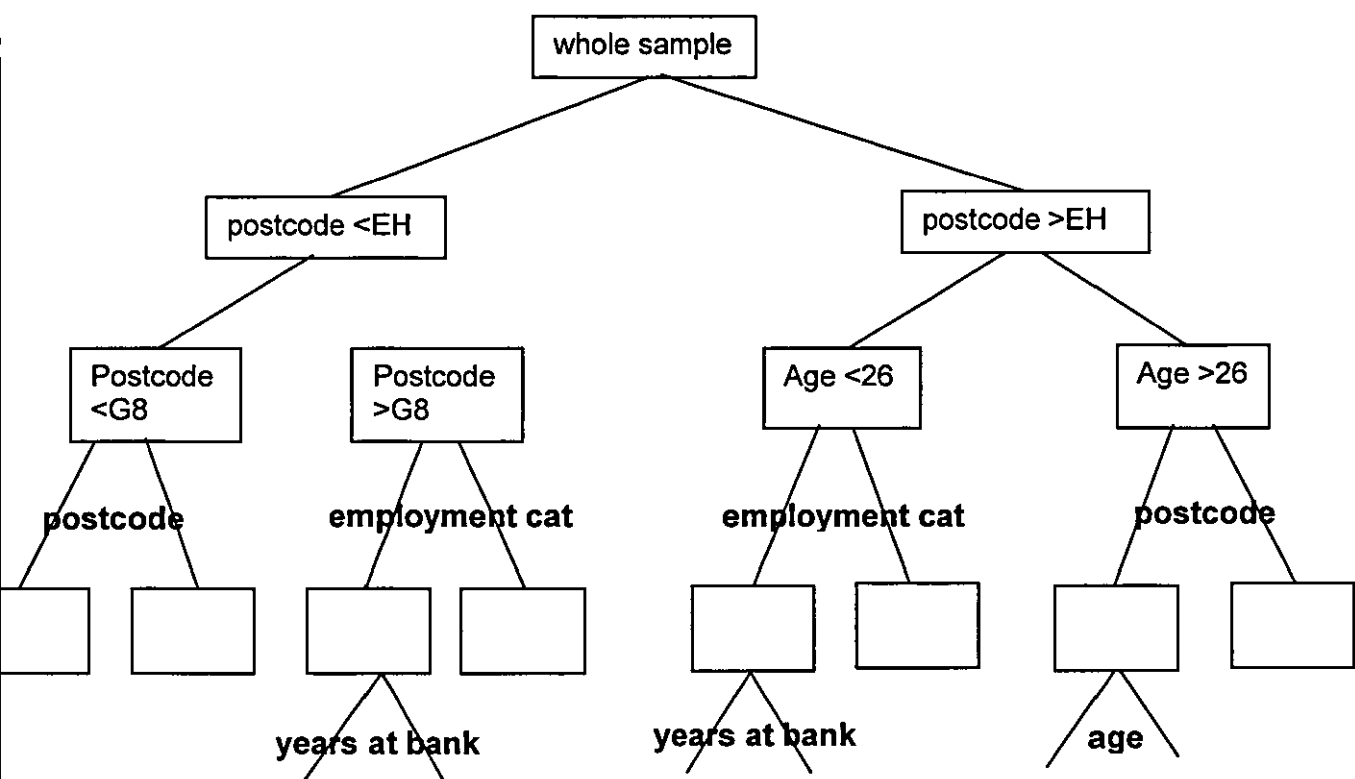
status, value of home, years at bank, years at present employment, hold a current account, and hold a major credit card. There was little change in the relative importance of the variables in the linear discriminant function obtained compared with their importance in the 24 variable case. However, the changes in the scores for the specific variables, varied from 4% to 250% - the larger changes affecting the variables highly correlated with a variable that had been removed. This is to be expected, as much of the discriminant function weight of the removed variable will be transferred to variables highly correlated with it.

3.3 RPA using 24 variables

The recursive partitioning methodology was used to build a binary tree to create a scoring system using all 24 variables. The top of the tree is given in Figure 3.

The tree actually had a depth of 11 nodes along one branch, but most branches were only 5 or 6 nodes deep.

Figure 3 Classification Tree



3.4 *RPA Hybrid using 11 variables*

The remaining three systems use both the discriminant analysis and recursive partitioning methodologies. In this system the discriminant analysis approach was used to identify the 11 variables that are most important in constructing the discriminant function and that have a low correlation with one another just as was done in method 3.2. A RPA tree was then built using splits on only these 11 variables. In fact the first three levels of the tree remain as in Figure 3 since postcode, age and employment category were three of the 11 variables. Changes do occur at the fourth level but the trees constructed are similar in size to those constructed by method 3.3.

3.5 *Hybrid DA using 2 compound and 20 other variables*

One of the disadvantages of linear discriminant functions is that they cannot deal with non-linear relationships between the variables, whereas this is one of the strengths of RPA. Therefore why not use RPA to identify which important variables are related and then introduce a new combined variable in the DA which expresses this relationship. From Figure 3 it is seen that the splitting variables at the top of the RPA tree are postcode, employment category, age and years at bank. Thus we introduce two new variables x_1 which is a function of postcode and employment category and x_2 – a function of age and years at bank. If postcode has m_1 values, employment category n_1 values, x_1 has $m_1 n_1$ values each corresponding to one value of the postcode and one of the employment category. These values are then modified to the corresponding $g_i/(g_i + b_i)$ values as described in section two with aggregation of values where necessary. The linear discriminant function is then constructed using these two compound variables, and the remaining 20 of the original 24 variables excluding postcode, employment category, age and years at bank. In fact x_1 and x_2 are by far the most important variables in the discriminant function.

3.6 *Hybrid DA using 2 compound and 7 other variables*

This scoring system is constructed in the same way as 3.5 except that only the two compound variables and the remaining 7 variables from the 11 identified in method II are used in the

discriminant function. Again x_1 and x_2 have the major impact on the discriminant function. The results of the six methods are given below. Table 1 describes the results of applying the system to the hold-out sample of 200

TABLE 1
Results of Applying the System to the Hold-Out Sample of 200

Method	Actual goods 152 cases		Actual slows 48 cases		% Correct
	Scored	Scored	Scored	Scored	
	good	bad	good	bad	
DA – 24 variables	150	2	43	5	77.5
DA – 11 variables	150	2	43	5	77.5
RPA – 24 variables	140	12	40	8	74.0
RPA Hybrid – 11 variables	143	9	41	7	75.0
Hybrid DA 2+20	149	3	39	9	79.0
Hybrid DA 2+7	150	2	41	7	78.5

“% Correct” is the percentage correctly classified in the sample, with no difference in weighting between the “goods” and the “slows” who are correctly classified. These compared with the percentage correct under a random decision of 63.5% and the percentage correct when classifying all as good of 76%. These results show that the hybrid systems do seem attractive. Trying to identify the slow payers among a set who have already been preselected under a non-defaulting criterion is unlikely to lead to impressive results. The best hybrid however identified 12 of the 200 in the sample as potential slow-payers and 9 of these were subsequently slow-payers. This suggests that such a procedure might be worthwhile even if it only identifies 20% of the slow-payers.

Table 2 shows the classification results that the system obtained on the 801 clients used to

build the system. It is well known that doing this gives results which are biased towards lower errors than the true errors for the systems

The dramatic improvement in the RPA results compared with the hold-out sample suggests that the trees are still over fitted and should be pruned back further.

TABLE 2
Results of Applying the System to the 801 Clients Used to Build the System

Method	Actual goods		Actual slows		% Correct
	662 cases		139 cases		
	Scored	Scored	Scored	Scored	
	good	bad	good	bad	
DA – 24 variables	650	12	122	17	83.3
DA – 11 variables	652	10	124	15	83.3
RPA – 24 variables	646	16	54	85	91.2
RPA Hybrid – 11 variables	643	19	73	66	88.5
Hybrid DA 2+20	649	13	111	28	84.5
Hybrid DA 2+7	646	16	114	25	83.7

4. CONCLUSIONS

Many other variants of the two main methodologies investigated in this paper are also appropriate for building credit scoring systems. The results obtained, however, imply that it does seem feasible to build systems to identify at an early stage, accounts which may become delinquent if not defaulting. As to the comparison between DA and RPA, the former seems marginally more satisfactory if only because of the care needed in pruning back the RPA trees sufficiently to prevent over fitting. The strength of the discriminant analysis is that it uses all

the data in all the scoring weightings it determines, however it does not deal satisfactorily with complex dependencies between the variables. The tree structure of RPA, on the other hand, allows the scoring system to incorporate complex dependencies between the variables, but at the lower nodes of the tree only a very small subset of the original data is being used to determine the next variable to split on. It does seem that systems can be built which benefit from the strengths of both methodologies. The hybrid methods outlined above use the RPA analysis to identify which of the important variables are dependent on another and then incorporates this dependency into the DA analysis by introducing compound variables.

ACKNOWLEDGEMENTS

This work was supported by a grant from the Economic and Social Research Council, which is gratefully acknowledged.

REFERENCES

- [1] Breiman, L., Friedman, J.H., Olshen, R.A and Stone, C J. (1984), Classification and Regression Trees Wadsworth International.
- [2] Chandler, G C. and Coffman, J.Y., (1979), "A Comparative Analysis of Empirical Vs Judgemental Credit Evaluation", *Journal of Retail Banking*, Vol 1, pp 15-26.
- [3] Choi, S G (1986), Discrimination and classification: Overview, Comp and Maths with Appls , Vol 12A, pp 173-177
- [4] Durand, D., (1941), Risk Elements in Consumer Instalment Financing; Financial Research Program, Study 8, Studies in Consumer Instalment Financing, National Bureau of Economic Research.

- [5] Eisenbeis, R.A., (1978), "Problems in Applying Discriminant Analysis in Credit Scoring Models", *Journal of Banking and Finance*, Vol 2, pp. 205-219.
- [6] Eisenbeis, R.A., Gilbert, G G and Avery, R B., (1973), "Investigating the Relative Importance of Individual Variables and Variable Subsets in Discriminant Analysis", *Communications in Statistics*, Vol.2 (3), pp. 205-219.
- [7] Friedman, J (1977) "A recursive partitioning decision rule for non-parametric classification". *IEEE Trans on Computers*, Vol.26, pp. 404-408.
- [8] Grablowsky, B J and Talley, W.K. (1981) "Probit and discriminant functions for classifying credit applicants a comparison" *Journal of Economics and Business*, Vol.33, pp.254-261.
- [9] Hsia, D C (1978) "Credit Scoring and the Equal Credit Opportunity Act" *The Hastings Law Journal*, Vol 30, pp.371-448
- [10] Krzanowski, W J., (1975), "Discrimination and Classification Using both Binary and Continuous Variables", *Journal of the American Statistical Association*, Vol 70, pp.782-790.
- [11] Myers, J.R. and Forgy, E.W., (1963), "The Development of Numerical Credit Evaluation Systems", *Journal of the American Statistical Association*, Vol. 58, pp. 799-806.
- [12] Reichert, A K., Cho, C C. and Wagner, G M., (1983), "An Examination of the Conceptual Issues involved in Developing Credit-Scoring Models", *Journal of Business & Economic Statistics*, Vol 1(2), pp. 101-114.
- [13] Srinivasan, V. and Kim, Y R , (1987), "Credit Granting: A Comparative Analysis of Classification Procedures", *Journal of Finance*, Vol 92(3), pp. 665-683.

- [14] Wiginton, J C , (1980), "A note on the Comparison of Logit and Discriminant Models of Consumer Credit Behaviour", *Journal of Financial and Quantitative Analysis*, Vol. 15(3), pp. 757-770.

CHAPTER 4

THE DEGRADATION OF THE SCORECARD OVER THE BUSINESS CYCLE

J. N. Crook and L. C. Thomas

(Edinburgh University Management School)

R. Hamilton

(Loughborough University Business School)

IMA Journal of Mathematics Applied in Business & Industry (1992) 4(1), pp. 111-123

The degradation of the scorecard over the business cycle

J. N. Crook and L. C. Thomas

Edinburgh University Management School

R. Hamilton

Loughborough University Business School

Abstract

The published literature on credit scoring has not compared the characteristics of those who default, nor the discriminating power of individual variables used to predict default, under different economic conditions. Similarly, scorecards constructed by credit-scoring agencies are estimated from data relating to two or three consecutive years for applications over three to five years before. The aim of this paper is to explore the effects of changes in a scoring function over time on the classification of applicants into those likely to default and those not likely to default.

Linear discriminant analysis is applied to a training sample of 26,043 applicants for a bank credit card to estimate empirically a model of their repayment behaviour in 1989 and 1990. The variables that have additional statistically significant discriminating power over others are broadly similar between the two years, although some differences exist. Using a holdout sample of 17,084 cases which are thought to be representative of a profile of applications to the data-supplying organisation, we cross-tabulate the number who would be accepted and rejected using the 1989 model with the corresponding predictions using the 1990 model. The characteristics of those who would be accepted using the 1989 model but rejected using the 1990 model are identified. Differences in the predicted classification of a case may be due to differences between the two years in the functions estimated and/or to difference in the prior probabilities of default. We consider the proportion of

applicants who would be accepted in one year but not in the other, if the prior probabilities are adjusted to give the same rejection rate in both years, and discuss their characteristics

1. Introduction

The literature on credit-scoring systems has concentrated on two issues. One is the predictive performance of different statistical techniques that may be used to distinguish between defaulters and non-defaulters (Myers & Forgy 1963; Wiginton 1980; Boyle *et al.*, 1991). The other issue is how to predict whether a person who has missed a given number of consecutive payments will subsequently miss more (Chandler & Coffman 1983-4; Bierman & Hausman 1970; Crook *et al.*, 1992a). However, the following questions have not been addressed: how do changes over time in default rates affect the ability of certain variables to predict default, and what are the characteristics of people who are predicted to be good in one year but bad in the other? The aim of this paper is to shed some light on these questions

The proportion of credit-card holders who default varies considerably over time, as does the importance of different characteristics of individuals that are used to predict defaulters and non-defaulters in a scoring rule. This means that an applicant for credit may be accepted (rejected) if (s)he is scored on a rule developed from payment performance in, say, an economic depression but rejected (accepted) if (s)he is scored on a rule developed from performances during an economic boom.

Credit grantors may react in a number of ways. One option is to develop a scoring rule over a number of years which includes a complete cycle of economic activity. A difficulty with this option is that it may involve so long a time period that the model is no longer accurate for the future period for which it is required to predict. There may be changes in culture, attitudes, and other factors that can affect repayment behaviour but which are not often included in score-cards. Another option is to develop and use a different scoring rule in different time periods. For example, a scoring rule may be developed and used for

periods of economic depression only, and another scoring rule developed and used in periods of economic prosperity. Since the state of the economy varies continuously, this policy may involve updating a scoring rule annually. A third option is to develop a scoring rule in a period of depression or prosperity, and vary the cut-off score to maintain the same reject rate.

In this paper, we estimate a scoring model in each of two years separately. The default rate differs between the two years. We consider how the discriminating power of different variables differs between the two years, and the characteristics of those who would be rejected using a model estimated for one year but accepted on the basis of a model estimated for a different year. We also consider the characteristics of those who may be affected by a change in the cut-off score from that indicated by the default rate in the observation period.

Following an explanation of our methodology in Section 2, Section 3 considers the relative discriminating power of each variable in the two years. Section 4 considers the effects of changes in the cut-off scores, Section 5 discusses the implications of the results for credit grantors in their policy decisions, and Section 6 concludes.

2. Methodology

2.1 *The data*

Data were acquired for two recent years which differed in terms of the state of the national economy. The years chosen were 1989 and 1990. Table 1 shows values of the Coincident Indicator of the state of the UK economy calculated from those published by the Central Statistical Office. It shows that the level of economic activity was clearly lower in 1990 than in 1989.

The initial sample consists of 37,213 individuals who held a bank credit card and who had used it since it was issued, and 6,444 individuals whose application for a card was rejected. Seventy percent of the accepted applications were randomly selected as a training sample. The remaining 30% were combined with an appropriate number of rejects to form a holdout sample such that the rejects made up 35% of the total holdout. This was the proportion that industry sources suggested were typically rejected. The holdout was therefore representative of a typical batch of applications to a bank credit-card issuer. Applicants aged under 18 in 1989 were deleted from the sample.

Table 1

Values of the Coincident Indicator for the UK economy*

1989	Q1	107.1	1990	Q1	103.9
	Q2	105.2		Q2	103.0
	Q3	104.5		Q3	98.5
	Q4	104.2		Q4	93.8

Long-term trend = 100

The Coincident Indicator is a weighted average of the following series:

- GDP (A) at factor cost, constant prices, 1985 = 100
- Output of the production industries, 1985 = 100
- CBI Quarterly Survey: below-capacity utilization (%)
- Index of volume of retail sales, 1985 = 100
- CBI Quarterly Survey: change in stocks of raw material (% balance)

*Calculated from 'Cyclical Indicators for the UK', *Economic Trends*, No.454, August 1991, page 72, Table A.

Many alternative definitions of ‘default’ by an individual could be adopted. In this paper we define *default* as the missing of three consecutive payments due on their credit-card debt outstanding. This definition was chosen because it is consistent with that used by the industry. Table 2 shows the division of the training and holdout samples into defaulters, non-defaulters, and rejected applications.

Data were available on 24 sociodemographic and economic variables which have been used in previous discriminant analysis scoring models (see Capon 1982) or for which an *a priori* reason why they may act as effective discriminators could be made. The 24 variables are shown in Table A1 of the appendix. All data, excluding repayment history data, were taken from each applicant’s application form

Table 2

<i>The samples</i>				
	1989		1990	
	Training sample	Holdout sample	Training sample	Holdout sample
Non-defaulters	25,070	10,744	24,135	10,381
Defaulters	973	420	1,908	783
Rejects	0	5,920	0	5,920
Total	26,043	17,084	26,043	17,084

2.2 Estimation

The methodology follows that of Crook *et al* (1992b). Briefly, many of the variables were measured at nominal level, whereas the estimation method used – linear discriminant analysis – requires data to be measured at least at interval level (see Klecka 1980). Additional information was used to derive interval-level data by ascribing to each predictor the values

$$X_j = 1n (g_k/b_k) + 1n (B_T/G_T) \quad (2.1)$$

where X_j = value of predictor for case j ,
 g_k = number of good payers in nominal category k , the category of which j was a member,
 b_k = number of poor payers in nominal category k , the category of which j was a member,
 G_T = number of good payers in the sample,
 B_T = number of poor payers in the sample.

The use of the X_j transformation means that X_j may not be monotone in the values of the original variable. High degrees of collinearity between predictor variables were removed by deleting cases where such collinearity had been detected in a different sample of 1001 cases who applied for a card around one year earlier than the cases in this study.⁽¹⁾

We were interested in variables which individually contributed additional statistically significant discriminatory power beyond that contributed by other variables. Therefore, in each discriminant analysis, predictors were selected for inclusion in the empirical function by a stepwise procedure.⁽²⁾

3. Changes In Discriminating Functions

Separate discriminant analyses were performed for 1989 and 1990, using the values of X_j for each respective year and the repayment behaviour of each individual in the relevant year. For both functions, the group centroids (goods and bads) are statistically different using a χ^2 test of the significance of Wilks' lambda. The classification matrices are shown in Table 3. These relate to the holdout sample. In each matrix the prior probability of group membership, i.e. the probability that a case is a member of a particular group when no

information about it is available, was calculated by treating the rejected cases (34.65% of the total holdout) as defaulters as well as the actual defaulters. That is

$$P_b = (B + R)/(G + B + R) \quad P_g = G/(G + B + R), \tag{3.1a,b}$$

where P_b = prior probability that a case is a bad, i.e. defaults,
 P_g = prior probability that a case is a good, i.e. does not default,
 G = number of goods, B = number of bads, R = number of rejects

Table 3 clearly shows that the empirical scoring systems predict group membership better than chance.

Table 3

Classification matrices

		Predicted group					
		1989			1990		
		Goods	Bads	Total	Goods	Bads	Total
Actual	Good	9,543	1,201	10,744	8,744	1,637	10,381
	Bad	319	101	420	535	248	783
	Rejects	4,399	1,521	5,920	3,463	2,457	5,920
	Total	14,261	2,823	17,084	12,742	4,342	17,084
% correct		65.35			67.02		
$C_{prop} = 100(P^2_b + P^2_g) (\%)$		53.32			52.32		

Table 4 shows the standardized canonical discriminant-function coefficients which indicate the relative discriminatory power that each variable has, given the other variables in the function.

Table 4

Standardized canonical discriminant function coefficients

1989			1990		
	Value	Rank		Value	Rank
Years at bank	0.45	1	Years at bank	0.43	1
Cheque card	0.33	2	Cheque card	0.32	2
Number of children	0.33	3	Outgoings	0.31	3
Appls empl status	0.27	4	Appls. empl status	0.25	4
Outgoings	0.25	5	Number of children	0.21	5
Years at pres. empl	0.21	6	Phone	0.20	6
Major credit card	0.20	7	Residential status	0.19	7
Phone	0.19	8	Years at pres. empl	0.19	8
Deposit account	0.11	9	Charge card	0.13	9
Store card	0.11	10	Store card	0.10	10
			Major credit card	0.09	11
			Mort. balance outs.	0.09	12
			Deposit account	0.08	13

Only those variables that have a significant amount of discriminatory power are included. While the discriminatory power of many variables was similar in both years, the relative discriminatory power of certain predictors was markedly different. First, 'number of children', 'major credit card', and 'deposit account' had relatively higher discriminatory power compared with the other included variables in 1989 (the year with the lower default rate) than in 1990, while 'outgoings' had relatively greater discriminatory power in 1990 than in 1989. In 1990, 'residential status', 'charge card', and mortgage balance outstanding' had statistically significant additional discriminatory power over that of other included variables, which they did not have in 1989, and so were not included in the estimated function for the latter year by the stepwise routine.

4. Effects Of Changes In Cut-Off Scores

Our data suggests that the behaviour of some individuals differed between the two years. Firstly, the overall default rate differs between the two years. This implies a difference in the prior probabilities of membership of a specific group. Secondly, the default rates for each value of each predictor variable differs between the two years. Therefore the X_j value of each group of values for a given variable differs between the two years. The second difference results in different standardized and unstandardized canonical discriminant-function coefficients between the two years, and in differences in the degree of separation between the two groups. This implies that there may be a difference between the two years in the conditional probability $P(S|G_i)$ that a case gains a score S , given that it is a member of a group i (see the appendix). A case is classified into the group in which the probability of its membership, given its score, is greater. That is

$$P(G_i | S) = P(S | G_i)P(G_i) / \sum_{i=1}^n P(S | G_i)P(G_i) \quad (4.1)$$

where $P(G_i | S)$ is the posterior probability that a case with score S is classified into group i , and $P(G_i)$ is the prior probability that a case is a member of group i . Therefore the difference in both the prior and conditional probabilities between two years implies that a case may be classified as a good (bad) in one year and a bad (good) in the other.

We now examine the effects that both the different empirical models and the different prior probabilities ('priors') together have on predicted applicant performance. Specifically we ask what the characteristics are of those who would be accepted in 1989 using the 1989 canonical function coefficients and priors but rejected in 1990 using the 1990 canonical function coefficients and priors.⁽³⁾ Table 5 shows the number of people affected. While the same decision would have been given to 88.3% of the holdout cases if either function and priors were used, the decision would have been different in 11.7% of cases. Approximately 10% of the holdout would have been accepted if the 1989 function and priors were used, but rejected if the 1990 function and priors were used instead, and 1.4% of cases would

have been accepted if the 1990 function and priors were used but rejected using the 1989 model.

Table 5

Total effect cross-tabulation

		Actual 1989 priors and function		
		Good	Bad	Total
Actual 1990 priors and function	Good	12,506 (73.2)	236 (1.4)	12,742
		1,755 (10.3)	2,587(15.1)	4,342
		14,261	2,823	17,084 (100)

Figures in parentheses are the number of cases in the cell as a percentage of the number of cases in the total holdout sample.

Table 6

Modal groups: total effects

The holdout sample in aggregate			Those predicted to be good on the 1989 function with 1989 priors but bad on the 1990 function with 1990 priors	
	Modal group	% of cases	Modal group	% of cases
Number of children	0,6,7,8	69	0,6,7,8	76
Number of dependants	0,3,4,5,24	98	0,3,4,5,24	98
Applicant's employment status	Private sector	65	Private sector	66
Deposit account	No	64	No	65
Loan account	No	95	No	96
Cheque card account	No	75	No	79
Current account	Yes	67	Yes	67
Major credit card	No	60	No	68
Charge card	No	76	No	86
Store card	No	78	Yes	78
Applicant's employment category	Services, Office, Sales, Labourer, Executive, Trades, Others	46	Services, Office, Sales, Labourer, Executive, Trades, Others	52
Age in 1990	18-24 years	27	25-30 years	31
Building society card	No	92	No	91
Phone	No	83	No	70
Spouse's income	£0	78	£0	82
Years at present employment	0, 1 years	28	0, 1 years	39
Years at same bank	0, 1 years	28	0, 1 years	35
Value of home	£0	64	£0	79
Applicant's income	£0-6000	24	£0-6000	26
Mortgage balance outstanding	£0	68	£0	92
Outgoings	£0	24	£1-99	30
Residential status	Owner	38	Tenant furnished	33
Spouse's employment category	No response	68	No response	74
Years at present address	0,1 years	28	0,1 years	38
Sample size	17084		1755	

Table 6 compares the characteristics of those who would be accepted on the 1989 model but rejected using the 1990 model⁽⁴⁾ with those of the holdout sample in aggregate. The table suggests that those for whom a different decision would be made depending on the year to which the model related are very similar to the holdout sample as a whole. The modal groups for both cells are the same for twenty characteristics. The differences in modal groups are whether or not a store card is possessed ('yes' for the 1990 rejects, 'no' for the holdout), age in 1990 (25-30 years for the 1990 rejects, 18-24 years for the holdout), outgoings (£1-99 for the 1990 rejects, £0 for the holdout), and residential status (tenants (furnished) for the 1990 rejects, owner for the holdout)

We now ask a second question. Suppose that we keep the proportion of cases who are predicted to be good (bad) the same in two years, years t and n . That is, we alter the priors in year n such that, when used with n 's canonical function coefficients, the same proportion of cases is rejected (i.e. predicted to be bad) as in year t . What, then, are the characteristics of those who would be predicted to be *bad* (good) by year t 's model (year t 's canonical function coefficients and actual priors) but who are predicted to be *good* (bad) using the model of a year n (year n 's canonical function coefficients, hypothetical priors)? Notice that the hypothetical priors applied in year n are not the priors used in year t 's classification matrix (Table 3). Instead they are the priors which, with year n 's canonical function coefficients would give the same proportion of cases predicted to be bad as predicted for year t . That is, they represent the 'cut-off score' that a credit granting agency would impose if they wished to use the current year's (n 's) function, but also wished the proportion of cases that are rejected to be the same as in another year (t)

This issue has been explored by performing two cross-tabulations. In both cases, we adjust the priors of 1990 so as to predict the same proportion of bads as were predicted for 1989. Firstly, we cross-tabulate the *numbers predicted to be good (bad) in 1990* with the numbers predicted to be good (bad) in 1990 had the priors been set so as to predict the same proportion of bads as predicted for 1989. Secondly, we cross-tabulate the *numbers*

predicted to be good (bad) in 1989 with those predicted to be good (bad) in 1990 again with the priors set to give the same proportion of bads as in 1989. The results are shown in Table 7.

Table 7
Two cross-tabulations

		Good	Bad	Total
(a) Actual 1989 priors, 1989 function				
1990 priors set to give same predicted proportion of bads as predicted in 1989, 1990 function	Good	13,629 (79.8)	640 (3.7)	14,269
	Bad	632 (3.7)	2,183 (12.8)	2,815
Total		14,261	2,823	17,084 (100)
(b) Actual 1990 priors, 1990 function				
1990 priors set to give same predicted proportion of bads as predicted in 1989, 1990 function	Good	12,742 (74.6)	1,527 (8.9)	14,269
	Bad	0 (0)	2,815 (16.5)	2,815
Total		12,472	4,342	17,084 (100)

Figures in parentheses are the number of cases in the cell as a percentage of the number of cases in the total holdout sample.

Table 7 shows that, if the priors of 1990 are adjusted to give the same reject rate in 1990 as in 1989, then 3.7% of 17,084 cases in the holdout sample would have been rejected using the 1990 rule, but accepted using the 1989 rule and cut-offs. On the other hand, 8.9% of cases would have been accepted using the 1990 system and adjusted cut-offs, but rejected if the 1990 function and cut-offs were used

Table 8 summarizes the characteristics of these two groups, and compares them with the characteristics of the total holdout sample. Firstly we compare the holdout with those accepted using the 1989 function and priors but rejected using the 1990 function with adjusted priors. The persons accepted on the 1989 model but rejected on the adjusted 1990 function are similar to the holdout in all respects except the following. They are older than the holdout (modal age group 25-30 years versus 18-24 years), they have a higher income (modal income range £13,000+ versus £0-6,000), they have greater outgoings (modal range £299 plus per month versus £0) and they typically have a different residential status (modal group 'tenant unfurnished' versus 'owner').

Table 8

Modal groups

	<u>The holdout sample in aggregate</u>		<u>Those members of the holdout sample predicted to be</u>			
	Modal group	% of cases	<u>Good on 1989 function but bad on 1990 function with adjusted priors</u>		<u>Bad on 1990 function but good on adjusted 1990 function</u>	
			Modal group	% of cases	Modal group	% of cases
Number of children	0,6,7,8	69	0,6,7,8	69	0,6,7,8	74
Number of dependants	0,3,4,5,24	98	0,3,4,5,24	98	0,3,4,5,24	98
Applicant's employment status	Private sector	65	Private sector	65	Private sector	65
Deposit account	No	64	No	64	No	66
Loan account	No	95	No	95	No	97
Cheque card account	No	75	No	75	No	84
Current account	Yes	67	Yes	67	Yes	68
Major credit card	No	60	No	60	No	78
Charge card	No	76	No	76	No	94
Store card	No	78	No	78	No	79
Applicant's employment category	Services, Office, Sales, Labourer, Executive, Trades, Others	46	Services, Office, Sales, Labourer, Executive, Trades, Others	46	Services, Office, Sales, Labourer, Executive, Trades, Others	55
Age in 1990	18-24 years	27	25-30 years	32	18-24 years	34
Building society card	No	92	No	91	No	92
Phone	No	83	No	56	No	77

Table 8 continued

Modal groups

	<u>The holdout sample in aggregate</u>		<u>Those members of the holdout sample predicted to be</u>			
	Modal group	% of cases	<u>Good on 1989 function but bad on 1990 function with adjusted priors</u>		<u>Bad on 1990 function but good on adjusted 1990 function</u>	
			Modal group	% of cases	Modal group	% of cases
Spouse's income	£0	78	£0	80	£0	83
Years at present employment	0, 1 years	28	0, 1 years	34	0, 1 years	41
Years at same bank	0, 1 years	28	0, 1 years	37	0, 1 years	41
Value of home	£0	64	£0	93	£0	81
Applicant's income	£0-6000	24	£13000 +	29	£0-6000	29
Mortgage balance outstanding	£0	68	£0	94	£0	83
Outgoings	£0	24	£299 +	28	£99-199	35
Residential status	Owner	38	Tenant (unfurnished)	43	With parents	31
Spouse's employment category	No response	68	No response	68	No response	75
Years at present address	0,1 years	28	0,1 years	44	0,1 years	34
Sample size	17084		632		1527	

We now turn to those cases that would be rejected on the 1990 function but would be accepted if the priors were adjusted to give the same reject rate as the 1989 model. These persons have the same modal values for characteristics as the holdout, except that they have greater outgoings (£99-199 versus £0) and they typically live with their parents as opposed to being owners.

5. Discussion

The holdout sample was constructed to have the same proportion of cases that were accepted and rejected by the organization supplying the data. Therefore, since the cases were also randomly selected by the organization for our sample, we believe that our holdout sample is representative of the applications that the organization would typically receive. We will interpret our results having made this assumption.

Table 5 shows that, even between the two adjacent years, changes in cut-off scores and canonical function coefficients can make a noticeable difference in the rejection rates yielded by a scoring model: 16.5% using the 1989 model against 25.4% using the 1990 model. A much greater proportion of applicants would have been rejected using the 1990 model but accepted on the 1989 model than vice versa: 10.3% compared with 1.4%. Since the prior probability of default in 1990 was much greater than in 1989, the cut-off score appears to have an effect on the classification of a case.

When we removed the effects of changes in the cut-offs, by adjusting them to give the same predicted proportion of cases rejected (when combined with the 1990 coefficients) in 1990 as was predicted using the 1989 priors and coefficients (Table 7), we found that 12.8% of cases would be rejected by both models, but 7.4% would be rejected by only one of the models. This gives some indication as to the effects of changes in the coefficients between the two years, since the priors – the other possible cause of a different

classification – have been adjusted to give the same rejection rate in both years. Furthermore, of the 10.3% of cases accepted using the 1989 model and rejected using the 1990 model (Table 5), 3.7 percentage points would still be rejected if the 1990 cut-off scores were adjusted (Table 7(a)). Therefore adjusting the cut-offs to maintain the same predicted rejection rate will not lead to the predicted group being invariant with respect to the year to which the data for the model relates. The different coefficients will result in some cases being classified differently between the two years.

If we change the 1990 cut-offs to give the same reject rate as in 1989 (Table 7(a)), we would accept 83.5% of cases rather than 74.6% without cut-off adjustment (Table 5). Of the 83.5% of cases, we would have rejected 8.9 percentage points (83.5% less 74.6%) of cases if the unadjusted 1990 model was used (Table 7(b)). Whether the 3.7% of cases that would be rejected in 1990 but accepted in 1989 (using the same proportion of rejects) should concern the credit grantor depends on the profit that these cases would have generated if they had been accepted. We have not built a profit model, but Table 6 shows the characteristics of such applicants. The same argument applies if the 1990 model was used, with 8.9% of cases rejected if the cut-offs indicated by 1990 behaviour were retained rather than the adjusted ones being used.

6. Conclusion

Our results suggest that changes in cut-off scores and in canonical function coefficients do result in sizeable differences in the proportion of applicants who would be rejected if the scoring model were based on a linear discriminant analysis estimated using data for one year rather than another, even if the years are adjacent to each other. Furthermore, changing the cut-off scores to maintain the same reject rate will not restore the same decision for each applicant. This suggests that credit grantors who build scoring models must be especially careful when choosing the years for which the data used in their model relates. They should attempt to estimate the profit that may be forgone by rejecting

applicants on one model when another suggests acceptance, and to estimate the increased loss that may result from accepting an applicant on one model when another suggests rejection. Only when the grantor has an accurate estimate of the financial cost of the errors involved in using one decision strategy rather than another will (s)he be able to evaluate different strategies accurately.

NOTES

1. Let A denote the earlier sample, and B the sample used for this study. Sample A contained data on exactly the same variables from the same bank as was used in sample B. To determine which variables to delete in sample B, it was assumed that the degree of collinearity detected in sample A applied to sample B also. Sample A consisted of 1001 cases, with data relating to applications in the period September 1986 to December 1987. To detect such collinearity, the tolerances were calculated for each variable, and the matrix of linear correlation coefficients was examined,
2. At each step, the variable that resulted in the greatest squared Mahalanobis distance D^2 was added. The significance of a change in D^2 when a variable was included was tested using a *partial-F* statistic. The probability that the *F*-to-enter value was significant was set equal to 5% regardless of the change in the degrees of freedom that occurred with the change in the number of included predictor variables. The same probability was adopted for the *F*-to-remove.
3. In the interests of brevity, the term 'differences in the canonical coefficients between the two years' will be taken to include differences between the two years in the variables included in the predictive models by the stepwise routines
4. We could examine the characteristics of those in any of the cells in Table 5. To save space, we consider only the one referred to.

Appendix

A case is classified into the group for which $P(G_i|x)$ is greatest, where

$$P(G_i | x) = P_i D_i^* / \sum_{i=1}^n P_i D_i^* ,$$

here n is the number of groups, P_i is the prior probability that a case is a member of Group i ,

$$D_i^* = (\det D_i)^{-1/2} \exp(-1/2 \chi_i^2) ,$$

And D_i is the covariance matrix of the canonical discriminant functions for group i , with

$$\chi_i^2 = (f - \bar{f}_i)^T D_i^{-1} (f - \bar{f}_i) ,$$

$$f = Bx + a ,$$

x = a $z \times 1$ vector of discriminant variables for a case,

B = the $m \times z$ matrix of unstandardized canonical discriminant function coefficients,

f = the $m \times 1$ vector of canonical discriminant function values,

$\bar{f}_i$ = the group centroids vector,

a = a vector of constants

Table A1

<i>The sociodemographic variables</i>	
Number of children	Building society card (yes/no)
Number of dependants	Phone (yes/no)
Applicant's employment status	Spouse's income
Deposit account (yes/no)	Years at present employment
Loan account (yes/no)	Years at same bank
Cheque guarantee card (yes/no)	Value of home
Current account (yes/no)	Applicant's income
Major credit card (yes/no)	Mortgage balance outstanding
Charge card (yes/no)	Outgoings
Store card (yes/no)	Residential status
Applicant's employment category	Spouse's employment category
Age in 1990	Years at present address

References

Bierman, H. & Hausman, W H , (1970) "The Credit Granting Decision", *Management Science*, Vol. 16(8), pp B519-B532.

Boyle, M., Crook, J N , Hamilton, R and Thomas, L C (1992) "Methods of Credit Scoring Applied to Slow Payers", in: *Credit Scoring and Credit Control* (Thomas, L.C , Crook, J.N., and Edelman, D., (eds.), Oxford University Press: Oxford, December, pp 75-90

Capon, N., (1982), "Credit Scoring Systems A Critical Analysis", *Journal of Marketing*, Vol. 46, pp. 82-91.

Chandler, G.C. and Coffman, J.Y., (1983-4) "Applications of Performance Scoring to Accounts Receivable Management in Consumer Credit", *Journal of Retail Banking*, Vol. 5(4), pp 1-10

Crook, J N., Hamilton, R and Thomas, L.C. (1992a) "A comparison of a credit scoring model with a credit performance model", *The Service Industries Journal*, Vol 12, No 4, October, pp.558-579.

Crook, J N., Hamilton, R , and Thomas, L.C. (1992b). "A comparison of discriminators under alternative definitions of default", in: *Credit Scoring and Credit Control* (Thomas, L C , Crook, J N , and Edelman, D., (eds.)). Oxford: Oxford University Press, December, pp.217-246

Eisenbeis, R A (1978), "Problems in Applying Discriminant Analysis in Credit Scoring Models", *Journal of Banking and Finance*, Vol 2, pp 205-219

Eisenbeis, R A., (1977), "Pitfalls in the Application of Discriminant Analysis in Business, Finance and Economics", *Journal of Finance*, Vol. 32 (3), pp. 875-900

Hair, J.F., Anderson, R.E. and Tatham, R.L., (1987), *Multivariate Data Analysis*, New York: Collier Macmillan.

Kshirsagar, A M (1972). *Multivariate Analysis*, Vol.2. Marcel Drekker.

Klecka, W .R. (1980), *Discriminant Analysis*, Sage University Paper No.19, Series Quantitative applications in the Social Sciences

Lindley, J T., Rudolph, P and Selby, E B. (1989) "Credit card possession and use: changes over time", *Journal of Economics & Business*, Vol 41, pp 127-142

Morrison, D.G , (1969), "On the Interpretation of Discriminant Analysis", *Journal of Marketing Research*, Vol. 4, pp 156-163

Myers, J R. and Forgy, E.W , (1963), "The Development of Numerical Credit Evaluation Systems", *Journal of the American Statistical Association*, Vol. 58, pp. 799-806

SPSS Inc (1988) *SPSS-X Advanced Statistics User's Guide*, Chicago: SPSS Inc.

Thomas, L.C., (1988) "Behavioural Scoring", Department of Business Studies, University of Edinburgh, Working Paper no 88/32, 12 pp

Wiginton, J.C., (1980) "A Note on the Comparison of Logit and Discriminant Models of Consumer Credit Behaviour", *Journal of Financial and Quantitative Analysis*, Vol 15(3), pp. 757-770

CHAPTER 5

A COMPARISON OF A CREDIT SCORING MODEL WITH A CREDIT PERFORMANCE MODEL

J. N. Crook and L. C. Thomas

(Department of Business Studies Edinburgh University)

R. Hamilton

(Business School, Loughborough University)

The Service Industries Journal, (1992), Vol. 12, No. 4 (October), pp. 558-579

A Comparison of a Credit Scoring Model with a Credit Performance Model

J. N Crook, R Hamilton and L C Thomas

Credit suppliers are interested in trying to predict which applicants are likely to default on repayments. They are also interested in predicting those who may miss one or two repayments rather than default by missing three. By considering a sample of 1001 bank credit card holders, this article compares those characteristics of borrowers which distinguish between (a) those who (in the sample period) never missed a repayment ('goods') and those who missed at least one ('slows'); (b) those who never missed a repayment and those who missed three consecutively ('defaulters'), and (c) those who missed one or two repayments ('bads') and those who missed three in succession

INTRODUCTION

Between 1981 and 1989 the real value of debt outstanding to UK consumers for other than house purchase increased by 122 per cent. To decide whether or not to grant credit to an individual, an increasing number of suppliers are adopting formal scoring techniques and Chandler and Coffman [1979] note that as early as 1970 such procedures were used by over 30 per cent of US credit grantors.

The aim of this article is to investigate whether the characteristics of individuals who miss three successive credit card repayments are the same as those who miss at least one payment and those who, having missed one or two payments, subsequently miss three. Models which predict whether an individual is likely to fit into the first two groups may be used to help the credit grantor to decide whether to issue credit or not. The third type of

model may allow the credit controller to score individuals whose repayment performance has been poor to decide whether they have the characteristics of those who miss three payments. Thus the former models relate to the credit granting decision, the third to predicting credit performance

Few, if any, studies have compared the ranking of predictors for these three groupings. There is some literature which compares the predictive performance of empirical models which have been constructed to distinguish between defaulters and non-defaulters. Thus Myers and Forgy [1963] compared the predictive performance of discriminant analysis, stepwise regression, and equal weights for all variables used, and found that equal weights were as effective as the other two methods. Wiginton [1980] compared the performance of a logit model with that of a discriminant model to find that the logit model predicted a greater proportion of cases relative to chance than did the discriminant analysis Boyle *et al.* [1991] compared the performance of linear discriminant analysis with a recursive partitioning algorithm to conclude that the predictive performance of the latter depended on the level of truncation of the tree. However, none of these studies compare the ranking of predictors of defaulters, slow payers and poor performers. One study [Crooke *et al.* 1991a] compared those of defaulters and slows, but not with those of poor performers.

Few published empirical performance scoring papers exist. One exception is that by Chandler and Coffman [1983-4], who applied discriminant analysis to accounts which were one month delinquent to distinguish between (a) accounts which were paid up and did not become delinquent again within six months and (b) accounts which became three or more months delinquent in the same six months. The model was shown to predict substantially better than chance, although the predictor variables are not mentioned. Recent contributions to behavioural scoring have constructed transition matrices of the probability that an account will move from being overdue by period *i* to period *j* for different risk classes of individuals and have indicated a rule to maximise expected profits given a maximum risk level [see Cyert, Davidson, Thompson 1962, Cyert and Thompson 1968, Fryman, Kallberg and Kao 1986]. Bierman and Hausman [1970] proposed a dynamic programming approach to maximise the present value of expected pay-off when the probability that an individual

will repay in a particular time period is estimated, given his past repayment history. These papers predict the probability of future defaults given the frequency of previous delinquency rather than predicting whether a person should be categorised as likely to go further delinquent on the basis of personal characteristics associated with such performance.

The following section describes the variables and methodology used in this study; and the results are then discussed.

VARIABLES AND METHODOLOGY

The Variables

To define precisely the elements of the sets of borrowers between which we wish to distinguish, consider the following definitions

Let $O = \{o_i \mid o_i = \text{an individual, } i, \text{ who has never missed even one payment in a given time period}\}$

$X = \{x_i \mid x_i = \text{an individual, } i, \text{ for whom the maximum number of consecutive missed payments in a given time period is 1}\}$

$Y = \{y_i \mid y_i = \text{an individual, } i, \text{ for whom the maximum number of consecutive missed payments in a given time period is 2}\}$

$Z = \{z_i \mid z_i = \text{an individual, } i, \text{ for whom the maximum number of consecutive missed payments in a given time period is 3}\}$

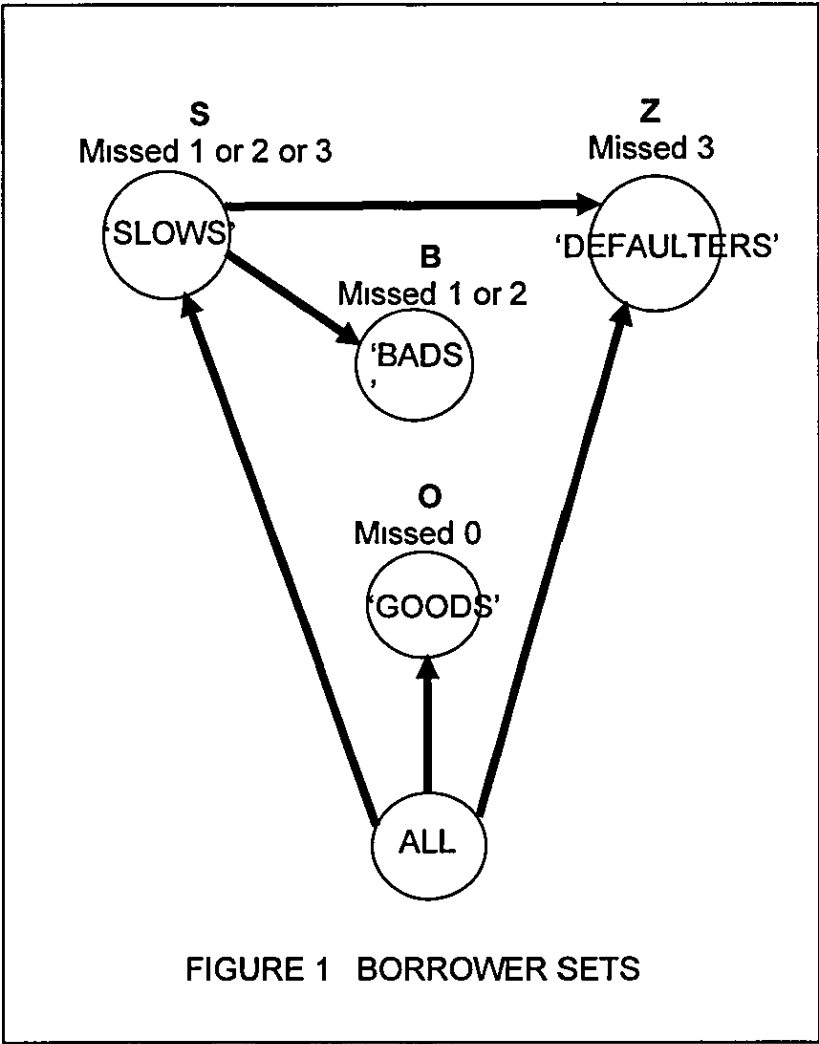
$S = X \cup Y \cup Z$

$B = X \cup Y$

We will call those in set O 'goods'. Those in set S will be called 'slows' because they have missed between one and three consecutive payments, but not necessarily three. Those in set B will be called 'bads'. Those in set Z will be called 'defaulters'. Casual evidence suggests that credit granters regard the failure to make three consecutive payments as

considerably worse than failure to make two consecutive payments, and some granters may pass the debt to a collection agency if three consecutive payments are missed.

In this article we wish to make three comparisons as follows. We wish to distinguish between sets: (1a) O and (1b) S; (2a) O and (2b) Z; and (3a) B and (3b) Z. Diagrammatically the sets are presented in Figure 1. Thus we ask: (1) can we distinguish between those who have never missed a payment and those who have missed at least one; (2) can we distinguish between those who have never missed a payment and those who have missed three consecutively; and (3) given that a person has missed at least one payment can we distinguish between those who miss only one or two consecutively and those who miss three consecutively?



The sample consists of 1001 individuals who held a bank credit card and who used it in the sample period. Data was available on 23 sociodemographic and economic variables which have either been used in previously published discriminant analysis scoring models [see Capon, 1982] or for which an *a priori* reason as to why they may act as effective discriminators could be made. The 23 variables are shown in Appendix 1. All data were taken from the applicants' application forms which they completed between September 1986 and December 1987

Estimation Methodology

The methodology follows that of Crook *et al.* [1991a]. Briefly, many of the variables were measured at nominal level, whilst the use of discriminant analysis requires data measured at least at interval level [see Klecka, 1980]. Additional information was used to derive interval level data by ascribing to each predictor the following values:

$$X_j = 1n \frac{g_i}{b_i} + 1n \frac{B_T}{G_T}$$

where

- X_j = value of predictor X for case j;
- g_i = number of good payers in nominal category i, the category of which j was a member,
- b_i = number of poor payers in nominal category i, the category of which j was a member;
- G_T = number of good payers in the sample;
- B_T = number of poor payers in the sample.

This use of the X_j transformation means that X_j may not be monotone in the values of the original variable. Thus X_j may not monotonically increase or decrease with, say, spouse's income. This will be considered subsequently.

As in Crook [op cit], for each discriminant analysis high degrees of collinearity between predictor variables were reduced by variable deletion. Predictors were selected for inclusion in the empirical function by a stepwise procedure. At each step the variable which results in the greatest Mahalanobis Distance (D^2)¹ was added. The significance of a change in D^2 when a variable was included was tested by the use of a partial F statistic. The F to enter and F to remove values were set equal to 1.00, this being a compromise between giving a high degree of predictive performance as well as including variables of a relatively high degree of statistically significant discriminatory power.

Turning to the assessment of the predictive performance of an estimated function, several methods are available [see Eisenbeis, 1977, Kschirsagar, 1972, Lachenbruch and Mickey, 1968]. Two commonly used alternative techniques are, first, to estimate the function from a sub-set of the total sample and to use this function to classify the remainder of the sample, and second, to delete one observation in turn, estimate the function and classify the deleted case. The former or hold-out sample method has the limitation of requiring a large sample size but the number of poor payers in two of our functions is very much lower than the number of good payers. The latter (or Jackknife, or U-method) does not have this limitation and in a comparison with nine other methods Eisenbeis [1977] argued that it was the best when used with small samples. For this reason we have used the Jackknife method.

Finally, note that because we are interested in the chance that we have correctly predicted group membership of a poor-paying individual given that he has been predicted to be a poor payer and the chance that we have correctly predicted group membership of a good payer given that he has been predicted as good, we will compare the proportion of cases correctly classified with C_{prop} where:

$$C_{prop} = \sum_{i=1}^2 P_i \alpha_i$$

P_i = proportion of cases in group i ,

α_i = proportion of cases predicted to be members of group i ;

$i = 1$ good payers $i = 2$ poor payers.

RESULTS

Significance and Predictive Performance

Table 1 shows that for each function separately the discrimination to be achieved by the appropriate set of predictor variables prior to the estimation of each function is statistically highly significant.

TABLE 1
SIGNIFICANCE OF THE ESTIMATED FUNCTIONS

	Wilks' Lambda	χ^2	d f.	Significance
SLOWS (Ever been at least one cycle delinquent)	0.8759	131.44	14	0.000
DEFAULTERS (Ever been 3 cycles delinquent)	0.9144	88.86	12	0.000
BADS (Maximum number of consecutive cycles delinquent is 3 not 1 or 2)	0.8295	69.72	12	0.000

Table 2 shows that the percentage correctly classified exceeded C_{prop} in all three cases. The greatest number of percentage points by which the proportion correctly classified by an

estimated function exceeded C_{prop} corresponded to the function predicting slow payers. However, it must be noted that there were only 5.5 and 14.8 percentage points between C_{prop} and 100 per cent which were available for improvement by the defaulters and 'bads' functions respectively.

TABLE 2				
CLASSIFICATION MATRICES				
(Jackknife Method)				
SLOWS				
		Predicted Group		
		Good	Bad	Total
<i>Actual Group</i>	Good	530	90	620
	Bad	226	155	381
	Total	756	245	1001
Percentage correctly classified		68.4%		
C_{prop}		46.8%		
DEFAULTERS				
		Predicted Group		
		Good	Bad	Total
<i>Actual Group</i>	Good	948	9	957
	Bad	41	3	44
	Total	989	12	1001
Percentage correctly classified		95.0%		
C_{prop}		94.5%		
BADS				
		Predicted Group		
		Good	Bad	Total
<i>Actual Group</i>	Good	326	11	337
	Bad	39	5	44
	Total	365	16	381
Percentage correctly classified		86.9%		
C_{prop}		85.2%		

For each of the three cases the proportion of good payers which were correctly classified considerably exceeded the proportion of poor payers. The proportion of individuals who were correctly predicted to miss no payments when predicting 'slows' was less than the corresponding proportion when predicting those who 'default', at 85.5 per cent and 99.1 per cent respectively. Alternatively, the proportion who were correctly predicted to miss at least one payment was greater than the proportion who were correctly predicted to miss three payments, at 40.7 per cent and 6.9 per cent respectively. Clearly, without knowledge of the opportunity costs of mis-classifying a 'poor' payer and those of mis-classifying a 'good' payer for each type of poor payer it is impossible to decide which function would be the most effective as a credit control device.

Ranking of Variables

First we will compare the credit granting models and second we will compare these with the performance scoring model. Table 3 shows the standardised coefficients for each variable which was included in the estimated function on the F statistic criteria of the stepwise procedure. For each function the rank order of variables in terms of their discriminating power is the same if the standardised coefficients are considered as if the partial F statistic is used. Therefore, the F statistics are not presented. The standardised coefficients represent the relative discriminatory power of each variable given the other variables in the function. On these criteria the rank descending order of the most powerful six predictors of those who miss at least one payment as opposed to no payments is applicant's employment status, number of children, years at the bank, mortgage balance outstanding, residential status and major credit card respectively. The rank descending order of the most powerful six predictors of those who miss three consecutive payments instead of none is applicant's employment status, spouse's income, years at bank, residential status, years at present employment and cheque account. The corresponding rank order of predictors which distinguishes between those who miss one or two consecutive payments and those who miss three is years at the bank, spouse's income, applicant's employment status, years at present employment, and deposit account and outgoings.

TABLE 3

STANDARDISED COEFFICIENTS

	Variable	Standardised Coefficient	Rank
Slows	Applicant's employment status	0.47	1
	Number of children	0.41	2
	Years at bank	0.37	3
	Mortgage balance outstanding	0.28	4
	Residential status	0.26	5
	Major credit card	0.22	6
	Years in present employment	0.22	7
	Outgoings	0.22	8
	Current account	0.20	9
	Charge card	0.15	10
	Applicant's income	0.13	11
	'Phone	0.11	12
	Estimated value of home	0.10	13
	Spouse's income	0.10	14
Defaulters	Applicant's employment status	0.41	1
	Spouse's income	0.40	2
	Years at bank	0.38	3
	Residential status	0.34	4
	Years in present employment	0.30	5
	Cheque card	0.27	6
	Outgoings	0.22	7
	Major credit card	0.20	8
	Number of other dependants	0.14	9
	Store card	0.14	10
	'Phone	0.13	11
	Deposit account	0.13	12

TABLE 3 (contd.)

STANDARDISED COEFFICIENTS			
	Variable	Standardised Coefficient	Rank
Bads	Years at bank	0.47	1
	Spouse's income	0.42	2
	Applicant's employment status	0.40	3
	Years in present employment	0.28	4
	Deposit account	0.25	5
	Outgoings	0.25	6
	Residential status	0.22	7
	Cheque card	0.20	8
	Estimated value of home	0.18	9
	Applicant's income	0.17	10
	Store card	0.15	11
	'Phone	0.14	12

But standardised coefficients may give an increasingly inaccurate indication of the discriminatory power of each variable individually, the greater is the degree of correlation between any predictor variables included in the function.

We therefore consider the rankings on the basis of the bivariate correlation coefficients between the discriminant scores and the values of each predictor variable. These are unaffected by other variables included in the function and are shown in Table 4.

TABLE 4

STRUCTURE MATRICES (POOLED WITHIN GROUPS CORRELATION COEFFICIENTS)			
	Variable	Correlation Coefficient	Rank
Slows	Applicant's employment status	0.53	1
	Years at bank	0.44	2
	Mortgage balance outstanding	0.43	3
	Number of children	0.33	4
	Years in present employment	0.32	5
	Residential status	0.25	6
	Current account	0.21	7
	Charge card	0.19	8
	Outgoings	0.19	9
	Estimated value of home	0.18	10
	'Phone	0.15	11
	Applicant's income	0.12	12
	Spouse's income	0.12	13
	Major credit card	0.10	14
Defaulters	Years at bank	0.52	1
	Applicant's employment status	0.43	2
	Cheque card	0.37	3
	Years in present employment	0.37	4
	Spouse's income	0.34	5
	Residential status	0.32	6
	'Phone	0.26	7
	Outgoings	0.18	8
	Deposit account	0.17	9
	Major credit card	0.16	10
	Store card	0.16	12
	Number of other dependants	0.14	14
Bads	Years at bank	0.51	1
	Applicant's employment status	0.40	2
	Cheque card	0.36	3
	Spouse's income	0.36	4
	Applicant's income	0.30	5
	Years in present employment	0.29	6
	Residential status	0.26	7
	Outgoings	0.23	9
	'Phone	0.22	10
	Estimated value of home	0.18	11
	Store card	0.17	12
	Deposit account	0.17	13

Variables not selected by stepwise routine

TABLE 4 (contd.)
STRUCTURE MATRICES (POOLED WITHIN GROUPS CORRELATION COEFFICIENTS)

	Variable	Correlation Coefficient	Rank
Slows	Store card	0 05	15
	Building society card	-0 02	16
	Loan account	-0 01	17
	Deposit account	0 01	18
	Number of other dependants	0 01	19
Defaulters	Mortgage balance outstanding	0 16	11
	Applicant's income	0 15	13
	Estimated value of home	0 10	15
	Charge card	0 05	16
	Loan account	0 05	17
	Number of children	0 04	18
	Building society card	0 001	19
Bads	Age	0 25	8
	Years at present address	0 16	14
	Mortgage balance outstanding	0 16	15
	Charge card	0 16	16
	Major credit card	0 15	17
	Number of other dependants	0 08	18
	Number of children	0 08	19
	Building society card	-0 06	20
	Loan account	0 05	21

In terms of the bivariate correlations four, applicant's employment status, years at bank, years at present employment and residential status, of the most important six predictors are

identical in the functions to predict slows and defaulters. Applicant's employment status and years at bank are ranked either first or second in the two functions. Residential status is ranked sixth on both cases and years at present employment is ranked fourth or fifth.

However, there are some noticeable differences in rankings. Some variables have a higher rank when used to predict defaulters as opposed to slows. Whilst spouse's income is ranked fifth in terms of its ability to discriminate those who miss three consecutive payments from the rest it is ranked only thirteenth in terms of its ability to predict those who miss at least one payment. The possession of a cheque card was the third most important predictor of defaulters but was not eligible for inclusion in the function which predicted slows due to correlation with years at bank and current account. However, the latter were ranked second and seventh respectively.

On the other hand, some variables have a higher rank when used to predict slows than when used to predict defaulters. For example, mortgage balance outstanding was the third most important discriminatory variable when predicting those who missed at least one payment, but it contributed no statistically significant additional discrimination (using $F = 1.0$ value) between those who did and those who did not miss three payments, and was ranked eleventh in terms of its bivariate correlation with the discriminant score. Similarly, number of children was the fourth most important discriminatory variable when used to predict slows but also contributed no statistically significant discrimination between defaulters and non-defaulters and was ranked eighteenth. The possession of a charge card was ranked eighth in the slows function but was not statistically significant (even at $F = 1.0$) and ranked sixteenth in the defaulters equation.

We now turn to a comparison of the rankings of the predictors in the performance model, which discriminates between those who missed one or two consecutive payments and those who missed three consecutive payments, with the rankings of the two scoring models. For reasons given earlier we confine our comparisons to be on the basis of the

bivariate correlation coefficients. Table 4 shows that the ranking of the variables which predict whether an individual will miss three rather than just one or two consecutive payments is very similar to that of the variables which predict the missing of three rather than no, one or two payments, but it has noticeable differences compared with the discriminators of those who miss one, two or three from those who missed zero payments.

We will compare the performance model firstly with the defaulters and secondly with the slows model. Five of the highest ranking six predictors are identical in the defaulters and performance models, the top three predictors being in the same rank order. These five predictors are years at bank, applicant's employment status, cheque card, spouse's income and years at present employment. Ten of the twelve variables which added a statistically significant amount of additional discrimination were the same in both estimated functions. One noticeable difference in ranking related to applicant's income, which was ranked fifth in the performance model but was not included in the defaulters function. Other major differences in rankings related to variables with relatively low correlation coefficients in both models. For example, estimated value of home is ranked eleventh in the performance model but fifteenth and not included in the defaulters model; possession of major credit cards and the number of other dependants were ranked tenth and fourteenth respectively in the defaulters model but seventeenth and eighteenth and not included in the performance model

Table 4 also shows that only three predictors applicant's employment status, years at bank and years at present employment, are amongst the top six for both the performance and the slows models. Certain other predictors of performance outside the top six have rankings which are within one rank of their rank in the slows function. These are residential status, outgoings, estimated value of home and possession of a phone. However, there the similarity ends. There are a number of relatively large differences in the ranks. Spouse's income is ranked fourth and thirteenth in the performance and slows models respectively and applicant's income is ranked fifth and twelfth respectively Mortgage balance

outstanding and number of children are ranked third and fourth in the slows model, but fifteenth and nineteenth respectively and not included in the performance function.

Interpretation Of Variables

As explained above, values of X_j which were ascribed to the predictor variables were not monotone in the values of those variables. Therefore to interpret the relationship between the discriminant score and the characteristics of individuals it is necessary to consider the relationship between X_j and these characteristics

Firstly, notice that $X_j = 1/n \cdot g_j/b_j + k$ where g_j and b_j are as defined earlier and k is a constant and so will not vary with the original values of the predictor variable. Therefore a higher value of X_j indicates a higher ratio of the number of 'goods' to 'bads' in a range of original values taken on by the predictor variable

Years at bank is ranked first or second in all three functions. In the analysis of defaulters years at bank and X_j are not monotonically related. Those with accounts at the bank for less than six months are better payers than those with accounts of one or two years of age. Thereafter the ratio of 'goods' to 'bads' increases with account age. The worst payers are those who have been with the bank for one or two years, the best are those who have had an account for over 11 years. The relationship between X_j and years at bank for the slows analysis is very similar, except those with accounts for under six months have just as low a ratio of 'goods' to 'bads' as those with accounts for one to three years. In the case of the performance model the relationship between X_j and years is almost identical to that of defaulters. Thus, of those who miss at least one payment the proportion of those who go on to miss three consecutive payments will be greater for those who have had an account for under two years than for those who have had an account for longer

Applicant's employment status was also ranked first or second in the three functions. In the case of defaulters, the proportion of each group who miss three consecutive payments rather than a zero, one or two, was greatest for housewives, members of the armed forces and the unemployed, followed by private sector employees. The proportion was lowest for public sector employees, the retired, government (non-military) employees, those with no response to the question for this data and those in 'other' groups³ Turning to the analysis of slows, the greatest percentage of those who missed one, two or three consecutive payments rather than none were in the 'others' and self-employed categories, followed by private sector. The lowest percentage was amongst public sector and retired employees. In short, everything else equal, those most likely to miss three consecutive payments rather than zero, one or two are housewives, members of the armed forces and the unemployed. Those most likely to miss at least one payment rather than none are the self-employed, 'others' and those in the private sector. The rankings of the X_j values for the defaulters model also apply to the performance model. Therefore, of those who have missed at least one payment, those most likely to miss three in succession are housewives, members of the armed forces and the unemployed.

The possession of a cheque card was ranked third in the defaulters and performance models but was not included in the slows model because it was correlated with years at bank and current account. Both those most likely to miss three consecutive payments rather than zero, one or two, and those likely to miss three consecutive payments rather than only one or two are those without a cheque card.

Years at present employment was ranked fourth or fifth or sixth in the three functions. The relationship between the values of X_j and this variable is similar for all three models. In each case the value of X_j decreases at first as years increase, reaches a minimum at a relatively small number of years and increases monotonically thereafter. Hence the proportion of individuals in each period grouping who miss three consecutive payments increases until they have been in the same employment for four years and decreases thereafter. The proportion who miss at least one payment rather than none increases for

one year and decreases thereafter, whereas the proportion who, having missed at least one payment, subsequently miss three in succession increases over three years and decreases thereafter.

Turning to residential status, the categories most likely to miss three consecutive payments rather than zero, one or two, those most likely to move from missing one or two to missing three, and those most likely to miss at least one are the same 'others' (i.e., not owners, living with parents or tenants). In the former two cases this is followed by tenants in unfurnished accommodation. Those least likely to default or to move from a one or two cycle delinquency to three-cycle delinquency are those living with parents. However, those least likely to miss at least one cycle rather than never to do so are tenants in unfurnished housing. This is consistent with the argument that tenants in unfurnished accommodation are relatively less likely to miss a payment than those on other types of accommodation, but if they do they have a greater chance of missing three in succession rather than just one or two.

We now turn to variables with large differences in rank between the three models.

Spouse's income was ranked fifth and fourth in the defaulters and performance models respectively but thirteenth in the slows model. The relationships between X_j and spouse's income are shown in Appendix 2. Remember that monetary values are at late 1986-87 prices. In the case of three-cycle delinquency, after a slight decrease the proportion who default increases as income rises to £5,000 to £7,500 and decreases thereafter. In the case of those who move from one or two to three-cycle delinquency, the pattern is broadly similar.⁴ For those who miss at least one payment there is no clear relationship. The data suggest that if a spouse has no income there is a relatively high chance that at least one payment will be missed but a relatively low chance that the individual will move from missing one or two payments to missing three in succession. We can also note that if a spouse earns over £15,000 the chance that at least one payment or three rather than zero, one or two payments is missed is relatively low. If the spouse earns over £10,000 the chance that an individual will move from missing one or two consecutively to missing three

is also reduced. The data also suggest that if a spouse earns between £5,000 and £7,500 (£10,000 in the case of the performance model) then, everything else equal, the chance that at least one payment and that three rather than zero, one or two payments are missed is greatest as is the chance that someone who is already delinquent will miss three successive payments.

Notice also that whilst all of these chances are relative to those at other income levels, spouse's income has greater discriminatory power compared with other discriminators when predicting three-cycle delinquency than when predicting at least one cycle delinquency. In short, having a spouse with no or a very high income significantly affects whether an individual misses three consecutive payments, whilst having such a spouse has little effect on predicting whether or not an individual misses at least one payment.

Mortgage balance outstanding was ranked third and number of children fourth in the function which identified those who missed at least one payment but neither had any statistically significant discriminatory power beyond the other variables in the other two models.⁵ Apart from being a non-owner, the chance of missing at least one payment monotonically increases as mortgage balance outstanding increases. Number of children was also negatively and monotonically related to the chance of missing at least one payment. The more children one has, the greater the chance that at least one payment will be missed.

Applicant's income is an especially interesting variable because it ranks fifth in terms of bivariate correlation when predicting those who will miss three rather than one or two payments but only twelfth in the case of predicting those who will miss at least one payment and is not included in the three cycle function at all. In all three functions the relationship with the proportion of 'goods' to 'bads' was a W shape as income increases. The chance of a person moving from two-cycle delinquency to three is lowest for those earning over £15,000, and greatest for those earning between £7,500 and £10,000 (late 1986-87 prices).

The same applies to those who miss three rather than zero, one or two payments. Thus, applicant's income has little effect on whether at least one or three rather than zero, one or two consecutive payments are missed. But it is strongly associated with whether an individual moves from two cycle into three cycle delinquency, with those earning most being least likely to do so.

CONCLUSION

We have estimated three discriminant functions. Two are credit scoring models which distinguish between bank credit card holders who miss at least one payment and those who miss none and between those who miss three consecutive payments and those who do not. The third is a credit performance model which distinguishes between those who miss one or two consecutive payments and those who miss three. All functions are statistically significant and all predict better than chance. Those most likely to miss at least one payment are those who have had an account with the bank for under three years, those who are self-employed or belong in the 'other' employment category, those who have been at their present employment for one year, who fit into the 'other' residential status group, those who have a large mortgage balance outstanding and those with four or more children. Those most likely to miss three consecutive payments rather than none, one or two are those who have been with the bank for one or two years, those who are members of the armed forces, unemployed people or housewives, those who have been in the same employment for four years, those without a cheque card, those in the 'other' residential status category, and those whose spouse earned between £5,000 and £7,500. Of those individuals who have missed at least one payment, those who are most likely to become three-cycle delinquent rather than miss only one or two payments are: those who have been with the bank for under two years, members of the armed forces, housewives, the unemployed, those who have been in their present employment for three years, those without a cheque card, those whose spouse earns between £5,000 and £7,500, and those who earn £7,500-£10,000 per year (late 1986-87 prices).

NOTES

The support of the Economic and Social Research Council (ESRC) is gratefully acknowledged. The work was funded by ESRC Award No.R000231152,

1. The Mahalanobis Distance statistic is defined as:

$$D^2_{g,b} = (n-k) \sum_{i=1}^m \sum_{j=1}^m w^*_{i,j} (\bar{x}_{i,j} - \bar{x}_{i,b})(\bar{x}_{j,g} - \bar{x}_{j,b})$$

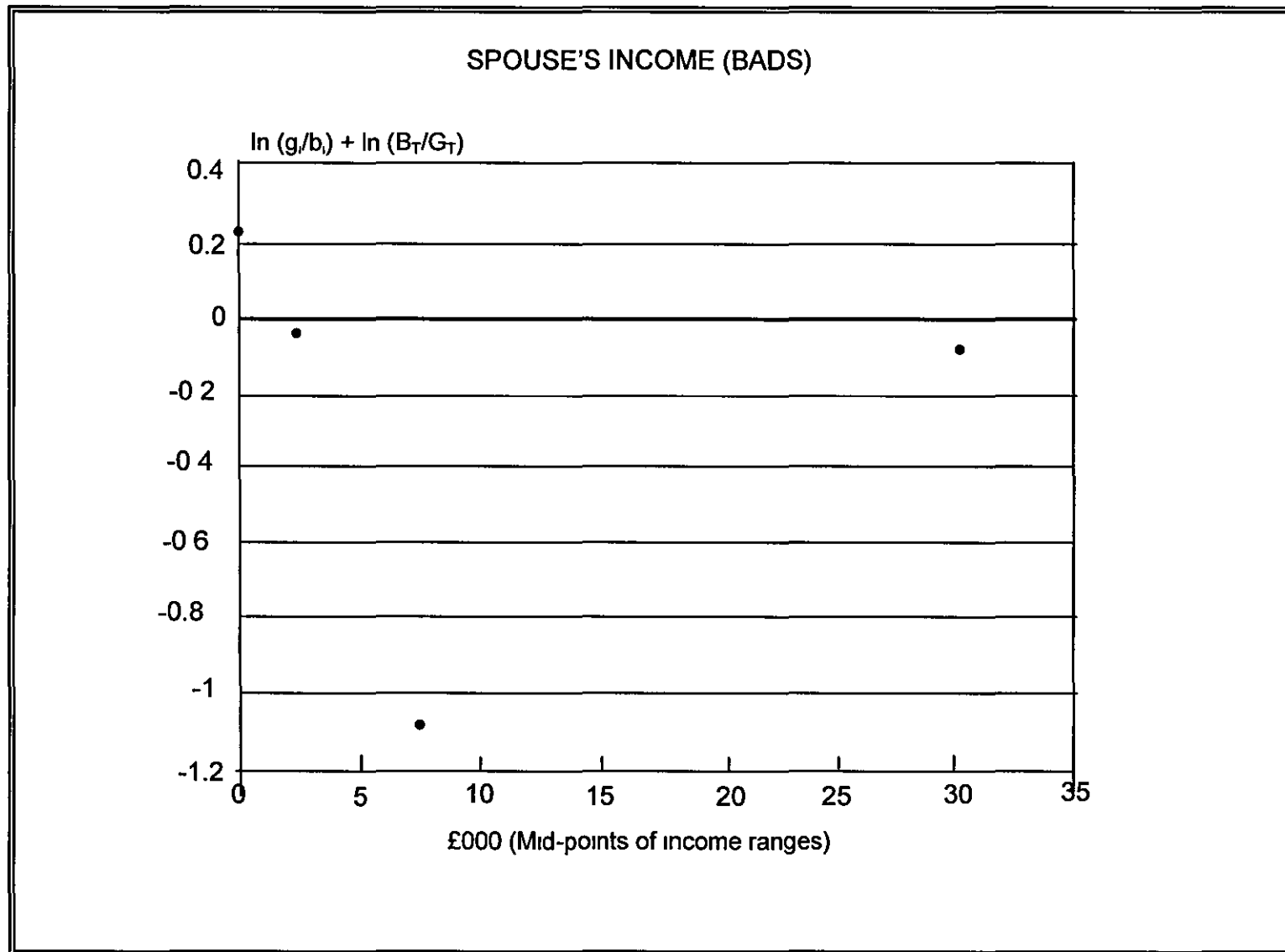
where m = number of predictor variables in the model;
 k = number of groups;
 g,b = the groups of 'good' and 'bad' cases respectively;
 $X_{i,g}$ = sample mean value of predictor i for group g ;
 $w^*_{i,j}$ = an element from the inverse of the within group's covariance matrix.

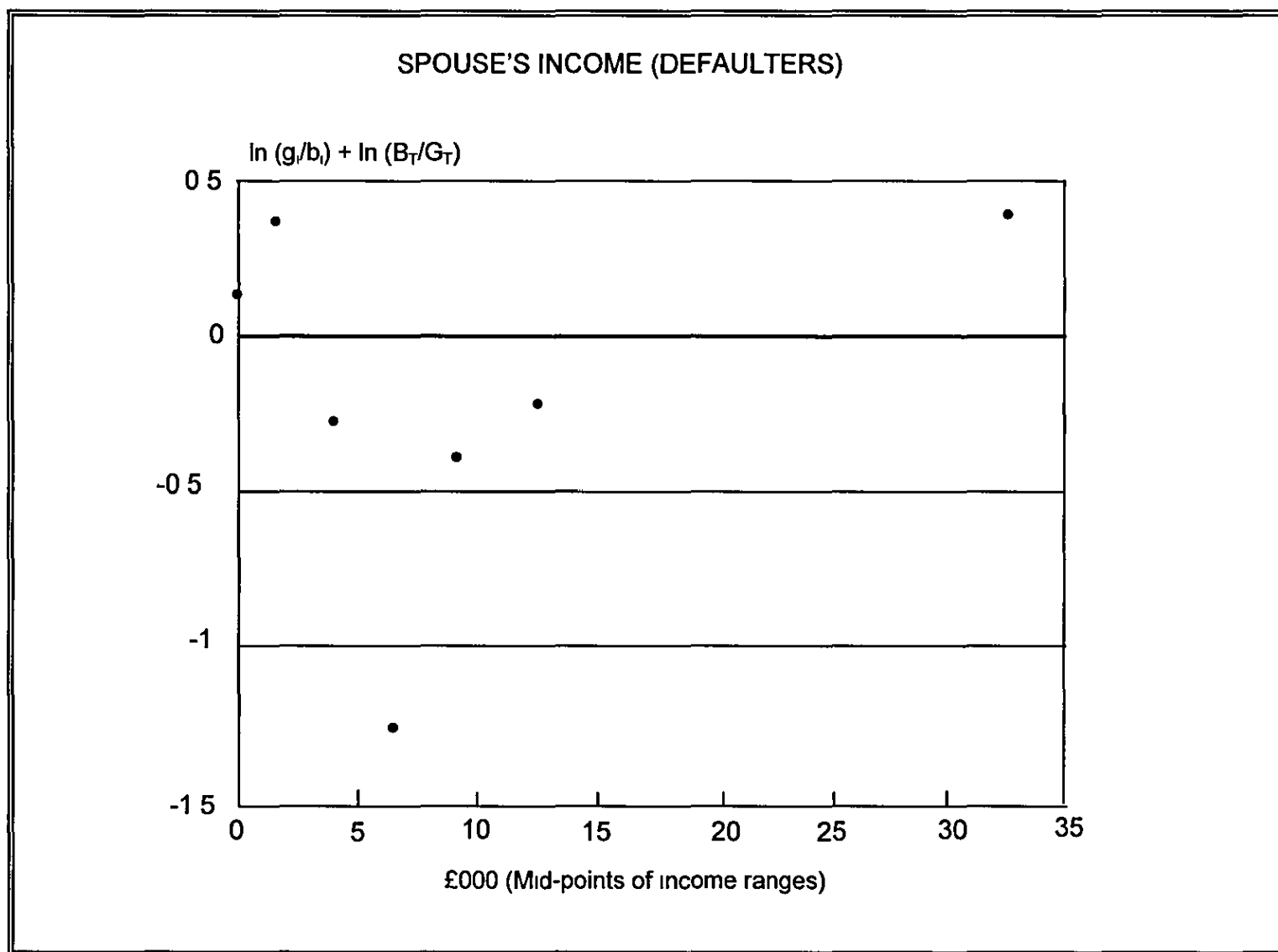
2. The values of X_i for each original value of the predictor variables is available from the authors on request.
3. The 'other' category includes all occupations except: public sector, retired, government (non-military), students, self-employed, private sector, housewife, military, and unemployed.
4. The income groupings differ between (a) the default and slow models and (b) the performance model due to the differing degree of homogeneity of the g/b_i values in each income range.
5. At an F value of 1 00.

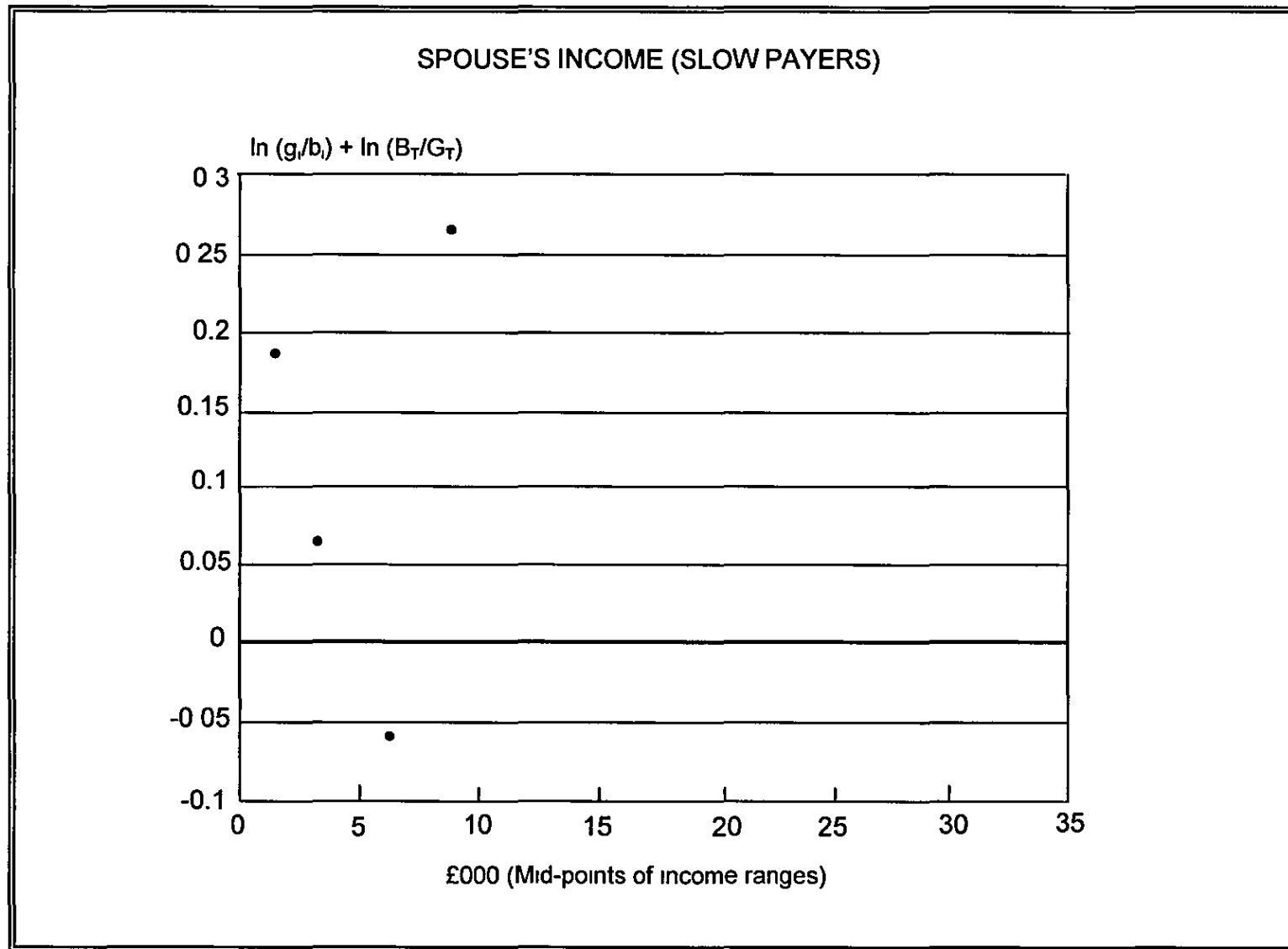
APPENDIX 1
THE ORIGINAL 23 PREDICTOR VARIABLES

Age
Number of children
Number of other dependants
Whether an applicant has a home 'phone
Spouse's income
Applicant's employment status
Applicant's employment category
Years at present employment
Applicant's income
Residential status
Years at present address
Estimated value of home
Mortgage balance outstanding
Years at bank
Whether a current account is held
Whether a deposit account is held
Whether a loan account is held
Whether a cheque guarantee card is held
Whether a major credit card is held
Whether a charge card is held
Whether a store card is held
Whether a building society card is held
Value of outgoings

APPENDIX 2







REFERENCES

Bierman, H. and Hausman, W.H. (1970), "The Credit Granting Decision", *Management Science*, Vol. 16(8), April, pp. B519-B532.

Boyle, M , Crook, J N., Hamilton, R. and Thomas, L.C., (1991), "Methods of Credit Scoring Applied to Slow Payers", in Crook, J N , Edelman, D and Thomas, L.C (eds.), *Credit Scoring and Credit Control*, Oxford. Oxford University Press, December, pp 75-90.

Capon, N., (1982), "Credit Scoring Systems A Critical Analysis", *Journal of Marketing*, Vol 46, Spring, pp. 82-91

Chandler, G C. and Coffman, J.Y., (1979), "A Comparative Analysis of Empirical Vs Judgemental Credit Evaluation", *Journal of Retail Banking*, Vol 1(2), September, pp. 15-25

Chandler, G.C. and Coffman, J.Y., (1983-4), "Applications of Performance Scoring to Accounts Receivable Management in Consumer Credit", *Journal of Retail Banking*, Vol. 5(4), Winter, pp. 1-10

Crook, J.N., Hamilton, R. and Thomas, L C., (1991a), "A Comparison of Discriminators under Alternative Definitions of Default", in Crook, J N , Edelman, D and Thomas, L C. (eds.), *Credit Scoring and Credit Control*, Oxford Oxford University Press, December, pp 217-246.

Cyert, R.M., Davidson, H.J. and Thompson, G.L. (1962) "Estimation of the Allowance for Doubtful Accounts by Markov Chains", *Management Science*, April.

Cyert, R.M., Davidson, H J and Thompson, G L., (1968), "Selecting a Portfolio of Credit Risks by Markov Chains", *Journal of Business*, Vol.1, January.

Eisenbeis, R A (1978), Problems in Applying Discriminant Analysis in Credit Scoring Models", *Journal of Banking and Finance*, Vol 2, pp 205-219.

Eisenbeis, R A (1977) "Pitfalls in the Application of Discriminant Analysis in Business, Finance and Economics", *Journal of Finance*, Vol 32 (3), pp 875-900.

Fryman, H., Kallberg, J .G and Kao, D.L., (1985), "Testing the Adequacy of Markov Chain and Mover-Stayer Models as Representations of Credit Behaviour", *Operations Research*, pp. 1203-1214.

Hair, J F , Anderson, R E and Tatham, R L , (1987), *Multivariate Data Analysis*, New York: Collier Macmillan.

Kshirsagar, A.M. (1972) *Multivariate Analysis*, Vol.2. Marcel Drekker

Klecka, W R. (1980), *Discriminant Analysis*, Sage University Paper No.19, Series: Quantitative Applications in the Social Sciences

Lachenbruch, P.A. and Mickey, M.R. 1968, "Estimation of Error Rates in Discriminant Analysis", *Technometrics*, Vol.10, No.1, February.

Morrison, D.G., (1969), "On the Interpretation of Discriminant Analysis", *Journal of Marketing Research*, Vol. 4, pp. 156-163.

Myers, J R and Forgy, E.W , (1963), "The Development of Numerical Credit Evaluation Systems", *American Statistical Association Journal*, September, pp. 799-806.

Thomas, L.C., "Behavioural Scoring", Department of Business Studies, University of Edinburgh, Working Paper no 88/32, 12 pp.

Wiginton, J.C., (1980), "A Note on the Comparison of Logit and Discriminant Models of Consumer Credit Behaviour", *Journal of Financial and Quantitative Analysis*, Vol. 15(3), September, pp. 757-770.

CHAPTER 6

CREDIT CARD HOLDERS: CHARACTERISTICS OF USERS AND NON-USERS

J N. Crook and L. C. Thomas

(Department of Business Studies, Edinburgh University)

R. Hamilton

(Business School, Loughborough University)

The Service Industries Journal, (1992), Vol 12, No 2 (April), pp. 251-262

Credit Card Holders: Characteristics of Users and Non-Users

J. N. Crook, R. Hamilton and L. C. Thomas

This paper aims to distinguish between those who hold and use bank credit cards and those who hold them but do not use them. Discriminant analysis is applied to a sample of 825 holders of a bank credit card. The most important discriminators were where a card holder lives, age, income, years for which an account has been held at the issuing bank, years at present address and residential status. The results suggest particular market segments towards which a bank may wish to target its promotion, product and pricing strategies if it wishes to attract users, non-users or to convert the latter into the former.

INTRODUCTION

The use of statistical techniques and mathematical models to assist financial institutions in the credit granting decision-making process has significantly increased in the last fifteen or twenty years. Such credit-scoring techniques are no longer used only in the simple 'reject or accept' situation but are applied in many other areas as well [Boyle, Crook, Hamilton, Thomas, 1989].

Exactly how the different techniques can and have been applied to the various decision-making situations has been well documented [Capon, 1982: 82-91, Bierman and Hausman, 1970. B519-532]. This paper, however, examines how one such technique – Discriminant Analysis – could be used notably by credit card companies especially in an area that has not as yet been addressed in any of the published literature.

Within the portfolio of any credit card company a number of distinct subsets can be identified: those accepted who default, those accepted who do not default and, finally, those accepted who never use the card issued. It is the members of this last group, especially in

the light of the various issues raised in the Monopolies and Mergers Commission (MMC) report (Monopolies and Mergers Commission, 1989), and the subsequent introduction of annual fees for credit cards, that this paper is particularly concerned with

If the MMC recommendations, such as the removal of any rules which force retailers to charge the same price to cash as to credit card customers, are introduced, then the usage of credit cards may decline. All card-issuing organisations may follow Lloyds, Barclays and other banks in charging an annual fee to all card-holders. Banks may need to reassess whether they wish to attract non-users of their cards or whether to target only users

Arguments concerning non-users may go either way. On the one hand, banks may argue that since non-users are not going to use their card they would not be willing to pay the fixed charge and so would yield no income. On the other hand, non-users may be viewed as an important source of revenue, albeit only for the fixed charge. In this case a supplier may wish to target non-users and potential non-user non-card holders with promotional messages which emphasise the card as a convenient and quick source of financial back-up. Furthermore, by holding a card the holder may, when requiring any new or additional financial service, think first of using the institution whose card (s)he holds. Then a supplier may wish to design promotional activity to target non-users to emphasise the product *brand*.

Clearly, users who pay interest on debt outstanding are attractive customers to acquire. But regardless of whether the company wishes to attract users or non-users or both, its promotion, product and pricing strategies could be more effectively targeted if the bank is able to predict those who use, as opposed to those who would not use, its card in terms of their socio-demographic and economic characteristics. This paper reports the results of a statistical analysis which indicates which socio-demographic and economic characteristics distinguish between these two groups and so presents the characteristics which segment the market. The second section of this paper describes the data and the methodology used. The third section discusses the results and the final section concludes

DATA, VARIABLES AND METHODOLOGY

The data were supplied by a UK clearing bank which must remain anonymous. The sample was selected from those who applied for and were granted the bank's credit card during the period 1 September 1986 to 31 December 1987 and who were recruited through a representative group of media. The selection procedure was random and based on account numbers.¹ Thus, 1,225 individuals were selected of whom 224 had never used their card ('non-users') and 1,001 who had used their card on at least one occasion (that is 'users')

Data were available on 24 socio-demographic and economic variables for which an *a priori* reason for their use as discriminators could be given. These variables are listed in Table 1 and it can be seen that most have been included in previously published discriminant analysis scoring models [Capon, 1982: 82-91].

Table 1

THE ORIGINAL 24 PREDICTOR VARIABLES

Postcode
Age
Number of children
Number of other dependants
Whether an applicant has a home 'phone
Spouse's income
Applicant's employment status
Applicant's employment category
Years at present employment
Applicant's income
Residential status
Years at present address
Estimated value of home
Mortgage balance outstanding
Years at bank
Whether a current account is held
Whether a deposit account is held
Whether a loan account is held
Whether a cheque guarantee card is held
Whether a major credit card is held
Whether a charge card is held
Whether a store card is held
Whether a building society card is held
Value of outgoings

An immediate difficulty can be seen in that many of the variables are measured only at nominal level whilst use of discriminant analysis requires that all predictor variables are measured at least at interval level [Klecka, 1980] To overcome this difficulty each such variable was replaced by one measured at interval or higher levels. This was done for each case, j , by replacing each nominal value, i , by a derived value $X^1_j(i)$:

$$X^1_j(i) = 1n \frac{u_i}{v_i} + 1n \frac{V_T}{U_T}$$

where u_i and v_i are the number of users and non-users respectively in the sample which take on the i^{th} nominal value, and U_T and V_T are the total number of users and non-users respectively in the sample.²

Turning to those variables which were measured at ratio level, it is sometimes the case that the proportion of non-users is not monotone in these variables. Since the primary objective of the model is to gain maximum discrimination and prediction, not to describe, the aggregation procedure was applied to these variables too, which meant that the derived values were not monotone in the original values³

Since using linear discriminant analysis to discriminate between users and non-users is particularly susceptible to any multi-collinearity between the predictor variables, any variables which are seriously inter-correlated were excluded from the analysis. The deleted variables were: applicant's employment status, applicant's employment category, years at present employment, estimated value of home, mortgage balance outstanding, whether a cheque guarantee card is held, and value of outgoings

To ensure that only those variables which contributed significantly to the discrimination were included in the final function, the predictors were selected by a step-wise procedure⁴ The selected variables are shown in Table 2.

Table 2

VARIABLES IN THE ANALYSIS

Variable	Step entered
Postcode	1
Age	2
Applicant's income	3
Years at bank	4
Years at present address	5
Residential status	6

To avoid bias in assessing the predictive performance of the model [Frank, Massy, Morrison, 1965 250-258], the analysis was carried out on a random sample of 825 from the 1,225 cases and the predictive accuracy assessed from the holdout sample of the remaining 400 cases.

To assess the predictive performance of the model, the proportion of the cases which is correctly classified by the function must be compared with the proportion which we would expect to be correctly classified by chance. In this paper we wish to classify correctly both users and non-users. Therefore, we use the proportional chance criterion (*Cprop*) which predicts the proportion of cases which one would expect to be correctly classified if we randomly allocate classes between the two groups given the proportions which are actually in each group. *Cprop* is given by the formula

$$C_{prop} = p^2 + (1 - p)^2$$

where *p* is the proportion of cases in one of the groups, for example, users.⁵

A limitation of our methodology should be acknowledged. Of the 84 postcodes for which data were available many had fewer than, say, five observations with consequently relatively high sampling variances for the value of X^1 . Since postcodes were aggregated by

similarity of the proportion of cases within a postcode who were users, it is possible that postcodes may have been inappropriately aggregated. Hence they may play an artificially significant role in the discriminating function.⁶

However, following the earlier work of Crook *et al* and, more importantly, the fact that the inclusion/exclusion of this variable makes very little difference to either the ranking of the other variables (only spouse's income enters the final function when postcode is excluded) or the predictive performance of the model, the following discussion of our results will refer to the analysis carried out with postcode included. For comparison purposes, Appendix 1 gives the results of calculations with postcode excluded.

RESULTS

Significance of the Function

Table 3 shows the significance of the estimated function. A common test of the null hypothesis that the group means differ is to consider whether, prior to the estimation of a function, the variables would be able to further discriminate between the two groups beyond the discrimination achieved by earlier functions (that is, we are examining the residual discrimination in the model) The statistic used is Wilks' Lambda, the significance of which is tested by a χ^2 .⁷

Table 3

SIGNIFICANCE OF THE ESTIMATED FUNCTION

Wilks' Lambda	χ^2	d.f	Significance
0.8547286	128.72	6	0 000

Using this statistic, it can be seen from Table 3 that the mean score for users is statistically different from the mean score for non-users

Predictive Performance

Table 4 shows the predictive performance of the final function.⁸ For both the holdout sample and the analysis sample the function out-performed the Cprop values as shown.⁹

Table 4

CLASSIFICATION MATRIX						
Actual Group	Hold-out Sample			Analysis Sample		
	Predicted Group			Predicted Group		
	Non-users	Users	Total	Non-users	Users	Total
Non-users	21	57	78	37	109	146
Users	8	314	322	19	660	679
Correctly Classified	83.75%			84 48%		
Cprop	68.61%			70 87%		

Another way of considering the predictive performance of the function is to examine the percentage of cases correctly classified within each group. In this case the function correctly classified 26.9 per cent of the non-users and 97.5 per cent of the users in the hold-out sample, and for the analysis sample the values were 25.3 per cent and 97.2 per cent respectively. Caution must be shown when examining the results for the analysis sample, as this will bias upwards the model's performance.

Ranking and Interpretation of the Variables

Table 5 shows the rankings of the variables in terms of the standardised coefficients, the bivariate correlations between each predictor variable and the discriminant function.

(structure coefficients) and the Partial-F statistics. Before we compare the rankings and interpret our findings, it is to be remembered that we are discussing the ability of values of $X_j^1 = 1n(u_i/v_i) + 1n(V_T/U_T)$ (see p.151) to distinguish between users and non-users and that for each ratio level variable the values of X_j^1 are often not monotonically related to the original X_j values

The first observation one can make is that on all three criteria the rankings of the final six variables are identical. This is to say, postcode is the variable which contributes most to determining the discriminant score (0.549) and has also most in common with the final function (0.584). The values for the other five variables provide the same information only in decreasing order of importance. The rankings on the basis of the Partial-F statistics indicate the significance of the discrimination which that variable contributes over that contributed by the other variables in the function.

Interestingly, several variables (for example, spouse's income, number of children and home 'phone indicator) were not included in the function because they did not contribute a significant amount of additional discriminating power beyond that contributed by the included variables. Since the degree of collinearity between the predictor variables was very low we can conclude that such variables have little discriminatory power in the context of users and non-users

Table 5

STANDARDISED COEFFICIENTS AND STRUCTURE MATRIX						
Variable	Standardised Coefficients		Pooled Within-Groups Correlations		Partial F (to remove)	
	Value	Rank	Value	Rank	Value	Rank
Postcode	0.549	1	0.584	1	37.22	1
Age	0.450	2	0.580	2	23.37	2
Applicant's income	0.388	3	0.488	3	17.77	3
Years at bank	0.358	4	0.425	4	15.18	4
Years at present address	0.139	5	0.315	5	2.19	5
Residential status	0.123	6	0.263	6	1.77	6
Children			0.147	7)	not in function	
Major credit cards			0.124	8)		
Store credit cards			0.120	9)		
Charge cards			0.115	10)		
Spouse's income			0.114	11)		
Home phone			0.108	12)		
Deposit account			0.068	13)		
Building Society cards			-0.048	14)		
Loan account			0.042	15)		
Current account			0.041	16)		
Other dependants			0.008	17)		

In order to interpret the variables we must examine the relationships between the X^1_j and the original X_j values, for each of the six variables. In terms of postcode, the areas of the country which give the greatest X^1_j values are so heterogeneous that few conclusions can be drawn. In the case of age of card holder (although there is not a monotonic relationship between X^1_j and age) we find that younger rather than older card holders are more likely to use their card, with the most likely users falling into the 30-40 age bracket. The least likely users are those aged 60 or over.

For applicant's income (at 1986/87 prices), the most likely users are to be found in the highest income band (that is, £14,700 and above) and the least likely in the less than £2,200 range. A monotonic relationship existed for this variable with the exception of those with an annual income of between £5,500 and £7,500. This group had the second highest X^1_j value and are hence the second most likely group to use their card.

Turning to the length of time for which an account was held at the bank, we find that those least likely to use their card fall into the less than six months bracket and the 19 years and over bracket. In contrast, the most likely card users are those who have held a bank account for four or five years. All the remaining groupings (that is, 1, 2, 3, 6-7, 8-10 and 11-18 years) had very similar X^1_j values and hence similar likelihoods of ever using their credit card

While there is no monotonic relationship between years at present address and the X^1_j values, longer term incumbents and those who have been in their present address for less than six months are by far the least likely to use their card. These two groupings are closely followed by those who have been at their present address for between four and nine years.

In terms of residential status, the most likely non-users were found to be either tenants in unfurnished accommodation or 'others' (that is, not falling into any of the other four categories) The latter group normally consists of people who live in the same accommodation as the owner but where the owner is not their parent. The X^1_j values for the remaining three categories, owners, with parents and tenants in furnished accommodation were very similar and significantly higher than the X^1_j values for the two 'least likely' groups

CONCLUSIONS

The results show that with the aid of Discriminant Analysis it is possible to discriminate significantly between those who hold a bank credit card and use it and those who hold such a card but do not use it. Apart from where the card holder lives those who are most likely to use their bank credit card are those aged 30-40 years, those with salaries of at least £14,700 (1986/87 prices), those with an account at the issuing bank for four to five years, those who have lived at the same address for two to three years and those who are owners of their home, who live in rented furnished accommodation or with their parents. Those least likely to use their card were those who were aged 60 or over, who had an income of less than £2,200 (1986/87 prices), who held an account with the bank for less than six months, who had lived at their present address for twenty or more years, and those who had lived in rented unfurnished accommodation

These results suggest where banks should target their promotional efforts if they wish to attract users and non-users, respectively, of their credit cards. These results also suggest which segments should receive different advertising messages. Thus, assuming that the main benefit of holding a card to non-users is that it provides a reserve source of immediate finance, promotional material which emphasises this aspect of a bank's card can be designed to appeal to the specific non-user groups above. Alternatively, assuming that the reason why users hold a card is the convenience with which credit can be extended, the above results show to whom banks should target their promotional messages which enhance these qualities of their card.

The results also point to possible pricing strategies. Thus, if the bank wishes to attract card users it may consider charging a lower fixed subscription rate and lower interest rates to those who are identified above as otherwise non-users. In addition new products may be introduced which are targeted at those on low incomes, and those who are aged over 60 years

But these policy suggestions typically require further information and so suggest further research. First, it would be useful to compare the attitudes of non-users towards *different types of credit* and to the *use of credit cards* to try to discover why such individuals are non-users. Similarly, it would be relevant to investigate what explains the amount of credit extended and debt outstanding which a user takes and maintains. Those who maintain a high level of debt outstanding whilst repaying the minimum amount each month are likely to be the most profitable customers to a credit granting agency, if also the most risky.

NOTES

1. Account numbers were allocated to individuals sequentially in order of their application. The values of the digits used to identify the sample were selected to be distributed throughout the ordering but otherwise randomly.
2. Hence, suppose a nominal variable takes on any of m possible values and let u_i and v_i be the number of users and non-users respectively in the sample which take on the i^{th} nominal value ($i \leq m$) such that

$$U_T = \sum_{i=1}^m u_i \text{ and } V_T = \sum_{i=1}^m v_i$$

that is, U_T and V_T are the total number of users and non-users respectively in the sample. Clearly, each of U_T , V_T , u_i and v_i are measured at ratio level. Therefore, we could replace the i^{th} value of a nominal variable by a combination of u_i , v_i , U_T and V_T and obtain a ratio level variable. The literature [Boyle, Crook, Hamilton, Thomas, 1989] describes several possible combinations which are related to the probability odds or log of the probability odds of the 'goods' and 'bads' taking on the i^{th} value of the nominal variable.

For reasons given in Boyle *et al* , the specific form of the predictor variables chosen was:

$$X^1_j = 1n \quad X_j = 1n \frac{u_i}{v_i} + 1n \frac{V_T}{U_T}$$

for case j, where

$$X_j = \frac{u_i}{U_T} \bigg/ \frac{v_i}{V_T}$$

Furthermore, for many variables, e.g. postcode, there were so many different values that the frequency distribution of cases left very few in certain categories – in some the number of non-users was zero. We therefore aggregated the values of the nominal variables according to similarity of $u_i/(u_i + v_i)$ and nominal categories for which there were no non-users were combined with those categories with the highest value of $u_i/(u_i + v_i)$.

3. However, in these cases the original values of each variable were aggregated with adjacent values because on *a priori* grounds it seems unlikely that the probability of non-users would vary considerably between very similar, say, spouses' income values, and such differences in estimated probabilities $u_i/(u_i + v_i)$ were ascribed to large sampling errors due to relatively small sample sizes associated with each ratio value.
4. The criterion for variable selection was the Mahalanobis Distance statistic (D^2). The Mahalanobis Distance is defined as

$$D^2_{u,v} = (n - g) \sum_{i=1}^m \sum_{j=1}^m w^*_{ij} (\bar{x}_{i,u} - \bar{x}_{i,v})(\bar{x}_{j,u} - \bar{x}_{j,v})$$

where m = number of predictor variables in the model
 g = number of groups

u, v = the groups of users and non-users respectively

$\bar{x}_{i,u}$ = sample mean value of predictor i for group u

w^*_{ij} = an element from the inverse of the within group's covariance matrix.

The F-to-enter and F-to-remove values were set equal to 1 0000.

5. Given the substantially different sample sizes for the two groups, it is possible that the covariance matrices for the two groups may not be equal, contrary to the assumptions of linear discriminant analysis. But it has been argued [Reichert, Cho and Wagner, 1983: 101-104] that the predictive ability of linear discriminant analysis in the credit-scoring context when covariance matrices differ between groups (and when rejected applications are excluded from the sample), is relatively robust. Moreover, if the covariance matrices differ between the two groups it has been shown that the appropriate method is quadratic discriminant analysis. But this is more difficult to use, because it is less robust to any interactions between the variables, and is less efficient as the number of predictors increases.
6. Given that postcodes were aggregated only by similarity of $u_i/(u_i + v_i)$, (without regard to geographical proximity), the variance of the population values of $u_i/(u_i + v_i)$ between postcodes within an aggregated group is likely to be relatively high compared to that between groups.
7. Wilks' Lambda is the ratio of the within group's sum of squares to the total sum of squares. Wilks' Lambda is inversely related to the degree of discrimination since a value close to zero (its minimum value) indicates that the group centroids are very different relative to the within group variation. When Lambda equals one (its maximum value) the group centroids are identical. The logarithm of the Lambda function has a chi-square distribution.
8. A case is classified into a group, j , if the conditional probability that the case is a member of group j , given a discriminant score, S , $P(G_j|S)$, is greater than the

conditional probability that it is a member of any other group $P(G_j|S)$ is estimated by:

$$P(G_j | S) = [P(S | G_j).P(G_j)] / [\sum_{j=1}^g P(S|G_j).P(G_j)].$$

The prior probability that a case belongs to group j , $P(G_j)$, was estimated as being equal to the proportion of users and non-users in the overall sample.

9. The proportion of cases correctly classified by the function also exceeded the $C_{\max}$ values of 80.5 per cent and 82.3 per cent for the hold-out and analysis samples respectively. The $C_{\max}$ value is the proportion which we would expect to be correctly classified if we allocated all cases into the group which has the larger number of cases in the sample.

ACKNOWLEDGEMENT

The support of the Economic and Social Research Council (ESRC) is gratefully acknowledged. The work was funded by ERSC award number R000231152.

REFERENCES

- Bierman, H. and Hausman, W. H. (1970), "The Credit Granting Decision", *Management Science*, Vol. 16(8), pp. B519-B532.
- Boyle, M., Crook, J.N., Hamilton, R. and Thomas, L. C. (1989), "Methods of Credit Scoring Applied to Slow Payers", Department of Business Studies, University of Edinburgh, Working Paper, 89/13.

Capon, N., (1982), "Credit Scoring Systems: A Critical Analysis", *Journal of Marketing*, Vol 46, pp 82-91

Chandler, G.C. and Coffman, J.Y., (1979), "A Comparative Analysis of Empirical Vs Judgemental Credit Evaluation", *Journal of Retail Banking*, Vol. 1(2), pp. 15-25.

Crook, J. N., Hamilton, R and Thomas L. C., (1989), "A Comparison of Discriminators Under Alternative Definitions of Credit Default", Department of Business Studies, University of Edinburgh, Working Paper, 89/12

Eisenbeis, R.A., Gilbert, G.G. and Avery, R.B., (1973), "Investigating the Relative Importance of Individual Variables and Variable Subsets in Discriminant Analysis", *Communications in Statistics*, Vol. 2(3), pp. 205-219.

Eisenbeis, R.A., (1978), "Problems in Applying Discriminant Analysis in Credit Scoring Models", *Journal of Banking and Finance*, Vol 2, pp. 205-219.

Frank, R.E., Massy, W.F. and Morrison, D.G., (1965), "Bias in Multiple Discriminant Analysis", *Journal of Marketing Research*, Vol. 2, pp. 250-258.

Hair, J.F., Anderson, R.E. and Tatham, R.L., (1987), *Multivariate Data Analysis*, New York Collier Macmillan

Klecka, W.R. (1980), *Discriminant Analysis*, Sage University Paper No 19, Series: Quantitative Applications in the Social Sciences.

The Monopolies and Mergers Commission, (1989) *Credit Card Services: A Report on the Supply of Credit Card Services in the United Kingdom*, London: HMSO.

Morrison, D.G., (1969), "On the Interpretation of Discriminant Analysis", *Journal of Marketing Research*, Vol. 4, pp. 156-163

Norusis, M J./SPSS Inc. (1988) *SPSS-X Advanced Statistics Guide*, Chicago: SPSS Inc

Reichert, A K., Cho, C C. and Wagner, G M (1983), "An Examination of the Conceptual Issues Involved in Developing Credit-Scoring Models", *Journal of Business and Economic Statistics*, Vol. 1(2), pp. 101-114.

Wiginton, J.C., (1980), "A note on the Comparison of Logit and Discriminant Models of Consumer Credit Behaviour", *Journal of Financial and Quantitative Analysis*, Vol. 15(3), pp. 757-770.

Appendix 1

RESULT FOR FUNCTION EXCLUDING POSTCODE

Table 1(a)

SIGNIFICANCE OF THE ESTIMATED FUNCTION

Wilks' Lambda	X ²	d f	Significance
8915316	94 15	6	0 000

Table 1(b)

CLASSIFICATION MATRIX

Actual Group	<i>Hold-out Sample</i>			<i>Analysis Sample</i>		
	Predicted Group			Predicted Group		
	Non-users	Users	Total	Non-users	Users	Total
Non-users	17	61	78	26	120	146
Users	8	314	322	28	651	679
Correctly classified	82.75%			82.06%		
C _{prop}	68.61%			70.87%		

Table 1(c)

STANDARDISED COEFFICIENTS AND STRUCTURE MATRIX

<i>Variable</i>	Standardised Coefficients		Pooled Within-Groups Correlations		Partial F (to remove)	
	Value	Rank	Value	Rank	Value	Rank
Age	0.538	1	0.685	1	24.88	1
Applicant's income	0.473	2	0.577	2	19.67	2
Years at bank	0.412	3	0.502	3	15.10	3
Years at present address	0.174	4	0.372	4	2.57	4
Residential status	0.169	5	0.311	5	2.48	5
Spouse's income	0.148	6	0.226	6	1.92	6
Children			0.156	7)		
Store credit cards			0.127	8)		
Home 'phone			0.111	9)		
Major credit cards			0.108	10)		
Loan account			0.090	11)	not in function	
Charge cards			0.069	12)		
Deposit account			0.057	13)		
Current account			0.049	14)		
Building Society cards			-0.045	15)		
Other dependants			0.018	16)		

CHAPTER 7

CREDIT CARDS: HAVES, HAVE-NOTS AND CANNOT-HAVES

J N. Crook and L. C. Thomas

(Department of Business Studies, Edinburgh University)

R Hamilton

(Business School, Loughborough University)

The Service Industries Journal, (1994), Vol. 14, No. 2 (April), pp. 204-215

Credit Cards: Haves, Have-Nots and Cannot-Haves

J N Crook, L. C. Thomas and R. Hamilton

Credit card ownership has grown enormously over the past twenty years. This article analyses two major data sets - the government's Family Expenditure Survey and a credit card grantor's database of clients - to investigate who has credit cards and, for those who do not have them, whether they would be given cards if they applied for them. The results show which sections of the population are averse to owning credit cards, and some surprisingly low levels of ownership among, for example, those who have bank accounts

INTRODUCTION

Over the past two decades credit cards have become of major importance in the financing of consumer purchases and as a method of money transmission. Credit cards were introduced into the UK in 1966. By 1978 there were 8 million cards issued and, as the Monopolies and Mergers Commission Report [1989] reported, this had grown to 25 million by 1988. This growth should be put in context. As far as consumer credit is concerned, credit cards only accounted for 16 per cent of the credit outstanding in 1988 (£6.7 billion out of £43 billion National Consumer Council [1990]). However, with the proportion of card holders paying off their balance each month increasing to above 50 per cent, the use of credit cards as a payment mechanism is substantial and remains so despite the introduction of annual charges by some card issuers in 1989.

This article addresses two questions: what sort of people have credit cards; and for those who do not have credit cards, is it because they cannot get them if they want them or that they do not want them? The methodology to answer these questions is based on two data sets - the Family Expenditure Survey results of 1986, and the application data and

subsequent performance of a sample of clients of a credit card issuer. The Family Expenditure Survey is a government-backed carefully sampled survey of the income and expenditure pattern of households in the UK. The 1986 survey, published in late 1988, was the first one to include data on the ownership of credit cards and thus enables one to distinguish between those who have or do not have credit cards.

Credit card companies use their experience with previous clients to score the various entries on the application form as well as considering a report from a credit reference agency on the applicant's credit worthiness. A new applicant will receive a credit card provided the cumulative score of his entries is higher than some specified cut-off. The data from the credit card company was used to construct a scoring system representation of those used in the industry, based on the methodologies outlined in Boyle, Crook, Hamilton and Thomas [1988]. All adults in the Family Expenditure Survey were scored using this scoring system and those with scores below the cut-off were considered to be at risk of being refused credit cards if they were to apply in reality. This splits the FES sample into four classes

W - those who have cards and would get cards under the scorecard constructed;

X - those who do not have cards but would get them under the constructed scorecard,

Y - those who have cards but would not get them under the constructed scorecard; and

Z - those who do not have cards nor would get them under the constructed scorecard

The ratio $X/(X+Z)$ suggests what fraction of those without credit cards could get them if they so wished. The numbers in Y ideally should be small as they indicate how much harsher the constructed scoring system is than some used in practice. However, low numbers in Y do not tell us whether the constructed scoring system is more generous than those used in practice.

Several papers have described the characteristics of holders of different types of credit and retailer cards but almost all in the US context. Mathews and Slocum [1969] compare social class and credit card usage on the East Coast of the US, Johnson [1975] describes the demographics of credit card usage nationally in the US; Martell and Fitts [1981] and Kinsey [1981] use quadratic discriminant analysis and tobit analysis respectively to analyse the characteristics of good users of credit cards. Lindley [1989] has considered how ownership and use of credit cards changes over time. There does not appear to be any previous investigation into whether those who do not have credit cards would be able to get them if they applied for such cards.

Section two outlines the methodology and variables used in constructing the scoring system. Section three analyses who owns credit cards, while section four looks at who could get credit cards under various rejection levels imposed by the credit card organisation. The final section highlights some of the results obtained.

METHODOLOGY

The Family Expenditure Survey obtained information on over 1,000 aspects of the members of 7,178 households in the UK which included 13,549 people aged 18 or over who are legally able to hold credit cards. This included the question - did they own a credit or charge card (e.g. American Express, Diner's, Gold cards). Since the latter are used in a similar way to credit cards, except for the credit facility and firms issuing them use similar scoring techniques to credit card issuers, we have treated them all as credit cards for the purpose of this article.

The data from the credit card issuer contained the application data - 24 sociodemographic and economic variables - and the subsequent performance history over several years of more than 1,000 clients. When examined it was possible to match exactly nine of these variables with corresponding data in the Family Expenditure Survey. These variables were residential status, length of residence at present address, outgoings on a monthly basis (i.e.

mortgage or rent plus other loans), phone ownership, age, occupational status, current account ownership, income and spouse's income.

A credit scoring system was built on these nine variables which gave a satisfactory discrimination between the good and bad client performance in the card issuers data set and which could then be used to score the members of the FES data set. A bad client performance was taken to be one where the client had defaulted on payment for three consecutive months during the performance period (see Crook, Hamilton and Thomas [1992] for discussion of the relationship between this and less severe definitions of bad performance) There are several techniques possible for developing a scoring system from such data: statistically based ones using discriminant analysis, log linear models; or recursive partitioning, mathematical programming ones; and also suggestions of methods based on artificial intelligence and neural networks Comparison of the different methods were made by Myers and Forgy [1963], Srinivasan and Kim [1987], Wiginton [1980], and Boyle, Crook, Hamilton and Thomas [1992].

Mathematical programming and statistical methods, particularly the ones based on discriminant analysis or log linear models are the norm in the industry. As outlined in Boyle *et al* [1992], it is necessary to translate both the quantitative independent variables such as age, and the qualitative ones such as residential status (e.g. owner occupier, unfurnished tenant, etc.) into categorical variables. The categories are chosen so that they both have some reasonable interpretation and that the ratios of bads to goods at each value of the variable in a category are fairly stable. The choice is then either to consider each category of a variable as a separate dummy variable in the analysis or to modify the variable, so that all the values in the same category are given the same modified value which is related to the odds or log odds of goods to bads in that block. Consider the example of age. If the categories were 18-24, 25-30, 31-40, 41-65, 65+, then in the former case age would have four dummy variables D1, D2, D3, D4 where $D1 = 1$ if client is aged 18-24, 0 otherwise, and $D4 = 1$ if client is 41-65, and 0 otherwise There is no need to put in a fifth variable D5 to represent the over 65s as then $D5 = (1 - D1 - D2 - D3 - D4)$ is a linear combination of the other variables. In the alternative approach, age is represented by one variable, but all those with

ages 18-24 would have the same value which is related to g/b , $g/g+b$ or $\log g/b$ where g is the number of good clients in the 18-24-year-old group and b is the number of bads. We chose this latter approach for the generic scoring system.

A discriminant function was built on the nine variables common to the two data sets modified as outlined above. The variables with the strongest impact on the discriminant function (highest standardised coefficients) were, respectively, current account ownership, spouse's income; residential status; occupation; phone ownership; and age.

Although not as good a predictor on a hold-out sample of the credit card data as a discriminant function built on all the 24 variables available in that data set, this discrimination function keeps more than two-thirds of the improved prediction over chance allocation, when both use the cut-off that minimises misclassification errors. 'Years at bank' is the only variable which has considerable significance in the discriminant function based on the 24 variables, which is not included in the nine common variables.

Having constructed a scoring function, the accept/reject decision depends on the cut-off score chosen; those with scores higher than this value would be accepted, those below, rejected. If L is the lost profit incurred by rejecting a client who is really good, and D is the debt that will need to be written off which is incurred by accepting a client who will default, choosing a cut-off score c gives an expected loss per client.

$$L \text{ Prob (good client has score } < c) + D \text{ Prob (bad client has score } > c) \quad (2.1)$$

Thus at the optimal cut-off score, this leads to the odds of goods to bads satisfying

$$\text{Prob (good client)}/\text{Prob (bad client)} = D/L$$

These odds ratios can either be calculated empirically by testing the scoring system on a representative sample of clients or theoretically using the form of the probability distribution of scores specified by discriminant analysis or log linear models. Different card issuers will choose different cut-off levels, and the same card issuer will change his cut-off over time.

depending on the business objectives sought and the current economic situation. In section four we analyse the FES survey using cut-off levels varying from odds ratios of 1:1 which minimise misclassification errors and give a 3 per cent rejection rate to 5:1 (i.e. D/L = 5), which is nearer the cut-off levels used by some card issuers and give a 13 per cent rejection rate.

Clearly, the calculation of a generic scorecard outlined above can be criticised on several grounds. There are substantial differences in the application characteristics of the subpopulations who apply for different cards, and this leads to significant differences in the scorecard used to score subsequent applicants. These differences cannot be reflected in a scorecard built on one such sub-population. The restriction to nine common variables may diminish the power of the card somewhat. Furthermore, most actual scoring systems use credit reference agency data as part of the scoring procedure either for all or a substantial number of the applications. However, credit reference data is strongly correlated to the score obtained without it and our contention is that ignoring credit references will not have a major effect on the broad outlines of the results. Lastly, it was not possible to use information on those clients who were rejected by the card issuer to modify the scoring system. Several commercial systems apply reject inference, which uses such information, by inferring a probability of 'badness' to each such rejected client to modify the initial scoring system. Despite these differences we would contend that the scoring system developed is able to give general indications of which types of people are most likely to be able or not able to acquire credit cards.

OWNERSHIP OF CREDIT CARDS

The Family Expenditure Survey (FES) of 1986 included returns from 13,549 adults of age 18 or over, who are legally entitled to hold credit cards. Of these, 31.8 per cent (4,306) reported that they had credit or charge cards. A smaller survey by the Monopolies and Mergers Commission [1988] gave a 38 per cent ownership rate. Since then there has been a 20 per cent rise in the number of credit cards in the UK (21 million to 25 million) between

1986 and 1988, and this result is in line with the FES findings. The rest of this section investigates which parts of the population comprise these credit card holders.

Males comprised 47.7 per cent of the sample population and had a card ownership rate of 37.1 per cent, while the ownership rate among females was 27.0 per cent. An even greater difference in ownership occurs between married people, who have an ownership rate of 36.8 per cent, and single people (including divorced and widowed), where the ownership rate was only 21.5 per cent.

Card ownership increases monotonically with income as might be expected. 18.3 per cent of those with incomes less than £2,500 have cards, 24.7 per cent of those with incomes between £2,500 and £7,500 have cards, 50.6 per cent of those with incomes between £7,500 and £15,000 have cards, while 76.6 per cent of those with incomes above £15,000 have credit cards.

For married couples the income of both spouses has an effect on the ownership of credit cards. The ownership rate increases with the card-holder's income irrespective of what the spouse earns except in the case where the spouse earns more than £15,000. In this case, there is a higher rate of card ownership among those who have no income than those whose income is between £2,500 and £5,000 pa. Examining these cases shows a high proportion of women cardholders, so suggests that wives who do not work or work only very little are more likely to hold cards than those with wages nearer the average for females. The trend is for increasing card ownership as the spouse's income increases, except when the person earns over £15,000 where the ownership levels drop until the spouse starts earning over £10,000 pa. In all cases ownership levels are higher among the higher earner of the partners, the difference in levels ranging from 7 per cent to 25 per cent. One can almost perfectly categorise the groups with card ownership level above 50 per cent as those who are earning at least £10,000, or where spouses earn at least £15,000. Similarly the 70 per cent card ownership level is those who earn at least £15,000 or who earn at least £10,000 and whose spouses earn at least £15,000. At the other extreme, if

neither partner earns more than £15,000 per annum, card ownership levels are below 20 per cent even though this group is one-third of the sample population.

Owning a phone and having a current account are positively related to credit card ownership. 36 per cent of phone owners have cards but less than 8 per cent of people without phones have cards. 45 per cent of those with current bank accounts have credit cards, while only 7.7 per cent of those without such accounts have cards. It is perhaps surprising that the level is as low as 45 per cent given that banks have been offering their own credit cards as alternative to cheque guarantee cards. Since this is the 1986 FES survey, it is possible that the impact of this was only beginning to be felt then. Alternatively, those surveyed may not have been aware that their cheque guarantee card was also a credit card. Putting current account and phone ownership together magnifies the difference in credit card penetration. Of those who have neither phone nor current accounts (10.6 per cent of the population) only 1.4 per cent have credit cards.

Credit card ownership increases with age from 18 to 40, and then decreases with age thereafter, peaking at 45.2 per cent in the age group 35-40 and dropping to 12.8 per cent in the over 70s. Comparing age and income together the highest level of ownership is the 30-35-year-old earning over £15,000 at 82.8 per cent, while those over 70 with an income of less than £2,500 have an ownership rate of 5.6 per cent. It is interesting to note that in the age ranges 24-40 the ownership level of those earning less than £2,500 is always higher than those earning between £2,500 and £7,500. One explanation might be that ownership among mothers with young children who can afford to earn less than £2,500 pa is higher than those who need to earn more.

Occupation also has a major effect on ownership of credit cards, but in some respects less than might be expected. The professional occupations have an ownership level of 61.8 per cent, not very dissimilar to administrators and managers at 60.2 per cent. Clerical workers have a 43.9 per cent ownership rate, skilled manual workers 32.7 per cent, semi-skilled 24.5 per cent, while unskilled manual workers have an ownership level of 12.2 per cent.

Those classified as unemployed have a similar rate to the retired - 18.3 per cent as against 20.5 per cent.

Thus, although credit card ownership is growing, there are some variations. Occupation, income and age play significant roles, but it is surprising how little is the penetration among those with bank accounts.

GRANTING OF CREDIT CARDS

Over 60 per cent of the population did not have a credit card in 1986. Was it because they would not have been awarded them if they applied for them, or did they not want them? Using the methodology of section two we constructed a credit scoring system based on the nine variables common between the FES survey and the credit company application form data. This gives each applicant a score and the company determines the acceptable cut-off level at which it will accept customers. Clearly we are unable to check the credit reference agency data to see which customers have unacceptable records. Private discussion with experts in the credit scoring industry suggest that although this will affect the proportion with particular characteristics who could get cards somewhat, the changes will be fairly minor.

Different companies will choose different cut-off levels of risk depending on their objectives, and even the same firm will change its cut-off levels depending on the time of year and the economic climate. To overcome this, we calculated who could obtain credit cards at various cut-off levels and report the results for two cut-off points - the results for intermediate points are close to a linear interpolation of the two results.

The low level, Level 1, represents the most lax situation of credit card organisations though it was the level which minimised overall misclassification of error in the credit card organisation data, i.e. minimised cost if $L = D = 1$ in (2.1). At this level, the type 1 error in our scoring system - those who have cards, whom we would refuse cards - is below 0.5 per cent. This suggests that this is around the lowest cut-off level in the past that credit card

organisations have employed. The higher cut-off point, level 2, is one where around 13 per cent of the population are rejected and represents a more realistic rejection rate for credit card organisations in recent years. It corresponds to the lowest misclassification of errors on credit card data if $D/L = 5$ in section 2.

The results show that the overall rejection rate at level 1 is 402 out of 13,549, i.e. 3.0 per cent and 1,804 rejections or 13.3 per cent at level 2. Of the extra 1,402 rejected between the two cut-off levels, 87 per cent were in the group who did not have credit cards. There is a significant difference in reject rates at all levels between those who already have cards and those who do not, but it is not startlingly so. 4.1 per cent of those without cards would not get them under the lax cut-off level, while 0.5 per cent of those with cards would not. At the harsher level, 17.3 per cent of those without cards would not get them, while 4.8 per cent of those with cards would not get them at this higher level. Thus it would appear that the vast majority do not have cards because they do not want them. Depending on the policies adopted by credit card organisations, between 4 per cent and 20 per cent of those without cards would not be able to obtain them.

One can also look at the types of people who fall into the various groups, using the characteristic variables described earlier. The most important discriminators in this sample on who could or could not get credit cards are phone ownership, current account ownership and income of spouse. At the higher rejection level, 91 per cent of those with phones will get credit cards but only 51 per cent of those without phones would get cards. At the lower level cut-off level, the reject rate is less than 0.9 per cent for those with phones and 13 per cent for those without.

Having a current account has a similar if slightly less decisive effect. 64 per cent of the population have current accounts. At the higher reject cut-off, 97 per cent of the people with current accounts would get credit cards, while only 67 per cent of those without would. At the lower reject level, only 2 per cent of those with accounts would be rejected while 8 per cent of those without would be. The results on income are also what would be expected, with acceptance rates at both low and high reject rates increasing with income,

though in both cases there is little difference in acceptance rates until incomes are above £15,000 pa.

Spouse's income is rather more interesting. The acceptance rate at all reject levels is a U-shaped function of spouses' incomes dropping sharply in the £5,000 to £7,500 band. In this band 16.8 per cent are rejected at the low reject level and 44 per cent at the high reject level. More careful examination shows that two-thirds of the group who would not get cards in this category are female (i.e. their husbands earn between £5,000 and £7,500) and 84 per cent of these women earn less than £5,000 themselves. The group with even lower spouse's income has a much higher proportion of males whose wives earn nothing or less than £2,500, but who have a high income themselves. For example, 83 per cent of the group whose spouse's income is between £0 and £2,500 are men.

The occupation of a person also has an effect on the ability of someone to get a credit card, but the variation is what one would expect and is perhaps less than expected. The one surprise may be that those who are retired were calculated to be risks as good as those in the professional classes, and hence were having equally high rates of being accepted for credit cards.

This ability of the retired to obtain credit cards is also reflected in the breakdown of age. Those aged over 61, although only having an ownership rate of 20 per cent, would find it very easy to obtain credit cards. At the high cut-off level, only 4.7 per cent would be rejected (5.6 per cent among those who do not already have cards), while at the low cut-off level less than 0.5 per cent would be rejected. The groups with the next highest rates for being accepted for cards are the 41-60 age group followed by the 18-24s. Those aged thirty-something have the highest ownership rates at 44 per cent, but at the high cut-off level the reject rate for those not having cards is 28 per cent. The least likely to get cards, however, are the 25-30-year-olds, who although having a card ownership rate of 37 per cent have a reject rate among non-owners of 34 per cent at the high rejection cut-off level and 10 per cent at the low rejection cut-off level. This suggests that with the higher rejection cut-off levels if you do not have a card by the time you are 25, it will be harder to

get until you are over 40. The figures also suggest that credit cards have most room for expansion among the young or retired sections of the population, who are also the best risks. This reflects the difference in the way credit and debit is viewed by those born before and after the Second World War.

The length of time at the present residence is much more predictable. Rejection rates stay fairly constant for all periods up to 10 years living at the present address around 20 per cent for the high cut-off level and 4 per cent for the low cut-off level and then drop slightly after ten years to 5 per cent and 1 per cent respectively in over 18 years at a present address category. The gentle n-shape of the credit card ownership rate reflects the correlation between this variable and age of the person.

CONCLUSIONS

It is obvious that the percentage of the population who get credit cards depends on what rejection rates the credit card organisations set. This varies between organisations and over time as the economic conditions and organisational strategy changes. However, the results of the last section imply that the vast majority of these without cards would be able to obtain cards if they applied. At the high rejection level, the reject rate of those who already have cards is around 5 per cent, while for those without it is 17 per cent. Thus we must conclude that most of those without credit cards either do not want them or are not yet financially sophisticated to require them. The older people in the community, especially those over 65, could come into the former category because almost all would be able to get cards, it appears, but the ownership rate is low. This would imply a natural increase in credit card ownership with the passing of time as younger generations with higher credit card ownership reach the age where even more of them are acceptable to credit card organisations. Phone ownership seems to be a very good indicator of whether one can get a credit card or not, but it is surprising that while almost all those with current accounts can obtain cards, only 45 per cent actually have cards. It must be remembered that this survey was made in 1986 and banks have made considerable efforts over the past four years to

increase credit card ownership among their customers. Such efforts have included the unsolicited direct mailing of credit cards and the badging of Connect cards as Visa cards

One group who appear to find it difficult to get credit cards are people whose spouses earn between £5,000 and £7,500 a year. On closer investigation this seemed to involve mostly women whose husband's wage was at this fairly low level. Again this raises the question that if different scores for men and women were allowed on the application scorecard, then the system might actually benefit women more. (The scorecard built in section two gives greater weight to spouse's income than applicant's income, for example.)

The results for occupations and residential status suggest that though we are right to consider owner-occupier professional people as typical credit card owners, the current reject rate is not that much lower among other categories of employment or those in rented accommodation

Thus, unless you are in your late twenties, unemployed with no phone or current account and married to someone earning less than £7,500 a year, it is likely that not having a credit card is a matter of choice rather than being refused. As for the credit grantors, what should they do to increase card ownership levels? One obvious point is to target the 55 per cent of their current account owners who still do not or do not realise that they have credit cards. The results also showed that those who live with their parents are good credit risks and could be wooed more vigorously while they seem to be in the financially more secure environment of their parental home. Lastly, and probably most difficult and with least long-term advantage, those of retirement age are far and away the largest group who do not have cards because they do not want them rather than because they do not have them.

ACKNOWLEDGEMENTS

The support of the Economic and Social Research Council (ESRC) is gratefully acknowledged. The work was funded by ESRC award number R000231152. The authors are also grateful to the credit card organisation who donated the data

REFERENCES

- Boyle, M , J.N. Crook, R Hamilton and L. C. Thomas (1992), "Methods for Credit Scoring Applied to Slow Payers", in L.C. Thomas, J N. Crook and D. Edelman (eds.), *Credit Scoring and Credit Control*, Oxford: Oxford University Press, December, pp.75-90
- Crook, J N , R Hamilton and L. C. Thomas (1992), "A Comparison of Discriminators under Alternative Definitions of Credit Default", pp 217-45, in L.C. Thomas, J.N. Crook and D. Edelman (eds.), *Credit Scoring and Credit Control*, Oxford: Oxford University Press.
- Johnson, R. (1975), "National Spanning Credit Cards", *Federal Reserve Bank of San Francisco Business Review*, Vol. 11, p 3.
- Kinsey, J (1981), "Determinants of Credit Card Accounts - An Application of Tobit Analysis", *Journal of Consumer Research*, Vol. 8, pp.172-82.
- Lindley, J.T. (1989), "Credit Card Possession and Use Changes over Time", *Journal of Economics and Business*, Vol 41, pp.127-42.
- Martell, T.F. and R.L. Fitts (1981), "A Quadratic Discriminant Analysis of Bank Credit Card User Characteristics", *Journal of Economics and Business*, Vol 33, pp.153-9.

- Mathews, A.L. and J.W. Slocum (1969), "Social Class and Commercial Bank Credit Cards Usage", *Journal of Marketing*, Vol. 33, pp.71-8.
- Monopolies and Mergers Commission, (1989), *Credit Card Services: A Report on the Supply of Credit Card Services in the United Kingdom*, London HMSO.
- Myers, J.H. and E W Forgy (1963), "The Development of Numerical Credit Evaluation Systems", *Journal of the American Statistical Association*, Vol. 58, pp.799-806.
- National Consumer Council, (1990), *Credit and Debt, the Consumer Interest*, London HMSO.
- Srinivasan, V and Y H. Kim (1987), "Credit Granting - A Comparative Analysis of Classification Procedures", *Journal of Finance*, Vol 92, pp.665-81.
- Wiginton, J.C. (1980), "A Note on the Comparison of Logit and Discriminant Models of Consumer Credit Behaviour", *Journal of Finance and Quantitative Analysis*, Vol. 15, pp.757-70.

CHAPTER 8

CUSTOMER RETENTION: A BEHAVIOURAL MODEL

R Hamilton, J. B Howcroft and J. Saunders

(Business School, Loughborough University)

IMA Journal of Mathematics Applied in Business & Industry (1995) 6(4), pp 333-342

Customer retention: a behavioural model

R HAMILTON, J.B. HOWCROFT and J SAUNDERS

One of the main problems currently facing credit-card issuers is the increasing number of cardholders who are using their cards less often (i.e. attrition) and/or returning their cards (closures). This problem is of particular concern as the total number of credit cards held by consumers is declining (by approximately 0.6 per cent per month in 1992) and the number of new applicants is also running at an all-time low (less than 1 per cent per month in 1992). Most of the published literature in the broad area of credit cards looks at credit scoring, rather than the need for card issuers to identify and retain a profitable portfolio of card customers. The overall objective of our research is 'segmentation for customer retention', and this paper aims to identify the characteristics of card customers who initiate the closure of their accounts. Linear discriminant analysis is applied to a sample of approximately 17,000 UK holders of bank credit cards, using various behavioural and sociodemographic variables, and tested on a holdout sample of 10,000 cases.

Introduction

In the 1980s the real value of consumer debt, excluding finance for house purchases, increased by 122 per cent in the UK (Crook *et al.* 1992a). At these rates of market growth, it was not surprising that the emphasis was placed on the development of credit-scoring models which assisted - and in some instances entirely determined - the allocations of credit facilities to prospective borrowers.

Research and academic literature on the use of credit cards not surprisingly reflected what was seen as the overriding need of the market at the time. Predictive models were consequently developed which concentrated on the use of statistical techniques that could

either (a) distinguish between defaulters or non-defaulters (Myers & Forgy 1963; Wiginton 1980; Boyle *et al* 1992) or (b) determine the likelihood of customers who miss a given number of consecutive payments (Chandler and Coffman 1983-84; Bierman and Hausman 1970, Crook *et al.* 1992a).

In the aftermath of the economic recession of the early 1990s, the credit-card industry is no longer growing at the rates typical of the previous decade in 1992. The total number of credit cards held by consumers was declining at a rate of approximately 0.6 per cent per month and the number of new applicants was also running at an all-time low of less than 1 per cent per month¹. The changing dynamics of the industry are also illustrated by the fact that, at its peak in 1990, Visa and Mastercard had 29.846 million cards in circulation, and value of turnover equalled £27,742 million; however, by 1992, even though value of turnover had increased to £31,272 million, the number of cards in circulation had declined to 26.458 million (Annual Abstract of Banking Statistics 1993). Recent changes in the marketplace therefore reflect an increasing number of cardholders returning their cards (closures) while the remainder apparently use their cards more often and/or for making larger purchases.

The changing behaviour of credit-card users suggests that a different approach is required by management which is less concerned with credit scoring and risk and more concerned with the identification and retention of a profitable portfolio of card customers (Lundy 1992). With these considerations in mind, the overall objectives of the research project were determined and can be summarized as being 'segmentation for customer retention'. This paper reports the initial stages of this research and is primarily concerned with identifying the characteristics of customers who close their accounts and developing a model which will predict this behaviour. By utilizing the existing customer base, the application of such a model could increase profitability by maximising customer retention. As such, the analysis represents the first tentative steps in identifying appropriate strategies, based upon customer behaviour, for reducing closures and encouraging greater usage from current and potential card-holders

Methodology

The data related to a 15-month period from 1 January 1992 to 31 March 1993, and consisted of 27,099 individuals who held a credit card as at 1 January 1992². The size of the database meant that it was possible to create randomly a holdout sample which was representative of the original sample, consisting of 10,000 individuals (approximately 37 per cent of the initial data), and therefore large enough to ensure stability of the coefficients³ (Klecka 1980).

As the primary objective of the research was to develop a behavioural model with the predictive ability to identify those customers most likely to close their credit-card accounts, it was important to establish an exact definition of the term 'closed'. However, a number of alternative meanings could be attached to the term, and so it was decided to adopt a definition which reflected the behaviour of card customers rather than the card issuers. As a consequence, *closed* within the context of this paper only refers to those instances where *cards are returned to the bank (for whatever reason) by customers of their own volition*. All other categories of 'external status'⁴ are referred to as *normal* - and this includes instances where, for example, the card has become inoperable because the customer has become bankrupt, lost the card, or had it stolen, or where the card was revoked by the bank.

The data originally contained over 70 variables, but eventually 22 predictor variables were identified (see Appendix 1) which tended to reflect the behaviour pattern of card customers, although some sociodemographic variables have also been used where on *a priori* grounds it was thought they had a discriminative effect on closures. Since a number of variables were measured at nominal level, whereas the use of linear discriminant analysis requires that all predictor variables are measured at least at interval level (Klecka 1980), the method used follows that of Crook *et al* (1992b). That is, the required interval-level data were derived using the formula

$$X'_j = \ln(n_i / c_i) + \ln(C_T / N_T),$$

where

X_j = value of the predictor variable X for case j ,

n_i = number of normal card accounts in nominal category i (the category of which j was a member),

c_i = number of closed card accounts in nominal category i (the category of which j was a member),

N_T = total number of normal card accounts in the sample,

C_T = total number of closed card accounts in the sample

By using the logarithmic values in the way described above, a linear relationship between the function and group variables was established, thereby facilitating the application of linear discriminant analysis in developing a predictive model of closures.

An important step in constructing the predictive model was to identify *a priori* those variables which are potentially the best at discriminating between accounts that will close and accounts that will continue to operate normally. In selecting these variables, it was essential to establish whether multicollinearity exists between the various predictor variables and to determine which of these variables should be omitted from the function. Unless this precaution is taken, there could be a high degree of correlation between the variables in the function which would reduce the reliability of the standardized coefficients as indicators of the relative importance of each predictor variable (Chandler & Coffman 1983-84)

To test for the existence of multicollinearity, each predictor variable was linearly regressed on all other predictors, and the tolerance $1 - R_i^2$ was calculated for each variable. Variables with a tolerance of ≤ 0.79 (Crook *et al.* 1992b) were considered for deletion. Next, with the existence of multicollinearity identified, the values of both the regression coefficients and the Pearson correlation matrix were examined to determine which variables to remove (i.e. which pair(s) of variables were highly correlated). In the case of the Pearson matrix, a value of ≥ 0.2 was taken as an indication of multicollinearity.

After this procedure, the number of predictor variables left in the analysis with a tolerance value ≥ 0.8 was reduced from 22 to 15. The seven rejected variables were account prefix (i.e. whether the customer has a Mastercard or Visa, etc), how long the card had been active; date when account was opened; credit limit; number of cash advances; number of purchases, and amount of purchases

While the remaining 15 variables may intuitively be good discriminators, a stepwise procedure had been adopted to ensure that all weak redundant variables were removed from the final discriminant function. The criterion for variable selection was the Mahalanobis Distance (D^2) where at each step the variable that maximizes the Mahalanobis distance⁵ is selected (SPSSX User's Guide), subject to the F-to-enter value being at least equal to 1 (note: the F-to-remove value was also set equal to 1).

In addition to using the classification matrix and the percentage correctly classified by the function to assess the predictive accuracy of the discriminant function, the results were also compared with the percentage correctly classified by chance. This may be calculated (Hair *et al* 1987) using either the maximum-chance criterion⁶ (this is used when the objective is to maximize the percentage correctly classified, regardless of group membership) or the proportional-chance criterion (C_{prop})

$$C_{prop} = p^2 + (1 - p)^2,$$

where p is the proportion of cases in group 1 and $(1 - p)$ is the proportion of cases in group 2. Since the latter criterion is most suited, and should be used, when the objective is to classify correctly membership of two or more unequal groups (e.g. 'closed' or 'normal'), we shall be comparing the percentage correctly classified by the function with C_{prop} .

Results

The statistical significance of the estimated function is shown in Table 1. Wilks' λ indicates the ability of predictor variables to discriminate among the groups beyond the discrimination achieved by the earlier function, i.e. residual discrimination (Klecka 1980) As λ decreases in value, it is indicating progressively greater discrimination. The significance of the function is tested by the χ^2 ; as Table 1 shows, the means for both 'closed' and 'normal' accounts are statistically different

TABLE 1
Residual discrimination and test of significance

Wilks' λ	χ^2_v	Degrees of freedom (v)	Significance
0.805 586 0	3694.5	15	0 0000

The results of the model incorporating the remaining predictor variables are shown in Table 2. This indicates that the proportion of grouped cases correctly classified by the model was 86.62 per cent for the analysis sample⁷ and 86.86 per cent for the holdout sample. Viewed in a slightly different way, the model was correctly predicting 90.9 per cent of the normal accounts and 34.5 per cent of the closed accounts for the analysis sample and 95.3 per cent of the normal accounts and 33.8 per cent of the closed accounts for the holdout sample.

TABLE 2

Classification of results (with corresponding percentages in parentheses)

Actual group	Analysis sample			Holdout sample		
	No of cases	Predicted group		No of cases	Predicted group	
		Normal	Closed		Normal	Closed
Normal	14,728	13,389 (90.9)	1,339 (9.1)	8,632	8,224 (95.3)	408 (4.7)
Closed	2,371	1,553 (65.5)	818 (34.5)	1,368	906 (66.2)	462 (33.8)
Percentage correctly classified		86.62			86.86	
C _{prop} (per cent)		76.0			76.0	

In assessing the behavioural model's efficacy, comparisons with C_{prop} indicate that the results are much better than those which would have been correctly classified by chance the model correctly classifies almost 87 per cent of the accounts, which is substantially greater than the 76 per cent expected by chance. In other words, the model is correctly classifying almost 11 percentage points above chance out of a possible total of 24. From the card issuers' perspective, they have a model which can correctly identify some 34 per cent of customers who are likely to close their accounts. The costs of misclassification are also less than with a credit-scoring model, where the purpose is to identify in advance the likelihood of bad as opposed to good customers. Misclassification with the latter model may well incur substantial costs and therefore lead to a reduction in profitability. On the other hand, with attrition and closures, the associated costs are relatively minimal - being typically related to the non-response of customers to direct mail shots

We turn now to the relative importance of each predictor variable in terms of its discriminatory power Table 3 shows the structure coefficients for each variable included in the estimated function The standardized coefficients are not shown because they represent the relative discriminatory power of each predictor variable, given the other variables in the function. As such, they can give an inaccurate indication of the

discriminatory power of each variable if there is a degree of correlation between any variables included in the function. Only the within-groups correlations are shown in Table 3, for this reason, and because (as simple bivariate correlations) they are not affected by other variables in the function and are in some respects a better guide (Klecka 1980).

Table 3
Within-groups structure coefficients

Variables	Within-groups	Rank
BEHSCORE	0 774 00	1
TOTALINT	0 413 04	2
PREVEXT	0.370 82	3
TYPCHAN	0 320 99	4
NPLASTIC	0 176 59	5
ACCTYP	0 168 95	6
AMCASHPM	0.154 86	7
SORTCODE	0.143 32	8
INSTAT	0 111 58	9
AGE	0 103 73	10
DIRECTDI	0 047 82	11
COCODE	0 037 43	12
SEX	0 007 06	13
AFF	-0 002 29	14
CREDITLF	0 000 27	15

Using this measure, the top four variables⁸ are BEHSCORE, TOTALINT, PREVEXT, and TYPCHAN. The other variables, all of which added significantly to the discriminatory power of the function (at F = 1.0), have noticeably lower values, which indicates that they contribute much less to the canonical discriminant function. This is particularly true for DIRECTDI, COCODE, SEX, AFF, and CREDITLF, all of which have a structure coefficient less than 0.05.⁹

In interpreting the results, emphasis has been placed on the ten most powerful discriminatory variables as indicated by the structure coefficients. It is important to note, however, that we are examining the ability of values $X'_j = 1/n (n_i / c_j) + 1/n (C_T/N_T)$ to

distinguish between 'normal' and 'closed'. We must, therefore, consider the relationships which exist between values of X_j' and X_j for each of the variables.

The BEHSCORE categories reveal that credit-card customers who have had a dormant account for longer than 12 months are most likely to close their accounts. Conversely, a BEHSCORE category indicating that an account is at least five cycles delinquent has the most important discriminatory effect on whether the account will operate normally¹⁰. Having regard to the definition of 'closed' that we have adopted, these five-cycle-delinquent customers are typical of those who will be closely controlled by the issuer in an attempt to reduce the arrears and bring the account under control. In this sense, therefore, those customers are arguably not in a position to 'close' their accounts and, in fact, run the distinct risk of having their accounts revoked by the issuer.

The categories relating to TOTALINT showed that those customers with no monthly outstanding interest were the most inclined to close their accounts. As outstanding monthly interest increased, however, there was a greater tendency to operate the account normally. This seems to add weight to the idea that whoever controls the account has an important influence on whether the account is operated 'normally' or 'closed'. If the customer is in control (in terms of regularly paying interest and principal), he at least places himself in a position to close the account. This is in direct contrast to a customer who is in arrears of either interest or principal, when the position is more likely to be controlled by the card issuer.

The various categories of PREVEXT indicate that, under circumstances where the credit card has been lost or stolen, the card is not likely to be returned to the issuer. Where the account operates normally, however, or where it has been revoked, or where the accrual of interest has been prohibited, etc., the account is more likely to be closed. This appears to follow the broad conclusions which were drawn from BEHSCORE and TOTALINT, as the exertion of some form of control over the account appears to determine, at least to some extent, whether the account will operate normally or not. By identifying the key

characteristics of the credit-card product, a distinct possibility arises to influence customer behaviour and therefore increase or decrease a customer's propensity to use the product.

The importance of control is also borne out by TYPCHAN. Where the credit limit is changed either automatically by the issuer or upon the instigation of the customer, the account is more likely to operate normally. However, where an increase in the credit limit has been permanently deferred, the account is more likely to be closed.

The remaining categories of NPLASTIC indicated that customers with one card were more inclined to close their accounts compared to customers with two cards, a conclusion which was also supported by an examination of ACCTYP. This indicated that customers who had a combination of credit cards, i.e. both VISA and MASTERCARD, were more inclined to operate the account normally compared to customers who had sole card accounts. Whether this reflects the greater need or the greater sophistication of the former customers is difficult to say but, when AMCASHPM was examined in closer detail, certainly the customers who had the largest monthly amounts of cash posted to their accounts had a tendency to operate normally, whereas customers with no cash posted were inclined to close their accounts.

SORTCODE was interesting too in the sense that customers who held a banking account with the card issuer were less inclined to close their card accounts compared to customers who banked elsewhere. This at least provides tentative evidence that established relationships with a financial institution reinforce the control element and possibly might reduce the likelihood of customers closing their card accounts.

INSTAT categories revealed that customers who were 'normal' or had a credit balance on their accounts were more inclined to close these accounts than customers who were at least one cycle delinquent, over the limit, or both. These points were also borne out by the final predictor variable AGE, which revealed that younger customers (under 40 years old) were more inclined to close their accounts. From about the age of 40 up to about the age

of 60, the accounts tended to operate normally, after which time the inclination to close increased.

An increase in mortality rates or a reduction in expenditure after retirement, and therefore a reduction in the need for credit, possibly explains the behaviour of the 60+ age group. However, at the other extreme, there may well be a very real need for credit, and therefore the issue of who controls the account and how this control is used arises once again. In the middle age ranges, 40-60 years old, control may be exercised more by the customer rather than the issuer. The behaviour of the customer, however, may also be more heavily influenced by the length and nature of the relationship with the card issuer.

The analysis of the categories relating to the important predictor variables suggests that the key determinants of whether an account will operate 'normally' or be 'closed' are (1) customer need, (2) how the account is controlled, and - closely related to this - (3) the relationship that the card holder has with the issuer. As such, the analysis represents the first tentative step in identifying appropriate strategies, based upon customer behaviour, for reducing closures and increasing profitability. In order to maximize the effectiveness of these strategies, however, it is important to target specific customer groupings by segmenting the customer portfolio.

Conclusion

Using linear discriminant analysis, this model was able to classify correctly 95 per cent of customers who operated their card account normally in the time period examined, and approximately 35 per cent of those who closed their account. Discussions with representatives of various card-issuing organizations suggest similarities between the performance of their models and our results.

On a less positive note, however, the research has also highlighted certain weaknesses of this type of approach. Firstly, the canonical discriminant function is explaining only 20 per

cent¹¹ of the variance in the dependent variable, and this suggests that additional predictor variables need to be considered, e.g. current account activity and the cost of this type of credit. Secondly, discriminant analysis is an *a priori* segmentation method and as such may be unable to differentiate between groups effectively (Frank *et al.* 1968). For instance, if we were to divide credit card users further into 'high-profit' and 'low-profit' segments, the variability within the groups could still remain high. For example, the 'low-profit' groups (i.e. for both 'normal' and 'closed') could contain both 'timids', who never or rarely use their cards, and 'spenders', who use their cards regularly but avoid paying any interest. This latter point suggests that an alternative segmentation model (e.g. a cluster-based model) should be used in any subsequent research.

NOTES

¹ Based on information provided by the card issuer sponsoring this research

² The majority of customers who closed their accounts in this period did so after June 1992

³ For a discussion of the predictive performance of our estimated model, see Eisenbeis (1977), Kschirsagar (1972), and Lachenbruch & Mickey (1968)

⁴ The dependent variable 'external status' has a variety of categories (e.g. normal, authorization prohibited, bankrupt, closed, revoked, frozen, interest accrual prohibited, lost, stolen, and charged off). For the purposes of this paper, however, all circumstances have been categorized as 'normal' unless the customer has returned the card to the issuer of his own free volition when it is categorized 'closed'

⁵ The distance between groups *a* and *b* is defined as

$$Dab^2 = (n - g) \sum_{i=1}^p \sum_{j=1}^p w_{ij} * (\bar{X}_{ia} - \bar{X}_{ib}) (\bar{X}_{ja} - \bar{X}_{jb}),$$

where g is the number of groups, p is the number of variables in the model, $\bar{X}_{ia}$ is the mean for the i th variable group a , and w_{ij}^* is an element from the inverse of the within-groups covariance matrix.

⁶ The maximum-chance criterion is defined as $C_{\max} = \max \{p, 1 - p\}$ where p is the proportion of cases in one of the groups, e.g. 'normal'. That is, if over half of the cases were 'normal', the greatest proportion correctly classified by chance would be obtained by placing every one in the 'normal' category

⁷ One would expect an upward bias with this classification (Hair *et al.* 1987)

⁸ The same was true using the F to remove criterion and the standardized coefficients

⁹ Consequently these variables have been excluded from the interpretation of the results

¹⁰ A customer who is five cycles delinquent will not be regarded as 'normal' by the card issuer but as 'delinquent', as indicated by the customer's internal status

¹¹ The canonical correlation equals 0.4409241

REFERENCES

- Bierman, H. and Hausman, W.H. (1970), "The credit-granting decision", *Management Science*, Vol.16, pp. 519-32.
- Boyle, M., Crook, J.N., Hamilton, R. and Thomas, L.C. (1992). "Methods of credit scoring applied to slow payers", In: *Credit scoring and control* (Thomas, L.C., Crook, J.N., & Edelman, D., (Eds)) Oxford: Oxford University Press, December, pp.75-90.
- British Bankers Association (1993), *Annual Abstract of Banking Statistics* 10.

- Chandler, G C., and Coffman, J.Y. (1983-1984), "Applications of performance scoring to accounts-receivable management in consumer credit", *Journal of Retail Banking*, Vol. 5, pp.1-10.
- Crook, J.N., Hamilton, R., and Thomas, L C (1992a), "A comparison of a credit scoring model with a credit performance model", *The Service Industries Journal*, Vol. 12(4), pp.558-79
- Crook, J.N., Hamilton, R., and Thomas, L C (1992b), "A comparison of discriminators under alternative definitions of default", In: *Credit scoring and credit control* (Thomas, L., Crook, J.N., and Edelman, D.B., (Eds.)). Oxford: Oxford University Press, December, pp 217-246.
- Eisenbeis, R A., (1977). "Pitfalls in the application of discriminant analysis in business finance and economics", *Journal of Finance*, Vol 32(3), pp 875-900.
- Frank, R.E., Massey, W.F , and Boyd, H., (1968), "The demographic segmentation of household products", In: *Application of science in marketing management* (Bass, W F , King, C W., and Pessemier, E M., (Eds)) New York Wiley
- Hair, J.F., Anderson, R E and Tatham, R L (1987) *Multivariate data analysis*. New York: Collier Macmillan.
- Klecka, W.R. (1980). *Discriminant Analysis*. Oxford: Sage University Paper No 19, Series: Quantitative Applications in the Social Sciences.
- Kschirsagar, A.M. (1972). *Multivariate Analysis*, Vol.2. Marcel Drekker.
- Lachenbruch, P.A. and Mickey, M.R. (1968). "Estimation of error rates in discriminant analysis", *Technometrics*, Vol 10, pp.1-11.

Lundy, M. (1992). "Cluster Analysis", In. *Credit scoring and credit control*, (Thomas, L.C , Crook, J.N., and Edelman, D. (Eds)), pp 91-107 Oxford: Oxford University Press

Myers, J.H. and Forgy, E.W. (1963). "The development of numerical credit evaluation systems", *American Statistical Association Journal*, Vol 58, pp.799-806.

SPSSX Inc, (1998) *SPSS-X User's Guide* 3rd edition.

Wiginton, J.C. (1980). "A note on the comparison of logit and discriminant models of consumer credit behaviour", *Journal of Financial and Quantitative Analysis*, Vol.15, pp 757-70

FURTHER READING

Crook, J.N., Hamilton, R , and Thomas, L.C., (1992c). "The degradation of the score card over the business cycle", *IMA Journal of Mathematics Applied in Business and Industry*, Vol 4, pp.111-23.

Morall, K., (1992). "Managing the deluge of data", *Credit Card Management*, March, pp.11-14.

SPSSX Inc, (1998) *SPSS-X Advanced Statistics Guide*, 2nd edition

Appendix: Twenty-two original variables

Variable	Description
SEX	Male or female
COCODE	Great Britain or others
AGE	Age in years
DIRECTDI	Whether charges are paid by direct debit
AFF	Whether the annual charge fee is to be waived
CREDITLF	Whether customer is in the cardholder repayment protector scheme
NPLASTIC	Number of credit cards held by customer
INSTAT	Whether customer is delinquent* or over the limit on credit balance or normal
PREVEXT	Relates to customer's previous [§] 'external status' and indicates whether the account operated normally, whether the card was returned by customer, or whether it was stolen or lost, etc
ACCPRE	Whether card is Mastercard, Visa, etc.
ACCTYPE	Whether card holder has combinations of different cards
SORTCODE	Where card holder has primary bank account
ACTIVEYY	How long the card has been active
LACCOPEN	How long the account has been open
CREDITLM	Credit limit
BEHSCORE	Score based on customer's behaviour in operating the account
TYPCHAN	Circumstances of last credit-limit change
AMCASHPM	Amount of cash posted in previous year (1992) - monthly average
NOCASHAD	Number of cash advances in previous year (1992) - monthly average
NOPURPM	Number of purchases in previous year (1992) - monthly average
AMPURPM	Amount of purchases in previous year (1992) - monthly average
TOTALINT	Total interest and service charge in previous year (1992) - monthly average

- Delinquency means 1 cycle default
- [§]'Previous' in this context means where, for example, the customer closed the account and then reopened it, or where the card issuer suspended the account and later re-opened it, or where a marital break-up resulted in a joint account becoming two separate accounts

CHAPTER 9

REVOLVING CREDIT CARD HOLDERS: WHO ARE THEY AND HOW CAN THEY BE IDENTIFIED?

Robert Hamilton and Mosahid Khan

(Business School, Loughborough University)

The Service Industries Journal, (2001), Vol 21, No. 3 (July), pp 37-48

Revolving Credit Card Holders: Who Are They and How Can They Be Identified?

Robert Hamilton and Mosahid Khan

All major credit card issuers, to a greater or lesser extent, are holding a portfolio consisting of three types of credit card holder: (i) non-active card holders; (ii) non-interest paying active card holders; and (iii) interest paying active card holders. This article, using two quantitative techniques more commonly associated with credit risk management or credit scoring, is concerned with identifying the characteristics of active card holders with the greatest propensity to revolve (i.e. pay interest).

The sample consists of 27,681 bank credit card holders who had held and used their card in the 14 month sample period. Data was available on 313 socio-demographic and behavioural variables for which, a priori, there was good reason to include so as to discriminate between users who paid interest on their outstanding balances (i.e. revolvers) and those who did not.

The main result of this research is that the most important discriminating variables are derived from the card holder's behaviour (i.e. cash advances, minimum payment due, interest paid in previous periods). This result is derived from and supported by the two competing techniques used for the analysis: Linear Discriminant Analysis and Logistic Regression.

INTRODUCTION

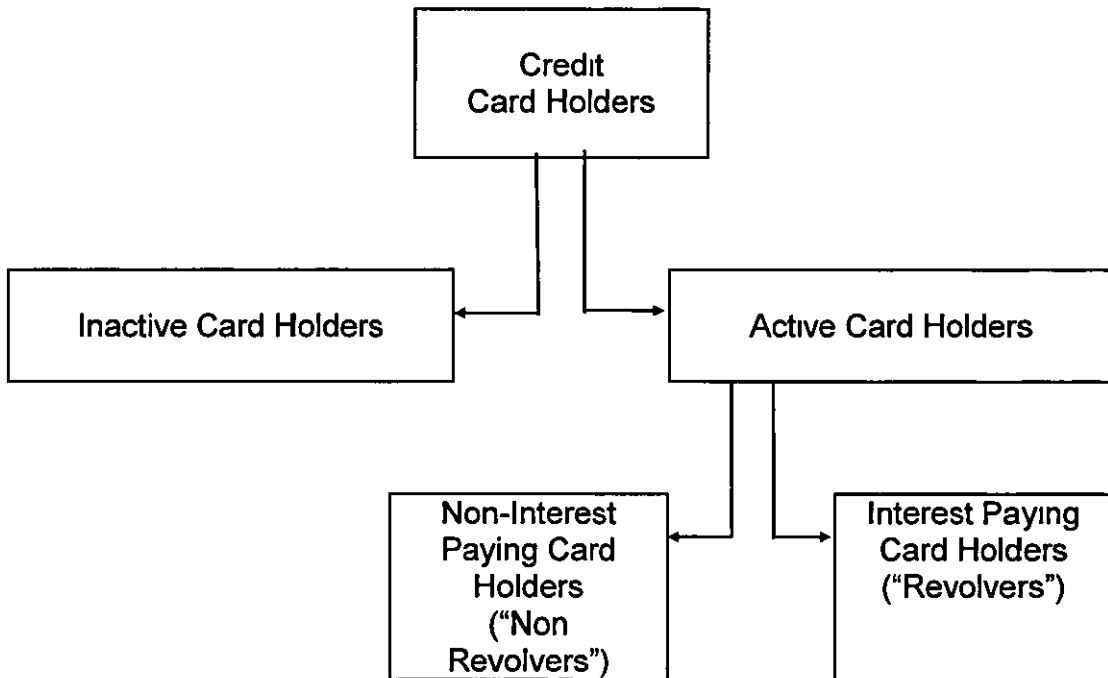
Rosenberg and Gleit (1994) and Frank (1996a) identify the many uses of quantitative techniques to assist decision-making in the broad area of credit (risk) management. *Inter alia*, such areas include: whether or not to offer an existing or potential customer credit in the first instance (credit scoring for the accept/reject situation); whether or not to change an existing credit limit (behavioural scoring); the collection possibilities of charged-off accounts; credit card fraud detection, and delinquency issues. This article looks at the use of two quantitative techniques more commonly associated with the areas of credit scoring and behavioural scoring, in the relatively new but fast growing area of database marketing or target marketing (Zahavi and Levin, 1997) in the UK credit card market.

Database or target marketing can be viewed as a means of segmenting a market which in the UK financial services sector has either (i) not previously played a key role in the marketing strategies of financial service providers or (ii) not appeared to any great extent in the published literature. A detailed review of various pieces of research in this area, mostly from the USA, was produced by Speed and Smith (1997).

Frank (1996a) argues that the increased use of such modelling techniques in this area can be explained with reference to the following developments in the credit card market:

- (i) increased competition to identify and retain profitable account holders;
- (ii) the proliferation of available card holder data;
- (iii) the falling cost of processing power and storage capacity,
- (iv) a rising industry comfort level with scoring;
- (v) recent increases in charge-offs;
- (vi) the increasing desire for credit card fraud detection.

FIGURE 1
CREDIT CARD ISSUER'S PORTFOLIO



All major credit card issuers, to a greater or lesser extent, are holding a portfolio consisting of three types of credit card holder (Figure 1). This paper, using linear discriminant analysis and logistic regression, is concerned with identifying the characteristics of active credit card holders with the greatest propensity to revolve (i.e. interest paying card holders). Logically, such customers, as they are paying interest plus any annual fee, are the most profitable to the card issuers and should, therefore, subject to credit status, be targeted for additional interest-charging services (e.g. loans, mortgages, additional credit cards, etc.) as their behaviour would suggest that they are the most comfortable with paying interest.

On the other hand, credit card holders less likely to pay interest (i.e. convenience users) could form another important segment of the card issuer's portfolio and might be targeted with alternative or differentiated products that would be more profitable or less costly for the card issuer. For example, a debit card, a gold card, a credit card differentiated on the basis of the annual fee or the interest rate charged (See Higgins, 1996).

Section 2 looks at the sample period, variable selection and methodology and Section 3 outlines the results with respect to the variables selected, the most powerful selected variables and the percentage correctly classified. Section 4 presents the conclusions of this research and considers further practical issues

SAMPLE PERIOD, VARIABLE SELECTION AND METHODOLOGY

Sample Period

Unlike with applicant credit scoring¹, this research is concerned with the likely behaviour of a credit card holder within a specific time period, i.e in this case three months. Furthermore, it was decided to try and explain this behaviour by examining the customers' behaviour over a period of time long enough to include both heavy and lighter periods of spending (e.g. Christmas, birthdays, Summer holidays). Therefore, a sample period of 14 months was selected (see Figure 2).

Figure 2

SAMPLE PERIOD

Period 1	January	94	}	
Period 2	February	94	}	
Period 3	March	94	}	Not included
Period 4	April	94	}	
Period 5	May	94	}	
Period 6	June	94		
Period 7	July	94		
Period 8	August	94	}	
Period 9	September	94	}	Customers did not pay any interest.
Period 10	October	94	}	
Period 11	November	94	}	
Period 12	December	94	}	Used to identify customers who paid interest.
Period 13	January	95	}	
Period 14	February	95	}	

Periods 1-5 inclusive were later omitted from the model (see "Variable Selection") as the association between the variables explaining the card holders' behaviour in these periods and their propensity to revolve was relatively weak. During periods 8-11 none of the 27,681 credit card holders paid any interest on their credit balances although they all had the opportunity, and periods 12-14 determined whether or not they were "revolvers", i.e. they were classified as a "revolver" if they had paid interest on their credit card balance at least once during periods 12-14 inclusive.

Variable Selection

For the random sample of 27,681 active credit card holders, 313 socio-demographic and behavioural predictor variables were made available for the research by a major UK bank. Because of the shortage of published research in this area, the 313 original variables were selected on the grounds that (i) they related either to the card holders' behaviour with respect to financial products held or they were demographic and (ii) most of the variables are readily available to a card issuer. Chi-square tests were initially used on all 313 variables to test the association between the dependent variable² and the independent variables.³ This exercise resulted in 55 variables being further considered on the grounds that (i) there was, *a priori*, justification for including them; and (ii) the chi-square test indicated a significant relationship between the likelihood that the customer will revolve their credit card balance and the independent variables selected.

The next stage involved utilising the stepwise method of variable selection available on SPSSX for both discriminant analysis and logistic regression. For discriminant analysis the criterion for variable selection (O'Gorman and Woolson, 1991) was the Mahalanobis Distance Statistic (D^2 , a generalised measure of the distance between the two groups), with the F-to-enter/remove criteria set, in order to maximise the discriminatory power of the model and minimise the number of variables included, at a relatively high value of 25 (the default values equal 1.00). Similarly, forward stepwise variable selection was used in the logistic regression model and again the criteria for variables entering or leaving the model were set so as to minimise the number of independent variables, but maximise the

predictive power of the model. In this respect, the probability of score statistic for variable entry was set at 0.05 and the likelihood ratio statistic to remove a variable was set relatively low to make it more difficult for a variable to stay in the model at 0.0005 (default = 0.10).

The final stage of variable selection involved checking for dependency between the independent variables left in the models. Multicollinearity, a situation where two or more independent variables are highly correlated, reduces the reliability of the estimated coefficients and would, therefore, make any further analysis of the relative importance of any single variable very unreliable. The approach adopted for dealing with multicollinearity was to remove all but one of the highly correlated variables so that all variables left in the model had a tolerance ($1 - R_i^2$)⁴ of at least 0.8 (Crook *et al.*, 1992, Hamilton, 1994).

Methodology

Rosenberg and Glerf (1994), when talking about the different approaches to credit management (e.g. quantitative and judgmental), argue that "credit management is currently as much of an art as a science". However, arguably one could also apply this dichotomy to the quantitative approaches alone with the science element being the techniques used and the art being the formation of meaningful classes (or categories) for each independent variable. Initially the discussion will centre briefly on the two techniques: linear discriminant analysis and logistic regression, and secondly on the forming of classes for each independent variable.

Linear discriminant analysis (LDA) is arguably the most commonly used technique in the broad area of credit risk management, (now being extended to database marketing), and as such has received wide coverage in the published literature.⁵ The linear discriminant function (equation 2.1), which is similar to the multiple regression equation, estimates the coefficients so as to provide the best discrimination between two or more groups.

$$Z = B_0 + B_{1x1} + B_{2x2} + \dots + B_{n x_n} \text{ (equation 1)}$$

where Z = discriminant score
 B 's = estimated coefficients
 x 's = values of the predictor variables

Despite the overwhelming acceptance of this technique, one must still be mindful of the assumptions (Gilbert, 1968; Eisenbeis, 1977; Klecka, 1980):

- (i) each case must be a known member of one or two or more mutually exclusive and exhaustive groups;
- (ii) discriminating variables must be measured at interval or ratio level of measurement;
- (iii) no discriminating variable may be a linear combination of other discriminating variables;
- (iv) the population covariance matrices are equal for each group,
- (v) each group is drawn from a population which has a multivariate normal distribution.

Logistic regression (LR) hypothesis testing, unlike LDA, does not require the same strict assumptions and one might suggest that, with the increased availability of powerful computers, the growing use of LR in a variety of situations is because LR requires only that; for each independent variable all of the observations are independent (Shott, 1991)

The formulae for LR, where one is directly estimating the probability of an event (e.g. revolving a credit balance) is given by:

$$\text{Probability (event)} = \frac{1}{1 + e^{-z}} \qquad \text{(equation 2)}$$

where $Z = B_0 + B_{1x1} + B_{2x2} + \dots + B_{n \times n}$

B's = estimated coefficients

e = base of the natural logarithms

and Probability (no event) = 1 - Probability (event)

The formation of groups or classes for each independent variable (i.e. the art) in this type of modelling should be viewed as a necessity rather than optional for two reasons. Firstly, for many variables some of the attributes will be under-represented (Lewis, 1994), e.g. very few people aged 70 will hold a credit card so it would, therefore, be dangerous to draw conclusions about the behaviour of people aged 70 based on only a few cases. Secondly, as more and more organisations are constructing their own decision system models in-house (Jost, 1993), classing helps the organisation to better understand the behaviour of their own customers especially if it is performed manually; something that is lost or ignored when the task is performed externally. Therefore in this research, for each independent variable classes were formed on the basis of similarity of $r_{ij} / r_{ij} + nr_{ij}$ (see equation 3) while paying attention to understanding the behaviour of the classes formed and also ensuring that no class was under-represented (see Crook *et al.*, 1992; Boyle *et al.*, 1992; Hand *et al.*, 1997).

Classing also provides two further benefits

- (i) LDA requires that all predictor variables be measured at interval or ratio level. Therefore, in this research, having formed classes for each and every independent variable⁶, each class was then given the value of their weight of evidence, W_{ij} (see Banasik *et al.*, 1995).

$$W_{ij} = \ln(r_{ij}NR_j / nr_{ij}R_j) \quad \text{(equation 3)}$$

where W_{ij} = weight of evidence for class i for variable j
 r_{ij} = number of revolvers for class i for variable j
 nr_{ij} = number of non-revolvers for class i for variable j
 R_j = total number of revolvers for variable j
 NR_j = total number of non-revolvers for variable j

- (ii) Classing as opposed to not classing will (a) render more meaningful results for the continuous variables and (b) for all variables the better the separation between classes, the better will be the model.

To obtain an unbiased estimate of the accuracy of the models (i.e. how well it predicts), the total sample of 27,681 cases was split 60 40 respectively into (i) a training sample to build the model and (ii) a holdout sample. The results presented in the next section relate to the holdout sample only.

RESULTS

Given the objectives of this modelling (i.e. to maximise the predictive power of the model while minimising the number of predictor variables), the results will be analysed in terms of (i) the variables selected by each model, the ranking of the selected variables and the interpretation of the models, and (ii) the classification tables

TABLE 1

DESCRIPTION OF SELECTED INDEPENDENT VARIABLES

AGE	Age of credit card holder
AMTDU (12)	Minimum payment due following previous period's activity.
AMTCSH (11)	Amount of cash advanced in period.
CLOAN	Whether or not the card holder has a loan(s).
DTE-OPN	The number of years the account has been open
INTCHG (7)	Amount of interest charged in period.

3.1 *Variables Selected, Ranking and Interpretation*

Table 1 lists the independent variables selected by both LDA and LR, with the former selecting all six variables and the latter selecting five out of the six (AMTCSH 11 was not selected). Table 2 gives an indication of the ranking of the selected variables for both techniques and for LDA this is based on the standardised coefficients (1) and the pooled within groups correlations (2). For LR, given that all our variables are categorical and LR creates a new variable for each class⁷, the ranking is based on when the variable entered the model (3). As shown, the ranking for selected variables is very similar; the only differences occur with the lower order variables

TABLE 2
RANKING OF SELECTED VARIABLES

Variable		Standardised Coefficients (1)		Pooled Within Groups Correlations (2)	Step Entered (3)
AMTDU (12)	1	0.66	1	0.67	1
INTCHG (7)	2	0.59	2	0.61	2
AGE	3	0.25	3	0.40	3
DTE-OPN	4	0.17	4	0.24	4
CLOAN	6	0.13	5	0.21	5
AMTSCH (11)	5	0.15	6	0.19	-

When it comes to interpreting the results⁸, both models show that

- the greater the amount spent on the credit card in the last month, the more likely the holder is to revolve. At first sight this may appear to be obvious, however it should be remembered that using the card is a necessary but not sufficient requirement for paying interest;

- the most likely revolvers paid interest on their credit balance in period 7;
- people aged under 35 were significantly more likely to become revolvers and the older one gets, the less likely they are to revolve;
- the longer one had held their card, the less likely they were to revolve, with the least likely “revolvers” having held their card for more than 14 years;
- people who held other interest-charging products (i.e. a loan) were more likely to become revolvers. This possibly indicates a positive attitude towards a buy now, pay later approach.

TABLE 3
CLASSIFICATION TABLE

		LDA			LR		
		Predicted Group			Predicted Group		
Actual Group		R	NR	Total	R	NR	Total
	R	297 (18.2%)	1337 (81.8%)	1634	165 (10.1%)	1469 (89.9%)	1634
	NR	298 (3.2%)	9140 (96.8%)	9438	81 (0.8%)	9357 (99.2%)	9438
Percentage correctly classified				85.2%	86%		
Percentage correctly classified by chance ⁹ :				74.8%			

Notes Linear Discriminant Analysis (LDA), Logistic Regression (LR), Revolvers (R), Non-Revolvers (NR)

Classification Tables

For ease of comparison (also see Harrell *et al* , 1985; Moore, 1973; Press *et al* , 1978), Table 3 shows the classification results for the two models^{10, 11}. The first observation to make is that the overall percentage correctly classified by both models is very good and much better than the chance measure. However, on closer examination one can see that both models perform poorly when it comes to correctly classifying cases belonging to the smaller group (i.e. revolvers), as both models tend to classify nearly all cases (particularly LR) into the larger of the two groups. This latter finding is a common problem with LDA and LR when you have one group much larger than the other (e.g. Morrison, 1969; Tansey *et al.*, 1996), however the justification of building the models with unequal groups is that the proportions used in this research are a reflection of card issuer's portfolio (population).

CONCLUSIONS AND FURTHER RESEARCH

This research, which has used two tried and tested quantitative techniques in a marketing situation, has shown that (i) logistic regression and linear discriminant analysis provide very similar results, although LR might be more acceptable to senior management since the results are presented (equation 2) in terms of the probability of revolving rather than simply a score (z), (ii) the most important discriminating variables are derived from the card holder's behaviour; and (iii) by forming classes for each independent variable the W_{ij} values indicate, for each of the selected variables, which class(es) are most likely to revolve their credit card balance.

This type of modelling should, therefore, be considered to further segment the card issuer's portfolio and also provide an input to profit models. However, on a less positive note even though the overall percentage correctly classified for each model is significantly better than the chance measure, the percentage correctly classified for the smaller group is really very poor. This finding would seem to signal the need for further research to analyse what would happen if equal size groups were used, an approach implied by Lewis (1994) and/or

an alternative technique was used (e.g. neural networks). Obviously if any significant differences did occur, this would have serious forecasting and planning implications for the organisation.

ACKNOWLEDGEMENTS

The authors wish to thank the bank for supporting this research and the anonymous referees for their helpful comments. However, all errors and omissions are the responsibility of the authors.

NOTES

1. With applicant credit scoring the model is trying to forecast whether or not the applicant is ever likely to be a "bad" risk based on the information provided on the application form
2. The dependent variable was derived from whether or not the credit card holder had paid interest on their credit card balance at least once during periods 12-14 inclusive. Therefore, this variable was binary in that the value was either 0 or 1 (i.e. "revolver" or "non-revolver").
3. The terms predictor variable, discriminating variable and independent variable are being used interchangeably to mean the right hand side variables of the relative function.
4. Where R_i^2 is the squared multiple correlation coefficient when the i^{th} independent variable is considered the dependent variable and the regression equation between it and the other independent variables is calculated (Norusis, 1990)

5. For a fuller coverage of (i) LDA, see Klecka (1980), Eisenbeis (1978), Lachenbruch, (1975), Rosenberg and Gleit (1994), Hair *et al* (1995), and (ii) LR, see Aldrich *et al* , (1984); Hosmer *et al* , (1989)
6. Norusis (1990) points out that when you have a mixed set of independent variables, LDA is not optimal.
7. The number of new variables created is one less than the number of classes.
8. For LDA, the values used to derive the model were the W_{ij} values not the original raw data.
9. $C_{prop} = p^2 + (1 - p)^2$

where p = the proportion of cases in group 1;
 $(1 - p)$ = the proportion of cases in group 2.
10. The two techniques use different classification rules For LDA, the classification rule is based on Bayes' rule and uses the prior probability, conditional probability and the posterior probability. For LR, if the probability is greater than 0.5 then it is predicted that the event will occur.
11. It could be argued that while we are interested in correctly classifying cases in both groups, neither of the two classification rules are satisfactory as they are assuming equal (opportunity) costs for all cases and constant opportunity costs within each, neither of which is generally true (see Rosenberg and Gleit, 1994)

REFERENCES

- Aldrich, J.H. and Nelson, F.D., (1984), "Linear Probability, Logit and Probit Models", Sage University Paper No 45, Series: Quantitative Application in the Social Sciences.
- Banasik, J , Crook, J N. and Thomas, L.C., (1995), "Does Scoring a Subpopulation Make a Difference?" Paper presented at the Credit Scoring and Credit Control Conference, University of Edinburgh, September.
- Boyle, M., Crook, J. N., Hamilton, R. and Thomas, L. C , (1992), "Methods of Credit Scoring Applied to Slow Payers", in Thomas, L C., Crook, J.N. and Edelman, D (eds.), *Credit Scoring and Credit Control*. Oxford. Oxford University Press, December, pp.75-90.
- Crook, J N , Hamilton, R. and Thomas, L.C , (1992), "A Comparison of Discriminators under Alternative Definitions of Credit Default", in Thomas, L.C., Crook, J.N. and Edelman, D (eds.), *Credit Scoring and Credit Control*. Oxford: Oxford University Press, December, pp217-246.
- Eisenbeis, R.A., (1977), "Pitfalls in the Application of Discriminant Analysis in Business, Finance and Economics", *Journal of Finance*, Vol. 32(3), pp. 875-900.
- Eisenbeis, R A., (1978), "Problems in Applying Discriminant Analysis in Credit Scoring Models", *Journal of Banking and Finance*, Vol. 2, pp 205-219.
- Frank, J N , (1996a), "Beyond Direct Mail", *Credit Card Management*, August, pp. 54-58
- Frank, J N , (1996b), "Scoring Takes on a New Meaning", *Credit Card Management*, September, pp 155-159.
- Gilbert, E.S , (1968), "On Discrimination Using Qualitative Variables", *Journal of the American Statistical Association*, Vol. 63, pp.1399-1412.

- Hair, J.F , Anderson, R E., Tatham, R C and Back, W.C., (1995), *Multivariate Data Analysis With Readings* 4th Edition, New Jersey: Prentice-Hall.
- Hamilton, R , (1994), "Credit Scoring Using Discriminant Analysis: A Teacher's Guide", *ABT Bulletin*, Vol 44, September, pp. 10-13
- Hand, D J. and Henley, W E , (1997), "Statistical Classification Methods in Consumer Credit Scoring: A Review", *Journal of the Royal Statistical Society*, Vol 160, Part 3, pp. 523-541.
- Harrell, F.E. and Lee, K.L , (1985), "A Comparison of the Discrimination of Discriminant Analysis and Logistic Regression Under Multivariate Normality", in Sen, P K (Ed), *Biostatistics: Statistics in Biomedical, Public Health and Environmental Sciences* North-Holland· Elsevier Science Publishers BV.
- Higgins, K.T., (1996) "Those Naughty Nonrevolvers", *Credit Card Management*, December, pp. 47-57.
- Hosmer, D W. and Lemeshow, S , (1989), *Applied Logistic Regression* New York· Wiley.
- Jost, A., (1993), "Neural Networks A Logical Progression in Credit and Marketing Decision Systems", *Credit World*, March/April, pp. 26-33.
- Klecka, W R , (1980), "Discriminant Analysis", Sage University Paper No 19, Series: Quantitative Application in the Social Sciences
- Lachenbruch, P.A., (1975), *Discriminant Analysis*, New York Hafner Press.
- Lewis, E.M., (1994), *An Introduction to Credit Scoring*, California: Athena Press
- Moore, D.H., (1973), "Evaluation of Five Discrimination Procedures for Binary Variables", *Journal of the American Statistical Association*, Vol. 68, No. 342, pp. 399-404.

Morrison, D G., (1969), "On Interpretation of Discriminant Analysis", *Journal of Marketing Research*, Vol VI, May, pp. 156-163.

Norusis, M J., (1988), *SPSS Advanced Statistics User's Guide*, Chicago: SPSS Inc.

O'Gorman, T.W. and Woolson, R F., (1991), "Variable Selection to Discriminate Between Groups: Stepwise Logistic Regression or Stepwise Discriminant Analysis?" *The American Statistical Association*, Vol. 45, August, pp. 187-193.

Press, S.J. and Wilson, S., (1978), "Choosing Between Logistic Regression and Discriminant Analysis", *Journal of the American Statistical Association*, Vol. 73, pp 699-705.

Rosenberg, E. and Gleit, A , (1994), "Quantitative Methods in Credit Management: A Survey", *Operations Research*, Vol. 42, No. 4, pp 589-614

Shott, S., (1991), "Logistic Regression and Discriminant Analysis", *JAVMA*, Vol 98, No. 11, pp. 1902-1905

Speed, R. and Smith, G., (1997), "Retail Financial Services Segmentation", reprinted in Meidan, A., Lewis, B. and Moutinho, L , (Eds.), *Financial Services Marketing: A Reader*, London: The Dryden Press

Tansey, R , White, M., Long, R.G. and Smith, M , (1996), "A Comparison of Loglinear Modeling and Logistic Regression in Management Research", *Journal of Management*, Vol. 22, No.2, pp. 339-358.

Zahavi, J. and Levin, N., (1997), "Applying Neural Computing to Target Marketing", *Journal of Direct Marketing*, Vol. 11, No.1, Winter, pp 5-22.

APPENDIX A

CREDIT SCORING USING DISCRIMINANT ANALYSIS: A TEACHER'S GUIDE

Robert Hamilton

(Business School, Loughborough University)

Association of Banking Teachers' Bulletin, (1994), (44), September, pp.10-13

Credit Scoring Using Discriminant Analysis: A Teacher's Guide

Robert Hamilton

Introduction

In 1983 the credit industry published the first 'Guide to Credit Scoring' and has, because of the increased use of more sophisticated techniques to make decisions about granting consumer credit, recently published a second 'Guide to Credit Scoring, 1993'. This second guide provides detailed principles and guidelines relating to the use of statistical techniques to make decisions about granting consumer credit and includes

- Principles of design
- Principles of implementation
- Principles of operation
- Principles of decision making
- Information to consumers
- Review of refusals
- Repeat applications
- Complaints procedures

Despite such developments, the teaching of the principles of credit scoring and the building of a credit scorecard is not commonly found in the syllabi of banking courses either at undergraduate, postgraduate or post-experience level. This article seeks to address this deficiency firstly by outlining the development path of credit scoring and secondly by briefly presenting some of the basic steps in the construction of a credit scorecard using one of the less sophisticated but more commonly used statistical techniques, multiple discriminant analysis.

The credit industry defines credit scoring as the use of statistical techniques to measure the likelihood that an application will be a good credit risk (Guide to Credit Scoring, 1993) and, while the widespread use of credit scoring in the credit evaluation situation did not gain prominence in this country until the late 1970s, it has its root in the USA as early as the 1940s and '50s. At that time the basic assumption underpinning the development of statistical analysis and computer technology in the consumer credit granting situation was that it should be possible to determine those facts about credit applicants that were associated with later satisfactory performance. This, it was argued, would present several distinct advantages over traditional judgemental decision making¹ (Lewis, IMA, 1992).

More recently the credit industry reinforced this earlier assumption by stating that 'it (credit scoring) is based on the fact that it is possible, using statistical techniques, to predict the future performance of groups with particular characteristics from the past performance of other groups with the same characteristics' and 'that it is one of the most consistent, accurate and fair forms of credit assessment available' (Guide to Credit Scoring, 1993).

Building a Bespoke Credit Scorecard

In this article we are going to look at the principles of design for the building of a bespoke credit scorecard, i.e. a scorecard based on information about the card issuer's own applications and experiences (as opposed to a generic scorecard), using information collected by a credit granter about previous accepted applicants². However, before discriminant analysis can be used to build a bespoke credit scorecard, the relevant groups and variables need to be specified.

Group Membership

As the main purpose of discriminant analysis is to determine whether or not it is possible to discriminate between two or more groups on the basis of the information collected, the first step is to specify what the groups are and the variable(s) that best determine group

membership. In this respect discriminant analysis is an *a priori* technique, that is each case must be a known member of one of two or more mutually exclusive and exhaustive groups. In what follows, we will assume for simplicity that each case is a member of one of two groups³ (the 'goods', those card holders who have never been more than two consecutive months' delinquent during the sample period and the 'bads', those card holders who have ever been three or more consecutive months' delinquent during the sample period) and that each case is fixed in the relevant group.

Variables and Validation

As we are concerned with credit scoring new applicants, the data used would normally be obtained from the card issuer's standard application form. In general, this will provide the following demographic and socio-demographic information⁴ (discriminating variables) about the applicants.

Table 1: Application Form Information

Postcode
Age
Number of children
Number of other dependants
Whether an applicant has a home 'phone
Spouse's income
Applicant's employment status
Applicant's employment category
Years in present employment
Applicant's income
Residential status
Years at present address
Estimated value of home
Mortgage balance outstanding
Years at bank
Whether a current account is held
Whether a deposit account is held
Whether a loan account is held
Whether a cheque guarantee card is held
Whether a major credit card is held
Whether a charge card is held
Whether a store card is held
Whether a building society card is held
Value of outgoings

Additionally, at this stage of development thought must be given to how the scorecard is going to be validated. In the context of this paper, validation refers to checking the predictive efficacy of the scorecard and ensuring that it correctly differentiates between the 'goods' and the 'bads' and that any predicted differences are not due to either chance or sampling methods. The most commonly used validation procedure involves the use of a

holdout sample, where the scorecard is constructed and the discriminant coefficients (see later) derived using a randomly selected proportion of the sample, say 80%. The discriminant coefficients are then used to predict group membership for each case in the holdout sample (the remaining 20%) and the results are then compared with the percentage classified by chance model (see later). While this method obviously requires a larger sample of data⁵, if such a validation procedure is not used it may lead to biased interpretations of any results (Frank, Massey and Morrison, 1995).

Using Discriminant Analysis

The application of discriminant analysis can be divided into three major stages (Hair *et al.*, 1987; Reichert *et al.*, 1983):

Derivation: Deriving a linear function that best discriminates between two or more groups

Validation: Classifying existing and new cases into predetermined groups

Interpretation⁶: Identifying the variable(s) that contribute most to the discrimination between the groups.

Derivation

In deriving the discriminant function, we will use the following notation (Morrison, 1969).

Let

X_{ji} be the i th individual's value of the j th discriminating variable

b_j be the discriminant coefficient for the j th variable

Z_i be the i th individual's discriminant score

Z_{crit} be the critical value for the discriminant score

where $Z_i = b_0 + b_1X_{1i} + b_2X_{2i} + \dots + b_nX_{ni}$

(n is the number of discriminating variables)

The classification procedure is:

if $Z_i > Z_{crit}$ classify individual i as belonging to group 1;

if $Z_i < Z_{crit}$ classify individual i as belonging to group 2.

NB The constant term is to ensure that the mean discriminant score is zero over all cases.

While discriminant analysis is frequently used to develop statistical credit scoring models, the adoption of this technique is not without criticism and such criticisms are generally levelled at the theoretical requirements of the model. Namely (Klecka, 1980):

- (i) Discriminating variables must be measured at the interval or ratio level of measurement (see later);
- (ii) The total number of cases must exceed the number of discriminating variables by more than two;
- (iii) No variable may be a linear combination of other discriminating variables (see later);
- (iv) The covariance matrices for each group must be equal,
- (v) Each group is drawn from a population which has a multivariate normal distribution.

A comprehensive examination of the aforementioned criticisms of discriminant analysis as used in the credit-granting situation is outside the scope of this paper (for example, see Eisenbeis, 1978; Frank *et al* , 1965), therefore only two of the more obvious problems will be examined and solutions suggested.

The first and possibly the most obvious difficulty stems from the information used to construct the scorecard. That is, most of the information is qualitative in nature (for example, postcode, residential status) rather than at the interval or ratio level, which is one of the more stringent requirements of discriminant analysis, i.e. assumption (1). Two alternative approaches to this problem are:

Create a variable with only two possible outcomes which may be given values 0 or 1 (a binary variable). For example, Table 2 looks at the variable residential status whose value may fall into one of five different categories: owner; with parents; tenant furnished; tenant unfurnished; other. With this approach (N-1), where N=number of categories, binary variables would be computed where one variable might take the value 1 if 'owner' and 0 if 'not owner', another variable might take the value 1 if 'with parents' and 0 if 'not with parents' and so on until the four new variables have been derived.

Note, only (N-1) binary or dummy variables are needed as the information provided by the last binary variable would be redundant (Hair *et al* , 1987). For example, with the variable 'whether a charge card is held' (assuming everybody responds with either a 'yes' or 'no' answer) when a respondent answers 'yes', let X1=1 and X2=0. When a respondent answers 'no', let X1=0 and X2=1. However, when X1=1 one already knows that X2 must equal 0, therefore X2 is providing redundant information and is not needed to represent the variable 'whether a charge card is held'.

Table 2: Residential Status

Category	'Goods'	'Bads'
Owner	493	22
With parents	205	5
Tenant furnished	103	5
Tenant unfurnished	117	6
Other	39	6
Total	957	44

The main drawback with this approach is that it will result in a large number of discriminating variables which are not normally distributed (Boyle *et al* , 1992)

The second approach is to replace all variables, both discrete and continuous, with variables measured at least at interval level. Again, using residential status (Table 2), let:

- g_i be the number of 'goods' in the sample who take the i th nominal value
- b_i be the number of 'bads' in the sample who take the i th nominal value
- G_i be the total number of 'goods' in the sample
- B_i be the total number of 'bads' in the sample

One can now replace the i th value of the nominal variable with a quantitative value depending on the values of g_i , b_i , G_i and B_i (Boyle *et al* , 1992).

For example, the quantitative value for someone who owns their property would equal $(X_j)^7$

$$X_j = \ln (g_i/b_i) + \ln (B_i/G_i)$$

$$X_j = \ln (493/22) + \ln (44/957)$$

$$X_j = 0.02985$$

The next stage in deriving the discriminant function involves selecting the variables that best discriminate between the groups and rejecting the variables that do not add significantly to the model. The three most commonly used selection procedures are.

- forward entry (starts with no variables in the function and enters the variables in order of their power of discrimination with the highest first);
- backward elimination (starts with all variables in the function and removes those variables that add least discrimination to the model);
- stepwise selection, which is in many respects a combination of the previous two selection procedures.

That is, at each step the variable with the greatest discriminating power, given the other variables in the function, is selected for inclusion and any variables already in the function

are considered for removal on the basis that the variable(s) does not add a statistically significant amount of discriminating power to the model. This process will continue until all variables in the equation satisfy both the inclusion and the removal criteria

The second problem stemming from the theoretical requirements of the model occurs after the selection process. Because the selection process is concerned solely with selecting the most powerful variables, it does not ensure that assumption (iii) has not been violated and one must therefore next check that the selected predictor variables are independent of each other and that high degrees of collinearity (i.e. relationships between the variables) do not exist. The possibility of multicollinearity occurs only in models with more than one predictor (or independent) variable and while its existence might not affect the predictive power of the model, it will affect the values of the coefficients assigned to any correlated variables (e.g. applicant's income and residential status) thus making the findings of the interpretation stage very suspect⁸

There are various statistical techniques available to identify variables that are highly correlated and to help decide what variables to omit in accordance with this assumption, for example, bivariate correlation matrix, tolerance tests (see Crook *et al.*, 1992)

Validation

Having calculated the discriminant coefficients, the model must now be evaluated. As discussed earlier, this will normally involve the use of a holdout sample to (i) compare predictions of group membership, and (ii) compare the percentage correctly classified by the model to that expected by chance. The required information is usually provided in the form of the following classification (or Confusion) matrix⁹ as illustrated in Table 3

With respect to (i), we must analyse the diagonal elements of the holdout sample matrix¹⁰ to determine how many cases are being correctly classified, i.e. 95.3% of the 'goods' and 33.8% of the 'bads'. Alternatively, the model is classifying 4.7% of the actual goods as

predicted 'bads' and 66.2% of the actual bads as 'goods'. In terms of costs to the card issuer, the card issuer must decide if such costs of misclassification are acceptable¹¹, that is, what are the costs associated with rejecting nearly 5% of all 'good' applicants and accepting 66% of all 'bad' applicants

To help answer the question of acceptability (ii), the card issuer should compare the predictions of the model with the chance model. However, two criteria might be considered for calculating the percentage correctly classified by chance (Morrison, 1969; Crook *et al* , 1992).

(a) *Maximum chance criterion*

$$C_{\max} = \max(p, 1-p)$$

where

p is the proportion of individuals in group 1

$(1-p)$ is the proportion of individuals in group 2

That is, place all the cases in the group with the greatest number of cases and in doing so maximise the percentage correctly classified by chance. For example, using the figures from Table 3, the percentage correctly classified by chance equals 86.32% giving the impression that the model is doing little better than the chance model. This, however, might not be the most appropriate criterion as the chance model is simply classifying every case as 'good'.

If the main objective of the scorecard is to maximise the percentage correctly classified, regardless of group membership and the costs of misclassification, then the appropriate chance criterion is $C_{\max}$. That is, if the discriminant function does not perform better than chance, then the card issuer should place all cases (including new applicants) in the group with the greatest membership.

Table 3: Classification of Results

Actual Group	Analysis Sample			Holdout Sample		
	No of Cases	Predicted Group		No of Cases	Predicted Group	
		Goods	Bads		Goods	Bads
Goods	14,728	13,389 (90.9)	1,339 (9.1)	8,632	8,224 (95.3)	408 (4.7)
Bads	2,371	1,553 (65.5)	818 (34.5)	1,368	906 (66.2)	462 (33.8)
Percentage correctly classified:		86.62%				86.86%
Cprop		76.0%				76.0%

(b) Proportional chance criterion

$$C_{prop} = p^2 + (1-p)^2$$

When the objective is to maximise the percentage correctly classified into both groups (and you have unequal sized groups) as in this case, then the percentage correctly classified by the model (87%) should be compared with the proportional chance criterion (76%). Using this criterion, the model is improving on the chance model by nearly 11 percentage points out of a maximum possible improvement of only 24 percentage points.

The model, if acceptable, could now be used to credit score new applicants. This involves using the new applicant's application form information and the derived discriminant function coefficients (b's) to derive a discriminant score for the new applicant, and

if $Z_i > Z_{crit}$ accept the application

if $Z_i < Z_{crit}$ reject the application

Notes

- 1 For a fuller discussion of this debate see Chandler and Coffman, 1979
- 2 Practitioners must also include an analysis of previously rejected applicants (Guide to Credit Scoring, 1993), otherwise any scorecard constructed solely on accepted applicants could be biased. The technique used to try to infer the true credit status of rejected applicants is known as reject inference. For further details about the techniques used, see Hand and Henley, 1993
- 3 The definitions of 'good' and 'bad' are very arbitrary. For example, a card issuer may wish to classify someone who has missed only one month minimum repayment as a 'bad'
- 4 In general, card issuers will use additional relevant information where applicable, for example credit reference agencies
- 5 In situations where only a relatively small sample is available an alternative validation procedure, the 'jackknife', may be used. This involves leaving out one of the cases in turn and deriving the discriminant function on $n-1$ cases and predicting group membership for the left-out case (SPSSX Advanced Statistics Guide)
- 6 This article examines only Stages I and II.
- 7 Alternatively, other combinations of g_i , b_i , G_t and B_t may be used. See Boyle *et al* , 1992
- 8 If two or more discriminating variables are highly correlated, only one of the variables should remain in the function otherwise the variances of the b_j 's will be unnecessarily large (Morrison, 1969). Additionally, one would get a false impression of the discriminating power of any such variables as any discrimination will be shared between the two (or more) variables.

- 9 Using the discriminant score SPSSX Discriminant (SPSSX, 1988) classifies each case using the Bayes' rule. The probability that a case with a discriminant score of D belongs to group i is estimated by

$$P(G_i/D) = \frac{P(D/G_i)P(G_i)}{\sum_{i=1}^g P(D/G_i)P(G_i)}$$

- 10 The classification matrix for the analysis sample is usually provided for comparison purposes only
- 11 The card issuer should also consider the 'interests of consumers' when considering the costs of misclassification

Robert Hamilton is a lecturer at the Business School, Loughborough. He thanks colleagues in the Business School for their most helpful comments regarding the article. Any errors, however, remain his responsibility.

References

- Boyle, M., Crook, J.N., Hamilton, R. and Thomas, L. C., (1992), "Methods of Credit Scoring Applied to Slow Payers", in Crook, J N , Edelman, D. and Thomas, L.C. (eds), *Credit Scoring and Credit Control*, Oxford: Oxford University Press.
- Chandler, G C. and Coffman, J.Y , (1979), "A Comparative Analysis of Empirical vs Judgmental Credit Evaluation", *Journal of Retail Banking*, Vol. 1, No 2, September
- Credit Industry, 'Guide to Credit Scoring, 1993', Credit Industry Working Party on Credit Scoring c/o FLA, JBL Printers Limited.
- Crook, J N , Hamilton, R. and Thomas, L.C., (1992), "A Comparison of Discriminators Under Alternative Definitions of Default", in Crook, J.N., Edelman, D. and Thomas, L.C. (eds.), *Credit Scoring and Credit Control*, Oxford: Oxford University Press.
- Eisenbeis, R A., (1978), "Problems in Applying Discriminant Analysis in Credit Scoring Models", *Journal of Banking and Finance*, Vol. 2.
- Frank, R.E., Massy, W F. and Morrison, D.G., (1965), "Bias in Multiple Discriminant Analysis", *Journal of Marketing Research*, Vol. II, pp. 250-258.
- Hair, J F., Anderson, R.E and Tatham, R L , (1987), *Multivariate Data Analysis*, New York: Collier MacMillan
- Hand, D J. and Henley, W.E , (1993), "Can Reject Inference Ever Work?" Paper presented at the Credit Scoring and Credit Control III Conference, University of Edinburgh, 8-10 September
- Klecka, W.R., (1980), *Discriminant Analysis*, Sage.

Lewis, E M., (1992), "Credit Scoring and Credit Control from Four Points of View", in Crook, J N , Edelman, D and Thomas, L.C. (eds.), *Credit Scoring and Credit Control*, Oxford: Oxford University Press.

Morrison, D G , (1969), "On the Interpretation of Discriminant Analysis", *Journal of Marketing Research*, Vol 4.

Reichert, A K , Cho, C.C and Wagner, G.M., (1983), "An Examination of the Conceptual Issues Involved in Developing Credit Scoring Models", *Journal of Business and Economic Statistics*, Vol. 1, No. 2, pp 101-104

SPSSX, Advanced Statistics Guide, 2nd edition, 1988.

APPENDIX B

A PRACTICAL APPROACH TO MAXIMIZING CUSTOMER RETENTION IN THE CREDIT CARD INDUSTRY

Robert Hamilton and J. Barry Howcroft

(Business School, Loughborough University)

Journal of Marketing Management (1995), **11**(1-3), pp. 151-163

A Practical Approach to Maximising Customer Retention in the Credit Card Industry

Robert Hamilton and J. Barry Howcroft

Abstract

One of the main problems currently facing credit card issuers is the increasing number of credit card holders who are using their cards less often (i.e., attrition) and/or returning their cards (closures). This problem is of particular concern as the total number of credit cards held by consumers is declining by approx 0.6% per month and the number of new applicants is also running at an all time low (less than 1% per month).

Most of the published literature in the broad area of credit cards looks at credit scoring, rather than the need for card issuers to identify and retain a profitable portfolio of credit card customers. The overall objective of this paper, therefore, is to construct a customer database model with the capacity to predict which customers are most likely to close their accounts and to identify certain customer characteristics which can be used by the card issuer as part of a marketing or relationship strategy to maximise retention and increase customer profitability.

The database model is constructed using linear discriminant analysis which is applied to a sample of approximately 17,000 UK bank credit card holders using various behavioural and socio-demographic variables and tested on a holdout sample of 10,000 cases.

Introduction

In the 1980's the real value of consumer debt, excluding finance for house purchases, increased by 122 per cent in the UK (Crook, *et al* 1992a). At these rates of market growth it was not surprising that research and academic literature focussed on evolving market structures (Worthington, 1990) and the changing patterns of competitive and consumer behaviour (Hirschman and Goldstucker, 1978; Bowers and Crosby, 1979; Hawes, 1987). Predictive models were also developed which concentrated on the use of statistical techniques which could either: distinguish between defaulters or non-defaulters (Myers and Forgy 1963; Wiginton 1980; Boyle, *et al* 1992), or determine the likelihood of customers who miss a given number of consecutive payments (Bierman and Hausman 1970, Chandler and Coffman 1983, 1984, Crook, *et al* 1992a)

In the aftermath of the economic recession of the early 1990s, the credit card industry is no longer growing at the rates typical of the previous decade. The total number of credit cards held by consumers is declining at a rate of approximately 0.6 per cent per month and the number of new applicants is also running at an all time low of less than 1 per cent per month.¹ The changing dynamics of the credit card industry are also illustrated by the fact that at its peak in 1990 Visa and Mastercard had 29.846 million cards in circulation and value of turnover equalled £27,742 million; however, by 1992, even though value of turnover had increased to £31,272 million, the number of cards in circulation had declined to 26.458 million (Annual Abstract of Banking Statistics, 1993). Recent changes in the marketplace have, therefore, been symptomised by an increasing number of credit card holders returning their cards (closures), and by the remainder apparently using their cards more often or for making larger purchases, or both.

The changing behaviour of credit card users suggests that a different approach is required by management which is less concerned with credit scoring and risk and more concerned with the identification and retention of a profitable portfolio of credit card customers (Lundy, 1992). With these considerations in mind, the overall objectives of the paper were determined and can be summarised as being concerned with database marketing, i.e.

managing the bank's or credit card issuer's existing database to maximise customer retention. As such, this paper is concerned with identifying the characteristics of credit card customers who close their accounts, and developing a model which will predict this behaviour. By utilising the existing customer base, such a model could be highly conducive to increasing customer profitability by maximising customer retention. As such, the analysis represents the first tentative steps in identifying appropriate marketing and relationship strategies based upon customer behaviour for reducing closures and encouraging even greater credit card usage from current and potential credit card holders

The Basic Elements of A Retention Strategy

Although the paper places emphasis on the development of a retention information system and the identification of appropriate strategies for maximising customer retention, it is important to recognise that such systems and strategies are only one part (albeit an important part) of a comprehensive approach to maximising retention.

The following four elements developed from Reichheld and Kenny's (1990) work on customer retention constitute the most important components of such an approach.

Senior Management Commitment

Improving customer retention involves sustained investments in both capital and management's time. Capital investment could, for example, include the upgrading of branch facilities, investment in information systems, etc., whereas management's investment in time could be taken up by the investigations necessary to uncover and address the multiple root causes of customer defections.

Senior management's commitment is also critical in establishing a corporate culture which is conducive to maximising customer retention. In this respect, the views and opinions of

senior management have got to be communicated within and throughout the organisation in such a way that they penetrate the attitudes and habits of all members of staff, thereby determining their business ethos. Much will depend upon the cultural assumptions already established, but if the assumptions already support customer retention the message will be effectively communicated and reinforce existing practices (Long, 1988)

Customer Focused Culture

Improvement seems to come when the value of developing customer relationships is clearly understood and when all employees focus their full attention on this objective. Customer retention based on enhancing relationships with customers is highly conducive to better customer service (Barlow, 1992) and improving bank revenue (Perrien *et al.*, 1993). As it is generally accepted that it is less expensive to market to existing rather than to new customers, a strong *prima facie* case can be made for banks and credit card issuers adopting a strategy which places emphasis on relationships which increase the sale of financial products to existing customers (Axon, 1992; Deutsch, 1992). This approach would also appear to be conducive to long-term market survival (Barrell, 1992), increased market share (Berry, 1983; Kotler, 1992) and increased profitability (Morgan and Chadha, 1993)

Front-Line Actions

Improving retention requires that front-line employees, i.e. those who have daily customer contact, have the power to take actions which provide immediate customer satisfaction and thereby reinforce customer retention. This necessitates that they also have the means in the form of appropriate information technology to access and interpret data as a sound basis for any such actions.

In an endeavour to improve service and maximise customer retention by focusing on good relationships with customers, emphasis should be placed on both internal and external considerations, i.e. on both employees and customers. This necessitates actively

managing the interactions between customers and staff and instigating improvements to the external quality of service by increasing the levels of internal service which staff receive from within the organisation from support departments and technology. The implicit assumption underlying this approach is that by satisfying the needs and wants of its own front-line staff, an organisation can better satisfy the needs of its customers. Available empirical evidence would seem to suggest that companies which promote the welfare of their customers and staff experience higher retention rates of both compared to companies which do not (Hunt *et al.*, 1985; Schneider and Brown, 1985). Similarly, there are grounds to believe that a strong relationship does exist between quality customer service, employee orientation and corporate success (Deal and Kennedy, 1982; Davis, 1985; Bank, 1988).

In addition to improving the quality and level of internal support for front-line staff within the organisation, emphasis should also be placed on continuous training and practice development. In this respect, it is critical that methods and systems for identifying and tracking good practice, especially those which affect the staff-customer or organisation-customer interchange, are introduced and disseminated throughout the organisation. In order to encourage and reinforce the introduction of these practices, incentive systems which reward staff on their ability to retain customers will be critical in sustaining the net growth of business based on a balance between acquiring new and retaining existing customers.

Retention Information Systems

Card issuers and banks already use their large databases in an attempt to strengthen relationships by sending out details of financial products to existing customers (Copulsky and Wolf, 1990), but the real issue is how to determine which customers would respond to such initiatives (Coogler, 1990). Irrespective of whether customers who respond to such approaches do so either because they are using the quality of the relationship with the financial institution as a surrogate for the quality of the product or simply because they want to reduce the search-buy costs associated with a purchase, there is a *prima facie* case for

attempting to identify and target those customers who are most likely to respond positively. As a consequence, there is a need to develop new and sophisticated methods of tracking and analysing the root causes of customer defection and using this information to strengthen customer relationships and thereby maximise customer retention.

These sorts of considerations are the essential cornerstones of a strategy aimed at closing a widening gap between competing financial institutions based on the differential capacity to improve customer retention. Those organisations which both manage and provide the means and incentives for their staff to bring about the greatest improvement in retention will undoubtedly establish themselves as both growth and profit leaders.

Whilst recognising the importance of all the key elements of a customer retention strategy, as stated earlier, this paper concentrates on just part of such a strategy, namely the development of a retention information system with the capacity to predict which customers are most likely to close their accounts. The retention information system is also conducive to the identification of characteristics which are symptomatic of those customers who are most likely to close their accounts, and this fact allows general conclusions to be drawn about how a card issuer could strengthen relationships with existing customers in an attempt to maximise customer retention.

Methodology

The data related to a 15-month period from 1 January 1992 - 31 March 1993 and consisted of 27,099 individuals who held a credit card as at 1 January 1992². The size of the data base meant that it was possible to create a holdout sample randomly, which was representative of the original sample, consisting of 10,000 individuals (approximately 37 per cent of the initial data), and, therefore, sufficiently large enough to insure stability of the coefficients³ (Klecka 1980).

As the primary objective of the research was to develop a behavioural model with the predictive ability to identify those customers most likely to close their credit card accounts, it

was important to establish an exact definition of the term "closed". A number of alternative meanings, could, however, be attached to the term and so it was decided to adopt a definition which reflected the behaviour of credit card customers rather than the credit card issuers. As a consequence, "closed" within the context of this paper only refers to those specific instances where credit cards are returned to the bank (for whatever reason) by customers of their own free volition. All other categories of "external status"⁴ are referred to as "normal", and this includes instances where, for example, the credit card has become non-operationable either because the customer has become bankrupt, lost the card, had it stolen or revoked by the bank.

The data originally contained over 70 variables, but eventually 22 predictor variables were identified (see Appendix 1) which tended to reflect the behaviour patterns of credit card customers, although some socio-demographic variables have also been used where on a *priori* grounds it was thought they had a discriminative effect on "closures".

As a number of variables were measured at nominal level, whereas the use of linear discriminant analysis requires that all predictor variables are measured at least at interval level (Klecka, 1980), the methodology used follows that of Crook *et al.* (1992b). That is, the required interval level data was derived using the following formula:

$$X'_j = \ln (n_i / c_i) + \ln (C_T / N_T);$$

where X'_j = value of the predictor variable X for case j ;

n_i = number of normal credit card accounts in nominal category i ;
the category of which j was a member;

c_i = number of closed credit card accounts in nominal category i ;
the category of which j was a member;

N_T = total number of normal credit card accounts in the sample;

C_T = total number of closed credit card accounts in the sample.

By using the log values in the way described above, a linear relationship between the function and group variables was established, thereby facilitating the application of linear discriminant analysis in developing a predictive model of "closures".

An important step in constructing the predictive model was to identify *a priori* those variables which are potentially the best at discriminating between those accounts which will close and those which will continue to operate normally. In selecting these variables it was essential to establish whether multicollinearity exists between the various predictor variables and to determine which of these variables should be omitted from the function. Unless this precaution is taken there could be a high degree of correlation between the variables in the function, which would reduce the reliability of the standardised coefficients as indicators of the relative importance of each predictor variable (Chandler and Coffman, 1983, 1984).

To test for the existence of multicollinearity, each predictor variable was linearly regressed on all other predictors and the tolerance ($1 - R^2_i$) was calculated for each variable. Those variables with a tolerance of ≤ 0.79 (Crook, *et al.*, 1992b) were considered for deletion. Next, having identified the existence of multicollinearity, in order to determine which variables to remove, i.e. which pair(s) of variables were highly correlated, the values of both the regression coefficients and the Pearson correlation matrix were examined. In the latter case a value of ≥ 0.2 was taken as an indication of multicollinearity.

Having applied this methodology, the number of predictor variables left in the analysis with a tolerance value ≥ 0.8 was reduced from 22 to 15. The seven variables which were rejected included: account prefix (i.e., whether the customer has a Mastercard or Visa etc.); how long the card had been active; date when account was opened, credit card limit, number of cash advances; number of purchases; and amount of purchases.

While the remaining 15 variables may intuitively be good discriminators a stepwise procedure had been adopted to ensure that all weak redundant variables were removed.

from the final discriminant function. The criterion for variable selection was the Mahalanobis Distance (D^2) where at each step the variable that maximises the Mahalanobis distance⁵ is selected (SPSSX User's Guide), subject to the F to enter value being at least equal to 1 (note the F to remove value was also set equal to 1).

In addition to using the classification matrix and the percentage correctly classified by the function to assess the predictive accuracy of the discriminant function, the results were also compared with the percentage correctly classified by chance. This may be calculated (Hair, *et al.*, 1987) using either the maximum chance criterion⁶ (this is used when the objective is to maximise the percentage correctly classified, regardless of group membership) or the proportional chance criterion (C_{prop}):

$$C_{prop} = p^2 + (1 - p)^2$$

where p = proportion of cases in group 1,

$(1 - p)$ = proportion of cases in group 2.

As this latter criterion is most suited and should be used when the objective is to correctly classify membership of two or more unequal groups (e.g. "closed" or "normal"), we shall be comparing the percentage correctly classified by the function with the C_{prop} .

Results

The statistical significance of the estimated function is shown in Table 1. Wilks' Lambda indicates the ability of predictor variables to discriminate among the groups beyond the discrimination achieved by the earlier function, i.e. residual discrimination (Klecka, 1980). As lambda decreases in value, it is indicating progressively greater discrimination. The

significance of the function is tested by χ^2 and, as Table 1 shows, the means for both "closed" and "normal" accounts are statistically different.

TABLE 1
Residual Discrimination and Test of Significance

<i>Wilks' Lambda</i>	χ^2	<i>d.f.</i>	<i>Significance</i>
0.8055860	3694.5	15	0.0000

The results of the model incorporating the remaining predictor variables are shown in Table 2. This indicates that the proportion of grouped cases correctly classified by the model was 86.62 per cent for the analysis sample⁷ and 86.86 per cent for the holdout sample. Viewed in a slightly different way, the model was correctly predicting 90.9 per cent of the normal accounts and 34.5 per cent of the closed accounts for the analysis sample, and 95.3 per cent of the normal accounts and 33.8 per cent of the closed accounts for the holdout sample.

TABLE 2
Classification of results (brackets denote percentages)

Actual group	No. of cases	Analysis sample Predicted group		No. of cases	Holdout Sample Predicted group	
		Normal	Closed		Normal	Closed
Normal	14,728	13,389 (90.9)	1,339 (9.1)	8,632	8,224 (95.3)	408 (4.7)
Closed	2,371	1,553 (65.5)	818 (34.5)	1,368	906 (66.2)	462 (33.8)
Percentage correctly classified		86.62%			86.86%	
C_{prop}		76.0%			76.0%	

In assessing the behavioural model's efficacy, comparisons with C_{prop} indicate that the results are much better than those which would have been correctly classified by chance: the model correctly classifies almost 87 per cent of the accounts, which is substantially greater than the 76 per cent expected by chance. Argued slightly differently, this means that the model is correctly classifying almost 11 percentage points above chance out of a possible total of 24. From the card issuer's perspective they have a model which can correctly identify some 34 per cent of customers who are likely to close their account. The costs of misclassification are also less than with a credit scoring model where the purpose is to identify in advance the likelihood of bad as opposed to good customers. Misclassification with the latter model may well incur substantial costs and, therefore, lead to a reduction in profitability, whereas with attrition and closures the associated costs are relatively minimal, being typically related to the non-response of customers to direct mail shots.

Turning now to the relative importance of each predictor variable in terms of their discriminatory power, Table 3 shows the structure coefficients for each variable included in the estimated function. The standardised coefficients are not shown because they represent the relative discriminatory power of each predictor variable given the other variables in the function. As such, they can give an inaccurate indication of the discriminatory power of each variable if there is a degree of correlation between any variables included in the function. For this reason, only the within-groups correlations are shown in Table 3, because as simple bivariate correlations, they are not affected by other variables in the function and are in some respects a better guide (Klecka, 1980).

TABLE 3

<i>Within groups structure coefficients</i>		
<i>Variables</i>	<i>Within-groups</i>	<i>Rank</i>
BEHSCORE	0.77400	1
TOTALINT	0.41304	2
PREVEXT	0.37082	3
TYPCHAN	0.32099	4
NPLASTIC	0.17659	5
ACCTYP	0.16895	6
AMCASHPM	0.15486	7
SORTCODE	0.14332	8
INSTAT	0.11158	9
AGE	0.10373	10
DIRECTDI	0.04782	11
COCODE	0.03743	12
SEX	0.00706	13
AFF	0.00229	14
CREDITLF	0.00027	15

Using this measure, the top four variables⁸ are. (1) BEHSCORE; (2) TOTALINT; (3) PREVEXT; (4) TYPCHAN. The other variables, all of which added significantly to the discriminatory power of the function (at $F=1.0$), have noticeably lower values, which indicates that they contribute much less to the canonical discriminant function. This is particularly true for DIRECTDI; COCODE, SEX; AFF; CREDITLF, all of which have a structure coefficient less than 0.05.⁹

In interpreting the results, emphasis has been placed on the ten most powerful discriminatory variables as indicated by the structure coefficients. It is important to note, however, that we are examining the ability of values $X'_j = \ln(n_i / c_j) + \ln(C_T / N_T)$ to distinguish between "normal" and "closed". We must, therefore, consider the relationships which exist between values for X'_j and X_j for each of the variables.

The BEHSCORE categories reveal that credit card customers who have had a dormant account for longer than 12 months are most likely to close their accounts. Conversely, a BEHSCORE category indicating that an account is at least five cycles delinquent has the most important discriminatory effect on whether the account will operate normally¹⁰. Having regard to the definition of "closed" in the paper, the latter customers are typical of those who will be closely controlled by the issuer in an attempt to reduce the arrears and bring the account under control. In this sense, therefore, those customers are arguably not in a position to "close" their accounts and, in fact, run the distinct risk of having their accounts revoked by the issuer.

The categories relating to TOTALINT showed that those customers with no monthly outstanding interest were the most inclined to close their accounts. As outstanding monthly interest increased, however, there was a greater tendency to operate the account normally. This seems to add weight to the idea that whoever controls the account has an important influence on whether the account is operated "normally" or "closed". If the customer is in control in terms of regularly paying interest (and principal), he at least places himself in a position to close the account. This is in direct contrast to a customer who is in arrears of either interest or principal, when the position is more likely to be controlled by the card issuer.

The various categories of PREVEXT indicate that under circumstances where the credit card has been lost or stolen, the card is not likely to be returned to the issuer. Where the account operates normally, however, or where it has been revoked or interest accrued prohibited, etc., the account is more likely to be closed. This appears to follow the broad conclusions which were drawn from BEHSCORE and TOTALINT, as the exertion of some form of control over the credit card account appears to determine, at least to some extent, whether the account will operate normally or not. By identifying the key characteristics of the credit card product, a distinct possibility arises to influence customer behaviour and, therefore, increase or decrease a customer's propensity to use the product.

The importance of control is also borne out by TYPCHAN. Where the credit limit is changed either automatically by the issuer or upon the instigation of the customer the account is more likely to operate normally. However, where an increase in the credit limit has been permanently deferred the account is more likely to be closed.

The remaining categories of NPLASTIC indicated that customers with one card were more inclined to close their accounts compared to customers with two cards, a conclusion which was also supported by an examination of ACCTYP. This indicated that customers who had a combination of credit cards, i.e. both VISA and MASTERCARD, were more inclined to operate the account normally compared to customers who had sole credit card accounts. Whether this reflects greater need or the greater sophistication of the former customers is difficult to say, but, when AMCASHPM was examined in closer detail, certainly the customers who had the largest monthly amounts of cash posted to their accounts had a tendency to operate normally, whereas customers with no cash posted were inclined to close their accounts.

SORTCODE was interesting too in the sense that customers who held a banking account with the card issuer were less inclined to close their credit card accounts compared to customers who banked elsewhere. This at least provides tentative evidence that established relationships with a financial institution reinforce the control element and possibly might reduce the likelihood of customers closing their credit card accounts

INSTAT categories revealed that customers who were "normal" or had a credit balance on their accounts were more inclined to close these accounts than customers who were at least one cycle delinquent, over the limit, or both. These points were also borne out by the final predictor variable AGE, which revealed that younger customers, under the age of 40 years, were more inclined to close their accounts. From about the age of 40-60 years, the accounts tended to operate normally, after which time the inclination to close increased.

An increase in mortality rates or a reduction in expenditure after retirement and, therefore, a reduction in the need for credit, possibly explains the behaviour of the 60 years+ age group.

At the other extreme, however, there may well be a very real need for credit, and, therefore, the issue of who controls the account and how this control is used arises once again. In the middle age ranges, 40-60 years, control may be exercised more by the customer rather than the issuer. The behaviour of the customer, however, may also be more heavily influenced by the length and nature of the relationship with the card issuer.

Conclusion

Using linear discriminant analysis, the customer base model was able to correctly classify 95% of customers who operated their card account normally, in the time period examined, and almost 35% of those who closed their account. Discussions with representatives of various card issuing organisations suggests similarities between the performance of their models and our results.

The analysis of the categories relating to the important predictor variables suggests that the key determinants of whether an account will operate "normally" or be "closed" are:

- customer need;
- how the account is controlled; and closely related to this;
- the relationship which the card holder has with the issuer.

The identification of these key determinants of customer behaviour and account activity have a number of important implications for management. In the first instance, they strongly suggest that management should be proactive in attempting to determine and influence customer need, or, at the very least, attempt to match more closely, appropriate financial products with the right "sort of customer". In other words, if the predictive model suggests that a particular customer is likely to close an account, management should be asking itself why, and, in the process, attempting to identify a more appropriate product which will encourage usage and retain business.

Retention information systems which utilise existing customer databases will, therefore, be critical in providing management with detailed information on the needs and behaviour patterns of customers which can be used to target identifiable customer segments with specific products. The same information can also be utilised to identify the essential cornerstones of an appropriate relationship strategy aimed at reinforcing customer loyalty with the organisation based on existing customer behaviour and perceived need. As such, the analysis represents the first tentative step in identifying appropriate strategies based upon customer behaviour, for reducing closures and increasing profitability. In order to maximise the effectiveness of these strategies, however, it is important to target specific customer groupings by segmenting the customer portfolio.

On a less positive note, the research has highlighted certain weaknesses of this type of approach. First, the canonical discriminate function¹¹ is explaining only 20% of the variance in the dependent variable, and this suggests additional predictor variables need to be considered, for example current account activity, the cost of this type of credit, etc. Second, discriminant analysis is an *a priori* segmentation method, and, as such, may be unable to differentiate between groups effectively. For instance, if we were to further divide credit card users into "high profit" and "low profit" segments, the variability within the groups could still remain high. For example, the "low profit" groups (i.e. for both "normal" and "closed") could contain both "timids" who never or rarely use their cards and "spenders" who use their cards regularly, but avoid paying any interest. In particular "timids" represent an interesting example because they do have a value to the card issuer in so much as they: at some point in time responded to an offer; have an established relationship with the bank, are familiar with the current bank card institution; use the account without prompting on an infrequent basis; respond to internal promotions and solicitations more readily than new customers; can be upgraded or downgraded, cross-sold other bank products and re-issued plastic without direct permission from the customer.

These considerations, therefore, suggest that significant advantages can be exploited by clever marketing organisations utilising knowledge based on customer behaviour. In an endeavour to introduce the necessary differentiation, an alternative segmentation model

(e.g. cluster-based model) should be used in any subsequent research. Indeed, in the extension of this study the aim will be to examine the impact of including a weighted "dependent variable", like profitability, in the clustering process.

NOTES

1. Based on information provided by the card issuer sponsoring this research
2. The majority of customers who closed their accounts in this period did so after June 1992
3. For a discussion of the predictive performance of our estimated model see:
Eisenbeis, R. A , (1977), "Pitfalls in the Application of Discriminant Analysis in Business Finance and Economics", *Journal of Finance*, Vol 32, No 30, June,
Kschirsagar, A. M., (1972), "Multivariate Analysis", Marcel Dekker Inc , Vol 2, and
Lachenbruch, P. A. and Mickey, M R , (1968), "Estimation of Error Rates in Discriminant Analysis", *Technometrics*, Vol. 10, No 1, February.
4. The dependent variable "external status" has a variety of categories (e.g. normal, authorisation prohibited, bankrupt, closed, revoked, frozen, interest accrual prohibited, lost, stolen and charged off). For the purposes of this paper, however, all circumstances have been categorised as "normal" unless the customer has returned the card to the issuer of his own free volition when it is categorised "closed"
5. The distance between groups a and b is defined as:

$$Dab^2 = (n - g) \sum_{i=1}^g \sum_{j=1}^g w_{ij}^* (\bar{X}_{ia} - \bar{X}_{ib}) (\bar{X}_{ja} - \bar{X}_{jb})$$

where g is the number of variables in the model, $\bar{X}_{ia}$ is the mean for the i th variable group a , and w_{ij}^* is an element from the inverse of the within-groups covariance matrix.

6 The Maximum Chance Criterion:

$$C_{\max} = \text{MAX} (p, 1 - p)$$

where p is the proportion of cases in one of the groups, e g "normal" That is, if over half of the cases were "normal", the greatest proportion correctly classified by chance would be obtained by placing every one in the "normal" category.

7. One would expect an upward bias with this classification (Hair, *et al* , 1987).
8. The same was true using the F to remove criterion and the standardised coefficients.
9. Consequently these variables have been excluded from the interpretation of the results
10. A customer who is five cycles delinquent will not be regarded as "normal" by the card issuer but as "delinquent" as indicated by the customer's internal status.
11. The canonical correlation equals 0.4409241

APPENDIX 1

Twenty Two Original Variables

Variable	Description
SEX	Male or female.
COCODE	Great Britain or others.
AGE	Age in years.
DIRECTDI	Whether charges are paid by direct debit.
AFF	Whether the annual charge fee is to be waived.
CREDITLF	Whether customer is in the cardholder repayment protector scheme.
NPLASTIC	Number of credit cards held by customer.
INSTAT	Whether customer is delinquent* or over the limit on credit balance or normal
PREVEXT	Relates to customer's previous [†] "external status" and indicates whether the account operated normally, whether the card was returned by customer or whether it was stolen or lost, etc.
ACCPRE	Whether card is Mastercard, Visa, etc

ACCTYPE	Whether card holder has combinations of different cards.
SORTCODE	Where card holder has primary bank account
ACTIVEYY	How long the card has been active
LACCOPEN	How long the account has been open
CREDITLM	Credit limit.
BEHSCORE	Score based on customer's behaviour in operating the account.
TYPCHAN	Circumstances of last credit limit change
AMCASHPM	Amount of cash posted in previous year (1992) - monthly average.
NOCASHAD	Number of cash advances in previous year (1992) - monthly average.
NOPURPM	Number of purchases in previous year (1992) - monthly average
AMPURPM	Amount of purchases in previous year (1992) - monthly average.
TOTALINT	Total interest and service charge in previous year (1992) - monthly average

* Delinquency means 1 cycle default

† "Previous" in this context means where, for example, the customer closed the account and then re-opened it, or where the card issuer suspended the account and later re-opened it, or where a marital break-up resulted in a joint account becoming two separate accounts

References

- Annual Abstract of Banking Statistics, (1993), *British Bankers Association*, Vol. 10, May.
- Axon, D. J , (1992), "A Return to Managing Customer Relationships", *International Journal of Bank Marketing*, Vol 10, No 7, pp. 17-25
- Bank, W., (1988), "Building Competitive Advantage Through Customer Service", *Marketing News*, pp. 16.
- Barlow, R. G., (1992), "Relationship Marketing - The Ultimate in Customer Service", *Retail Control*, Vol 60, No 3, pp. 29-37.
- Barrell, A., (1992), "Relationship Marketing: Way Ahead for the 90s?", *Business Marketing Digest*, Vol 17, No 3, pp. 49-54.
- Berry, L. L., (1983), "Relationship Marketing" in: *Emerging Perspectives on Services Marketing* (Berry, L. L., Shostack, G. L. and Upah, G. D , (Eds)), Chicago: American Marketing Association, pp. 25-28.
- Bierman, H and Hausman, W. H., (1970), "The Credit Granting Decision", *Management Science*, Vol. B16, pp. 519-32.
- Bowers. J. S and Crosby K R , (1979), "Credit Card Use by Low-Income Consumers", *The Journal of Consumer Credit*, Vol. 10, Winter, pp. 80-87
- Boyle, M., Crook, J N , Hamilton, R. and Thomas, L. C., (1992), "Methods of Credit Scoring Applied to Slow Payers". In: *Credit Scoring and Credit Control* (Thomas, L. C., Crook, J. N., and Edelman, D , (Eds.)), Oxford. Oxford University Press.

- Chandler, G C and Coffman, J. Y , (1983-84), "Applications of Performance Scoring to Accounts Receivable Management in Consumer Credit", *Journal of Retail Banking*, Vol. 5, pp. 1-10.
- Coogle, J., (1990), "Research: Database Marketing", *Marketing and Media Decisions*, Vol. 25, No 1, pp. 75-76
- Copulsky, J R and Wolf, M J., (1990), "Relationship Marketing for the Future", *Journal of Business Strategy*, Vol. 11, No 4, pp. 16-20.
- Crook, J N., Hamilton, R. and Thomas, L. C., (1992a), " A Comparison of Credit Scoring Model with a Credit Performance Model", *Services Industries Journal*, Vol 12, No 4, October.
- Crook, J. N., Hamilton, R. and Thomas L C , (1992b), "A Comparison of Discriminators Under Alternative Definitions of Default". In *Credit Scoring and Credit Control* (Thomas, L , Crook, J N , and Edelman, D. B , (Eds.)), Oxford: Oxford University Press
- Davis, S I., (1985), "Excellence in Banking", London. Macmillan Press.
- Deal, T. and Kennedy, A , (1982), "Corporate Cultures", Reading: Addison-Wesley.
- Deutsch, B. I., (1992). "Supplier Relationships That Make the Difference", *Bank Marketing*, Vol. 24, No 7, pp. 20-23
- Frank, R E., Massey, W. F. and Boyd, H. (1968), "The Demographic Segmentation of Household Products". In: *Application of Science in Marketing Management*, (Bass, W. F , King, C. W. and Pessemier, E. M., (Eds.)), New York: Wiley

- Hair, J. F., Anderson, R. E. and Tatham, R. L., (1987), "Multivariate Data Analysis", New York: Collier Macmillan.
- Hawes, D. K., (1987), "Profiling Visa and Mastercard Holders", *Journal of the Academy of Marketing Science*, Vol. 15, No. 1, Spring, pp. 62-69
- Hirschman, E. C. and Goldstucker, J., (1978), "Bank Credit Card Usage in Department Stores", *Journal of Retailing*, Vol. 54, Summer, pp. 3-9
- Hunt, S. D., Chanko, L. B. and Wood, Van R., (1985), "Organisational Commitment and Marketing", *Journal of Marketing*, Winter, pp. 112-126.
- Kotler, P., (1992), "Marketing's New Paradigm: What's Really Happening Out There", *Planning Review*, Vol. 20, No. 5, Sept-Oct, pp. 50-52.
- Klecka, W. R., (1980), *Discriminant Analysis*, Oxford: Sage
- Long, R. H., (1988), "High Performance Bank Culture", *Journal of Retail Banking*, Vol. 10, No. 3, pp. 17-25
- Lundy, M., (1992) "Cluster Analysis" In: *Credit Scoring and Credit Control*, (Thomas, L. C., Crook, J. N., and Edelman, D., (Eds.)), Oxford: Oxford University Press, pp. 91-107.
- Morgan, R. E. and Chadha, S., (1993), "Relationship Marketing at the Service Encounter: The Case of Life Insurance", *Service Industries Journal*, Vol. 13, No. 1, pp. 112-125
- Myers, J. H. and Forgy, E. W., (1963), "The Development of Numerical Credit Evaluation Systems", *American Statistical Association Journal*, Vol. 58, pp. 799-806.

Perrien, J , Filiatrault, P., and Ricard, L , (1992), "Relationship Marketing and Commercial Banking: A Critical Analysis", *International Journal of Bank Marketing*, Vol. 10, No. 7, pp. 25-29.

Reichheld, F. F., and Kenny, D. W , (1990), "The Hidden Advantages of Customer Retention", *Journal of Retail Banking*, Vol. 12, No. 4, pp. 19-23.

SPSSX Inc , (1988), *SPSS-X User's Guide*, 3rd Edition.

Schneider, B. and Brown, D. E., (1985), "Employee and Customer Perceptions of Service in Banks: Replications and Extensions", *Journal of Applied Psychology*, August, pp 423-433.

Wiginton, J. C., (1980), "A Note on the Comparison of Logit and Discriminant Models of Consumer Credit Behaviour", *Journal of Financial and Quantitative Analysis*, Vol. 15, pp. 757-70.

Worthington, S., (1990), "Retailer Credit Cards: A Competitive Threat", *International Journal of Bank Marketing*, Vol. 8, No 4, pp. 3-9.

FURTHER READING

Crook, J. N., Hamilton, R and Thomas, L. C., (1992c), "The Degradation of the Score Card Over the Business Cycle", *IMA Journal of Mathematics Applied in Business and Industry*, Vol 4, No 1, pp 111-123.

Morall, K., (1992), "Managing the Deluge of Data", *Credit Card Management*, March, pp 11-14

SPSSX Inc, (1988), *SPSS-X Advanced Statistics Guide*, 2nd Edition.

