Credit Scoring, Response Modelling and Insurance Rating

A Practical Guide to Forecasting Consumer Behaviour

Steven Finlay

First published 2010 by
PALGRAVE MACMILLAN

Palgrave Macmillan in the UK is an imprint of Macmillan Publishers Limited, registered in England, company number 785998, of Houndmills, Basingstoke, Hampshire RG21 6XS.

Palgrave Macmillan in the US is a division of St Martin's Press LLC, 175 Fifth Avenue, New York, NY 10010.

Palgrave Macmillan is the global academic imprint of the above companies and has companies and representatives throughout the world.

Palgrave® and Macmillan® are registered trademarks in the United States, the United Kingdom, Europe and other countries

ISBN 978-0-230-57704-6 hardback

This book is printed on paper suitable for recycling and made from fully managed and sustained forest sources. Logging, pulping and manufacturing processes are expected to conform to the environmental regulations of the country of origin.

A catalogue record for this book is available from the British Library.

Library of Congress Cataloging-in-Publication Data

Finlay, Steven, 1969–
Credit scoring, response modelling and insurance rating : a practical guide to forecasting consumer behaviour / Steven Finlay.
p. cm.
Includes bibliographical references.
ISBN 978-0-230-57704-6 (alk. paper)
1. Credit analysis. 2. Consumer behavior–Forecasting. 3. Consumer credit. I. Title.

HG3701.F55 2010
658.8'342–dc22 2010033960

10 9 8 7 6 5 4 3 2 1
19 18 17 16 15 14 13 12 11 10

Printed and bound in Great Britain by
CPI Antony Rowe, Chippenham and Eastbourne

Acknowledgements

I am indebted to my wife Sam and my parents Ann and Paul for their support. I would also like to thank Dr Sven Crone, of the Management Science Department at Lancaster University, for his input to some of the research discussed in the book. Also my thanks to Professor David Hand of Imperial College London for bringing to my attention a number of papers that helped form my opinion on a number of matters. I'm also grateful to the Decision Analytics Division of Experian UK (Simon Harben, Dr Ian Glover and Dr John Oxley in particular) for their support for my research over a number of years, as well as Dr Geoff Ellis for providing information about SPSS during a number of interesting discussions about the merits/drawbacks of different statistical packages.

Contents

List of Tables

List of Figures

1
Introduction

As IT systems have evolved, the amount of information financial services organizations maintain about individuals and the speed at which they can process information has increased dramatically. One consequence of this development is a paradigm shift in the way organizations manage their relationships with consumers. At one time, when dealing with large populations, standard practice was to segment people into relatively few homogenous groups, and then apply an identical relationship management strategy to everyone in each group. Take the case of a provider of home insurance wanting to launch a promotional campaign to recruit new customers. The marketing department may have decided, somewhat arbitrarily, that its target audience were white collar families with children. Its promotional strategy would be to send identical mail shots to all households in middle class suburbs where it was believed the majority of its target audience resided. Some mailings would be received by the intended audience, but many would be wasted because they would be sent to singles, the unemployed or those on low incomes who just happened to live in middle class suburbs. Similarly, white collar families living in inner cities or rural communities represented a missed opportunity because they lived outside the target areas.

At the other end of the spectrum, where customer relationships needed to be managed on a case by case basis, deciding how to deal with someone was a laborious, time consuming and expensive process. A classic example is when someone wanted a loan from their bank. The process would begin with them making an appointment to see the manager of their local branch, and they may have had to wait several weeks before an appointment was available. The customer would arrive at the meeting in their best clothes, in order to make a good impression, and the bank manager would question them about their personal and financial circumstances in order to form an opinion about their creditworthiness. If the bank manager felt they were likely to repay what they borrowed and represented a good investment, then they would be granted a loan. However, if the bank manager didn't want to grant the loan then they were under no obligation to do so. Many people had their loan requests declined due to their gender, because they belonged to a minority group, or simply because the bank manager was in a bad mood and didn't fancy granting any loans that day. Even if the loan was granted, the time taken between the initial enquiry and receipt of funds could be considerable.

Within the financial services industry today, most decisions about how to deal with people are taken automatically by computerized decision making systems. These assess each person on a case by case basis using geo-demographic information that is known about them. Human involvement in such decisions is very much the exception rather than the rule. At the heart of these decision making systems lie mathematically derived forecasting models that use information about people and their past behaviour to predict how they are likely to behave in future. Decisions about how to treat people are made on the basis of the predictions generated by the forecasting model(s). These days, the insurer would use a response model to predict the likelihood of someone on their mailing list responding to a promotional mail shot. If the model predicted that the person was likely to respond by requesting an insurance quotation, then they would automatically be mailed with an individually tailored communication – regardless of where they lived or which socio-economic group they belonged to. Likewise, a loan provider will use a credit scoring model (a model of creditworthiness) to predict how likely an applicant is to repay the loan they are applying for. Loans will only be offered to those the model predicts are creditworthy and likely to make a positive contribution towards profits. Those that the credit scoring model predicts are uncreditworthy will be declined.

The application of automated decision making systems brings many benefits. One is that they make better decisions, more quickly and more cheaply than their human counterparts. This allows decisions to be made in real time while the customer is in store, on the phone or on the internet. A second benefit is they allow tailored decisions to be made, based on an individual's propensity to behave in a certain way, instead of vague estimates about the general behaviour of large populations. Another advantage is that they are consistent in their decision making. Given the same information twice, they will always arrive at the same decision. This is something that cannot be guaranteed with human decision makers. Developed correctly, these systems also display no unfair bias in terms of gender, race or any other characteristic deemed undesirable by society. Automated decision making systems also facilitate centralized decision making. This means that changes to an organization's strategy for dealing with customers can be made quickly and easily at the site where the decision making system is controlled. This removes the need to coordinate changes in decision strategy across many different branches/offices/regions.

Not surprisingly, the value of automated decision making systems to the organizations that employ them is substantial. A large financial services organization will make billions of dollars worth of decisions each year, based solely on predictions made by their forecasting models. There is therefore, considerable effort expended to ensure that the forecasting models an organization employs perform in an optimal capacity.

1.1 Scope and content

The goal of this book is to convey an understanding of how forecasting models of consumer behaviour are developed and deployed by major financial services organizations such as banks, building societies (saving and loan companies) and insurers. However, before going any further, I want to make two things clear. First, this is a book about the development and application of forecasting models of consumer behaviour within business environments. It is not a book about forecasting techniques. What's the difference? In the classroom it's not unusual to spend 80–90 percent of the time learning about the mathematical/statistical processes underpinning methods such as logistic regression and discriminant analysis that are used to construct models of consumer behaviour. Rarely is more than 10–20 percent of teaching time (and sometimes none at all) spent discussing wider issues that drive model development and usage in real world environments. Within the commercial sector the opposite is true. On average, well over 80 percent of the effort involved in the successful delivery of a modelling project is concerned with business issues. This includes spending time doing things such as: drawing up a project plan to determine how long the project will take and what resources are required, working out how much the project will cost, deciding what behaviour the model should predict, agreeing where the data to construct the model will come from, complying with audit and legal requirements, producing documentation, deploying the model within the business and determining how the performance of the model, once deployed, will be monitored to ensure it continues to work as intended. Only a few percent of a project's resources will actually be spent working with the statistical techniques that are used to construct the model. A typical bank, for example, will take anywhere between four and 12 months to develop and implement a new suite of credit scoring models to estimate the creditworthiness of people applying for a product such as a credit card. Yet, no more than a few days or weeks will be required for the modelling part.

The second point I want to make is that this is a book about practice. It is not a theoretical text, nor is the intention to provide a comprehensive literature review of the subject. References are made to theory and relevant academic material, but the primary objective is to explain, in simple terms, the steps required to deliver usable, high quality forecasting models within realistic timeframes, based on my personal experience of working in this area of financial services. There are very many complex data analysis/modelling/forecasting techniques and practices (and more are being developed all the time), that in some situations may achieve marginally better results than the methods discussed here, or which are more appropriate from a theoretical perspective. However, the effort required to develop, implement and maintain such solutions means that in practice few organizations employ them – the cost/benefit case just doesn't add up. Where such methods are employed the justification for doing so is often questionable. In some cases the decision to use a given modelling technique is driven by political pressures, originating from people who want to say that they and the organization that employs them are working at the cutting edge, not because there is a good business case for doing so. I have also come across many examples where one method of model construction appears to work much better than another, but on closer inspection the difference is found to be due to errors in the way models have been constructed and/or poor methodology when validating results. When the errors are corrected the differences, more often than not, disappear or are so small as to have no practical significance. Remember – Occam's razor applies – all other things being equal, simple solutions are best.

Given the aforementioned points, there are two groups I have written this book for. The first are those who are not model builders themselves, but who would like to know more about how models of consumer behaviour are constructed and used. Perhaps you work with model builders, or manage them, and feel you should understand more about what they do. For this audience I would like to emphasize that a degree in statistics or mathematics is not a requirement for understanding the material in this book. There are some formulas and equations, but the maths is kept to a minimum, and where it appears I have attempted to explain it in simple terms without assuming that the reader has much prior knowledge. If you can get through this introductory chapter, then you should not have any trouble with any material in subsequent chapters. The second group are those with a more technical background who may be, or about to be, involved in model construction, but who have little practical experience of how models are constructed within business environments. Maybe you are a graduate working in your first job or an experienced academic who wants to know more about the development and application of models within the financial services sector. For readers in this group, I don't claim to offer much that is new in terms of theory, but I do hope to provide some useful guidance about the practical aspects of model development and usage.

With regard to the structure of the book, in the remainder of this chapter the idea of a forecasting model is introduced, which going forward I shall simply refer to as "a model". This covers the type of behaviour that models are used to predict, the different forms of model that can be developed and the main stages that comprise a typical modelling project. Chapters 2 through 8 look at the key processes involved in developing a model, starting with project planning, then moving on to consider sampling, preparing data, data analysis, data pre-processing and modelling. Chapter 9 looks at the problem of sample bias. Sample bias is when behavioural information is missing for some cases, due to previous decisions made about them. This can lead to sub-optimal (biased) models, unless appropriate corrective action is taken. Chapter 10 discusses implementation and how models are monitored, post implementation, to see how well they are performing. The final chapter discusses a number of topics, in particular, small sample validation methodologies and multi-model 'fusion' systems, which generate a forecast of behaviour by combining several different forecasts together.

1.2 Model applications

There are many different behaviours that financial services organizations are interested in forecasting. Table 1.1 summarizes the most common behaviours that models are used to predict.

I do not claim that Table 1.1 provides an exhaustive list of every type of model used within the financial services industry, but it probably covers over 95 percent of the models in use today.

Table 1.1 Model applications

Model	Behaviour the model predicts
Classification models	
Response	The likelihood someone responds to direct marketing activity such as a mail shot. Marketing activity is only targeted at people the model predicts have a high likelihood of responding.
Conversion	The likelihood someone becomes a customer. For example, the likelihood that someone who responded to a mail shot by asking for an insurance quote, subsequently accepts the quote they are given.
Creditworthiness (Credit scoring/ Probability of default)	The likelihood someone will repay money that they owe. Models of this type are used to decide whether to offer credit in the form of personal loans, credit cards, mortgages and motor finance.
Fraud	The likelihood a credit application or an insurance claim is fraudulent. Fraud models are also widely used to predict if a credit card transaction is fraudulent.
Attrition (Retention/ Churn)	The likelihood that a customer defects to a rival or fails to renew a relationship.
Revolver	The likelihood a credit card customer revolves the balance (does not pay the balance in full) on their account each statement period.
Recovery (Collections)	The likelihood someone pays the arrears owing on a credit agreement when they have already missed one or more scheduled repayments.
Insurance risk	The likelihood someone will make a claim against their insurance policy.
Regression models	
Response time	The time it takes someone to respond to a marketing communication.
Affordability	Someone's disposable income after bills, rent, mortgage and so on have been taken into account. This type of model is used to check that people who are applying for a new line of credit have sufficient income to be able to meet their repayments.
Revenue	The income generated from a customer over a given time horizon.

Table 1.1 Model applications – *continued*

Model	Behaviour the model predicts
Customer lifetime value	The financial contribution a customer makes over the lifetime of the relationship.
Exposure at default	The amount someone owes when they default on a credit agreement. For a product such as a credit card, this takes into account any further advances that are made prior to default occurring.
Loss given default	The loss incurred from a customer who defaults on a credit agreement, taking into account the exposure at the time of default and any recoveries that are made after default has occurred.
Loss given claim	The loss incurred from a customer who makes a claim against their insurance policy.

You will note from Table 1.1 that the models have been segmented into two types: classification models and regression models. Classification models are used to predict how likely a customer behaviour is to occur. A response model, for example, predicts how likely someone who is targeted with a marketing communication, encouraging them to buy a product or service, will respond to it. A credit scoring model predicts whether or not someone will make their loan repayments, and a fraud model predicts the likelihood that someone acts fraudulently. In each case the output generated by the model – the model score – can be interpreted as a probability that someone will or will not exhibit the behaviour in question.

Regression models are used to predict quantities; that is, the magnitude of something. Typically, this will be a financial measure such as how much the customer is likely to spend (a revenue model), how much you expect to make from a customer over the lifetime of the relationship you have with them (a customer lifetime value model) or the loss you can expect to incur when someone defaults on a credit agreement (a loss given default model).

With regard to the split between classification and regression, the most popular usage is classification. Credit scoring models were the first classification models to be widely used in a commercial setting, with their origins dating back to the 1940s (Durand 1941; Wonderlic 1952). As better technology became available, so there was a corresponding growth in the application of classification and regression