Quantitative Models in Marketing Research

Philip Hans Franses
and
Richard Paap

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York NY 10011–4211, USA
10 Stamford Road, Oakleigh, VIC 3166, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http://www.cambridge.org

© Philip Hans Franses and Richard Paap 2001

This book is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without
the written permission of Cambridge University Press.

First published 2001

Printed in the United Kingdom at the University Press, Cambridge

Typeface $10\frac{1}{2}$/13pt Times System

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

Library of Congress Cataloguing in Publication data

Franses, Philip Hans, 1963–
Quantitative models in marketing research / by Philip Hans Franses
and Richard Paap.
p. cm.
Includes bibliographical references and index.
ISBN 0-521-80166-4
1. Marketing research–Mathematical models. I. Paap, Richard. II. Title.

HF5415.2 .F714 2001
658.8′3′015118–dc21 2001025424

ISBN 0 521 80166 4 hardback

Contents

List of figures page ix
List of tables xi
Preface xiii

1 Introduction and outline of the book 1
1.1 Introduction 1
1.1.1 On marketing research 2
1.1.2 Data 4
1.1.3 Models 5
1.2 Outline of the book 6
1.2.1 How to use this book 6
1.2.2 Outline of chapter contents 7

2 Features of marketing research data 10
2.1 Quantitative models 10
2.2 Marketing performance measures 12
2.2.1 A continuous variable 13
2.2.2 A binomial variable 15
2.2.3 An unordered multinomial variable 18
2.2.4 An ordered multinomial variable 19
2.2.5 A limited continuous variable 21
2.2.6 A duration variable 24
2.2.7 Summary 26
2.3 What do we exclude from this book? 26

3 A continuous dependent variable 29
3.1 The standard Linear Regression model 29
3.2 Estimation 34
3.2.1 Estimation by Ordinary Least Squares 34
3.2.2 Estimation by Maximum Likelihood 35

3.3 Diagnostics, model selection and forecasting 38
3.3.1 Diagnostics 39
3.3.2 Model selection 41
3.3.3 Forecasting 43
3.4 Modeling sales 44
3.5 Advanced topics 47

4 A binomial dependent variable **49**
4.1 Representation and interpretation 49
4.1.1 Modeling a binomial dependent variable 50
4.1.2 The Logit and Probit models 53
4.1.3 Model interpretation 55
4.2 Estimation 58
4.2.1 The Logit model 59
4.2.2 The Probit model 60
4.2.3 Visualizing estimation results 61
4.3 Diagnostics, model selection and forecasting 61
4.3.1 Diagnostics 62
4.3.2 Model selection 63
4.3.3 Forecasting 65
4.4 Modeling the choice between two brands 66
4.5 Advanced topics 71
4.5.1 Modeling unobserved heterogeneity 71
4.5.2 Modeling dynamics 73
4.5.3 Sample selection issues 73

5 An unordered multinomial dependent variable **76**
5.1 Representation and interpretation 77
5.1.1 The Multinomial and Conditional Logit models 77
5.1.2 The Multinomial Probit model 86
5.1.3 The Nested Logit model 88
5.2 Estimation 91
5.2.1 The Multinomial and Conditional Logit models 92
5.2.2 The Multinomial Probit model 95
5.2.3 The Nested Logit model 95
5.3 Diagnostics, model selection and forecasting 96
5.3.1 Diagnostics 96
5.3.2 Model selection 97
5.3.3 Forecasting 99

5.4 Modeling the choice between four brands 101
5.5 Advanced topics 107
5.5.1 Modeling unobserved heterogeneity 107
5.5.2 Modeling dynamics 108
5.A EViews Code 109
5.A.1 The Multinomial Logit model 110
5.A.2 The Conditional Logit model 110
5.A.3 The Nested Logit model 111

6 An ordered multinomial dependent variable 112
6.1 Representation and interpretation 113
6.1.1 Modeling an ordered dependent variable 113
6.1.2 The Ordered Logit and Ordered Probit models 116
6.1.3 Model interpretation 117
6.2 Estimation 118
6.2.1 A general ordered regression model 118
6.2.2 The Ordered Logit and Probit models 121
6.2.3 Visualizing estimation results 122
6.3 Diagnostics, model selection and forecasting 122
6.3.1 Diagnostics 123
6.3.2 Model selection 124
6.3.3 Forecasting 125
6.4 Modeling risk profiles of individuals 125
6.5 Advanced topics 129
6.5.1 Related models for an ordered variable 130
6.5.2 Selective sampling 130

7 A limited dependent variable **133**
7.1 Representation and interpretation 134
7.1.1 Truncated Regression model 134
7.1.2 Censored Regression model 137
7.2 Estimation 142
7.2.1 Truncated Regression model 142
7.2.2 Censored Regression model 144
7.3 Diagnostics, model selection and forecasting 147
7.3.1 Diagnostics 147
7.3.2 Model selection 149
7.3.3 Forecasting 150
7.4 Modeling donations to charity 151
7.5 Advanced Topics 155

8 A duration dependent variable **158**
8.1 Representation and interpretation 159
8.1.1 Accelerated Lifetime model 165
8.1.2 Proportional Hazard model 166
8.2 Estimation 168
8.2.1 Accelerated Lifetime model 169
8.2.2 Proportional Hazard model 170
8.3 Diagnostics, model selection and forecasting 172
8.3.1 Diagnostics 172
8.3.2 Model selection 174
8.3.3 Forecasting 175
8.4 Modeling interpurchase times 175
8.5 Advanced topics 179
8.A EViews code 182
8.A.1 Accelerated Lifetime model (Weibull distribution) 182
8.A.2 Proportional Hazard model (loglogistic distribution) 183

Appendix **184**
A.1 Overview of matrix algebra 184
A.2 Overview of distributions 187
A.3 Critical values 193

Bibliography 196
Author index 202
Subject index 204

Figures

2.1 Weekly sales of Heinz tomato ketchup *page* 14
2.2 Histogram of weekly sales of Heinz tomato ketchup 14
2.3 Histogram of the log of weekly sales of Heinz tomato ketchup 15
2.4 Histogram of the choice between Heinz and Hunts tomato ketchup 17
2.5 Histogram of the choice between four brands of saltine crackers 18
2.6 Histogram of ordered risk profiles 20
2.7 Histogram of the response to a charity mailing 22
2.8 Histogram of the amount of money donated to charity 23
2.9 Histogram of the number of days between two liquid detergent purchases 25
3.1 Density function of a normal distribution with $\mu = \sigma^2 = 1$ 30
3.2 Scatter diagram of y_t against x_t 31
3.3 Scatter diagram of $\log S_t$ against $\log P_t$ 45
3.4 Scatter diagram of $\log S_t$ against $\log P_t - \log P_{t-1}$ 45
4.1 Scatter diagram of y_i against x_i, and the OLS regression line of y_i on x_i and a constant 51
4.2 Scatter diagram of y_i^* against x_i 53
4.3 Graph of $\Lambda(\beta_0 + \beta_1 x_i)$ against x_i 56
4.4 Probability of choosing Heinz 69
4.5 Quasi price elasticity 70
5.1 Choice probabilities versus price 106
5.2 Price elasticities 107
6.1 Scatter diagram of y_i^* against x_i 115
6.2 Quasi-elasticities of type 2 funds for each category 128
6.3 Quasi-elasticities of type 3 transactions for each category 129
7.1 Scatter diagram of y_i against x_i given $y_i > 0$ 136

7.2 Scatter diagram of y_i against x_i for censored y_i 138
8.1 Hazard functions for a Weibull distribution 163
8.2 Hazard functions for the loglogistic and the lognormal distributions with $\alpha = 1.5$ and $\gamma = 1$ 163
8.3 Empirical integrated hazard function for generalized residuals 178

Tables

2.1 Characteristics of the dependent variable and explanatory variables: weekly sales of Heinz tomato ketchup *page* 16
2.2 Characteristics of the dependent variable and explanatory variables: the choice between Heinz and Hunts tomato ketchup 17
2.3 Characteristics of the dependent variable and explanatory variables: the choice between four brands of saltine crackers 19
2.4 Characteristics of the dependent variable and explanatory variables: ordered risk profiles 21
2.5 Characteristics of the dependent variable and explanatory variables: donations to charity 23
2.6 Characteristics of the dependent variable and explanatory variables: the time between liquid detergent purchases 26
2.7 Characteristics of a dependent variable and the names of relevant models to be discussed in chapters 3 to 8 27
4.1 Estimation results for Logit and Probit models for the choice between Heinz and Hunts 67
5.1 Parameter estimates of a Conditional Logit model for the choice between four brands of saltine crackers 101
5.2 Parameter estimates of a Nested Logit model for the choice between four brands of saltine crackers 103
6.1 Estimation results for Ordered Logit and Ordered Probit models for risk profiles 126
6.2 Estimation results for two binomial Logit models for cumulative risk profiles 127
7.1 Estimation results for a standard Regression model (including the 0 observations) and a Type-1 Tobit model for donation to charity 152

7.2 Estimation results for a Type-2 Tobit model for the charity donation data 152
7.3 A comparison of target selection strategies based on the Type-1 and Type-2 Tobit models and the Probit model 155
8.1 Some density functions with expressions for their corresponding hazard functions 161
8.2 Parameter estimates of a Weibull Accelerated Lifetime model for purchase timing of liquid detergents 176
8.3 Parameter estimates of a loglogistic Proportional Hazard model for purchase timing of liquid detergents 177
A.1 Density functions (pdf), cumulative distribution functions (cdf), means and variances of the univariate discrete distributions used in this book 188
A.2 Density functions, means and variances of J-dimensional discrete distributions used in this book 189
A.3 Density functions (pdf), cumulative distribution functions (cdf), means and variances of continuous distributions used in this book 190
A.4 Density functions, cumulative distribution functions, means and variances of continuous distributions used in chapter 8 192
A.5 Some critical values for a normally distributed test statistic 194
A.6 Some critical values for a $\chi^2(\nu)$ distributed test statistic 194
A.7 Some critical values for an $F(k, n)$ distributed test statistic 195

1 Introduction and outline of the book

Recent advances in data collection and data storage techniques enable marketing researchers to study the characteristics of many actual transactions and purchases, that is, revealed preference data. Owing to the large number of observations and the wealth of available information, a detailed analysis of these data is possible. This analysis usually addresses the effects of marketing instruments and the effects of household-specific characteristics on the transaction. Quantitative models are useful tools for this analysis. In this book we review several such models for revealed preference data. In this chapter we give a general introduction and provide brief introductions to the various chapters.

1.1 Introduction

It is the aim of this book to present various important and practically relevant quantitative models, which can be used in present-day marketing research. The reader of this book should become able to apply these methods in practice, as we provide the data which we use in the various illustrations and we also add the relevant computer code for EViews if it is not already included in version 3.1. Other statistical packages that include estimation routines for some of the reviewed models are, for example, LIMDEP, SPSS and SAS. Next, the reader should come to understand (the flavor of) the latest methodological developments as these are put forward in articles in, for example, *Marketing Science*, the *Journal of Marketing Research*, the *Journal of Consumer Research* and the *International Journal of Research in Marketing*. For that matter, we also discuss interesting new developments in the relevant sections.

The contents of this book originate from lecture notes prepared for undergraduate and graduate students in Marketing Research and in Econometrics. Indeed, it is our intention that this book can be used at different teaching levels. With that aim, all chapters have the same format, and we indicate which sections correspond with which teaching level. In section 1.2, we will provide more details. For all readers, however, it is necessary to have a basic knowledge of elementary regression techniques and of some matrix algebra. Most introductory texts on quantitative methods include such material, but as a courtesy we bring together some important topics in an Appendix at the end of this book.

There are a few other books dealing with sets of quantitative models similar to the ones we consider. Examples are Maddala (1983), Ben-Akiva and Lerman (1985), Cramer (1991) and Long (1997). The present book differs from these textbooks in at least three respects. The first is that we discuss the models and their specific application in marketing research concerning revealed preference data. Hence, we pay substantial attention to the interpretation and evaluation of the models in the light of the specific applications. The second difference is that we incorporate recent important developments, such as modeling unobserved heterogeneity and sample selection, which have already become quite standard in academic marketing research studies (as may be noticed from many relevant articles in, for example, the *Journal of Marketing Research* and *Marketing Science*). The third difference concerns the presentation of the material, as will become clear in section 1.2 below. At times the technical level is high, but we believe it is needed in order to make the book reasonably self-contained.

1.1.1 On marketing research

A useful definition of marketing research, given in the excellent introductory textbook by Lehmann et al. (1998, p. 1), is that "[m]arketing research is the collection, processing, and analysis of information on topics relevant to marketing. It begins with problem definition and ends with a report and action recommendations." In the present book we focus only on the part that concerns the analysis of information. Additionally, we address only the type of analysis that requires the application of statistical and econometric methods, which we summarize under the heading of quantitative models. The data concern revealed preference data such as sales and brand choice. In other words, we consider models for quantitative data, where we pay specific attention to those models that are useful for marketing research. We do not consider models for stated preference data or other types of survey data, and hence we abstain from, for example, LISREL-type models and various multivariate techniques. For a recent treatment of combining revealed and stated preference data, see Hensher et al. (1999). Finally, we assume that the data have already been collected and that the research question has been clearly defined.

The reasons we focus on revealed preference data, instead of on stated preference data, are as follows. First, there are already several textbooks on LISREL-type models (see, for example, Jöreskog and Sörbom, 1993) and on multivariate statistical techniques (see, for example, Johnson and Wichern, 1998). Second, even though marketing research often involves the collection and analysis of stated preference data, we observe an increasing availability of revealed preference data.

Typical research questions in marketing research concern the effects of marketing instruments and household-specific characteristics on various marketing performance measures. Examples of these measures are sales, market shares, brand choice and interpurchase times. Given knowledge of these effects, one can decide to use marketing instruments in a selective manner and to address only apparently relevant subsamples of the available population of households. The latter is usually called segmentation.

Recent advances in data collection and data storage techniques, which result in large data bases with a substantial amount of information, seem to have changed the nature of marketing research. Using loyalty cards and scanners, supermarket chains can track all purchases by individual households (and even collect information on the brands and products that were not purchased). Insurance companies, investment firms and charity institutions keep track of all observable activities by their clients or donors. These developments have made it possible to analyze not only what individuals themselves state they do or would do (that is, stated preference), but also what individuals actually do (revealed preference). This paves the way for greater insights into what really drives loyalty to an insurance company or into the optimal design for a supermarket, to mention just a few possible issues. In the end, this could strengthen the relationship between firms and customers.

The large amount of accurately measured marketing research data implies that simple graphical tools and elementary modeling techniques in most cases simply do not suffice for dealing with present-day problems in marketing. In general, if one wants to get the most out of the available data bases, one most likely needs to resort to more advanced techniques. An additional reason is that more detailed data allow more detailed questions to be answered. In many cases, more advanced techniques involve quantitative models, which enable the marketing researcher to examine various correlations between marketing response variables and explanatory variables measuring, for example, household-specific characteristics, demographic variables and marketing-mix variables.

In sum, in this book we focus on quantitative models for revealed preference data in marketing research. For conciseness, we do not discuss the various issues related to solving business problems, as this would require an entirely different book. The models we consider are to be viewed as helpful practical tools when analyzing marketing data, and this analysis can be part of a more comprehensive approach to solving business problems.

1.1.2 Data

Marketing performance measures can appear in a variety of formats. And, as we will demonstrate in this book, these differing formats often need different models in order to perform a useful analysis of these measures.

To illustrate varying formats, consider "sales" to be an obvious marketing performance measure. If "sales" concerns the number of items purchased, the resultant observations can amount to a limited range of count data, such as $1, 2, 3, \ldots$. However, if "sales" refers to the monetary value in dollars (or cents) of the total number of items purchased, we may consider it as a continuous variable. Because the evaluation of a company's sales may depend on all other sales, one may instead want to consider market shares. These variables are bounded between 0 and 1 by construction.

Sales and market shares concern variables which are observed over time. Typically one analyzes weekly sales and market shares. Many other marketing research data, however, take only discrete (categorical) values or are only partially observed. The individual response to a direct mailing can take a value of 1 if there is a response, and 0 if the individual does not respond. In that case one has encountered a binomial dependent variable. If households can choose between three or more brands, say brands A, B, C and D, one has encountered a multinomial dependent variable. It may then be of interest to examine whether or not marketing-mix instruments have an effect on brand choice. If the brands have a known quality or preference ranking that is common to all households, the multinomial dependent variable is said to be ordered; if not, it is unordered. Another example of an ordered categorical variable concerns questionnaire items, for which individuals indicate to what extent they disagree, are indifferent, or agree with a certain statement.

Marketing research data can also be only partially observed. An example concerns donations to charity, for which individuals have received a direct mailing. Several of these individuals do not respond, and hence donate nothing, while others do respond and at the same time donate some amount. The interest usually lies in investigating the distinguishing characteristics of the individuals who donate a lot and those who donate a lesser amount, while taking into account that individuals with perhaps similar characteristics donate nothing. These data are called censored data. If one knows the amount donated by an individual only if it exceeds, say, \$10, the corresponding data are called truncated data.

Censoring is also a property of duration data. This type of observation usually concerns the time that elapses between two events. Examples in marketing research are interpurchase times and the duration of a relationship between a firm and its customers. These observations are usually collected for panels of individuals, observed over a span of time. At the time of the first observations, it is unlikely that all households buy a product or brand at the same time, and hence it is likely that some durations (or relationships) are already ongoing. Such interpurchase times can be useful in order to understand, for example, whether or not promotions accelerate purchasing behavior. For direct marketing, one might model the time between sending out the direct mailing and the response, which perhaps can be reduced by additional nationwide advertising. In addition, insurance companies may benefit from lengthy relationships with their customers.

1.1.3 Models

As might be expected from the above summary, it is highly unlikely that all these different types of data could be squeezed into one single model framework. Sales can perhaps be modeled by single-equation linear regression models and market shares by multiple-equation regression models (because market shares are interconnected), whereas binomial and multinomial data require models that take into account that the dependent variable is not continuous. In fact, the models for these choice data usually consider, for example, the probability of a response to a direct mailing and the probability that a brand is selected out of a set of possible brands. Censored data require models that take into account the probability that, for example, households do not donate to charity. Finally, models for duration data take into account that the time that has elapsed since the last event has an effect on the probability that the next event will happen.

It is the purpose of this book to review quantitative models for various typical marketing research data. The standard Linear Regression model is one example, while the Multinomial Logit model, the Binomial Logit model, the Nested Logit model, the Censored Regression model and the Proportional Hazard model are other examples. Even though these models have different names and take account of the properties of the variable to be explained, the underlying econometric principles are the same. One can summarize these principles under the heading of an econometric modeling cycle. This cycle involves an understanding of the representation of the model (what does the model actually do? what can the model predict? how can one interpret the parameters?), estimation of the unknown parameters, evaluation of the model (does the model summarize the data in an adequate way? are there ways to improve the model?), and the extension of the model, if required.

We follow this rather schematic approach, because it is our impression that studies in the academic marketing research literature are sometimes not very explicit about the decision to use a particular model, how the parameters were estimated, and how the model results should be interpreted. Additionally, there are now various statistical packages which include estimation routines for such models as the Nested Logit model and the Ordered Probit model (to name just a few of the more exotic ones), and it frequently turns out that it is not easy to interpret the output of these statistical packages and to verify the adequacy of the procedures followed. In many cases this output contains a wealth of statistical information, and it is not always clear what this all means and what one should do if statistics take particular values. By making explicit several of the modeling steps, we aim to bridge this potential gap between theory and practice.

1.2 Outline of the book

This book aims to describe some of the main features of various potentially useful quantitative models for marketing research data. Following a chapter on the data used throughout this book, there are six chapters, each dealing with one type of dependent variable. These chapters are subdivided into sections on (1) representation and interpretation, (2) the estimation of the model parameters, (3) model diagnostics and inference, (4) a detailed illustration and (5) advanced topics.

All models and methods are illustrated using actual data sets that are or have been effectively used in empirical marketing research studies in the academic literature. The data are available through relevant websites. In chapter 2, we discuss the data and also some of the research questions. To sharpen the focus, we will take the data as the main illustration throughout each chapter. This means that, for example, the chapter on a multinomial dependent variable (chapter 5) assumes that such a model is useful for modeling brand choice. Needless to say, such a model may also be useful for other applications. Additionally, to reduce confusion, we will consider the behavior of a household, and assume that it makes the decisions. Of course this can be replaced by individuals, customers or other entities, if needed.

1.2.1 How to use this book

The contents of the book are organized in such a way that it can be used for teaching at various levels or for personal use given different levels of training.

The first of the five sections in each of chapters 3 to 8 contains the representation of the relevant model, the interpretation of the parameters, and sometimes the interpretation of the full model (by focusing, for example, on elasticities). The fourth section contains a detailed illustration, whose content one should be able to grasp given an understanding of the content of the first section. These two sections can be used for undergraduate as well as for graduate teaching at a not too technical level. In fact, we ourselves have tried out these sections on undergraduate students in marketing at Erasmus University Rotterdam (and, so far, we have not lost our jobs).

Sections 2 and 3 usually contain more technical material because they deal with parameter estimation, diagnostics, forecasting and model selection. Section 2 always concerns parameter estimation, and usually we focus on the Maximum Likelihood method. We provide ample details of this method as we believe it is useful for a better understanding of the principles underlying the diagnostic tests in section 3. Furthermore, many computer packages do not provide diagnostics and, using the formulas in section 2, one can compute them oneself. Finally, if one wants to program the estimation routines oneself, one can readily use the material. In many cases one can replicate our estimation results using the relevant standard routines in EViews (version 3.1). In some cases these routines do not exist, and in that case we give the relevant EViews code at the end of the relevant chapters. In addition to sections 1 and 4, one could consider using sections 2 and 3 to teach more advanced undergraduate students, who have a training in econometrics or advanced quantitative methods, or graduate students in marketing or econometrics.

Finally, section 5 of each chapter contains advanced material, which may not be useful for teaching. These sections may be better suited to advanced graduate students and academics. Academics may want to use the entire book as a reference source.

1.2.2 Outline of chapter contents

The outline of the various chapters is as follows. In chapter 2 we start off with detailed graphical and tabulated summaries of the data. We consider weekly sales, a binomial variable indicating the choice between two brands, an unordered multinomial variable concerning the choice between four brands, an ordered multinomial variable for household-specific risk profiles, a censored variable measuring the amount of money donated to charity and, finally, interpurchase times in relation to liquid detergent.

In chapter 3 we give a concise treatment of the standard Linear Regression model, which can be useful for a continuous dependent variable. We assume some knowledge of basic matrix algebra and of elementary statistics. We discuss the representation of the model and the interpretation of its parameters. We also discuss Ordinary Least Squares (OLS) and Maximum Likelihood (ML) estimation methods. The latter method is discussed because it will be used in most chapters, although the concepts underlying the OLS method return in chapters 7 and 8. We choose to follow the convention that the standard Linear Regression model assumes that the data are normally distributed with a constant variance but with a mean that obtains different values depending on the explanatory variables. Along similar lines, we will introduce models for binomial, multinomial and duration dependent variables in subsequent chapters. The advanced topics section in chapter 3 deals with the attraction model for market shares. This model ensures that market shares sum to unity and that they lie within the range [0, 1].

The next chapter deals with a binomial dependent variable. We discuss the binomial Logit and Probit models. These models assume a nonlinear relation between the explanatory variables and the variable to be explained. Therefore, we pay considerable attention to parameter interpretation and model interpretation. We discuss the ML estimation method and we provide some relevant model diagnostics and evaluation criteria. As with the standard Linear Regression model, the diagnostics are based on the residuals from the model. Because these residuals can be defined in various ways for these models, we discuss this issue at some length. The advanced topics section is dedicated to the inclusion of unobserved parameter heterogeneity in the model and to the effects of sample selection for the Logit model.

In chapter 5 we expand on the material of chapter 4 by focusing on an unordered multinomial dependent variable. Quite a number of models can be considered, for example the Multinomial Logit model, the Multinomial Probit model, the Nested Logit model and the Conditional Logit model. We pay substantial attention to outlining the key differences between the various models in particular because these are frequently used in empirical marketing research.

In chapter 6 we focus on the Logit model and the Probit model for an ordered multinomial dependent variable. Examples of ordered multinomial data typically appear in questionnaires. The example in chapter 6 concerns customers of a financial investment firm who have been assigned to three categories depending on their risk profiles. It is the aim of the empirical analysis to investigate which customer-specific characteristics can explain this classification. In the advanced topics section, we discuss various other models for ordered categorical data.

Chapter 7 deals with censored and truncated dependent variables, that is, with variables that are partly continuous and partly take some fixed value (such as 0 or 100) or are partly unknown. We mainly focus on the Truncated

Regression model and on the two types of Tobit model, the Type-1 and Type-2 Tobit models. We show what happens if one neglects the fact that the data are only partially observed. We discuss estimation methods in substantial detail. The illustration concerns a model for money donated to charity for a large sample of individuals. In the advanced topics section we discuss other types of models for data censored in some way.

Finally, in chapter 8 we deal with a duration dependent variable. This variable has the specific property that its value can be made dependent on the time that has elapsed since the previous event. For some marketing research applications this seems a natural way to go, because it may become increasingly likely that households will buy, for example, detergent if it is already a while since they purchased it. We provide a discussion of the Accelerated Lifetime model and the Proportional Hazard model, and outline their most important differences. The advanced topics section contains a discussion of unobserved heterogeneity. It should be stressed here that the technical level both of chapter 8 and of chapter 5 is high.

Before we turn to the various models, we first look at some marketing research data.