Recent advances in data collection and data storage techniques enable marketing researchers to study the characteristics of a large range of transactions and purchases, in particular the effects of household-specific characteristics and marketing-mix variables.

This book presents the most important and practically relevant quantitative models for marketing research. Each model is presented in detail with a self-contained discussion, which includes: a demonstration of the mechanics of the model, empirical analysis, real-world examples, and interpretation of results and findings. The reader of the book will learn how to apply the techniques, as well as understand the latest methodological developments in the academic literature.

Pathways are offered in the book for students and practitioners with differing statistical and mathematical skill levels, although a basic knowledge of elementary numerical techniques is assumed.

Philip Hans Franses is Professor of Applied Econometrics affiliated with the Econometric Institute and Professor of Marketing Research affiliated with the Department of Marketing and Organization, both at Erasmus University Rotterdam. He has written successful textbooks in time series analysis.

Richard Paap is Postdoctoral Researcher with the Rotterdam Institute for Business Economic Studies at Erasmus University Rotterdam. His research interests cover applied (macro-)econometrics, Bayesian statistics, time series analysis, and marketing research.
Quantitative Models in Marketing Research

Philip Hans Franses
and
Richard Paap
# Contents

| List of figures | page ix |
| List of tables | xi |
| Preface | xiii |

1 Introduction and outline of the book  
1.1 Introduction  
\hspace{1cm} 1.1.1 On marketing research  
\hspace{1cm} 1.1.2 Data  
\hspace{1cm} 1.1.3 Models  
1.2 Outline of the book  
\hspace{1cm} 1.2.1 How to use this book  
\hspace{1cm} 1.2.2 Outline of chapter contents  

2 Features of marketing research data  
2.1 Quantitative models  
2.2 Marketing performance measures  
\hspace{1cm} 2.2.1 A continuous variable  
\hspace{1cm} 2.2.2 A binomial variable  
\hspace{1cm} 2.2.3 An unordered multinomial variable  
\hspace{1cm} 2.2.4 An ordered multinomial variable  
\hspace{1cm} 2.2.5 A limited continuous variable  
\hspace{1cm} 2.2.6 A duration variable  
\hspace{1cm} 2.2.7 Summary  
2.3 What do we exclude from this book?  

3 A continuous dependent variable  
3.1 The standard Linear Regression model  
3.2 Estimation  
\hspace{1cm} 3.2.1 Estimation by Ordinary Least Squares  
\hspace{1cm} 3.2.2 Estimation by Maximum Likelihood  

© Cambridge University Press  www.cambridge.org
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 Modelingsales 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
Contents

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
Contents

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 interpurchasetimes 175
  8.5 Advanced topics 179
  8.A EViews code 182
    8.A.1 Accelerated Lifetime model 182
       (Weibull distribution)
    8.A.2 Proportional Hazard model 183
       (loglogistic distribution)

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
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 $y_i$ against $x_i$ 31
3.3 Scatter diagram of $\log S_i$ against $\log P_i$ 45
3.4 Scatter diagram of $\log S_i$ against $\log P_i - \log P_{i-1}$ 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
List of figures

7.2 Scatter diagram of $y_i$ against $x_i$ for censored $y_j$ 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
List of tables

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 $\chi^2(1)$ distributed test statistic 194
A.7 Some critical values for an $F(k, n)$ distributed test statistic 195
Preface

In marketing research one sometimes considers rather advanced quantitative (econometric) models for revealed preference data (such as sales and brand choice). Owing to an increased availability of empirical data, market researchers can choose between a variety of models, such as regression models, choice models and duration models. In this book we summarize several relevant models and deal with these in a fairly systematic manner. It is our hope that academics and practitioners find this book useful as a reference guide, that students learn to grasp the basic principles, and, in general, that the reader finds it useful for a better understanding of the academic literature on marketing research.

Several individuals affiliated with the marketing group at the School of Economics, Erasmus University Rotterdam, helped us with the data and presentation. We thank Dennis Fok and Jedid-Jah Jonker for making some of the data accessible to us. The original data were provided by ROBECO and the Astmafonds and we thank these institutions for their kind assistance. Paul de Boer, Charles Bos, Bas Donkers, Rutger van Oest, Peter Verhoef and an anonymous reviewer went through the entire manuscript and made many useful comments. Finally, we thank the Rotterdam Institute for Business Economic Studies (RIBES) and the Econometric Institute for providing a stimulating research environment and the Netherlands Organization for Scientific Research (NWO) for its financial support.

PHILIP HANS FRANSES
RICHARD PAAP
Rotterdam, November 2000