Quantitative Approaches for
Profit Maximization in Direct Marketing

Hiek van der Scheer

To Ellen
The research for this thesis was supported by:

Marketing Support Group Groningen
Rabofacet

Its publication has been made possible by the financial support of:

PTT Post Mediaservice

ISBN 90-9011264-2
QUANTITATIVE APPROACHES FOR
PROFIT MAXIMIZATION IN DIRECT MARKETING

Proefschrift

ter verkrijging van het doctoraat in de
Economische Wetenschappen
aan de Rijksuniversiteit Groningen
op gezag van de
Rector Magnificus dr. F. van der Woude
in het openbaar te verdedigen op
donderdag 8 januari 1998
des namiddags te 2.45 uur

door

Hielke Roelof van der Scheer

geboren op 24 februari 1969
te Tweede Exloërmond
Promotores:  Prof. dr. T.J. Wansbeek
Prof. dr. P.S.H. Leeflang
Acknowledgements

This thesis reflects the research I did at the Faculty of Economics of the University of Groningen in the last four years. I am indebted to many persons who contributed to this research.

I am especially grateful to my promotor Tom Wansbeek. It was a privilege working with such a stimulating person. He was always willing to discuss my research and think about all kinds of questions. I have benefited tremendously from his support, knowledge and critical notes.

I am also grateful to my second promotor Peter Leeflang. He has provided me with well-considered criticism, which was often complementary to that of Tom Wansbeek. His comments on various aspects of this thesis have been very useful.

Further I would like to thank the members of the graduation committee, Ton Steerneman, Michel Wedel and Dick Wittink for reading the thesis and for their useful comments. Moreover, I am grateful to Ton Steerneman for his contributions to various statistical problems throughout this thesis.

Special thanks are due to Jan Roelf Bult, Janny Hoekstra, Lars Muus, Pieter Otter and Marco Vriens for their contributions to various parts of this thesis. It was a pleasure working with colleagues with different interests. It goes without saying that I retain full responsibility for any errors.

I would like to acknowledge the hospitality of the Department of Statistics of the Australian National University at Canberra, where part of this research was done.

I thank Gerard van den Hooff for correcting my English and Erik Frambach for his \LaTeX-support.

Financial support for this research was provided by the Faculty of Economics, the AIO-school, the NWO (Dutch organization of scientific research), the Marketing Support Group and Rabofac. The publication of this thesis has been made possible by the financial support of PTT Post Mediaservice.

Finally, I would like to thank Ellen for many reasons mainly beyond the scope of this thesis.
Contents

1 Introduction 1
  1.1 Direct marketing 1
  1.2 Objective and outline of the thesis 4

2 Target selection techniques 9
  2.1 Introduction 9
  2.2 Objective of the direct marketing campaign 10
  2.3 Selection variables 11
  2.4 Selection techniques 12
  2.5 Model comparison 24
  2.6 Discussion and conclusion 27

3 Target selection by joint modeling of the probability and quantity of response 29
  3.1 Introduction 29
  3.2 Description of the model 31
  3.3 Optimal selection 34
  3.4 Approximations 36
  3.5 Operationalization 38
  3.6 Tobit model 41
  3.7 Data 43
  3.8 Empirical results 45
  3.9 Discussion and conclusion 47

4 Bayesian decision rule approach for target selection 51
  4.1 Introduction 51
  4.2 The decision theoretic framework 54
  4.3 The case of a normal posterior 57
  4.4 Numerical evaluation of the optimal Bayesian rule 61
7.7 Discussion and conclusion 165

8 Summary, future research and management implications 169

8.1 Summary 169
8.2 Future research 173
8.3 Management implications 180

Bibliography 183

Author index 193

Subject index 197
Chapter 1

Introduction

1.1 Direct marketing

In the last few decades, direct marketing has become an important field of marketing. Various definitions of direct marketing have been proposed from the late seventies (e.g. Direct Marketing Association 1981, and Kobs 1979) till more recent times (e.g. Bauer and Miglautsch 1992, Beukenkamp and Leeflang 1988, Hoekstra and Raaijmaakers 1992, Murrow and Hyman 1994, and Schofield 1995). On the one hand this reflects a natural development in a relatively young discipline. On the other hand it indicates that neither practitioners nor academics are unanimous in their opinion about the exact domain of direct marketing. Recent discussions on the definition mainly focus on the question what aspects should be emphasized, or stated explicitly, in the definition. For example, Bauer and Miglautsch (1992) strongly emphasize the information process and Hoekstra and Raaijmaakers (1992) stress the long-term relationship. A thorough overview and comparison of the various proposed definitions can be found in Murrow and Hyman (1994).

Although there may be some disagreement about the exact definition there is a clear consensus about the essential characteristics of direct marketing. First, direct marketing is a marketing strategy rather than a mere use of specific marketing instruments. Secondly, direct marketing is an interactive system, which means that there is a two-way communication between an organization and its customers. The customers can, for example, communicate (respond) through inserts in magazines, postal reply cards or by telephone. Thirdly, the results of direct marketing promotion are generally better measurable than in traditional marketing. That is, the costs of the promotion and the actual response can be measured at the individual level. This precise information
enables the direct marketing organization to evaluate the effectiveness and efficiency of specific direct marketing campaigns, which can be used for a precise targeting for future campaigns. Consequently, a database that registers all the communications with the customers, a so-called mailing list, plays a central role in direct marketing. The usefulness of a database can be further enhanced by other customer characteristics such as age, income and geographic information.

Direct marketing may be adopted at several levels in the distribution chain: producers, wholesalers as well as retailers may choose direct marketing (e.g., Marshall and Vredenburg 1988). The transaction can consist of tangible goods, services, ideas etc. Direct marketing can be employed in consumer marketing and in business-to-business marketing, while suppliers can direct their activities at domestic as well as foreign markets. The media that can be employed to communicate directly with specific individuals and/or households in order to transmit direct marketing offers and messages include:

- Direct mail: an addressed or unaddressed printed message that is delivered at the addressees by a postal service or by private carriers such as door-to-door distributors. This can be done either in the form of a solo mailing, or by mailing a mailpack or a catalog.
- Telephone: either inbound (customers call to obtain information or service, or to order merchandise) or outbound calls (service and sales calls directed at the customers).
- Interactive devices like interactive TV and internet, which make electronic communications available.

(Baier 1985, p. 333, and Roberts and Berger 1989, p. 14.)

The importance of direct marketing can be gleaned from the Direct Marketing Association’s Statistical Factbook 1995-1996, which contains information about direct marketing in the USA. The importance is also reflected in the expanding number of articles that have recently been published. Developments in information technology, individualization tendencies, rising distribution costs and the increase in dual-income households have been identified as the factors responsible for the increased reliance on direct marketing (Pettit 1987).

Direct mail is the most important medium of the various direct marketing media. Advertising expenditures on direct mail increase annually. In the USA it increased by 7.4 percent point from 1992 to 1993. Of the total direct response marketing budget 36.7% was spent on direct mail. Direct mail is even the third largest category (after newspapers and television) of all advertising expenditures. (See the Direct Marketing Association’s Statistical Factbook
Table 1.1: Total advertising expenditures in the Netherlands (in millions of NLG)

<table>
<thead>
<tr>
<th></th>
<th>1980</th>
<th>1990</th>
<th>1995</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newspaper &amp; magazines</td>
<td>2 886</td>
<td>4 673</td>
<td>5 049</td>
</tr>
<tr>
<td>Broadcast media</td>
<td>264</td>
<td>830</td>
<td>1 569</td>
</tr>
<tr>
<td>Direct mail</td>
<td>1 131</td>
<td>2 374</td>
<td>3 542</td>
</tr>
<tr>
<td>Other media</td>
<td>256</td>
<td>998</td>
<td>1 779</td>
</tr>
<tr>
<td>Total expenditures</td>
<td>4 537</td>
<td>8 875</td>
<td>11 939</td>
</tr>
</tbody>
</table>

Source: VEA 1996, and AdfoDirect Gebruikersgids 1997

1995-1996.) In Europe, more than 50% of the expenditures on direct marketing is used for direct mail. This percentage differs markedly between European countries. For example, it is 72.4% in France and 33.9% in the United Kingdom (Leefflang and Van Raaij 1995). As for the Netherlands, table 1.1 shows the growth of direct mail. The share of direct mail of total advertising expenditures in this country has increased from 24.9% in 1980 to 26.7% in 1990 and to 29.7% in 1995. Direct mail is the second largest category after newspapers and magazines. Please note that table 1.1 does not include the expenditures of other direct marketing media such as telephone and internet.

Examples of Dutch companies using direct mail to sell tangible products are catalog retailers such as Neckermann, Otto and Wehkamp. Services are offered by insurance companies (e.g. OHRA, DELA), financial institutions (e.g. Rabobank, Postbank), and credit card companies (e.g. American Express, VISA). Furthermore, many nonprofit organizations (e.g. Word Wildlife Foundation, Greenpeace) use direct mail to raise funds for their activities. Direct mail for foreign markets is used by many academic publishers (e.g. Macmillan, Addison-Wesley). The shares over the various branches in the Netherlands, in numbers of addressed direct mailings, for 1993 and 1996 are depicted in table 1.2. It shows that financial institutions have a large share, which increased in the three years. In several empirical applications in this thesis, we consider mailings of charity foundations, which also have a fairly large share.
Table 1.2: Shares of direct mailing of various branches in the Netherlands

<table>
<thead>
<tr>
<th>Branch</th>
<th>1993</th>
<th>1996</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transport</td>
<td>6%</td>
<td>5%</td>
</tr>
<tr>
<td>Clothing &amp; shoes</td>
<td>8%</td>
<td>9%</td>
</tr>
<tr>
<td>Catalog retailer</td>
<td>16%</td>
<td>15%</td>
</tr>
<tr>
<td>Media</td>
<td>9%</td>
<td>5%</td>
</tr>
<tr>
<td>Tourism (travel/restaurant)</td>
<td>13%</td>
<td>8%</td>
</tr>
<tr>
<td>Financial institutions (incl. insurance)</td>
<td>18%</td>
<td>24%</td>
</tr>
<tr>
<td>Charity foundations</td>
<td>12%</td>
<td>15%</td>
</tr>
<tr>
<td>Other branches</td>
<td>18%</td>
<td>19%</td>
</tr>
<tr>
<td></td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Source: Wie mailt wat?, Rotterdam

1.2 Objective and outline of the thesis

Direct mail has many advantages over other media. For instance, direct mail enables an organization to achieve a more precise targeting, it offers the opportunity to personalize to any desired degree and there is a high flexibility with regard to formats, timing and testing. However, the relatively high cost per potential customer (particularly in the case of addressed direct mail), compared to alternative media, requires sufficient response rates to ensure profitable implementation. Hence, it is important to develop methods to improve the effectiveness of direct mail campaigns.

According to conventional wisdom, the important components for the development of a direct marketing campaign are:

- The list of customers to be targeted. Elements that influence this list are (a) the quality of the available list of potential customers and (b) the choice of the selection technique.
- The characteristics of the offer, which include price and the elements to position the product.
- The creative elements or the design of the mailing, the so-called communication elements.
- The timing and sequencing of the mailing(s).

(Roberts and Berger 1989, p. 5.) An effective direct marketing campaign aims at selecting those target groups, offer and communication elements (at the right
time) that maximize the net profits (Kobs 1979). Of these four components the list of customers to be selected is considered to be the most important. Therefore, a large amount of direct marketing research focuses on list segmentation or target selection techniques. Moreover, it is the aspect that receives most of the attention of practitioners (Hoekstra and Vriens 1995). It is therefore not surprising that many target selection techniques are available.

The objective of the thesis is to propose several modifications to existing methods and to introduce new approaches in order to improve the effectiveness of addressed direct mailing campaigns. The ultimate goal, which should be the result of the improved effectiveness, is an increase in the net profits. The emphasis of this thesis lies on the modeling aspects. Chapters 3, 4 and 5 relate to the list of customers to be targeted. Chapter 6 deals with the communication elements, and chapter 7 with the frequency of the mailings. With the exception of chapter 7, we concentrate on short-run profit maximization. Furthermore, we consider the use of addressed direct mail on the consumer market, although in principle several of the techniques can also be applied to other types of media and to other markets.

We denote a household, addressee, customer or potential customer by individual, who - for the sake of convenience - is henceforth referred by he/his. The most promising individuals are called targets.

In chapter 2 we give an overview of the various target selection techniques, which have either been proposed in the literature and/or have commonly been employed by direct marketers. In contrast to the other subjects that will be discussed in this thesis, we give the overview in a separate chapter because three chapters focus on aspects regarding selection techniques. Furthermore, the literature on this aspect is more extensive.

Nearly all the selection techniques proposed so far deal with the case of a fixed revenue to a positive reply and hence concentrate on binary choice modeling (response versus nonresponse). Although many direct mailing campaigns do not generate simple binary response but rather a response where the quantity differs between individuals, it is hard to find publications that take this aspect into account. The quantity can be, for example, the total revenues of purchases from a catalog retailer, or the amount of money donated to a charitable foundation. In chapter 3 we propose several methods to incorporate the quantity of response in the selection process. The underlying profit function indicates that there are two relevant decisions to be taken by the individuals. The two decisions are: (a) whether to respond or not, and, in the case of response, (b)
the quantity. Whereas in traditional selection techniques only the first decision is considered, our models take both decisions into account.

Generally, the parameters of a selection model are unknown and have to be estimated. On the basis of these estimates, the potential individuals are ranked in order to select the targets. All of the proposed selection methods consider the estimation and selection step separately. Since by separation of these two steps the estimation uncertainty is neglected, these methods generally lead to a suboptimal decision rule and hence not to the realization of optimal profits. In chapter 4 we formulate a decision theoretic framework which integrates the estimation and decision step. This framework provides an optimal Bayesian decision rule that follows from the organization’s profit function and which explicitly takes the estimation uncertainty into account. One of the difficulties of such an approach is the evaluation of the high-dimensional integral resulting from the Bayesian decision rule. We show that this integral can either be approximated through a normal posterior, or evaluated numerically by a Laplace approximation or by Markov chain Monte Carlo integration.

There are two aspects which determine the quality of the mailing list. First of all, the list should contain sufficient promising targets. Furthermore, it should supply sufficient information to identify these targets. The degree to which targets can be identified indicates the value of information on the list. In chapter 5 we analyze the relation between the value of information and profits. Of particular interest is the value of postal code information. This is information of the individuals at the postal code level, which in the Netherlands comprises 16 households on average. Thus, instead of having information at the individual level, we have so-called proxy information. The selection procedures can still be employed but the maximally attainable profit will be lower. However, the profit could still be higher than in the situation where no information is available at all. We attach a monetary value to proxy information and assess whether such information is worth its costs. Although our approach is rather stylized and mainly theoretical, it gives a clear insight in the essential ingredients and the problems that arise when information is valued.

In chapter 6 we focus on the communication elements of direct marketing, viz. the characteristics of a mailing. We propose several methods to improve the effectiveness of a campaign by determining the optimal mailing design. A traditional approach to test the effect of many mailing characteristics is by analyzing them separately; generally speaking, this is very inefficient. We discuss two approaches, based on conjoint analysis, to determine efficiently the optimal design. We illustrate these approaches with three applications.
An extension of these approaches, which is discussed in the remainder of chapter 6, is to combine target selection with design optimization. That is, we simultaneously optimize the mailing design and select the targets. In other words, the interaction between target characteristics and mailing characteristics is taken into account.

All the methods considered so far are concerned with an optimal selection or design for one particular mailing campaign. One of the most challenging aspects in direct marketing is the extension of these methods to a series of campaigns. There are various aspects that come into play. The most important is the lifetime value (LTV) of an individual, denoting the net present value of future expected profits. Maximizing the lifetime value indicates that an organization accepts negative profits in the beginning if this is outweighed by positive profits later on. In chapter 7 we discuss the LTV concept in somewhat more detail. Then we focus on the question of the optimal number of mailings in a given time period. The basis of our proposed method is a conceptual framework that describes the consumers’ purchase behavior. By simulating the various aspects of this purchase behavior, an organization can determine, at least in principle, the optimal mailing frequency.

Chapter 8 contains a summary and various suggestions for future research.
Chapter 2

Target selection techniques

2.1 Introduction

Target selection, also called list segmentation, can be seen as the process of dividing the market, i.e. the mailing list, into two distinct groups, viz. a group that should receive a mailing and a group that should not receive a mailing. It is obvious that target selection is a crucial component for the development of a direct mailing campaign since a campaign can only be effective if the mailing reaches the proper targets. Therefore, direct marketers have expended considerable effort towards list segmentation techniques. By eliminating the least promising targets of the mailing list, the direct marketing organization can increase its profits.

Generally speaking, in order to achieve an effective segmentation for a promotional campaign, the distinct segments or groups must satisfy at least the following criteria (cf. Leeflang 1994, p. 41, and Wedel 1990, p. 17): first, the size and composition of the selected group must be known. Secondly, it must be possible to reach the selected group effectively in a communicative manner. Thirdly, the selected group should be economically large enough for the development of a promotional campaign. Usually, these conditions are met in direct marketing when a mailing list is available. The size is simply measured by counting the number of targets and the composition is determined by the variables available on the list. Since a mailing list contains at least names and addresses, each individual can be reached by mail. The last criterion is met when the number of individuals in the group that should receive a mailing is sufficiently large. This number implicitly depends on the revenues to a positive reply, the costs of a single mailing piece and the fixed costs. Hence, a mailing
list offers an organization a perfect possibility for effective segmentation, which coincides with target selection here.

The objective of this chapter is to review the selection techniques that have either been proposed in the literature or have commonly been employed by direct marketers. There are at least two aspects that have a large impact on the target selection process, viz. the objective of the direct marketing campaign and the (availability of) selection variables. Since it is important to keep these two aspects in mind, we start in section 2.2 and 2.3, respectively, with a brief discussion on these aspects. Then, in section 2.4, we present the selection techniques. In section 2.5 we discuss the various criteria on which the models can be compared and briefly discuss the literature that compares some of these techniques. Section 2.6 states our conclusion.

2.2 Objective of the direct marketing campaign

Each direct marketing campaign should, of course, have a specific objective. We distinguish three classes of objectives for the consumer market (e.g. Roberts and Berger 1989, p. 9):

1. The accomplishment of the sale of a product or service. This is the most frequently used objective.
2. The generation of leads, denoting the request of an individual for additional information about the subject. In the next stage, an organization may try to convert these leads into sales. Expensive or complex products, like insurance policies, are often sold by this two-stage process. For a successful campaign the organization should not only focus on generating leads but also in the conversions of these leads.
3. The maintenance of customer relationships. The mailing under consideration does not have the purpose to sell the product or to generate leads, but to strengthen the relation (e.g. a newsletter of an insurance company which is sent to all the policy holders).

When the target selection strategy is directed to profit maximization the first two objectives, roughly speaking, focus on short-run profit maximization, whereas the latter focuses on long-term profit maximization. This could imply that an organization aims at maximizing the lifetime value of the individuals, which can be described as “the net present value of future contributions to profit expected from the individual”. Different direct marketing objectives, in particular the difference between short-run and long-run profit maximization,
Selection variables

lead to different selections. However, the critical issue in each case is to select those individuals whose expected revenues are sufficiently high to offset the costs. The difference in selection is caused by the way the costs and expected revenues are determined. Note that, when a profit maximization approach is used, the number of individuals that will receive a mailing is determined by a cost-benefit analysis and not by a fixed budget figure. Sometimes, however, direct marketing campaigns are based on a fixed budget. In that case the same selection techniques can be used but the number of individuals that will receive a mailing is specified a priori by the size of the budget.

2.3 Selection variables

Selection variables, often called segmentation variables in the direct marketing literature, are of course a crucial element for target selection. If the available variables do not have any explanatory or discriminatory power to distinguish respondents from nonrespondents, segmentation is not very useful. Four basic categories of selection variables can be distinguished (Roberts and Berger 1989, p. 106): (1) geographic variables, (2) demographic variables, (3) psychographic or lifestyle variables and (4) behavioral variables. Within direct marketing the behavioral variables, in particular variables related to the individuals’ purchase history, are considered the most important. They include the recency of the last purchase, the frequency of purchases and the monetary amount spent on the purchases (e.g. Roberts and Berger 1989, p. 106); the so-called RFM-variables. Recency includes the number of consecutive mailings without response and the time period since the last order. Frequency measures include the number of purchases made in a certain time period. Monetary value measures include the amount of money spent during a certain period. It is clear that this kind of information solely relates to individuals previously mailed. A database with this type of customer information is called the internal list (e.g. Bickert 1992). This list often also includes individuals who have never bought a product but only made an inquiry or to whom a mailing has been directed.

Other information on the individuals is often available on external lists. These lists contain geographical variables (region, degree of urbanization), demographic variables (age, sex, family size) and lifestyle characteristics (habits, leisure interests). External lists may be put together for direct marketing support or for other reasons (e.g. telephone directories). It is important to realize that external lists can be used as additional information to the internal list but
Target selection techniques may also be the only source of information. The latter is often the case when the mailing is directed to acquire new customers. External lists can be rented from commercial companies. In the United States these lists are offered by, among others, R.L. Polk, National Demographics and Lifestyles and Database America. In the Netherlands lists can be rented from Geo-Markt pro®el, Mosaïc and recently from Calyx (in cooperation with Centrum voor Consumenten Informatie). A comprehensive discussion on the use of commercial databases can be found in Lix et al. (1995).

Often information can be rented either at an individual level or at a postal code level. The latter means that the average of a group of individuals is known. Consequently, the selection takes place at the aggregate level rather than at the individual level. As a result, the maximally attainable profit decreases. In the literature it is shown (e.g. Lix et al. 1995, Roberts and Berger 1989, pp. 99-103) that an organization should make a trade-off between the cost of renting additional information (e.g. an external list or information at the individual level instead of at the postal code level) and a more effective selection. However, to the best of our knowledge, there is not a single paper that discusses the analytical and statistical aspects involved. In chapter 5 we introduce a comprehensive framework for analyzing the value of information.

Before we turn to the discussion of the selection techniques, we would like to stress that an important aspect of modeling, whichever technique is used, consists of processing, auditing and understanding the data. This includes the selection of variables from a database and transformation of variables. An example of the former is whether a step-wise procedure should be used to select the explanatory variables or a so-called forced entry, i.e. including all the variables (chosen a priori). Transformation of variables is employed, for instance, in the following two situations: (1) recode a categorical variable that has (too) many categories; (2) incorporate a continuously measured variable in a flexible, i.e. nonlinear, way. A straightforward way to do so is by dividing this variable into a number of categories and using a dummy variable for each category; this is often called binning (Lix and Berger 1995).

2.4 Selection techniques

Generally, selection takes place on the basis of some measure of response. For direct mail, three kinds of responses can be distinguished, depending on the offer submitted in the mailing. The first kind concerns mailings with fixed
Selection techniques

revenues (given a positive reply), such as subscriber mailings of a magazine, membership mailings and single-shot mailings offering just one product, for example a book. A second kind concerns mailings where the response is the number of units ordered, e.g. the number of compact discs ordered by direct mail selling or the subscription time (a quarter of a year, half a year, a full year) when a magazine is offered through direct mail. Third, there are mailings with a response that can take on all positive values. This may involve total revenues in purchases from a catalog retailer, or the monetary amount donated to a charitable foundation raising funds by mail.

Nearly all the selection techniques that have been proposed deal with the case of fixed revenues to a positive reply. Although it is recognized in most of the papers published that many direct mail campaigns generate a response of which the revenues vary across individuals, it is hard to find publications that take this aspect into account. So, in the case that the revenues vary across individuals one should use the average of the individuals who responded as the fixed revenues. In chapter 3 we present several methods to incorporate the quantity of response into the selection process.

Before we turn to the discussion of the selection techniques, we want to describe an important characteristic of a technique, namely its parameterization. Since we will regularly refer to this characteristic, we introduce the parameterization issue here. The intention is to give the basic idea of the possible parameterizations rather than a thorough discussion. Three approaches can be distinguished: parametric, semiparametric and nonparametric.

In a parametric approach, it is typically assumed that the dependent variable functionally depends on the independent variables and an unobservable error, in accordance with a fixed relation. That is, the functional form between the dependent variable and the independent variables is known as well as the functional form of the distribution of the error term. The error term is introduced to account for, among other things, the lack of a perfect fit of this relation, omitted variables, or badly measured variables (i.e. measurement error). An example of a parametric specification is the classical linear regression model. It assumes that the relation between the dependent variable and independent variables is linear and that the errors are independent and identically normally distributed. The main advantage of the parametric approach is its ease of interpretation and implementation.

A nonparametric approach hardly imposes any restrictions on the functional form of the relations. Only some very weak conditions have to be set on the model. For instance, the aim of the nonparametric approach in a regression
model is to determine a reasonable approximation for the curve that describes the relation between the dependent variable and independent variables. The problem is to determine a smooth curve in such a way that it shows the structure of the data without showing all the details. In other words, one has to find a balance between fitting and smoothing (Härdle 1990). The advantage of the nonparametric approach is that there is little room for misspecification. However, the disadvantage is that even with a large data set the precision of the estimates is often poor (Powell 1994). This is caused by the slow rate of convergence; a problem which is exacerbated by many explanatory variables, the so-called curse of dimensionality.

A semiparametric approach is a hybrid of parametric and nonparametric approaches. Broadly speaking, a semiparametric approach maintains the structure of a model that is most useful for interpreting the results, but hardly imposes restrictions on the other aspects of the model. A linear regression model with an unspecified functional form of the distribution of the error term is an example of a semiparametric model. Its advantage is that it is more flexible than a parametric approach but avoids the curse of dimensionality. Estimation of semiparametric and nonparametric models can be quite awkward since the function that has to be optimized does not behave neatly. Furthermore, the results are often sensitive to the degree of smoothing.

The parametric-nonparametric distinction loosely characterizes the difference of empirical analysis by statisticians and econometricians over the last few decades. Statistical analysis of data is in essence a descriptive activity. That is, the aim is to describe the data patterns; the value of the results rests on the notion that new data sets will exhibit the similar patterns. The nonparametric approach emphasizes this view. In the words of Lehmann (1975): “Why assume anything if you don’t have to?” In contrast, econometric modeling makes use of economic theory and reasoning in two ways. First, economic reasoning suggests certain kinds of variables that have to be included in the empirical analysis. Second, and more fundamental, is that econometric modeling regards the process generating the data as a behavioral regularity. That is, an econometric model specifies an economic behavioral process, as well as the connection between that process and the observed data. The aim of the analysis is the interpretation of this process, and the value of the results depends on the question whether the specified model adequately represents new data sets. Although it may seem that this discussion is somewhat beyond the scope of this thesis, it should be realized that it indicates a fundamental difference
underlying the various models presented in the thesis. We revert to this issue in section 2.5.

We start the overview with the RFM-model, which is the pioneering method of list segmentation. Subsequently, we discuss two techniques, the contingency table and CHAID, which are sometimes classified as clustering-like methods (Cryer et al. 1989), since the analysis generates clusters as output. Then we present several specifications that focus on the binary choice, like probit and logit models, discriminant analysis, the beta-logistic model and latent class analysis. Finally, we present the neural network approach, which is an alternative to conventional statistical methods. We assume that a test mailing is sent to a relatively small sample of individuals of the mailing list. Then, one of the techniques is employed to link the (non)response to the segmentation variables. Next, the technique is used to select the most promising targets on the mailing list.

**RFM-model** Traditionally, the so-called RFM-model is most frequently used for target selection. As its name suggests, it is solely based on the RFM-variables. Each RFM-variable is split into a number of categories, which are chosen by the researcher. To each of those categories a score is assigned, which is the difference between the average response rate and the observed response rate in that category. For an individual we obtain a score by adding his category scores. This enables us to rank the individuals on the mailing list. The advantage of this method is its simplicity. The drawback is that the model does not account for multicollinearity unless the RFM-variables are mutually independent. Furthermore, it is not clear which individual should receive a mailing. Bauer (1988) extended this method by using the RFM-model in a functional relationship in which the unknown parameters are determined a priori.

**Contingency table** The simplest method of modeling response is by a contingency table. It is a two- or more dimensional matrix of which the dimensions are determined by the number of variables involved. Each of these variables is divided into a number of categories. Each specific combination of categories, i.e. one category for each variable, defines a cell. For target selection contingency tables can be used as follows: the researcher chooses the variables to be taken into account. If a variable is continuous, it has to be categorized. On the basis of these categories the individuals are scanned and assigned to a cell.
In each cell the response rate is computed, which is used to decide whether the individuals belonging to that cell should be selected. It is a nonparametric method which can be classified as non-criterion based (Magidson 1994), denoting that the clusters, i.e. cells, are not explicitly derived on the basis of a dependent variable. This implies that the continuous variables, or variables with too many categories that have to be recategorized, are categorized in a rather ad-hoc manner. The major drawback of this method is that it becomes very disorderly in the case of many variables, with the risk of having not enough observations per cell, which is the curse of dimensionality. It can, however, be very useful for a first data analysis in order to choose selection variables for further analysis.

**Chi-square automatic interaction detection (CHAID)** Automatic interaction detection (AID), the predecessor of CHAID, is based on a stepwise analysis-of-variance procedure (Sonquist 1970). The dependent variable is interval-scaled and the independent variables are categorical. AID consists of the following steps: (a) for every independent variable the optimal binary split is determined in such a way that the within-group variance of the dependent variable is minimal. That is, AID merges the categories of an independent variable which are, more or less, homogeneous with respect to the dependent variable. For each independent variable this results in an optimal binary split; (b) the variable that generates the lowest within-group variance is used to subdivide the dependent variable in accordance with its optimal binary split; (c) given this binary split, two groups are constructed. Each group is analyzed in the same manner as defined by (a) and (b). This process continues until some stopping criterion is satisfied. This criterion can be either a number of segments or number of individuals in a segment. The result is a treelike group of segments; it is a so-called tree-building method. The individuals that belong to a segment with a sufficiently high response rate should be selected.

The CHAID, developed by Kass (1980), improves the AID in several respects. CHAID merges categories of an independent variable that are homogeneous with respect to the dependent variable, but does not merge the other categories. The result of this merging process, unlike AID, is not necessarily a dichotomy. Hence, steps (a) and (b) may result in more than two subgroups. On the other hand, also in contrast with AID, only those independent variables are eligible to be split into subgroups for which the subgroups differ significantly from each other; the significance is based on a $\chi^2$-statistic. This explicitly defines a stopping rule. Note that it is a nonparametric technique since it does
not impose a functional form for the relation between the dependent and independent variables. In contrast with cross-tabulation, it may be classified as criterion-based since the splits are derived on the basis of a dependent variable. The main advantage of (CH)AID is that the output is easy to understand and therefore easy to communicate to the management. In practice it is often seen as a useful exploratory data analysis technique, in particular to discover interaction effects among variables, of which output may be used as input for other techniques with more predictive power (Shepard 1995). Magidson (1994) provides an extensive examination on CHAID.

A closely related technique is the so-called classification and regression trees (CART), which was developed by Breiman et al. (1984). The result of this technique is also a treelike group of segments. In contrast with CHAID, the independent variables may be categorical and continuous. The splits are based on some homogeneity measure between the segments formed. Like CHAID, it is a useful method for explorative data analysis and also easy to communicate. A comprehensive discussion on CART can be found in Haughton and Oulabi (1993) and Thrasher (1991).

**Linear probability model (LPM)** Regression analysis with a dichotomous dependent variable (response/nonresponse) is called a linear probability model. The predicted value can be interpreted as the response probability. This probability is assumed to depend on a number of characteristics of the individuals receiving the mailing. That is

\[ y_i = x_i \beta + u_i, \]

where \( y_i \) is a random variable that takes the value one if the \( i \)th individual responds and zero otherwise, \( x_i \) is a vector of characteristics of \( i \), i.e. RFM-variables, geographic and demographic variables etc.; \( u_i \) is a random disturbance, and \( \beta \) is an unknown parameter. The unknown parameters can be estimated by ordinary least squares (OLS). On the basis of the estimated response probability the potential targets can be ranked in order to make a selection. This model has several readily apparent problems (Judge et al. 1985, p. 757). For example, the use of OLS will lead to inefficient estimates and imprecise predictions. The relation is very sensitive to the values taken by the explanatory variables, and the estimated response probability can lie outside the \([0, 1]\)-interval.
Probit and logit model An alternative way to handle discrete response data is by the probit or logit model. These models overcome the disadvantages of the linear probability model. Probit and logit models assume that there is a latent variable, $y^*_i$, which measures the inclination to respond, which depends on a vector of individual characteristics. That is,

$$y^*_i = x'_i \beta + u_i.$$  \hspace{1cm} (2.1)

As we cannot observe the inclination but only whether or not $i$ responded, we define the dependent variable which we observe, as

$$y_i = \begin{cases} 1 & \text{if } y^*_i \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$  \hspace{1cm} (2.2)

The probit model assumes that the distribution of $u_i$ is normal and the logit model assumes that $u_i$ is logistic. With these fully parametric specifications, the model can easily be estimated by maximum likelihood (ML). The individuals for which the predicted response probabilities is sufficiently high are selected for the direct mailing campaign. In chapters 3, 4 and 5 we elaborate on the probit model in more detail.

Bult and Wansbeek (1995) argue that the selection procedure rests heavily upon the functional form of the disturbance, which could cause inconsistent estimates and would consequently lead to a suboptimal selection rule. Therefore, they use the semiparametric method developed by Cosslett (1983) to estimate the parameters. This method not only maximizes the likelihood function over the parameter space, $\beta$, but also over the space which contains all the distribution functions. Of course, other semiparametric estimators of the binary choice model, e.g. Klein and Spady (1993), could be used as well.

Beta-logistic model Rao and Steckel (1995) consider a beta-logistic model for the binary choice to allow for heterogeneity. Let $p_i$ be the probability of obtaining a response from $i$. The heterogeneity in $p_i$ is assumed to be captured by a beta distribution with parameters $a_i$ and $b_i$. That is,

$$f(p_i) = \frac{1}{B(a_i, b_i)} p_i^{a_i-1}(1-p_i)^{b_i-1}, \quad 0 \leq p_i \leq 1,$$

where $B(a_i, b_i)$ is the beta function defined by

$$B(a_i, b_i) = \int_0^1 t^{a_i-1}(1-t)^{b_i-1} \, dt.$$
Rao and Steckel favor this distribution mainly because it is flexible so that it can take on a variety of shapes (flat, unimodel, J-shaped), depending on its parameters. The mean, $a_i/(a_i + b_i)$ could serve as the response probability for individual $i$. The variation in the response probability among similar individuals is thus described by the beta distribution. The parameters $a_i$ and $b_i$ are modeled as exponential functions of the selection variables (Heckman and Willis 1977), i.e. $a_i = \exp(\beta'_1 x_i)$ and $b_i = \exp(\beta'_2 x_i)$. Loosely speaking, the $a_i$s and $b_i$s can be used to define segments. The differences in response probabilities within a segment due to variables not included in the analysis are captured by the beta distribution. Although the heterogeneity can be modeled in this way, it cannot be used for making specific predictions. Therefore the mean of the distribution is used to determine the response probability of a individuals, i.e. 0

$$p_i = \frac{\exp(\beta'_1 x_i)}{\exp(\beta'_1 x_i) + \exp(\beta'_2 x_i)}.$$  

Note that this expression is equivalent to a logit model with parameters $\beta_1 - \beta_2$, hence the name beta-logistic. These parameters can be estimated by ML.

To sum up, in the beta-logistic model a beta distribution is used to allow for heterogeneity in the response probabilities. However, there is still one parameter vector that describes the relation between the selection variables and the response probability; thus we have a logit model with parameter vector $\beta_1 - \beta_2$. Rao and Steckel (1995) propose this model since it enables them to determine the response probability for an individual that received $T$ mailings (for the same offer) but did not respond to any of these. This so-called posterior response probability, $p_{Ti}$, is given by

$$p_{Ti} = \frac{\exp(\beta'_1 x_i)}{\exp(\beta'_1 x_i) + \exp(\beta'_2 x_i) + T}.$$  

Note that this probability is decreasing in $T$. The parameters can be estimated by ML if the organization followed the strategy for at least a sample of the list, of sending individuals follow-up mailings (up to a certain maximum) until they responded. The estimated model can then be used to decide whether or not an individual should receive a follow-up mailing.

**Discriminant analysis**  In discriminant or classification analysis we try to find a function that provides the best discrimination between the respondents
and nonrespondents (see Maddala 1983, p. 16). That is, the following loss function is minimized:

\[ L = \sum_{i=1}^{n} | \hat{y}_i - y_i |, \]

where \( y_i \) is one if \( i \) responded and zero otherwise, \( \hat{y}_i = 1 \) if \( i \) is classified as a respondent and \( \hat{y}_i = 0 \) otherwise. On the basis of the estimated discriminant function, individuals can be classified in one of the two groups. Most widely used is the linear discriminant function, \( x'\beta \), which is based on a normality assumption of the explanatory variables. Hence it is a fully parametric specification. Individual \( i \) is classified in the group of nonrespondents if

\[ |x_i'\hat{\beta} - \hat{x}_0| > |x_i'\hat{\beta} - \hat{x}_1|, \]

where \( \hat{x}_0 \) and \( \hat{x}_1 \) are the averages of the explanatory variables of the nonrespondents and respondents, respectively; \( \hat{\beta} = S^{-1}(\hat{x}_0 - \hat{x}_1) \), with \( S \) the overall variance of \( x \) (see Maddala 1983, pp. 15-18). Shepard (1995) shows that if the response is 10% or less, the analysis is very sensitive to the violations of the normality assumptions. This problem can be overcome by using a logit or probit model as the discriminant function (Sapra 1991). Then \( \hat{y}_i = 1 \) if \( P(y_i = 1) \geq 0.5 \) and \( \hat{y}_i = 0 \) if \( P(y_i = 1) < 0.5 \), with \( y_i \) defined in (2.2). It is clear that the misclassification of an individual that responded is penalized as much as the misclassification of an individual that did not respond. In practice, however, the economic costs of these two kinds of misclassification are very different. Hence, the asymmetry in costs of misclassification should be taken into account. This can be accomplished by employing the following classification rule: \( \hat{y}_i = 1 \) if \( P(y_i = 1) \geq \tau \), where \( \tau \) is the ratio between the costs of a mailing and the returns to a positive reply. Note that, although this classification rule captures the asymmetry in the costs of misclassification, the estimated parameters are still based on symmetric costs of misclassification.

Manski (1975) proposes a semiparametric approach to estimate a classification model, the so-called maximum score method. This method yields consistent estimates under weak distributional assumptions. It enables one to incorporate, in a very straightforward manner, the asymmetry of misclassification (Manski 1985). Bult (1993) employs this method in a direct marketing application. The function Bult considers is

\[
L(\beta) = \sum_{i=1}^{n} \left\{ \rho \ | \ y_i - \text{sgn}(x_i'\beta) | \ I(y_i = 1) \\
+ (1 - \rho) \ | \ y_i - \text{sgn}(x_i'\beta) | \ I(y_i = -1) \right\}, \tag{2.3}
\]
where \( y_i \) is the observed response, \( y_i = 1 \) if \( i \) responded and \( y_i = -1 \) otherwise, \( \rho = (a - c)/a \), with \( a \) the revenues to a positive reply and \( c \) the cost of a mailing; \( I(\cdot) \) is an indicator function which is one if the argument is true and zero otherwise, and \( \text{sgn}(\cdot) \) is an indicator function which is one if the sign of the argument is positive and minus one otherwise. If \( \rho \) is close to one, it is economically more important to correctly classify the outcome \( y_i = 1 \) than the outcome \( y_i = -1 \). The vector \( \beta \) is estimated by minimizing \( L(\beta) \). Note that, in contrast with the discriminant function described above, the estimated parameters are based on the asymmetry of misclassification.

Model (2.3) can be estimated in LIMDEP (Greene 1992) in a straightforward manner. It can, however, be quite hard to find the optimum since the function generally only reaches a local optimum. Problems in the optimization may also arise by the fact that at least one explanatory variable must have an infinite support (Manski 1985), which is often not available. Moreover, many observations are needed since the estimator has a slow rate of convergence (Manski and Thompson 1986). This means that even 2000 observations, which is the maximum number of observations LIMDEP can handle, might not be sufficient.

Bult and Wittink (1996) apply this method to mailings with response that can take all positive values. On the basis of past purchase behavior individuals are a priori classified in segments. For each segment a classification model is estimated, which is used for the selection of individuals. The disadvantage of the approach is the fact that it is generally difficult to specify the segments a priori in a satisfactory way.

**Latent class analysis** A crucial assumption in the (binary choice) models considered so far is that individuals respond in a homogeneous way. This means that a single model, i.e. one vector of parameters, describes the relation between the selection variables and the dependent variable. This can be misleading if there is substantial heterogeneity in the sample with respect to the magnitude and direction of the response parameters across the individuals (DeSarbo and Ramaswamy 1994). We illustrate this with a very simple example. Assume that there are two regressors, \( x_1 \) and \( x_2 \) to predict \( y^* \), as defined in (2.1). Furthermore, assume that there are two segments of approximately equal size where the models are given by

\[
y^*_i = \beta_1 x_{1i} - \beta_2 x_{2i} + u_i \quad \text{for individuals in segment 1;}\\
y^*_i = -\beta_1 x_{1i} + \beta_2 x_{2i} + u_i \quad \text{for individuals in segment 2.}
\]
Thus the two regressors have an opposite effect on the inclination to respond in the two segments. The use of a single response model would yield regression coefficients $\beta_1 \approx \beta_2 \approx 0$. This shows that a single response model is of little benefit for selection if the given underlying segments drive the response behavior.

Therefore DeSarbo and Ramaswamy (1994) propose a (parametric) method that accounts for heterogeneity. Their method simultaneously derives market segments and estimates response models in each of these segments. Thus, instead of specifying a single model, this approach allows for various models. The method provides a (posterior) probability of membership of a certain segment, so-called probabilistic segmentation. Consequently, each individual can be viewed as having a separate response model whose parameters are a combination, based on the posterior probabilities, of the model parameters in the various segments.

In the validation sample it is somewhat tricky to compute these posterior probabilities. Ideally, this should be based on information that is not used to estimate the response models, but useful information to do so is not always available. Another drawback of this approach is that it often has several locally optimal solutions. Furthermore, a lot of effort is needed to specify and estimate the whole model. Since latent class models are developed for descriptive rather than predictive purposes, it would be worthwhile to compare the predictive power with other methods; this is is not provided by DeSarbo and Ramaswamy (1994).

A similar approach is given by Wedel et al. (1993). Instead of binary response data they focus on count data, i.e. the number of units ordered in a certain time period. In their application this number ranges from 0 to 75 with an average of 5.2. They use a latent class poisson regression model. This is a parametric specification which allows for heterogeneity across individuals in two ways. First, the mean number of ordered units has a discrete mixture distribution, i.e. it varies over a finite number of latent classes. Second, the mean number of ordered units varies within each segment, depending upon the explanatory variables. Like DeSarbo and Ramaswamy (1994), the model provides a posterior probability that an individual belongs to a certain segment.

**Neural Networks (NNs)** Neural Networks have been developed rapidly since the mid-eighties and are now widely used in many research areas. Also in the field of business administration, there has been a tremendous growth in recent years; see Sharda (1994) for a bibliography of applications in this field. Interest
among researchers in applying neural networks to marketing applications did not start until a few years ago (Kumar et al. 1995). With respect to target selection for direct marketing campaigns, NNs have been mentioned in several theoretical and applied journals as a promising and effective alternative to conventional statistical methods (e.g. Zahavi and Levin 1995, 1997). What makes a NN particularly attractive for target selection is its ability of pattern recognition for automatically specifying and estimating a relation between independent variables and a dependent variable. This property is especially useful when the relationship is complex.

Neural networks can be thought of as a system connecting a set of inputs to a set of outputs in a possible nonlinear way. The links between inputs, the selection variables, and outputs, the response, are typically made via one or two hidden layers. The number associated with a link from node $i$ to node $j$, $w_{ij}$, is called a weight. Figure 2.1 shows a neural network with $k$ input nodes, one hidden layer, which consists of three nodes, and one output node. The input to a node in the hidden layer or output layer is some function of the weighted (using $w_{ij}$) combinations of the inputs. There are several methods to determine the optimal weights for a given NN; the most commonly used method is the back-propagation algorithm.

In a NN, the individual’s likelihood of purchase is expressed by means of a NN score which, although it cannot be interpreted as response probabilities, can be used to rank individuals. Thus, the higher the score, the higher the likelihood of response (Zahavi and Levin 1995). Since we may not interpret the NN score as response probability, it is difficult to determine an optimal selection rule, i.e. a critical score. The individuals with a score above this critical score should receive a mailing and the other individuals should not. Zahavi and Levin (1995, 1997) suggest to use the score that maximizes the profits for the estimation sample as the critical score.

Although NNs are positioned as a method differing from existing statistical methods, which is emphasized by the completely different terminology, there is a close link between these two. Cheng and Titterington (1994) point out the similarities between statistical methods and NNs. Ripley (1994) shows the analogy between nonlinear discrimination analysis and NNs. Thus, many special cases of the neural networks are similar to statistical models. Consequently, in these cases the NN score can be interpreted in the statistical sense, i.e. as (estimated) response probabilities. Hence, an optimal selection rule can be determined. Similarly, a frequently used criticism on NNs is that it is more or less a ‘black box’, and that as a consequence the resulting specification is
 Target selection techniques

difficult to interpret. Given the link between NNs and statistical models, it is clear that this is not always a valid criticism.

The difficulty with a NN is that when the data are ‘noisy’, there will be a tendency during the modeling process to correlate noise in the input data with noise in the output data (e.g. overfitting). Unfortunately, in (direct) marketing application this is very often the case. This is the main reason why NNs did not turn out to be the expected breakthrough in target selection for direct marketing (e.g. Chauhan 1995, Zahavi and Levin 1995, 1997).

2.5 Model comparison

From the various possible selection techniques, we want to select the most appropriate. However, there are many - sometimes conflicting - criteria to evaluate and compare the techniques. Four broad classes of criteria are:

1. The theoretical foundations of the model. This basically refers to the parametric-nonparametric distinction. In the nonparametric approach the aim is to describe the data adequately whereas in the parametric approach an (economic) theory specifies the behavioral process and hence the model (see page 14). This includes the statistical assumptions and the choice of variables.

2. Goodness-of-fit measures. For example, the (adjusted) $R^2$ can be used to evaluate and compare regression based techniques.
3. **Predictive validity statistics.** These include measures such as the mean squared error and median relative absolute error. These statistics are based on the so-called split-sample analysis. That is, part of the data set is used for estimation, the estimation sample; the other part is used for validation, the validation set. It is a classic cross-validation technique which is commonly employed for models that are built for prediction. Among other things, the analysis guards against overfitting the data.

4. **Practical aspects of the model.** These include the ease of use, interpretability of the output, intuitive appeal, computer time and so on.

   Although these criteria may determine the choice of a technique and may be useful for specifying a model, the bottom-line is to compare the models on the basis of their list segmentation performance. That is, which model generates the highest net profits? Hypothetically, if a method exists about which nothing is known but which guarantees to provide the superior selection at any time, it “would be a sufficient argument to use that method” (Lix and Berger 1995). Note that this argument coincides with that of Lehmann (1975) with respect to the choice between a nonparametric and a parametric approach (see page 14). The argument of Lix and Berger, however, does not mean that theory or ‘understanding’ is unimportant. On the contrary, the best predictive models are often those that are generated or guided by theoretical concerns (e.g. Leahy 1992). In addition, the reliability of such models is also enhanced when the processes by which the observed data are generated are understood. Similarly, practical aspects, goodness-of-fit measures and predictive measures may play an important role in model choice and variable selection.

   A method that is often used to analyze the performance with respect to net profits is the gains charts or Pareto curves (e.g. Banslaben 1992, Magliozzi and Berger 1993). This method consists of the following steps: first, the individuals are ranked on the basis of the estimated model; second, groups of equal size are defined. Usually, ten groups are considered. The first group contains the first 10% of the rank-ordered individuals, the second group the following 10% of the rank-ordered individuals, etc. Finally, the average actual and cumulative response probabilities in each of the groups are determined. The gains chart or Pareto curve gives particular insights into the performance of the model. A gains chart should be downward sloping, ideally quite steep. The gains chart of the estimation sample provides useful information to judge the performance of the model on the validation sample. The gains chart of the validation sample could be used to compare various models. It can be rather difficult, however, to compare various models on the basis of this method, since the differences are
Target selection techniques
generally small and the curves may intersect. That is, one model is better able
to identify the top-segments whereas the other model better identifies the mid-
range. It is clear that the former is preferred when the organization wants to
select the top-segment but it is not necessarily the preferred model when many
more individuals are selected. Hence, the gains chart is helpful in comparing
models but still does not answer the bottom-line of which method generates
the highest net profits. This question can only be answered by specifying a
selection rule a priori. Therefore, Bult and Wansbeek (1995) compare models
on the basis of an optimal cutoff point, which is defined as the point where the
marginal costs equal the marginal revenues; they call this the profit maximizing
approach. We take this approach as the validation criterion for our models.

To summarize, there are various ways to compare selection techniques. Our
view, which lies at the basis of the models developed in this dissertation, entails
that the models should be compared on the basis of a selection rule, which is
obtained by equating marginal cost and marginal returns. It does not mean,
however, that we believe that the other ways to compare the techniques are
useless. Given the diverse criteria for model selection it is, unfortunately, im-
possible to come up with a model that is universally preferred. In the remainder
of this section we will review the scarce literature on model comparisons.

Magidson (1988) compares the CHAID with log-linear modeling. Al-
though the paper reports a successful application of the CHAID, there is
no comparison with different techniques using real data. Bult and Wansbeek
(1995), on the basis of the profit maximizing approach, show that CHAID
performs better than the LPM, parametric, and semiparametric approach of
the binary choice model (i.e. logit model and Cosslett’s approach) on the es-
timation sample. However, as expected, CHAID performs considerably worse
than the three other methods on the validation sample. Surprisingly, the para-
metric approach of the binary choice model outperforms the semiparametric
approach. In contrast, Bult (1993) shows that the semiparametric approach of
the discriminant model performs better than the parametric approach (logit
model).

Within the field of direct marketing, neural networks are not more success-
ful than traditional methods. For instance, Kumar et al. (1995) are inconclusive
as to the relative performance of NN in comparison with log-linear regression.
Lix and Berger (1995) show that a particular NN performed better than regres-
sion based methods, whereas an other network performed worse. In a rather
comprehensive study, Zahavi and Levin (1997) demonstrate that logistic re-
gression models achieve approximately the same results as NN. They conclude
that “these results are not encouraging for the neural network approach since
the process of configuring and setting up a NN is not straightforward”.

With respect to the question whether it is preferred to force variables to enter
in the regression equation versus stepwise regression, two studies have been
performed. Both compare the models on the basis of a gains chart. Magliozzi
and Berger (1993) show that stepwise regression performs better whereas Lix
and Berger (1995) show the opposite.

Altogether, three general conclusions may be made. First, data-exploration
methods such as CHAID are less useful for selection than regression type
models (e.g. probit or discriminant model). Second, regression models per-
form reasonably well and, with step-wise entry, have the advantage of clear
interpretation. Third, neural networks did not turn out to be the revolution in
target selection since the results are comparable to the conventional statistical
techniques.

2.6 Discussion and conclusion

We have discussed various techniques that can be employed to select indi-
viduals from a mailing list. Unfortunately, it is not possible to draw general
conclusions as to the best model for target selection. The main reason for this
is that many - sometimes conflicting - criteria play a role in qualifying a tech-
nique. Furthermore, even from a theoretical point of view we cannot define a
ranking of the models. Moreover, there is no literature that compares all the
proposed methods on all these criteria.

As to future research, it would be very interesting to obtain a general con-
clusion with respect to the generated net profits, given a certain selection rule.
It should be realized, however, that there is a number of aspects which could
affect the performance of the models. For instance, the kind of information
that is available, number of individuals in the estimation sample and the re-
sponse rate. It is likely that these aspects will make it harder to obtain such a
conclusion.

All the techniques discussed in this overview deal with the case of fixed
revenues to a positive reply. That is, the approaches focus on binary choice
modeling. However, most direct marketing campaigns do not generate simple
binary response but rather a response of which the revenue varies between the
individuals. Hence, a striking lack in the literature is that hardly any attention
has been given to this kind of response. This is not only theoretically a matter
of concern but it can be of major practical importance. In chapter 3 we fill this gap by incorporating the quantity of response in the selection process.

In the selection process considered so far, the organization specifies a model and formulates an (optimal) selection rule in order to select the targets. This selection rule is implicitly based on the assumption that the parameters are known. Of course, the parameters are unknown and have to be estimated. These estimates are then plugged into the selection rule, assuming that they are the true values. It is, however, well known from the statistical literature on decision making that separation of estimation and decision making, as is the case here, generally yields lower profits. The reason for this is that estimation uncertainty is neglected. In chapter 4 we show how these two steps can be integrated. This leads to a better selection rule and hence to higher expected profits.
Chapter 3

Target selection by joint modeling of the probability and quantity of response

3.1 Introduction

For direct mail, three kinds of responses can be distinguished, depending on the offer submitted in the mailing. The first kind concerns mailings with fixed revenues (given a positive reply), such as subscriber mailings of a magazine, membership mailings, and single-shot mailings offering just one product, e.g. a book. A second kind concerns mailings where the number of units ordered can vary, e.g. the number of compact discs ordered by direct mail selling or the subscription time (a quarter of a year, half a year, a full year) when a magazine is offered through direct mail. Third, there are mailings with a response that can take on all positive values. This may involve total revenues in purchases from a catalog retailer, or the monetary amount donated to a charitable foundation raising funds by mail.

Nearly all of the target selection techniques that have been proposed (see chapter 2) deal with the case of fixed revenues to a positive reply and hence concentrate on binary choice modeling. Thus, the quantity of response is implicitly assumed to be equal across individuals. Although the literature recognizes that most direct mail campaigns do not generate simple binary response but rather response of which the quantity varies over the individuals, it is hard to find publications that take this aspect into account. Simon (1987), for example, suggests to take the average quantity of purchases from a random sample of the customers on the mailing list over a couple of years and use this as the expected value of an individual. Then the response to a positive reply is considered fixed as yet and the response can be modeled again by a binary choice model. Rao and Steckel (1995) suggest to use an ordinary least squares
(OLS) model to determine the expected revenues, and obtain the total revenue just as the expected revenues times the probability of response. However, their empirical example is just a binary choice model.

Recently, Bult and Wittink (1996) proposed a method to incorporate the quantity of response. On the basis of the quantity of response in the last year, individuals are classified in segments a priori. Within each segment the quantity of response is assumed to be this fixed value. Then, for each segment a binary response model is estimated, which is used for selection. Note that (1) this approach is only applicable if information on past behavior is available and (2) it heavily depends on the assumptions that this year's quantity of response is equal to that of last year.

The purpose of this chapter is to present a unified framework for modeling response to a direct mailing, in order to optimally select individuals for a mailing campaign. Our framework specifies the relevant decisions taken by the individuals. These decisions are (a) whether to respond or not, and, in the case of response, (b) the quantity of response. As is argued by Courtheoux (1987), higher profits can be obtained when both decisions are modeled jointly. We specify a model that takes both decisions into account. This model constitutes the basis for a selection rule which can be used to optimize profits. In order to make the model operational we distinguish three approximations for this selection rule. For reasons of comparison we also consider simplified versions of this method that concentrate on one of the individuals’ decision dimensions. In addition, we specify a model that assumes that both dimensions are driven by the same structure, i.e. the tobit model. An empirical application shows considerably higher profits when both decisions are modeled explicitly relative to modeling the probability of response only.

The chapter, which is based on Otter, Van der Scheer, and Wansbeek (1997), is structured as follows: in section 3.2 we present a simple response model that consists of two components. The first component models the probability of response and the second component is used to model the quantity of response. This allows us to formulate a profit maximizing selection rule that takes both dimensions into account. Section 3.3 is devoted to this. Some simplifying approximations to this rule are presented in section 3.4. Section 3.5 adds details on practical operationalization. In section 3.6 we discuss the tobit model. To show how the various approaches behave in practice, we describe the data underlying the empirical illustration in section 3.7. The results are presented in section 3.8. Section 3.9 summarizes and concludes.
3.2 Description of the model

Consider a direct marketing organization that has to make the decision whether to send an individual a mailing or not. In case a mailing is sent, the profit of the organization, $\Pi$, is given by

$$\Pi = AR - c,$$

where $R$ is the random variable given by

$$R = \begin{cases} 1 & \text{if the individual responds} \\ 0 & \text{if the individual does not respond} \end{cases}$$

$A$ is the random variable that denotes the quantity of response and $c$ is the cost of a mailing. We assume that $R$ is driven by a probit model. We denote the inclination to respond by the latent variable $R^*$ that satisfies a linear model,

$$R^* = x'\beta + v,$$

where $x$ is a $k \times 1$ vector of explanatory variables, $\beta$ is a $k \times 1$ vector of regression coefficients, and $v \sim N(0, 1)$, independently from $x$; $x$ is assumed to be a random variable with unknown distribution. Response is indicated by the observed dummy variable $R$ that is related to $R^*$ in the following way: $R = 1$ if $R^* \geq 0$ and $R = 0$ otherwise. Hence the response probability of an individual is given by

$$P(R = 1 | x) = \Phi(x'\beta),$$

with $\Phi(\cdot)$ is the standard normal distribution function. If $R = 1$ the quantity of response also satisfies a linear model,

$$A = x'\gamma + u,$$

with in particular the assumption

$$E(u | R = 1, x) = 0.$$  

We call this the quantity model. For convenience of notation we assume the same $x$ in both relations but this is innocuous since elements of $\gamma$ and $\beta$ can a priori be set at zero. The disturbance terms in both relations, $v$ and $u$, may correlate but this will play no role in what follows. This way of modeling
Target selection by joint modeling of the probability and quantity of response

probability and corresponding quantity is called a two-part model, henceforth TPM (cf. Duan et al. 1983).

On the basis of a test mailing we derive estimates of the model parameters. We assume that the sample used for the estimation is so large that we can neglect the differences between estimators and the true values; hence we assume for the moment that \( \gamma \) and \( \beta \) are known. We define

\[
\begin{align*}
    p & \equiv \Phi(x' \beta) \\
    a & \equiv x' \gamma,
\end{align*}
\]

which are random variables with a joint density function. We denote the marginal density (with respect to \( p \)) of \( a \) conditional on \( R = 0 \) and \( R = 1 \) by \( f_0(a) \) and \( f_1(a) \), respectively, and the corresponding distribution function by \( F_0(a) \) and \( F_1(a) \), respectively.

**Sample selection model**

One of the advantages of modeling the response with the TPM, i.e. with (3.2) and (3.3), is that the two parts can be estimated separately. Note, however, that the error terms, \( v \) and \( u \), may very well be correlated but that this correlation does not affect the separability (e.g. Duan et al. 1983). The sample-selection model or, as Amemiya (1985, chapter 10) classifies it, a type-2 tobit model (henceforth SSM), is an alternative model that has been proposed in the econometric literature for related problems in which there are two decisions to be taken. Examples include job-search models (Blundell and Meghir 1987, Heckman 1976, and Mroz 1987), insurance claims models (Hsiao et al. 1990), modeling charitable donations (Garner and Wagner 1991, and Jones and Posnett 1991), and sales promotions models (Chintagunta 1993, and Krishnamurthi and Raj 1988).

The SSM assumes that the probability and quantity of response are based on a bivariate normal distribution. Unlike the TPM, the quantity of response is modeled unconditionally on the probability of response. That is, (3.3) is assumed to be the underlying latent model for all the individuals; \( A \), however, is only observed when \( R = 1 \). Hence, the main difference between the TPM and SSM is that the latter assumes \( E(u \mid R = 1 \text{ or } R = 0, x) = 0 \) instead of \( E(u \mid R = 1, x) = 0 \). Consequently, least squares using only the sample with positive values, i.e. the individuals of which \( R = 1 \), will provide biased estimates of
the unconditional response if the errors are correlated (see Heckman 1979),
the so-called sample-selection bias. That is

\[ E(A \mid R = 1, x) = E(x' \gamma + u \mid x' \beta + v > 0) \]
\[ = x' \gamma + E(u \mid v > -x' \beta) \]
\[ = x' \gamma + \rho \frac{\phi(x' \beta)}{\Phi(x' \beta)}, \quad (3.5) \]

where \( \rho \) is the correlation between \( u \) and \( v \), and \( \phi(\cdot) \) is the standard normal
density. Hence, if \( \rho \neq 0 \), regression estimators of the parameters of (3.3) based
on the sample with positive values omit the second term of (3.5) as a regressor.
Thus, the bias in the TPM model results from the problem of omitted variables.
There is, however, a number of reasons why we favor the TPM over the SSM.

First, in our case there is no obvious reason to use the SSM model since
our problem is not to account for the sample-selection bias but to model the
nonresponse. A standard example of the SSM is a wage equation combined
with a binary employment choice equation. The wages for someone who
does not work have a clear interpretation: potential earnings if he would find
a job. In our case, however, purchases of nonrespondents are by definition
zero. Whereas a continuous distribution of positive potential wage rates in the
population of workers and nonworkers makes sense, the concept of positive
potential quantity of response of individuals who do not respond does not seem
very useful. The same argument is used, among others, by Duan et al. (1983),
and by Melenberg and Van Soest (1996) to model health care expenditures and
vacation expenditures, respectively.

Secondly, the TPM is far less susceptible of misspecification of the distribu-
tional assumptions (e.g. Hay et al. 1987, and Manning et al. 1987). Arabmazar
and Schmidt (1982), and Goldberger (1981) show that even the maximum
likelihood estimator (MLE) of a simple form of the SSM, i.e. the standard tobit
model, suffers from a substantial inconsistency under non-normality and het-
eroscedasticity. Thus, only if the distributional assumptions are correct should
the SSM - at least theoretically - be preferred. Unfortunately, the assumption in
the sample-selection model regarding the censored part is untestable because
the censored data are not observed. Apart from using a more robust model, we
can specify a model with less strict distributional assumptions and employ, for
example, a semiparametric estimation method to estimate the parameters. For
the SSM, semiparametric estimators have been proposed by Ahn and Powell
(1993), and Cosslett (1991), among others. A drawback of these methods is
that estimation can be quite awkward since the function that has to be optimized does not behave neatly and the results are often sensitive to the choice of the smoothing parameter.

Thirdly, we use the model only for predictive purposes and we are not interested in the parameter values per se. This in contrast with the wage equation, for example, in which the researcher and policy maker are interested in the parameters as well as the sample-selection effect. Moreover, a direct marketing organization prefers using the information with the largest predictive power. If available, this is usually information on past purchase behavior of the individuals (e.g. Bauer 1988). Consequently, the model specifies a statistical relation rather than a structural relation. Generally speaking, it is not really necessary to specify the full data generating process, which may be obscured by lagged (censored) dependent variables. Furthermore, Duan et al. (1983), Hartman (1991), Hay et al. (1987), and Manning et al. (1987) in extensive Monte Carlo studies show that even if the SSM is the true model, i.e. if the errors have a bivariate normal distribution, the TPM works very well.

Fourthly, the second part of the TPM can be estimated in a straightforward manner, in contrast with the SSM which may suffer from optimization problems. Instead of estimating the second part by OLS, several semiparametric techniques can be used (e.g. Melenberg and Van Soest 1996). Again, these may give computational difficulties. Computational advantages are important in direct marketing since the model has to be used regularly and some experimentation is needed to decide which variables to include in the model. Furthermore, when the second part is estimated by OLS traditional residual analyses can be used to evaluate the appropriateness of the specification.

3.3 Optimal selection

We now turn to the selection problem. We assume the presence of a mailing list containing information on $x$s, hence on the implied $p$s and $a$s. We wish to determine the subset of the $\{p, a\}$ space such that selection of individuals from this space maximizes expected profit. We follow the strategy of conditioning on $p$ and determine the threshold $a^\ast (:= a^\ast(p))$ above which a mailing is sent. We determine $a^\ast$ by maximizing the expected profit given $p$.

So we are interested in

$$E \equiv E(\Pi \mid p, a \geq a^\ast)P(a \geq a^\ast \mid p)$$

$$= E(AR - c \mid p, a \geq a^\ast)P(a \geq a^\ast \mid p)$$
This should be maximized with respect to $a^*$. At this point we introduce a simplifying approximation that greatly improves the analytical tractability of the maximization just defined. In (3.6) there are conditional expectations and probabilities where the condition involves both $p$ and $R = 1$ or $R = 0$. Both types of conditioning overlap to a certain extent. Intuitively, the optimal profit to be obtained may not be affected too much if we omit the conditioning with respect to $p$. This yields the following approximation

$$E \approx E(A \mid R = 1, a \geq a^*)P(a \geq a^* \mid R = 1)p$$

$$- c \{P(a \geq a^* \mid R = 1)p + P(a \geq a^* \mid R = 0)(1 - p)\}$$

$$= E(A \mid R = 1, a \geq a^*)P(a \geq a^* \mid R = 1)p$$

$$- c \{(1 - F_1(a^*))p + (1 - F_0(a^*))(1 - p)\}$$

$$= E(A \mid R = 1, a \geq a^*)P(a \geq a^* \mid R = 1)p$$

$$- c \{1 - F_1(a^*)p - F_0(a^*)(1 - p)\}$$

$$= E(a \mid R = 1, a \geq a^*)P(a \geq a^* \mid R = 1)p$$

$$- c \{1 - F_1(a^*)p - F_0(a^*)(1 - p)\}$$

(3.7)

where the third strict equality is based on (3.4). In order to get some insight in the key aspect of our method we rewrite (3.7) as

$$E(a - c \mid R = 1, a \geq a^*) \left(1 - F_1(a^*)\right) p - c \left(1 - F_0(a^*)\right) (1 - p).$$

(3.8)
This expression indicates the two groups that have to be distinguished in order to derive the optimal selection rule, i.e. the threshold $a^*$. The first group consists of the correctly selected individuals, i.e. the individuals who responded ($R = 1$) and will be selected ($a \geq a^*$). The probability of being correctly selected is given by $1 - F_1(a^*)$. The second group consists of the individuals who will be selected ($a \geq a^*$) while they did not respond ($R = 0$). The probability of incorrect selection is $1 - F_0(a^*)$. The expected profits of these two groups are $E(a - c \mid R = 1, a \geq a^*)$ and $-c$, respectively. By weighing the expected profits with the probability of occurrence, $1 - F_1(a^*)$ and $1 - F_0(a^*)$, the optimal selection rule can be derived. Note that we are able to distinguish these two groups since the selection rule is explicitly incorporated in the expression for the expected profit. As we will demonstrate in section 3.4, this provides a better selection rule than in an approach in which the selection rule is not incorporated in the expected profit.

The first-order condition of (3.8) with respect to $a^*$ is

$$\frac{\partial E}{\partial a^*} = -a^* f_1(a^*) p + c \left\{ f_1(a^*) p + f_0(a^*) (1 - p) \right\} = 0$$

or

$$a^* p = c \left\{ 1 + \left( \frac{f_0(a^*)}{f_1(a^*)} - 1 \right) (1 - p) \right\}.$$  

(3.9)

This is an implicit equation in $a^*$, which can be solved numerically since the densities $f_0(\cdot)$ and $f_1(\cdot)$ are known functions in the sense discussed above. The result is a curve in the $(p, a)$ space separating the profitable from the non-profitable individuals. For simplicity of notation we omit the asterisk superscript to $a$ when further discussing this curve below. The mailing region, denoting the individuals to whom a mailing should be sent, in the $(p, a)$ space, is given by

$$M \equiv \left\{ (p, a) \mid a \geq \frac{c}{p} \left\{ 1 + \left( \frac{f_0(a)}{f_1(a)} - 1 \right) (1 - p) \right\} \right\},$$

(3.10)

which follows directly from (3.9).

### 3.4 Approximations

In order to make (3.9) operational we distinguish three increasingly precise but complex approximations to the solution of (3.9). The first one, further
on referred to as approximation I, neglects the difference between the two densities. Hence

\[ a = \frac{c}{p} \quad (3.11) \]

This is simply an orthogonal hyperbola in the \((p, a)\) space. It coincides with the approach in which the selection rule, \(a \geq a^*\), is not explicitly incorporated in the expected profit. That is, the mailing region is simply defined by the \((p, a)\) space for which \(E(\Pi \mid p, a) = ap - c \geq 0\).

The second approximation (II) does more justice to the difference between the two densities. We make the (evidently crude) working hypothesis that both densities are normal with the same variance \(\sigma^2\) but with different means, \(\mu_0\) and \(\mu_1\), in obvious notation. Let \(\tilde{\mu} \equiv (\mu_0 + \mu_1)/2\) and \(\delta \equiv \sigma^{-2}(\mu_1 - \mu_0)\). Then

\[
\frac{f_0(a)}{f_1(a)} - 1 = \exp \left\{ -\frac{1}{2} \left( \frac{a - \mu_0}{\sigma} \right)^2 + \frac{1}{2} \left( \frac{a - \mu_1}{\sigma} \right)^2 \right\} - 1 \\
\approx \frac{1}{2} \left\{ \left( \frac{a - \mu_1}{\sigma} \right)^2 - \left( \frac{a - \mu_0}{\sigma} \right)^2 \right\} \\
= -\frac{1}{2} \left( \frac{\mu_1 - \mu_0}{\sigma} \right) \left( \frac{2a - (\mu_1 + \mu_0)}{\sigma} \right) \\
= -\delta (a - \tilde{\mu}),
\]

so

\[ ap = c \{ 1 - \delta (a - \tilde{\mu})(1 - p) \} \]

or

\[ a = c \frac{1 + \tilde{\mu} \delta (1 - p)}{p + c \delta (1 - p)} \quad (3.12) \]

Again, this is an orthogonal hyperbola in the \((p, a)\) space; for \(\delta = 0\), which holds when \(\mu_0 = \mu_1\) and when \(\sigma^2 \to \infty\), this reduces to the result of approximation I.

The third approximation (III) is obtained by employing a nonparametric technique to approximate the densities; we denote these by \(\hat{f}_0(\cdot)\) and \(\hat{f}_1(\cdot)\). Then

\[ a = \frac{c}{p} \left( 1 + \left( \frac{\hat{f}_0(a)}{\hat{f}_1(a)} - 1 \right) (1 - p) \right) \quad (3.13) \]

defines the boundary of the mailing region, which is a curve in the \((p, a)\) space.
Now we have specified three approximations, (3.11), (3.12), and (3.13), which can all be represented by curves in the \((p, a)\) space. Examples of these curves can be found in figure 3.2 (p. 46), which will be discussed in section 3.8. It is illuminating to look at these three curves and the corresponding mailing regions in more detail. The three curves come together in \((1, c)\), denoting that with a response probability of one the expected quantity should be at least the cost. Obviously, this holds for the three approximations. The effect of the approximations on the mailing region depends on the ratio \(f_0(a)/f_1(a)\). This ratio equals one if approximation I holds, and if \(a = \bar{a}\) in approximation II. Hence, the first two curves intersect in \((c/\bar{a}, \bar{a})\). As is obvious from (3.13), this is also a point on the third curve if \(f_0(\bar{a}) = f_1(\bar{a})\). This is the case when e.g. the two densities are symmetric, as in the second approximation, but will in general only hold approximately. If \(f_0(a)/f_1(a) > 1\), i.e. \(-\delta(a - \bar{a}) > 0\) in approximation II and \(f_0(a)/f_1(a) > 1\) in approximation III, the mailing region can be written as \(M = \{(p, a) \mid a > \gamma \lambda\}\), with \(\lambda > 1\). Hence, the mailing regions of approximations II and III are smaller than that of approximation I. Consequently, fewer individuals should be selected. For approximation II this holds if \(a < \bar{a}\). If \(f_0(a)/f_1(a) < 1\), the opposite holds. Thus, the mailing regions of approximations II and III expand with respect to approximation I, which implies that more individuals should be selected. For approximation II this holds if \(a > \bar{a}\).

### 3.5 Operationalization

The three approximations defined in the previous section specify three methods to select individuals from the mailing list. In order to make the methods operational we need estimates of \(\gamma\) and \(\beta\) using the results of a test mailing on a subset of the mailing list. The test mailing produces respondents and nonrespondents and for respondents a response quantity. We follow the simplest approach, and estimate \(\gamma\) by OLS on (3.3) using the data on the respondents, and probit on (3.2) using the results for respondents and nonrespondents. Given these estimates we impute for all individuals, \(\hat{a}\) as \(x'\hat{\gamma}\) and \(\hat{p}\) as \(\Phi(x'\hat{\beta})\). The values of \(\hat{a}\) are used to estimate \(\mu_0\), \(\mu_1\) and \(\sigma^2\) to operationalize approximation II.

In order to operationalize approximation III we further need nonparametric estimates of \(f_0(a)\) and \(f_1(a)\). We use a simple approach and employ the
Gaussian kernel (see e.g. Silverman 1986). Let $\phi(\cdot)$ denote the standard normal density, then

$$
\hat{f}_0(a) = \frac{1}{n_0 h} \sum_{i=1}^{n_0} \phi \left( \frac{\hat{a}_{a_0} - a}{h} \right),
$$

$$
\hat{f}_1(a) = \frac{1}{n_1 h} \sum_{i=1}^{n_1} \phi \left( \frac{\hat{a}_{a_1} - a}{h} \right),
$$

where the first subscript to $\hat{a}$ is 0 for the $n_0$ nonrespondents in the test mailing and 1 for the $n_1$ respondents; $n \equiv n_1 + n_0$. For the smoothing parameter $h$ we choose $h = 1.06 \omega n^{-1/5}$, where $\omega$ is the standard deviation of $\hat{a}$ (Silverman 1986, p. 45). Since we estimate two functions we have two smoothing parameters. In order to have only one smoothing parameter, we use the weighted average of these two.

We have now operationalized three methods for selection, which are straightforward to use. We consider each individual in its turn to check whether its value $\hat{a}$ falls in the mailing region. Of course, the mailing regions of the three methods differ (e.g. figure 3.2). We revert to this issue when we discuss the empirical application. First, however, we want to describe four additional methods that we will employ to put the results of the three methods introduced so far into perspective. These additional methods are all simplifications of the methods discussed above since they set either $\hat{a}$ or $\hat{p}$ at a fixed value.

The first of these additional methods is based on substituting the average quantity of response from the respondents in the test mailing, denoted by $N_a$, for $\hat{a}$ for all individuals. That is, we neglect the variation in the quantity of response. The selection rule is then based on approximation I, equation (3.11). Hence, an individual is selected solely according to his value $\hat{p}$ and selection takes place if $\hat{p} \geq p_c \equiv c/\hat{a}$, where $p_c$ is called the optimal cutoff probability. This method is interesting since it comes closest to current practice: select an individual if the probability of response exceeds the ratio of cost to (average) yield. Thus, the response probability is modeled but the quantity of response is not.

The other three additional methods have the opposite point of departure and are based on modeling the quantity of response but not the response probability. The response fraction $\hat{p}$ from the test mailing is assigned to all individuals. An individual is selected $\hat{d} \geq a_c \equiv c/\hat{p}$, where $a_c$ is called the optimal cutoff quantity. In other words, in the second additional method, we confront $\hat{a}$ with the first approximating curve, i.e. equation (3.11), at the point
Table 3.1: Summary of the methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Approximation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( p \hat{a} )</td>
</tr>
<tr>
<td>2</td>
<td>( \tilde{p} \ a )</td>
</tr>
<tr>
<td>3</td>
<td>( \tilde{p} \ a )</td>
</tr>
<tr>
<td>4</td>
<td>( \tilde{p} \ a )</td>
</tr>
<tr>
<td>5</td>
<td>( \hat{p} \ a )</td>
</tr>
<tr>
<td>6</td>
<td>( \hat{p} \ a )</td>
</tr>
<tr>
<td>7</td>
<td>( \hat{p} \ a )</td>
</tr>
</tbody>
</table>

\( p = \tilde{p} \). Two variations of this method are obtained by confronting \( \hat{a} \) also with the other two, more sophisticated approximating curves, also at \( p = \tilde{p} \). Hence, the third additional method is approximation II, i.e. equation (3.12), with \( p = \tilde{p} \). Similarly, the fourth additional method is approximation III, i.e. equation (3.13), with \( p = \tilde{p} \).

Table 3.1 summarizes the seven methods thus obtained, ordered in increasing degree of sophistication. The first column labels the methods; the second column has \( p \) if the response probability is modeled and has \( \tilde{p} \) if the response fraction from the test mailing is used. The third column has analogous entries as to \( a \). The fourth column indicates which of the three approximating curves is used. Thus, methods 5, 6 and 7 correspond to equations (3.11), (3.12) and (3.13), respectively. Methods 2, 3 and 4 are based on modeling the quantity of response only, and method 1 only models the probability of response.

Illustration of the mailing regions of approximation I

In figure 3.1 we illustrate the mailing regions for the three methods of approximation I (methods 1, 2 and 5). The mailing region of method 5 is bounded by the curve, defined by \( a = c/p \) (equation (3.11)). All the individuals to the north-east of this curve should receive a mailing.

In order to compare the mailings regions of methods 1 and 2, we split the potential mailing region into seven parts. These parts are divided by the straight lines and are denoted by \( \mathcal{M}_j \), \( j = 1, \ldots, 7 \). In method 1, the individuals for whom \( p \geq p_c \) should be selected, hence \( \mathcal{M}_2, \mathcal{M}_4, \mathcal{M}_6 \) and \( \mathcal{M}_7 \) define the mailing region. In method 2, an individual is selected if \( a \geq a_c \), so the mailing region consists of \( \mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3 \) and \( \mathcal{M}_4 \). Thus, methods 1 and 2 both include
Figure 3.1: Illustration of the mailing regions of approximation I. The area to the north-east of the curve, $a = c/p$, is the mailing region for method 5. To the right of $p = p^*$ is the mailing region of method 1; above the line $a = a_*$ the mailing region of method 2.

$\mathcal{M}_3$ and $\mathcal{M}_4$ and do not include $\mathcal{M}_6$. Note that the expected profits of $\mathcal{M}_3$ and $\mathcal{M}_6$ are negative since the maximally attainable profits in these regions are $\tilde{a} p_c - c$ and $a_c \tilde{p} - c$, respectively, which both are zero. Furthermore, $\mathcal{M}_1$ and $\mathcal{M}_7$ have positive expected profits only in the region to the north-east of the curve. Thus, the difference between methods 1 and 2 is determined by $\mathcal{M}_1$, $\mathcal{M}_3$, $\mathcal{M}_6$, and $\mathcal{M}_7$. Apart from estimating $a$ and $p$, the difference in the actual net profits depends of course on the number of individuals in each region.

Note that the mailing region of method 5 includes the profitable part of $\mathcal{M}_1$ and $\mathcal{M}_7$ and excludes a part of $\mathcal{M}_6$. Since this method also excludes the non-profitable regions $\mathcal{M}_3$ and $\mathcal{M}_6$, it should generate higher profits.

### 3.6 Tobit model

In the models considered so far, we implicitly assumed that the two decisions are driven by a different structure. Alternatively, we can assume that the same underlying structure drives both decisions. That is, we treat the term $AR$ of
(3.1) as a single variable, \( y \), which depends on \( x \). The dependent variable \( y \), which also specifies the quantity of response, can only take non-negative values. Since \( y \) is zero for a large number of individuals, OLS is not appropriate here. A model that accounts for the zeros is the tobit model (Tobin 1958). This model is defined by

\[
y^* = x'\theta + e
\]

(3.14)

with

\[
y = \begin{cases} 
    y^* & \text{if } y^* > 0 \\
    0 & \text{otherwise}, 
\end{cases}
\]

where \( e \) is independent and identically distributed as \( N(0, \sigma^2) \). Instead of observing the latent variable \( y^* \), we observe the variable \( y \) that is either zero or positive. Like the SSM, the tobit model is sensitive to misspecification of the error term. That is, the MLE is inconsistent under non-normality or heteroscedasticity (e.g. Arabmazar and Schmidt 1982, and Goldberger 1981).

The expected profit for an individual that receives the mailing is given by

\[
E(\Pi \mid x) = E(y-c \mid y^* > 0, x)P(y^* > 0 \mid x) + E(y-c \mid y^* \leq 0, x)P(y^* \leq 0 \mid x)
\]

\[
= E(y \mid y^* > 0, x)P(y^* > 0 \mid x) - c
\]

\[
= E(x'\theta + e \mid x'\theta + e > 0, x)P(x'\theta + e > 0 \mid x) - c
\]

\[
= x'\theta P(e > -x'\theta \mid x) + E(e \mid e > -x'\theta)P(e > -x'\theta \mid x) - c
\]

\[
= x'\theta \left(1 - \Phi \left(\frac{-x'\theta}{\sigma}\right)\right) + \frac{\phi \left(\frac{-x'\theta}{\sigma}\right)}{1 - \Phi \left(\frac{-x'\theta}{\sigma}\right)} \left(1 - \Phi \left(\frac{-x'\theta}{\sigma}\right)\right) \sigma - c
\]

\[
= x'\theta \Phi \left(\frac{x'\theta}{\sigma}\right) + \phi \left(\frac{x'\theta}{\sigma}\right) \sigma - c. \tag{3.15}
\]

Evidently, the selection rule is that an individual should receive a mailing if (3.15) is larger than zero. The first term on the right hand-side of the last term can be interpreted as the expected quantity of response, \( x'\theta \), times the probability of response, \( \Phi(x'\theta/\sigma) \). Thus, the two decisions can be distinguished in this expression. Note that the implicit assumption that the same characteristics have the same influence on both decisions becomes clear by the fact that \( x'\theta \) determines both parts. In contrast, in the TPM these terms are \( x'\gamma \) and \( \Phi(x'\beta) \), respectively. The second term of (3.15), which increases in \( \sigma \), can be interpreted as a factor that accounts for the uncertainty. The uncertainty is most prevalent for large values of \( \sigma \), and close to the threshold of response/nonresponse, i.e.
when \( x'\theta \) is in the neighborhood of zero. If the uncertainty about the quantity of response increases, reflected by an increase of \( \sigma \), more individuals should be selected. This can be seen by considering the derivative of \( E(\Pi \mid x) \) with respect to \( \sigma \):

\[
\frac{\partial E(\Pi \mid x)}{\partial \sigma} = - \frac{(x'\theta)^2}{\sigma^2} \phi \left( \frac{x'\theta}{\sigma} \right) + \frac{(x'\theta)^2}{\sigma^3} \phi \left( \frac{x'\theta}{\sigma} \right) \sigma + \phi \left( \frac{x'\theta}{\sigma} \right)
\]

which is obviously larger than zero. Hence, the expected profit for given \( x \) and \( \theta \) increases in \( \sigma \). This implies that for more individuals the expected profit is positive and hence that the fraction of selected individuals increases.

Since there is no separation of the two decisions, we cannot straightforwardly show the mailing region in figure 1. However, if the \( x \)s have a multivariate normal distribution, both the probit estimates and the OLS estimates of the quantity model are a (known) scalar times the true parameter (Goldberger 1981, and Greene 1981, 1983). Hence, when we set \( x'\beta \) of the probit model on the \( x \)-axes and \( x'\gamma \) of the quantity model on the \( y \)-axes, the tobit results should lie on a straight line with a slope which is defined by these scalars. A specific point on this line defines the optimal cutoff point. Note that normally distributed \( x \)s also imply that the ordering of the population is similar for the probit model, tobit model and even for linear regression as well. Thus, if the organization employs a selection rule of mailing the best \( k \% \), for a given \( k \), these three models should give identical results.

### 3.7 Data

We illustrate and compare the different methods with an application based on data from a Dutch charitable foundation. This foundation heavily rests on direct mailing. Every year it sends mailings to almost 1.2 million individuals in the Netherlands.

The data sample consists of 40 000 observations. All individuals on the list have donated at least once to the foundation since entry on the mailing list. The dependent variable in (3.3) is the amount of donation in 1991 and in (3.2) the response/nonresponse information. From 58 potential explanatory variables, the following variables were selected after a preliminary analysis. The amount of money donated in 1990 (A90), the amount of money donated
Table 3.2: Estimated parameters of the three models. Standard errors in parentheses, which are based on the 500 bootstrap samples

<table>
<thead>
<tr>
<th></th>
<th>Probit model</th>
<th>Quantity model</th>
<th>Tobit model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-0.760</td>
<td>7.594</td>
<td>-21.418</td>
</tr>
<tr>
<td></td>
<td>(0.47)</td>
<td>(7.09)</td>
<td>(11.37)</td>
</tr>
<tr>
<td>A90</td>
<td>0.00240</td>
<td>0.105</td>
<td>0.100</td>
</tr>
<tr>
<td></td>
<td>(0.0015)</td>
<td>(0.078)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>A89</td>
<td>0.00627</td>
<td>0.259</td>
<td>0.235</td>
</tr>
<tr>
<td></td>
<td>(0.0032)</td>
<td>(0.11)</td>
<td>(0.10)</td>
</tr>
<tr>
<td>INT</td>
<td>-0.00255</td>
<td>-0.0267</td>
<td>-0.0540</td>
</tr>
<tr>
<td></td>
<td>(0.0031)</td>
<td>(0.13)</td>
<td>(0.11)</td>
</tr>
<tr>
<td>ENTRY</td>
<td>-0.00181</td>
<td>0.0483</td>
<td>-0.0195</td>
</tr>
<tr>
<td></td>
<td>(0.0050)</td>
<td>(0.068)</td>
<td>(0.11)</td>
</tr>
<tr>
<td>FS</td>
<td>-0.120</td>
<td>0.944</td>
<td>-2.23</td>
</tr>
<tr>
<td></td>
<td>(0.16)</td>
<td>(2.66)</td>
<td>(3.87)</td>
</tr>
<tr>
<td>CHAR</td>
<td>0.107</td>
<td>-0.916</td>
<td>1.824</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.99)</td>
<td>(1.26)</td>
</tr>
<tr>
<td>( \sigma )</td>
<td>12.897</td>
<td>25.063</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.03)</td>
<td>(3.75)</td>
<td></td>
</tr>
</tbody>
</table>

in 1989 (A89), the interaction between these two (INT), the year of entry on the mailing list (ENTRY), family size (FS; dummy variable: one if the size is larger than 3), own opinion on charitable behavior in general (CHAR; four categories: donates never, donates sometimes, donates regularly and donates always).

The overall response rate \( \bar{p} \) is 33.9%, which is rather high but not really surprising since charitable foundations generally have high response rates (Statistical Fact Book 1994-1995), and the mailing list only contains individuals that had donated to the foundation before. The average amount donated \( \bar{a} \) was NLG 17.04 and the cost of a mailing \( c \) was NLG 3.50.
3.8 Empirical results

We employ the cross-validation technique to examine the performance of the various methods. In order to have a sufficiently large validation set we use 1 000 observations for estimation and 39 000 observations for validation. The estimation sample can be interpreted as the test mailing and the validation sample as the mailing list of which the targets have to be selected. Note that this ratio between the estimation and validation sample is not the optimal ratio to examine the appropriateness of the proposed model (Steckel and Vanhonacker 1993).

In order to obtain more insight into the performance of the various methods we use the bootstrap method (e.g. Efron 1982, and Efron and Tibshirani 1993) instead of a single estimation and validation sample. To generate a (bootstrap) estimation sample we draw with replacement 1 000 observations from the data set of 40 000 observations. We then draw 39 000 observations, again with replacement, to generate a (bootstrap) validation sample. The estimation sample is used to estimate $\gamma$, $\beta$, $\mu_0$, $\mu_1$, $\sigma^2$, and hence $\bar{a}$ and $\bar{p}$ for all observations, and finally $f_0(a)$ and $f_1(a)$. Then, for the various methods, we employ the selection rule to compute on the validation sample the actual profits that would have been obtained. We generated 500 bootstrap replications.

Table 3.2 gives the estimated parameters of the probit model, quantity model and tobit model. Although these parameters are not of our interest per se, we give a few comments. As expected, the donations in 1990 and 1989 are positively related with the response probability. The negative sign of the interaction term can be interpreted as a correction for overestimation of the response probability if an individual responded in 1990 and 1989. ENTRY is negative in the response model, indicating that the response probability increases in the number of years on the list. The negative sign of the FS indicates that larger families have a lower response probability. Surprisingly, the coefficients for these variables have the opposite sign for the quantity model. The opposite signs of CHAR for the probit and quantity model indicate that an individual who indicated that he ‘donates always’ has a higher response probability but a smaller expected amount of donation. The opposite signs of the various coefficients also indicate that it is not likely that the tobit model is the appropriate specification.

Figure 3.2 depicts the resulting selection rules. It shows the three curves, based on the three approximations (equations (3.11), (3.12) and (3.13)), separating the ($p$, $a$) combinations that should or should not be selected. Thus the
mailing region for the three approximations is the \((p, a)\) space to the north-east of the curves. Selection according to the curve labeled I characterizes method 5 as given in table 1. Analogously, the curves labeled II and III define methods 6 and 7, respectively. The other, simpler methods can also be characterized in this figure. Methods 2–4 are based on fixing \(p\) at its average value, \(\bar{p}\). Hence, the intersection points of these curves with the vertical line at \(p = \bar{p}\) determine values of \(a\) beyond which selection should take place. This characterizes methods 2–4. Method 1, based on fixing the quantity, is characterized by the intersection of the horizontal line at \(a = \bar{a}\) with curve I and determines values of \(p\) beyond which selection should take place.

Table 3.3 contains the bottom line results. The last column shows the number of individuals selected when the various methods are applied. The preceding column gives the profit obtained by this selection by considering the amounts actually donated by the selected individuals. Both columns contain the average over the 500 bootstrap replications. We consider the current practice in direct marketing as the benchmark, i.e. method 1. To make the results more transparent, we present them in figure 3.3 graphically, using the percentage of individuals excluded from the mailing list instead of the number of selected individuals. A great gain results from modeling the quantity of response, even if the probability of response is not modeled (methods 2–4). A relatively minor
but not negligible further gain results from joint modeling the probability and quantity of response (methods 5–7). Within the array of these methods the added value of increased sophistication seems to be marginal. However, if the probability as well as the quantity are modeled, the incremental cost of implementing method 6 is relatively small. Surprisingly, solely modeling the quantity of response, i.e. methods 2–4, gives better results than the tobit model.

A further analysis of the performance of the seven methods and the tobit model relative to each other is given in table 3.4. Figure 3.3 may be too evocative as to a unique ordering of the profits to be obtained by the methods. Since our analysis is based on 500 bootstrap samples we can simply count the number of cases, out of these 500, in which one method yields a higher profit than another method. The table shows that modeling the response probability only generally gives highly suboptimal profits, and that methods 5, 6 and 7 are more or less equivalent, although an increase in sophistication in approximation will on average pay off. This table also shows the unexpectedly moderate performance of the tobit model.

3.9 Discussion and conclusion

We have introduced an approach to joint modeling of response probability and quantity that leads to selection methods that can be applied in practice in a straightforward way. The outcomes of the empirical illustration suggest that adding quantity modeling to probability modeling, which is the current
Target selection by joint modeling of the probability and quantity of response

Figure 3.3: Profits and percentage excluded from the mailing list for the various approximations.

practice, can be highly rewarding. Even the simplest approach to joint modeling can add significantly to profitability.

There are various limitations of this approach that should be addressed in future work. The results of the empirical illustration are highly evocative, especially the qualitative impression given by figure 3.3. The figure suggests that only modeling response probabilities, the focus of nearly all work in target selection, misses a dominant feature in striving for optimality: the gain when the quantity of response is taken into account is large. This may be an idiosyncratic result and we do not claim generality. The example concerns charitable donations, and the picture may be qualitatively different if the proposed methods are applied to other cases where quantity of response varies across individuals, e.g. money amounts involved in mail order buying.

An implicit assumption of model is that the parameters are constant across individuals. This assumption may be unrealistic in practice. It runs for example counter to the idea of trying to customize promotions through direct marketing. An organization could deal with this kind of heterogeneity by using e.g. latent class analysis (e.g. DeSarbo and Ramaswamy 1994, and Kamakura and Russell 1989). Since the focus is not heterogeneity but modeling the quantity of
Discussion and conclusion

Table 3.4: Relative performance of methods

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td></td>
<td>97</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>98</td>
<td>70</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>98</td>
<td>78</td>
<td>71</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>100</td>
<td>82</td>
<td>73</td>
<td>63</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>100</td>
<td>81</td>
<td>75</td>
<td>71</td>
<td>63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>100</td>
<td>85</td>
<td>81</td>
<td>76</td>
<td>62</td>
<td>59</td>
<td></td>
</tr>
<tr>
<td>tobit</td>
<td>92</td>
<td>21</td>
<td>13</td>
<td>10</td>
<td>3</td>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>

Entry (i, j) is the percentage of cases (in 500 bootstrap samples) where method i outperforms method j.

response, and since this assumption is often made in direct marketing research, we have incorporated this assumption in the analysis.

In section 3.3 we assumed that $E(A \mid R = 1, p) \approx E(A \mid R = 1)$. The advantage of this simplifying assumption is that it improves the analytical tractability of the analysis. It would be interesting to examine the effect of this assumptions on the profit implications. This should be accomplished by incorporating the joint distribution of $a$ and $p$, conditional on $R = 0$ and $R = 1$, in the analysis.

Although we gave various reasons in favor of the TPM over the SSM, it would still be worth examining the performance of the SSM. Unfortunately, we encountered a serious problem by employing this model. The estimated correlation between the disturbance terms of the response and quantity model is very close to one. Though we expect this correlation to be positive, we do not believe that this could be the true value for the underlying model. Possible explanations for this phenomenon are a non-normal distribution of the error terms, e.g. the distribution is skewed, and the fact that we consider a statistical relation instead of a structural model. The latter means that our model specification of the SSM does not give the appropriate data generating process, which may imply that the distributional assumptions are violated. Given the non-robustness of the SSM to these violations, this may result in the encountered estimation problem.

As a final issue, our approach is limited in the sense that the underlying model is static and does not take into account behavior over time. This issue
has two aspects. In the first place, the behavioral model should be improved to a panel data model where a central role is played by the individual effect; individual response vis-à-vis direct mailing will have a strong, persistent component largely driven by unobservable variables. The other aspect concerns the optimality rule to be applied by the direct mailing organization, which is essentially more complicated than in the one-shot, static case considered in this chapter.
Chapter 4

Bayesian decision rule approach for target selection

4.1 Introduction

In order to decide which individuals should be selected it is of crucial importance for the organization to assess how the individual’s response probability depends on his characteristics (demographic variables, attitudes, etc.). If the effect of the characteristics on the response probability are known, the individuals can be ranked and the targets can be selected. Of course, these effects are unknown and have to be estimated. Typically, an organization specifies and estimates - at the individual level - a model based on a test mailing to obtain some prior knowledge on the effects of the characteristics on the response probability (see chapter 2). The estimates obtained are then used to formulate a decision rule to select households from the mailing list.

By and large this separation of parameter estimation and formulation of a decision rule does not lead to optimal profits since a suboptimal decision rule is specified (Klein et al. 1978). The reason for this is that estimation usually takes place by considering (asymptotic) squared-error loss, which puts equal weight at overestimating and underestimating the parameters. However, while a squared-error loss function may be useful when summarizing properties of the response function, it completely ignores the economic objectives of the direct marketing organization. Rather, the inferential process should be embedded in the organization’s decision making framework, taking explicitly into account the organization’s objective of maximizing expected profit. Put differently, the decision maker should take the estimation uncertainty into account when formulating a decision rule regarding which individuals to select. The resulting loss structure is, broadly speaking, asymmetric in contrast to the
Bayesian decision rule approach for target selection

traditional squared-error loss structure. Consequently, the traditional methods yield suboptimal decision rules.

The purpose of this chapter is to formulate a strict decision theoretic framework for a direct marketing organization. In particular, we derive an optimal Bayesian rule deciding when to send a mailing to an individual with a given set of characteristics. This formal approach has a number of advantages. First of all, a rigorous decision theoretic framework clarifies the essential ingredients entering the organization’s decision problem. By deriving the optimal Bayesian rule based on an expected profit loss function, the present framework yields admissible decision rules with respect to the organization’s economic objective. Furthermore, the estimation uncertainty resulting from the organization’s assessment of the characteristics of the population of targets is explicitly taken into account as an integral part of the optimal decision procedure. Thus, the decision theoretic procedure provides a firmer theoretical foundation for optimal decision making on the part of the organization. Equally important, the present framework provides decision rules yielding higher profits to the organization.

It is important to realize that even in a non-Bayesian framework, it is generally not correct to ignore the effect of estimation uncertainty on the decision. This follows directly from the Von Neumann-Morgenstern (Von Neumann and Morgenstern, 1967) paradigm which says that a decision maker chooses a decision that maximizes the expected utility, e.g. profits. It is, however, difficult for a non-Bayesian decision maker to take the uncertainty into account in a non ad-hoc way (Klein et al. 1978).

Integration of the estimation and decision steps has been studied thoroughly in statistics (e.g. Berger 1985, DeGroot 1970). This formal decision theoretic framework has been employed in a number of economic decision making situations (e.g. Cyert and DeGroot 1987).

An illustrative economic application is given by Varian (1975). He uses the approach for a real estate assessor whose decision is to estimate the value of a house on the basis of a linear regression model. If the assessor underestimates the value of a house, the loss equals the amount of underestimation. If the house is overestimated, the homeowner can appeal to an Assessment Board, which evaluates the assessment. If the assessor loses this case, the loss consists not only of the amount of overassessment but also of the costs involved in the appeal procedure. It is clear that the loss is asymmetric. Varian shows that the average loss resulting from the loss minimizing Bayesian regression is only 72% of the loss resulting from applying ordinary least squares.
In the finance literature estimation uncertainty is recognized to be an important aspect in portfolio theory. Recently, Chopra and Ziemba (1993) stressed the importance of estimation uncertainty in portfolio selection. They show that in particular the estimation uncertainty of expected returns causes large differences in the optimal portfolio choice. We illustrate the idea and effect of incorporating estimation uncertainty in the portfolio choice by the following very simple example (Bawa et al. 1979). The technicalities are omitted.

Consider an individual who chooses to invest the proportion \( D \) of his capital in a single risky asset yielding return \( r_a \). The other part of his capital is in the form of riskless securities yielding a known rate of return \( r_f \). The individual chooses \( D \) in such a way that the variance of the portfolio, the so-called risk, equals a value specified by the individual, \( v^2 \). For reasons of simplicity, we assume that \( r_a \sim N(\mu_a, \sigma^2) \), where \( \mu_a \) is unknown to the individual and \( \sigma^2 \) is known. There are \( n \) observations on \( r_a \) and we assume that the prior information is negligible. The rate of return for a given value of \( D \) is given by

\[
D = (1-D)r_f + Dr_a,
\]

which, conditional on \( \mu_a \), is distributed as \( N((1-D)r_f + D \mu_a, D^2 \sigma^2) \). Let

\[
p(\mu_a) \propto \text{constant} \quad -\infty < \mu_a < \infty
\]

be the uninformative prior information on \( \mu_a \). Zellner (1971, p. 20) showed that, conditional on the sample size, the posterior distribution for the mean, \( \mu_a \), is \( N(\bar{r}_a, \sigma^2/n) \), where \( \bar{r}_a \) is the sample mean. This posterior distribution can be used to derive the unconditional distribution, of \( r_a \), which is \( N((1-D)r_f + D \bar{r}_a, D^2(1+1/n)\sigma^2) \) (Zellner 1971, pp. 29-30). If the individual ignores the estimation uncertainty, \( D \) is chosen such that \( D^2 \sigma^2 = v^2 \). Hence, \( D = v/\sigma \). However, if the individual does take into account the estimation uncertainty \( D \) is chosen such that \( D^2(1+1/n)\sigma^2 = v^2 \), or \( D = v/(\sigma \sqrt{1+1/n}) \). Hence, a different decision is made, namely, less is invested in the risky asset when estimation uncertainty is taken into account. A rigorous overview on estimation uncertainty in portfolio analysis is provided by Bawa et al. (1979).

To the best of our knowledge, only a few papers on optimal decision making under uncertainty have been applied to marketing questions. Blattberg and George (1992) consider an organization whose goal it is to maximize profits by setting the optimal price. They conclude that the organization is better off by charging a higher price than the price resulting from traditional methods which are based on an estimated price elasticity. However, in contrast with
our approach, they consider a loss function that results from a rather ad-hoc specified model, with only one unknown parameter. Kalyanam (1996) extends this model by taking into account the uncertainty about the specified model. Rossi et al. (1996) employ the decision theoretic approach to determine the optimal face value of a coupon, i.e. the price discount provided by the coupon. They conclude that separation of the estimation and decision step generates a larger face value than integration of the two steps. Unfortunately, they do not demonstrate the effect of neglecting estimation uncertainty on the net revenue.

This chapter, which is based on Muus, Van der Scheer, and Wansbeek (1996), is organized as follows. In the section below we formulate the decision theoretic framework and derive the optimal Bayesian decision rule. We demonstrate that the decision rule crucially depends on the estimation uncertainty facing the organization. The estimation uncertainty can be incorporated through a posterior density. An important aspect of our approach is the evaluation of the integral resulting from the Bayesian decision rule. In section 4.3 we derive a closed form for this integral by approximating the posterior by the asymptotically normal density of the maximum likelihood (probit) estimator. In section 4.4 we discuss the Laplace approximation and Markov chain Monte Carlo integration that can be used to calculate the integral resulting from the optimal decision rule. In section 4.5 we discuss an empirical example, using data provided by a charity foundation. Applying the formal decision framework appears to generate higher profits indeed. In this section we also show the asymmetric loss. For illustrative purposes we show - in section 4.6 - how point estimates can be derived within the proposed framework. Finally, in section 4.7, we demonstrate how the above approach can be used in the case that the response elicited from the campaign varies over the individuals, as discussed in chapter 3. We conclude in section 4.8.

### 4.2 The decision theoretic framework

Consider a direct marketing organization that has the option whether or not to mail to potential targets. In case a mail is sent to a given individual the profit to the organization, $\Pi$, is given by

$$\Pi = aR - c,$$
The decision theoretic framework

where \( a \) is the revenue from a positive reply, \( c \) is the mailing cost, and \( R \) is a random variable given by

\[
R = \begin{cases} 
1 & \text{if the individual responds} \\
0 & \text{if the individual does not respond.}
\end{cases}
\]

Clearly, \( c < a \) if the organization has to obtain positive profits at all. We assume that the response is driven by a probit model. Hence, the response probability of an individual is

\[
P(R = 1 \mid x, \beta) = \Phi(x'\beta),
\]

where \( \Phi(\cdot) \) is the standard normal distribution function, \( x \) is a \( k \times 1 \) vector of regressors of which the first element equals one, and \( \beta \equiv (\alpha, \beta_2, \ldots, \beta_k) \) is a \( k \times 1 \) vector of regression coefficients (\( \beta \in \mathcal{B} \subseteq \mathbb{R}^k \)), where \( \alpha \) denotes the intercept. In case a mail is sent, the expected profit given \( x \) and \( \beta \) is

\[
E(\Pi \mid x, \beta) = aE(R \mid x, \beta) - c = a\Phi(x'\beta) - c. 
\quad (4.1)
\]

With an unknown \( \beta \) the organization has to make a decision whether to send a mail \( (d = 1) \) to a given individual or not \( (d = 0) \). The loss function considered in the following is given by

\[
L(d, \beta \mid x) = \begin{cases} 
a\Phi(x'\beta) - c & \text{if } d = 1 \\
0 & \text{if } d = 0.
\end{cases}
\quad (4.2)
\]

Note that the above loss function is naturally induced by the organization’s economic profit maximization objective. In this sense, the present decision theoretic framework naturally encompasses the phenomena of estimation uncertainty, without introducing rather ad hoc statistical criteria.

Inference on the parameter vector \( \beta \) is obtained through a test mailing, resulting in the sample

\[
\mathcal{S}_n \equiv \{(x_1, R_1), \ldots, (x_n, R_n)\}.
\]

The posterior density, using Bayes’ rule, is given by

\[
f(\beta \mid \mathcal{S}_n, \theta) = \frac{L(\beta \mid \mathcal{S}_n) f(\beta \mid \theta)}{f(\mathcal{S}_n \mid \theta)},
\quad (4.3)
\]

where \( L(\beta \mid \mathcal{S}_n) \) is the likelihood function corresponding to the sample,

\[
L(\beta \mid \mathcal{S}_n) = \prod_{i=1}^{n} \Phi(x'_i\beta)^{R_i}(1 - \Phi(x'_i\beta))^{1-R_i}.
\]
and \( f(\beta \mid \theta) \) denotes the prior density; \( \theta \in \Theta \subseteq \mathbb{R}^p \) is a \( p \times 1 \) vector of so-called hyperparameters. Finally, \( f(S_n \mid \theta) \) denotes the predictive density given by,

\[
f(S_n \mid \theta) = \int L(\beta \mid S_n) f(\beta \mid \theta) \, d\beta.
\] \hspace{1cm} (4.4)

The posterior risk corresponding to the loss function (4.2) is then given by

\[
\mathcal{R}(d \mid x) = \mathbb{E}(L(d, \beta \mid x) \mid S_n) = \begin{cases} a \int \Phi(x' \hat{\beta}) f(\beta \mid S_n, \theta) \, d\beta - c & \text{if } d = 1 \\ 0 & \text{if } d = 0. \end{cases} \] \hspace{1cm} (4.5)

The Bayesian decision rule corresponding with the posterior risk (4.5) is the decision variable \( d \) maximizing \( \mathcal{R}(d \mid x) \). It is obvious that this decision rule is given by

\[
d = 1 \quad \text{if and only if} \quad \int \Phi(x' \hat{\beta}) f(\beta \mid S_n, \theta) \, d\beta \geq \frac{c}{a}. \] \hspace{1cm} (4.6)

Note that this decision rule explicitly takes into account the estimation uncertainty inherent when the organization does not know the parameter vector \( \beta \).

The Bayesian optimal mailing region, denoting the individuals to whom a mail should be sent, is given by

\[
\mathcal{M}_n = \left\{ x \in \mathbb{R}^k \mid \int \Phi(x' \hat{\beta}) f(\beta \mid S_n, \theta) \, d\beta \geq \frac{c}{a} \right\}.
\]

Generally speaking, the structure of the mailing region may be quite complicated.

It is often recommended to base the organization’s mailing decision on the point estimates obtained from the test mailing. These point estimates are typically derived by implicitly assuming a squared-error loss function, resulting from the use of standard estimation procedures. As this squared-error loss does not reflect the actual loss suffered by the organization, the use of the point estimate motivated by squared-error loss will be inappropriate. If the organization neglects the estimation uncertainty it would specify a decision rule on the basis of a point estimate of \( \beta \), say \( \hat{\beta} \), e.g. the probit estimator based on \( S_n \). The point estimate is then used as if it were the true parameter value. The resulting decision rule, which we call the naive decision rule, is thus given by

\[
d = 1 \quad \text{if and only if} \quad \Phi(x' \hat{\beta}) \geq \frac{c}{a}. \] \hspace{1cm} (4.7)
This rule evidently ignores the estimation uncertainty surrounding \( \hat{\beta} \). By a second order Taylor series expansion of \( \Phi(x' \beta) \), we obtain

\[
\Phi(x' \beta) \approx \Phi(x' \hat{\beta}) + (\beta - \hat{\beta})'x\Phi(x' \hat{\beta}) - \frac{1}{2} x' \hat{\beta} \phi(x' \hat{\beta}) x'(\beta - \hat{\beta})(\beta - \hat{\beta})'x,
\]

using the fact that the derivative of \( \phi(t) \) is \(-t\phi(t)\), where \( \phi(\cdot) \) is the standard normal density. Hence, an approximate Bayesian decision rule is given by

\[
\Phi(x' \hat{\beta}) - \frac{1}{2} x' \hat{\beta} \phi(x' \hat{\beta}) x'Mx \geq \frac{c}{d},
\]

(4.8)

since \( E(\beta - \hat{\beta}) = 0 \); \( M \equiv E(\hat{\beta} - \beta)(\hat{\beta} - \beta)' \) denotes the mean square error matrix of the estimator \( \hat{\beta} \). The major difference between the (approximate) Bayesian rule (4.8) and the naive rule (4.7) is that estimation uncertainty is explicitly taken into account in the former. Evidently, if the estimation uncertainty is small, i.e. \( M \) is small, the approximate Bayesian rule (4.8) is adequately approximated by the naive decision rule (4.7). Note that the mailing region for the naive rule is the space given by

\[
M_N \equiv \{ x \in \mathbb{R}^k \mid \Phi(x' \hat{\beta}) \geq \frac{c}{d} \}.
\]

The result of applying the naive decision rule is thus an approximation of the mailing region \( M_B \) by the space \( M_N \). As will be demonstrated below this approximation may be rather crude, resulting in a suboptimal level of profits.

In order to implement the optimal decision rule (4.6), we need to evaluate the expectation of \( \Phi(x' \beta) \) over the posterior density of \( \beta \). If the posterior admits a closed form solution and is of a rather simple analytical form, this expectation can be solved analytically. Otherwise, numerical methods need to be implemented in order to assess the decision rule (4.6). In section 4.4 we explore various numerical strategies to evaluate the decision rule. However, it is instructive to consider the case where the posterior density is normal, in which case we can fully characterize the mailing region.

### 4.3 The case of a normal posterior

If the posterior density is normal with mean \( \mu \) and covariance matrix \( \Omega \), we can obtain a closed form expression for (4.6), viz.¹:

¹. We are indebted to Ton Steerneman for bringing the result to our attention and for providing this derivation.
58 Bayesian decision rule approach for target selection

\[
\int \Phi(x') f(\beta \mid S_n, \theta) \, d\beta
\]

where \( \Phi(x') \) is an indicator function that is one if its argument falls in the region specified by \((\cdot, \cdot)\) and is zero otherwise. Hence, the mailing region is given by

\[
\mathcal{M}_b = \left\{ x \in \mathbb{R}^k \mid \Phi \left( \frac{x'\mu}{(1 + x'\Omega x)^{1/2}} \right) \geq \frac{c}{\alpha} \right\}
\]

where

\[
\tau \equiv \Phi^{-1} \left( \frac{c}{\alpha} \right).
\]

Since in any practical situation \( c \ll a \), we assume \( \tau < 0 \) whenever the sign of \( \tau \) is relevant. Notice that, if \( \Omega_1 > \Omega_2, 1 + x'\Omega_1 x > 1 + x'\Omega_2 x \). Thus, since \( \tau < 0 \), greater uncertainty as to \( \beta \), indicated by the larger \( \Omega \), implies that the mailing region expands.

Expression (4.9) enables us to show explicitly that the Bayesian decision rule generates higher expected profits than the naive decision rule with \( \hat{\beta} = \mu_0 \). The expected profit (cf. (4.1)), in case mail is sent, is

\[
q(x) \equiv E_{\hat{\beta}}(E(\Pi \mid x, \beta))
\]

\[
= \alpha \Phi \left( \frac{x'\mu}{(1 + x'\Omega x)^{1/2}} \right) - c.
\]

For all \( x \) in \( \mathcal{M}_b \) there holds, by definition, that \( q(x) > 0 \). Since \( \mathcal{M}_N \subseteq \mathcal{M}_b \) it follows that the expected profit is lower for the naive decision rule.

We consider the mailing region \( \mathcal{M}_b \) in somewhat more detail. The boundary of this mailing region is given by

\[
\left\{ x \in \mathbb{R}^k \mid x'\mu = \tau (1 + x'\Omega x)^{1/2} \right\}.
\]
We assume that $\Omega > 0$. By squaring and rewriting the argument of (4.11) we obtain
\[ x' (\mu \mu' - \tau^2 \Omega) x = \tau^2, \] (4.12)
which can be written as
\[ x' \Omega^{1/2} (\Omega^{-1/2} \mu \mu' \Omega^{-1/2} - \tau^2 I_k) \Omega^{1/2} x = \tau^2, \] (4.13)
where $I_k$ is a $(k \times k)$ identity matrix. Let
\begin{align*}
A_1 & \equiv \frac{\Omega^{-1/2} \mu \mu' \Omega^{-1/2}}{\mu \Omega^{-1} \mu} \\
A_2 & \equiv I_k - A_1 \\
\lambda & \equiv \mu' \Omega^{-1} \mu - \tau^2;
\end{align*}
$A_1$ and $A_2$ are idempotent matrices of rank 1 and $k-1$, respectively, $A_1 A_2 = 0$ and $A_1 + A_2 = I_k$. Hence, we can write (4.13) as
\[ x' \Omega^{1/2} (\lambda A_1 - \tau^2 A_2) \Omega^{1/2} x = \tau^2. \]
Let $A_1 = z_1 z_1'$ and $A_2 = Z_2 Z_2'$, so $(z_1, Z_2)$ is orthonormal. Then
\[
G \equiv \lambda A_1 - \tau^2 A_2
= \lambda z_1 z_1' - \tau^2 Z_2 Z_2'
= \begin{pmatrix} z_1' & Z_2 \end{pmatrix} \begin{pmatrix} \lambda & 0 \\ 0 & -\tau^2 I_{k-1} \end{pmatrix} \begin{pmatrix} z_1' \\ Z_2' \end{pmatrix}.
\]
Hence, the eigenvalues of $G$ are $-\tau^2$ with multiplicity $k-1$, and $\lambda$ with multiplicity one. The sign of $\lambda$ depends on $\Omega$ and $\mu$. Informally speaking, for small values of $\Omega$, $\lambda > 0$, and for large values, $\lambda < 0$. In the first case $G$ has one positive and $k-1$ negative eigenvalues. Due to ‘Sylvester’s law of inertia’ (e.g. Lancaster and Tismenetsky 1985, p. 188), the same holds for $\mu \mu' - \tau^2 \Omega$. Hence, the matrix is indefinite and the boundary is a hyperboloid in the $x$-space. When the uncertainty as to $\beta$ is so large that $\lambda < 0$, all eigenvalues of $G$ are negative and (4.12) does not have a solution. Hence, all individuals should be included in the mailing campaign.

We illustrate the mailing region for the case with two explanatory variables ($k = 2$), $\mu' = (1, 1)$, $\tau = -1$, and
\[
\Omega = \begin{pmatrix} \sigma^2 & \sigma_{12} \\ \sigma_{12} & \sigma^2 \end{pmatrix}.
\]
Then, from (4.10), the mailing region is

\[ M_B = \left\{ x_1, x_2 \mid x_1 + x_2 \geq -\sqrt{1 + \sigma^2(x_1 + x_2) + 2\sigma_{12}x_1x_2} \right\}, \]

which reduces to the space \( x_1 + x_2 \geq -1 \) if \( \sigma^2 = \sigma_{12} = 0 \). The matrix in (4.12) becomes

\[
\begin{align*}
\mu\mu' - \tau^2\Omega &= \begin{pmatrix} 1 - \sigma^2 & 1 - \sigma_{12} \\ 1 - \sigma_{12} & 1 - \sigma^2 \end{pmatrix} \\
&= \frac{1}{2} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 2 - \sigma^2 - \sigma_{12} & 0 \\ 0 & -(\sigma^2 - \sigma_{12}) \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}.
\end{align*}
\]

Hence, the matrix \( \mu\mu' - \tau^2\Omega \) has one negative eigenvalue, \(-(\sigma^2 - \sigma_{12})\) and one eigenvalue that is positive if \( \sigma^2 + \sigma_{12} < 2 \). Using (4.14), (4.12) can be rewritten as

\[
(2 - \sigma^2 - \sigma_{12})(x_1 + x_2)^2 - (\sigma^2 - \sigma_{12})(x_1 - x_2)^2 = 2,
\]

which is a hyperbola in \( \mathbb{R}^2 \). Its asymptotes are found by putting the left-hand side equal to zero. On letting

\[
\varphi \equiv \sqrt{\frac{2 - \sigma^2 - \sigma_{12}}{\sigma^2 - \sigma_{12}}},
\]

these asymptotes are found to be

\[
\varphi(x_1 + x_2) = \pm (x_1 - x_2),
\]
or

\[
\frac{x_2}{x_1} = \frac{1 - \varphi}{1 + \varphi} \quad \text{and} \quad \frac{x_2}{x_1} = \frac{1 + \varphi}{1 - \varphi}.
\]

Figure 4.1 illustrates the boundary for \( \sigma_{12} = 0 \), and \( \sigma^2 = 0, 0.5, 1.5 \) and 1.95, respectively. If \( \sigma^2 = 0 \) we have a straight line. This bounds the mailing region of the naive method. The mailing region increases as \( \sigma^2 \) increases; the arrows indicate the direction of the increase. When \( \sigma^2 \geq 2 \), the mailing region is simply \( \mathbb{R}^2 \). The distance between the straight line corresponding with \( \sigma^2 = 0 \) and the hyperbola is larger when the \( x \)-value is larger. This reflects the fact that the uncertainty as to \( x' \beta \) increases by the (absolute) value of \( x \).
4.4 Numerical evaluation of the optimal Bayesian rule

Numerical implementation of the optimal Bayesian decision rule (4.6) requires the evaluation, for each value of $x$, of the integral

$$Q(x) = \int \Phi(x' \beta) f(\beta \mid \mathcal{S}_n, \theta) \, d\beta$$

using (4.3) and (4.4) in the last step. We will now explore various methods to evaluate this integral. Henceforth, we denote the probit estimate of $\beta$, based on $\mathcal{S}_n$, by $\hat{\beta}$ and covariance matrix by $\hat{\Omega}$ (e.g. the inverse of the Fisher information matrix evaluated in $\hat{\beta}$).
Normal posterior approximation

It is well-known that under suitable regularity conditions the posterior density converges to a normal distribution, with mean $\hat{\beta}$ and covariance matrix $\hat{\Omega}$, if the sample size is sufficiently large (Jeffreys 1967, p. 193; Heyde and Johnstone 1979). Obviously, the approximation may be rather crude, since it is solely based on the asymptotic equivalence of the Bayes and maximum likelihood estimator. Thus, this approximation completely ignores the prior distribution $f(\beta | \theta)$. However, as we showed in section 4.3, this property appears to be very valuable since it enables us to obtain a closed form expression for (4.15), which is given in (4.9) by substitution of $\hat{\beta}$ for $\mu$ and $\hat{\Omega}$ for $\Omega$. Moreover, Zellner and Rossi (1984) show that, for moderate sample sizes ($n = 100$), the normal posterior approximation works well for the logit model.

Laplace approximation

A more refined asymptotic approximation is the Laplace approximation proposed by Tierney and Kadane (1986) (see also Kass et al. 1988, and Tierney et al. 1989). The Laplace approximation of (4.16) is given by

$$
\hat{Q}(x) = \frac{\Psi_1(\hat{\beta}_1)|H_1(\hat{\beta}_1)|^{-1/2}}{\Psi_0(\hat{\beta}_0)|H_0(\hat{\beta}_0)|^{-1/2}}
$$

where $\hat{\beta}_0$ and $\hat{\beta}_1$ are the maximizers of $\Psi_0(\cdot)$ and $\Psi_1(\cdot)$, respectively, and

$$
\Psi_0(\beta) \equiv L(\beta | S_n) f(\beta | \theta)
$$

$$
\Psi_1(\beta) \equiv \Phi(x' \beta) L(\beta | S_n) f(\beta | \theta),
$$

and

$$
H_0(\beta) \equiv -\frac{\partial^2 \ln \Psi_0(\beta)}{\partial \beta \partial \beta'}
$$

$$
H_1(\beta) \equiv -\frac{\partial^2 \ln \Psi_1(\beta)}{\partial \beta \partial \beta'}.
$$

By means of the Laplace approximation, the integral $Q(x)$ is thus evaluated without any need for numerical integration. Instead the Laplace approximation requires maximization, in order to determine $\hat{\beta}_0$ and $\hat{\beta}_1$, and differentiation, in order to find $H_0(\cdot)$ and $H_1(\cdot)$. For $\hat{\beta}_0$ and $\hat{\beta}_1$ we use the values obtained by a single Newton-Raphson step from $\beta$ when maximizing $\ln \Psi_0(\beta)$ and $\ln \Psi_1(\beta)$, which does not affect the rate at which the approximation error vanishes. As demonstrated by Tierney and Kadane (1986), Kass et al. (1988), and Tierney...
et al. (1989), the general error of the approximation vanishes at rate $n^{-2}$. As these authors demonstrate, this approximation is often very accurate.

We apply this approximation for an informative prior and an uninformative prior. As to the former, for $f(\beta \mid \theta)$ we choose the normal density with mean $\hat{\beta}$ and covariance matrix $\hat{\Theta}$. This is called empirical Bayes since the prior is derived from the data. Since $\partial \ln f(\beta \mid \theta)/\partial \beta = -\hat{\Theta}^{-1}(\beta - \hat{\beta})$, we have $\hat{\beta}_0 = \hat{\beta}$, and

$$
\hat{\beta}_1 = \hat{\beta} + \xi(\hat{\beta})H_0(\hat{\beta})^{-1}x,
$$

where

$$
\xi(\beta) \equiv \frac{\zeta}{1 + \zeta(\zeta + x'\beta)x'H_0(\beta)^{-1}x}, \quad (4.17)
$$

with $\zeta \equiv \phi(x'\beta)/\Phi(x'\beta)$, the inverse of Mills’ ratio.

The proof of (4.17) follows below. Let

$$
g_0(\beta) \equiv \frac{\partial \ln \Psi_0(\beta)}{\partial \beta} = \frac{\partial \ln L(\beta \mid S_n)}{\partial \beta} - \hat{\Theta}^{-1}(\beta - \hat{\beta})
$$

and

$$
g_1(\beta) \equiv \frac{\partial \ln \Psi_1(\beta)}{\partial \beta} = \frac{\phi}{\Phi}x + g_0(\beta) = \zeta x + g_0(\beta),
$$

where $\phi \equiv \phi(x'\beta)$ and $\Phi \equiv \Phi(x'\beta)$. Notice that $g_0(\hat{\beta}) = 0$. Further,

$$
H_0(\beta) = \frac{-\partial^2 \ln L(\beta \mid S_n)}{\partial \beta \partial \beta'} + \hat{\Theta}^{-1},
$$

$$
H_1(\beta) = \frac{\Phi(\phi + x'\beta \Phi)}{\Phi^2}xx' + H_0(\beta) = \zeta(x + \hat{\beta})xx' + H_0(\beta).
$$

Then $\hat{\beta}_1$ follows from the Newton-Raphson step

$$
\hat{\beta}_1 = \hat{\beta} + H_1(\hat{\beta})^{-1}g_1(\hat{\beta})
$$

$$
= \hat{\beta} + \left(\zeta(\zeta + x'\hat{\beta})xx' + H_0(\hat{\beta})\right)^{-1}g_1(\hat{\beta})
$$

$$
= \hat{\beta} + \frac{1}{1 + \zeta(\zeta + x'\hat{\beta})x'H_0(\hat{\beta})^{-1}x}H_0(\hat{\beta})^{-1}g_1(\hat{\beta})
$$

$$
= \hat{\beta} + \xi(\hat{\beta})H_0(\hat{\beta})^{-1}x,
$$

where $\xi(\cdot)$ is defined in (4.17), and $\hat{\zeta}$ denotes $\zeta$ evaluated in $\hat{\beta}$.

For the uninformative prior we use Jeffreys’ prior (e.g. Berger 1985, pp. 82-89, and Zellner 1971, pp. 41-53), given by

$$
f(\beta \mid \theta) = -\text{E}\left(\frac{\partial^2 \ln L(\beta \mid S_n)}{\partial \beta \partial \beta'}\right)^{1/2}.
$$
Note that no hyperparameters are involved here. Within the context of binary response models this prior has been examined by, among others, Ibrahim and Laud (1991), and Poirier (1994). These authors support the use of Jeffreys’ prior as an uninformative prior but note that it can be quite cumbersome to work with, analytically as well as numerically. However, for the Laplace approximation we only need to derive the first and second derivative analytically, which is rather straightforward. The computation of these derivatives is quite a tedious task indeed but not really problematic with high-speed computers.

The first and second derivative of Jeffreys’ prior can be derived as follows. Let

\[ A \equiv \sum_{i=1}^{n} \frac{\phi_i^2}{D_i} x_i x'_i, \]

with \( \phi_i \equiv \phi(x'_i \beta) \) and \( D_i \equiv \Phi_i(1 - \Phi_i) \), where \( \Phi_i \equiv \Phi(x'_i \beta) \), hence \( f(\beta | \theta) = |A|^{1/2} \). Using some well-known properties of matrix differentiation (e.g. Balestra 1976), we obtain the logarithmic first derivative

\[
\frac{\partial \ln |A|^{1/2}}{\partial \beta} = \frac{1}{2} \frac{\partial \ln |A|}{\partial |A|} \frac{\partial |A|}{\partial \beta} = \frac{1}{2} |A|^{-1} (\text{vec}(|A|^{-1})' \otimes I_k) \text{vec} \left( \frac{\partial A}{\partial \beta} \right) = \frac{1}{2} (\text{vec}(A^{-1}) \otimes I_k) \text{vec} \left( \frac{\partial A}{\partial \beta} \right).
\]

Let

\[ M \equiv \text{vec} I_k \otimes I_k, \]

then, using the product rule for matrices, we can write the second derivative as

\[
\frac{\partial^2 \ln |A|^{1/2}}{\partial \beta \partial \beta'} = \frac{1}{2} \left[ \left( (\text{vec} A^{-1})' \otimes I_k \right) \left( I_k \otimes \frac{\partial^2 A}{\partial \beta \partial \beta'} \right) - M' \left( I_k \otimes (A^{-1} \otimes I_k) \frac{\partial A}{\partial \beta} A^{-1} \frac{\partial A}{\partial \beta'} \right) \right] M.
\]

Finally, to complete the derivatives we need an expression for \( \frac{\partial A}{\partial \beta} \) and \( \frac{\partial^2 A}{\partial \beta \partial \beta'} \), which are given by

\[
\frac{\partial A}{\partial \beta} = - \sum_{i=1}^{n} \left( \frac{2x'_i \Phi_i^2}{D_i} + \frac{\Phi_i^3 (1 - 2\Phi_i)}{D_i^2} \right) x_i x'_i \otimes x_i.
\]
Numerical evaluation of the optimal Bayesian rule

\[ \frac{\partial^2 A}{\partial \beta \partial \beta'} = \sum_{i=1}^{n} \left( \frac{2\phi^2_i (2(x_i' \beta)^2 - 1)}{D_i} + \frac{5x_i' \beta \phi^i (1 - 2 \Phi_i) + 2 \phi^i}{D_i} \right) \]

\[ + \left( \frac{2 \phi^i (1 - 2 \Phi_i)^2}{D_i} \right)x_i x_i' \otimes x_i x_i' \]

which enables us to calculate the derivatives of Jeffreys’ prior.

**Monte Carlo integration**

The recent development of Markov chain Monte Carlo (MCMC) procedures has revolutionized the practice of Bayesian inference. See, for example, Chib and Greenberg (1995), Gilks et al. (1995), and Tierney (1994) for expositions of basic Markov chain Monte Carlo procedures. These algorithms are easy to implement and have the advantage that they do not require evaluation of the normalizing constant of the posterior density, given by (4.4). As a candidate density it is natural to select the asymptotic approximation,

\[ q(\beta) \sim N(\beta, \Omega) \]

The density of interest, the so-called target density, is given by

\[ h(\beta) \equiv L(\beta \mid S_n) f(\beta \mid \theta) \]

The independence sampler (e.g. Tierney 1994), a special case of the Hastings-Metropolis algorithm, is used to generate random variates \( \beta_j, j = 1, \ldots, J \) from the (unnormalized) density \( h(\beta) \) through the following algorithm, where \( \beta_0 \) is arbitrarily selected:

1. draw a candidate point, \( \beta_j^* \), from \( q(\cdot) \)
2. draw \( u_j \) from the uniform density on \( (0, 1) \)
3. if \( u_j \leq \alpha(\beta_{j-1}, \beta_j^*) \), then \( \beta_j = \beta_j^* \), else \( \beta_j = \beta_{j-1} \).

Here

\[ \alpha(\beta_{j-1}, \beta_j^*) \equiv \begin{cases} \min \left( \frac{h(\beta_j^*) q(\beta_{j-1})}{h(\beta_{j-1}) q(\beta_j^*)}, 1 \right) & \text{if } h(\beta_{j-1}) q(\beta_j^*) > 0 \\ 1 & \text{else} \end{cases} \]

The generated \( \beta_j \)'s, \( j = 1, \ldots, J \) are used to evaluate the integral by

\[ \hat{Q}(x) = \frac{1}{J} \sum_{j=1}^{J} \Phi(x' \beta_j) \]

We use this algorithm instead of more advanced MCMC procedures like the Gibbs sampler (e.g. Albert and Chib 1993), since we have a candidate density that is a good approximation of the target distribution (Roberts 1995). Again, we apply this algorithm for the (informative) normal prior and for the (uninformative) Jeffreys’ prior.
Bayesian decision rule approach for target selection

Table 4.1: Probit estimates and results of the independence sampler

<table>
<thead>
<tr>
<th></th>
<th>Probit Estimates¹</th>
<th>Independence Sampler²</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Normal prior</td>
<td>Jeffreys’ prior</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.3938</td>
<td>-0.3948</td>
</tr>
<tr>
<td></td>
<td>(0.4511)</td>
<td>(0.4539)</td>
</tr>
<tr>
<td>A90</td>
<td>0.0052</td>
<td>0.0053</td>
</tr>
<tr>
<td></td>
<td>(0.0014)</td>
<td>(0.0014)</td>
</tr>
<tr>
<td>A89</td>
<td>0.0074</td>
<td>0.0072</td>
</tr>
<tr>
<td></td>
<td>(0.0030)</td>
<td>(0.0030)</td>
</tr>
<tr>
<td>INT</td>
<td>-0.0056</td>
<td>-0.0053</td>
</tr>
<tr>
<td></td>
<td>(0.0029)</td>
<td>(0.0027)</td>
</tr>
<tr>
<td>ENTRY</td>
<td>-0.0063</td>
<td>-0.0063</td>
</tr>
<tr>
<td></td>
<td>(0.0048)</td>
<td>(0.0048)</td>
</tr>
<tr>
<td>FS</td>
<td>-0.1526</td>
<td>-0.1513</td>
</tr>
<tr>
<td></td>
<td>(0.1408)</td>
<td>(0.1397)</td>
</tr>
<tr>
<td>CHAR</td>
<td>0.0683</td>
<td>0.0685</td>
</tr>
<tr>
<td></td>
<td>(0.0537)</td>
<td>(0.0530)</td>
</tr>
</tbody>
</table>

¹ Asymptotic standard errors in parentheses
² Standard deviation, based on 10 000 $\beta_i$’s, in parentheses

4.5 Application

We illustrate our approach with an application based on data from a charitable foundation in the Netherlands (see chapter 3). The dependent variable is the response/nonresponse in 1991. The explanatory variables are the amount of money (in NLG) donated in 1990 (A90) and 1989 (A89), the interaction between these two (INT), the date of entry on the mailing list (ENTRY), the family size (FS), own opinion on charitable behavior in general (CHAR; four categories: donates never, donates sometimes, donates regularly and donates always). The data set consists of 40 000 observations. All the individuals on the list donated at least once to the foundation since entry on the mailing list. In order to have a sufficiently large validation sample we used 1 000 observations for estimation.
Table 4.1 gives the probit estimates and the average of the coefficients based on the independence sampler with the normal and Jeffreys’ prior, respectively. These results are based on one estimation sample, in contrast with the estimation results given in table 3.2 of chapter 3. The donations in 1990 and 1989 are, as expected, positively related with the response probability. The negative sign of the interaction term can be interpreted as a correction for overestimation of the response probability if an individual responded in 1990 and 1989. The other three coefficients do not significantly differ from zero. As expected, the average values of the coefficients for the independence sampler are similar to the probit estimates. The standard deviations, however, of the normal prior are much smaller.

The basic difficulty in MCMC procedures is the decision when the generated sequence of parameters has converged to a sample of the target distribution. Many diagnostic tools to address this convergence problem have been suggested in the recent literature (see Cowles and Carlin (1996) for an extensive overview). Following the recommendations of these authors, we generated six parallel sequences of parameters with starting points chosen systematically from a large number of drawings from a distribution that is overdispersed with respect to the target distribution. We inspected the sequences of each parameter by displaying them in a common graph and in separate graphs. We used the Gelman-Rubin statistics (Gelman and Rubin 1992) to analyze the sequences quantitatively. The results of these diagnostics are satisfying, indicating an almost immediate convergence of the sample.

Table 4.2 shows the profit implications for the various approaches to determine the posterior risk function and the naive approach for the validation sample. As a benchmark we also give the situation in which the foundation sends all individuals a mailing. Of these 39 000 individuals, 13 274 responded, generating a net profit of NLG 91 007. If the foundation would have used the naive selection approach they would have selected 87.03% (33 946) of the individuals, with a net profit of NLG 93 290. Using the Bayesian decision rule, the foundation would have selected more individuals, as expected. This ranges from 34 018 of the Laplace approximation with the normal prior to 34 271 of the independence sampler with Jeffreys’ prior. Except for the Laplace approximation with the normal prior, the additional selected individuals generate sufficient response to increase the net profits, thus reinforcing the importance of the Bayesian decision rule. Net profits increase by 2.5% if the naive selection is used instead of selecting all the individuals. This percentage increases to 3.3% if we apply the normal posterior approximation,
Table 4.2: Target selection and profit implications

<table>
<thead>
<tr>
<th></th>
<th># selected</th>
<th>Response</th>
<th>Profit (NLG)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Selection</td>
<td>39 000</td>
<td>13 274</td>
<td>91 007</td>
</tr>
<tr>
<td>Naive approach</td>
<td>33 946</td>
<td>12 236</td>
<td>93 290</td>
</tr>
<tr>
<td>Normal posterior</td>
<td>34 240</td>
<td>12 337</td>
<td>93 967</td>
</tr>
<tr>
<td>Laplace approximation:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal prior</td>
<td>34 018</td>
<td>12 250</td>
<td>93 266</td>
</tr>
<tr>
<td>Jeffreys’ prior</td>
<td>34 256</td>
<td>12 341</td>
<td>93 796</td>
</tr>
<tr>
<td>Independence sampler:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal prior</td>
<td>34 153</td>
<td>12 310</td>
<td>93 479</td>
</tr>
<tr>
<td>Jeffreys’ prior</td>
<td>34 271</td>
<td>12 347</td>
<td>94 119</td>
</tr>
</tbody>
</table>

and to 3.4% when using the independence sampler with Jeffreys’ prior. Given that the foundation’s database contains 1.2 million targets, these increases turn out to be quite substantial. Note that the figures of the Laplace approximation and independence sampler with the normal prior are much closer to those of the naive approach than those with Jeffreys’ prior. This makes intuitive sense since informative priors put more weight to values of \( \beta \) near \( \hat{\beta} \). In the case of the posterior density degenerating at \( \hat{\beta} \), i.e. perfect prior information on \( \beta \), the decision rule is equivalent to the naive rule.

In order to illustrate the asymmetry in the loss function we examine the average loss, given by

\[
\tilde{L}(d, \beta \mid \mathcal{X}_n) \equiv \frac{1}{N-n} \sum \left( a \Phi(x' \beta - c) \right) \left( I(x' \beta \geq \tau) - d \right), \quad (4.18)
\]

where \( \tau \equiv \Phi^{-1}(c/a) \), \( \mathcal{X}_n \equiv \{x_{n+1}, \ldots, x_N\} \), i.e. the regressors of the individuals for whom the organization has to make a decision, and \( I(\cdot) \) is an indicator function that is one if its argument is true and zero otherwise. The argument of this indicator function is the optimal decision rule if \( \beta \) is known to the foundation. If this optimal decision rule and the actual decision \( d \) are equal, the contribution to the loss is zero; if these decisions differ this contribution is positive.

Figures 4.2 and 4.3 give the average loss for the estimated intercept (\( \hat{\alpha} \)) and the parameter estimate of A90 (\( \hat{\beta}_{A90} \)), respectively, where we use the probit estimates as the ‘true’ (known) parameters. Thus, the average loss is zero.
Figure 4.2: Relation between the average loss and the estimated intercept, \( \hat{\alpha} \) if the estimated parameter is equal to the probit estimates. The figures make clear that the loss is asymmetric, i.e. overestimation generates less loss than underestimation. Hence, it is preferable to reduce the probability of underestimation. This means that on average it is better to overestimate the parameter, which implies that a larger fraction of the individuals should be selected. This coincides with figure 4.1 and the empirical illustration. Note that when the estimated intercept is larger than, say, -0.14 the average loss hardly increases in \( \hat{\alpha} \) since nearly all the individuals on the list are already selected. Then the average loss is the difference between the naive approach and ‘no selection’ (divided by 39,000).

For the other parameters similar figures can be drawn. By combining two of these figures we obtain a three-dimensional graph with the average loss on the z-axes and estimated coefficients on the x- and y-axes. Since the average loss, in absolute value, is much larger for changes in the estimated intercept than for changes in the estimated coefficient of A90, a three-dimensional graph that combines figure 4.2 and 4.3 would be completely dominated by the former. Therefore we illustrate it with the estimated coefficients of A90 and A89 (see figure 4.4). Again, we assume that probit estimates are the ‘true’ parameters, hence the average loss is zero if \( (\hat{\beta}_{A90}, \hat{\beta}_{A89}) = (0.0052, 0.0074) \). As expected, underestimation causes larger losses than overestimation. The ‘peak’ is obtained when both parameters are underestimated. An interesting aspect of the figure is that the difference between overestimation and underestimation of e.g.
$\hat{A}_{A90}$ is much larger if $\hat{A}_{A89}$ is underestimated than if $\hat{A}_{A89}$ is overestimated. This can be explained as follows. If $\hat{A}_{A89}$ is underestimated, it implies that insufficient individuals are selected, which causes a (relatively) large loss. This loss can be partly compensated by overestimating the other parameter. Compensation is less important if $\hat{A}_{A89}$ is overestimated, since sufficient individuals are selected.

Figures 4.2–4.4 give clear evidence that we have an asymmetric loss function. Note, however, that asymmetry in each parameter separately or in two parameters does not necessarily mean that the loss function is asymmetric over the joint parameter space.

### 4.6 Point estimation

In the approach described above we did not explicitly derive point estimates of $\beta$. This is not necessary since an organization is not interested in the parameters per se. This is in contrast with, for example, Blattberg and George (1992), who use the point estimates to set the optimal price. However, it is interesting to show the aspect involved in deriving point estimates. Moreover, we are able to derive a close form expression for the intercept term. Of course, for practical purposes the mode of the posterior distribution can be used as point estimates.
Point estimation

Following the traditional approaches, the objective is to find an estimator \( \hat{\beta} \) such that an individual receives a mailing if and only if \( x'\hat{\beta} \geq \tau \); this defines the following mailing region:

\[
\mathcal{M}_p \equiv \left\{ x \in \mathbb{R}^k \mid x'\hat{\beta} \geq \tau \right\}.
\]

Since the profit function equals the risk function defined in (4.5), the total expected profits, \( \Pi_\tau(\cdot) \), are obtained by integrating the profit function over the density of \( x \). For the sake of simplicity we consider the case of a normal posterior, so that we can use (4.9). Then the total expected profits are given by

\[
\Pi_\tau(\hat{\beta}) = \int_{\mathcal{M}_p} \left\{ a \Phi(x'\hat{\beta} - c) \right\} f(\beta \mid \mathcal{S}_n, \theta) g(x) \, d\beta \, dx
\]

\[
= \int_{\mathcal{M}_p} \left\{ a \Phi \left( \frac{x'\mu}{\sqrt{1 + x'\Sigma x}} \right) - c \right\} g(x) \, dx, \tag{4.19}
\]

where \( g(x) \) is the density function of \( x \), assumed to be positive for all \( x \). Since, by definition, the expected profits are nonnegative if and only if \( x \in \mathcal{M}_n \), and \( \mathcal{M}_p \) is not equal to \( \mathcal{M}_n \), it follows that a point estimate does not lead to an
optimal decision rule. In terms of Figure 4.1, the estimator $\beta$ defines a boundary of the mailing region by a straight line, in contrast with the optimal decision rule. Presumably, this line differs from the line of the naive method.

To determine $\beta$ we have to take the derivative to $\Pi_1(\cdot)$ with respect to $\beta$, and set it equal to zero. In general this is very complicated to solve. Therefore, we consider a simple case to illustrate the approach. We assume that only the first element of $\beta$, i.e. the intercept $\alpha$, is unknown. As before, we substitute the probit estimates for $\mu$ and $\Omega$, but only the first element. Thus, $\mu = (\hat{\alpha}, \mu_2, \ldots, \mu_k)$, and

$$\Omega = \begin{pmatrix} \hat{\alpha}^2 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}.$$  

Define, $m \equiv x_{-1}^\prime \mu_{-1}$, where $x_{-1}$ is $x$ without the first element, and $\mu_{-1} = (\mu_2, \ldots, \mu_k)$. Assume that $m \sim h(m)$, $h(m) > 0 \ \forall m$, then (4.19) can be written as

$$\Pi_1(\tilde{\alpha}) = \int_{\tau - \tilde{\alpha}}^{\infty} \left\{ a \Phi \left( \frac{\hat{\alpha} + m}{\sqrt{1 + \hat{\alpha}^2}} \right) - c \right\} h(m) \, dm$$

since

$$M_\nu = \{ m \in \mathbb{R}^1 \mid m \geq \tau - \tilde{\alpha} \}.$$  

The first order condition is

$$\frac{\partial \Pi_1(\tilde{\alpha})}{\partial \tilde{\alpha}} = - \left\{ a \Phi \left( \frac{\hat{\alpha} + \tau - \tilde{\alpha}}{\sqrt{1 + \hat{\alpha}^2}} \right) - c \right\} h(\tau - \tilde{\alpha}) = 0,$$

from which it immediately follows that

$$\tilde{\alpha} = \hat{\alpha} + \tau \left( 1 - \sqrt{1 + \hat{\alpha}^2} \right).$$

This expression explicitly shows how the (traditional) estimator $(\hat{\alpha})$ should be adjusted in the presence of estimation uncertainty in order to increase expected profits. That is, the intercept has to be updated in an upward direction (the second term on the right-hand side is positive), hence more addresses have to be mailed than the estimator $\hat{\alpha}$ implies. Thus, the number of selected individuals increases in $\hat{\alpha}^2$, which coincides with our earlier findings, i.e. the mailing region
Figure 4.5: The loss for different values of the ‘true’ and estimated intercept increases when the uncertainty increases. Note that the distribution $h(m)$ of $m$ plays no role in the estimator.

Finally, we graphically illustrate the loss function, as we have done in the figures 4.2 and 4.3. Again, we consider the case that only the intercept is unknown to the organization. As in (4.18) we write the average loss as

$$
\bar{L}_{T}(\tilde{\alpha}, \alpha) = \int_{\alpha - \tilde{\alpha}}^{\alpha + \tilde{\alpha}} \{a \Phi(\alpha + m) - c\} h(m) \, dm.
$$

where $\alpha$ is the true intercept. Figure 4.5 shows the average loss for $a = 10$, $c = 1$, and $m \sim N(0, 3)$, for a range of $(\alpha, \tilde{\alpha})$ values from which it is immediately clear that we have an asymmetric loss function. The figure shows that overestimation generates a smaller loss than underestimation; a result that corresponds with the conclusions from figures 4.2–4.4.
4.7 Modeling the quantity of response

The results in table 4.2 clearly favor the Bayesian decision rule approach. However, we considered only a binary choice model which (see chapter 3) turned out to generate fewer profits than models which also take the quantity of response into account. Hence, the next step to be taken is to specify the decision theoretic framework for those models. For the two-part model (TPM) it turns out to be rather straightforward. Please recall that the TPM specifies a model for the quantity conditional on response, i.e. \( A = x' \gamma + u \) for \( A > 0 \). Therefore, the estimators \( \hat{\beta} \) and \( \hat{\gamma} \) are independent. This can be shown as follows. Note that independence of \( A \) and \( R \) implies that \( \hat{\beta} \) and \( \hat{\gamma} \) are independent. Independence of the former means that

\[
P(A \in \mathcal{F}, R \in \mathcal{G} | A > 0) = P(A \in \mathcal{F} | A > 0)P(R \in \mathcal{G} | A > 0),
\]

for two nonempty sets \( \mathcal{F} \) and \( \mathcal{G} \). In our case, \( \mathcal{G} \) has only two relevant elements, viz. 0 and 1. If \( \mathcal{G} = \{0\} \), then

\[
P(A \in \mathcal{F}, R \in \mathcal{G} | A > 0) = 0 = P(A \in \mathcal{F} | A > 0)P(R \in \mathcal{G} | A > 0),
\]

since \( P(R \in \mathcal{G} | A > 0) = 0 \). If \( \mathcal{G} = \{1\} \), then

\[
P(A \in \mathcal{F}, R \in \mathcal{G} | A > 0) = P(A \in \mathcal{F} | A > 0),
\]

since \( P(R \in \mathcal{G} | A > 0) = 1 \). Thus \( A \) and \( R \) are independent.

The important implication is that we may apply the decision theoretic framework to the response model and quantity model separately. Write (4.1), using (3.1), as

\[
E(\Pi | x, \beta, \gamma) = x' \Phi(x' \beta) - c.
\]

Thus the loss function, cf. (4.2), is given by

\[
\mathcal{L}(d, \beta, \gamma | x) = \begin{cases} 
  x' \Phi(x' \beta) - c & \text{if } d = 1 \\
  0 & \text{if } d = 0.
\end{cases}
\]

Denote the posterior density of \( \gamma \) by \( g(\gamma | \mathcal{S}_n, \lambda) \), where \( \lambda \) a vector of hyperparameters. Then, the posterior risk corresponding to the loss function is

\[
\mathcal{R}(d | x) 
\equiv E(\mathcal{L}(d, \beta, \gamma | x) | \mathcal{S}_n)
\]
Modeling the quantity of response

\[
= \begin{cases} 
\int x'\gamma \Phi(x'\beta) f(\beta \mid S_n, \theta) g(\gamma \mid S_n, \lambda) \, d\beta \, d\gamma - c & \text{if } d = 1 \\
0 & \text{if } d = 0
\end{cases}
\]

\[
= \begin{cases} 
\int x'\gamma g(\gamma \mid S_n, \lambda) \, d\gamma \int \Phi(x'\beta) f(\beta \mid S_n, \theta) \, d\beta - c & \text{if } d = 1 \\
0 & \text{if } d = 0
\end{cases}
\]

The Bayesian decision rule corresponding to the posterior risk (4.5) is the decision variable \( d \) maximizing \( R(d \mid x) \). It is obvious that this decision rule is given by

\[
d = 1 \quad \text{if and only if} \quad \int x'\gamma g(\gamma \mid S_n, \lambda) \, d\gamma \int \Phi(x'\beta) f(\beta \mid S_n, \theta) \, d\beta \geq c.
\]

The first integral defines the risk in the quantity. Since the second-order term of the Taylor series expansion of \( x'\gamma \) is zero, so \( x'\gamma = x'\hat{\gamma} + (\gamma - \hat{\gamma})'x \), it is immediately clear that we may neglect the estimation uncertainty of \( \gamma \) if no prior information is available. Since we will consider the case of a posterior density with mean \( \hat{\gamma} \), we may simply use this point estimate. Thus, the Bayesian decision rule reduces to

\[
d = 1 \quad \text{if and only if} \quad x'\gamma \int \Phi(x'\beta) f(\beta \mid S_n, \theta) \, d\beta \geq c.
\]

This decision rule makes clear that the estimation uncertainty problem reduces to the one we considered in section 4.2, viz. the uncertainty in the response probability. Hence, we may determine \( a(= x'\gamma) \) as before and the response probability by one of the approximations described in section 4.4. For the selection of individuals we proceed as in chapter 3. That is, a given individual should receive a mailing if \( (\hat{\beta}, \hat{a}) \) falls in the mailing region, which is defined by one of the three approximations described in section 3.4. Please remember that in approximation I it is assumed that the density of \( a \) is similar for the respondents and nonrespondents. In approximation II it is assumed that both densities are normal with different means and the same variance. In approximation III the densities are estimated by a kernel. For the TPM these approximations correspond with methods 5–7. The Bayesian approach of the methods 1, 5–7 are denoted by method 1B, 5B–7B, respectively.

Table 4.3 shows the results of the various approaches, based on the 500 bootstrap samples of chapter 3. We limit the bootstrap analysis for model comparison to the normal posterior approach, due to computer time constraints.
Bayesian decision rule approach for target selection

Table 4.3: Performance of methods

<table>
<thead>
<tr>
<th>method</th>
<th>approx.</th>
<th>profit</th>
<th>#selected</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$p \tilde{a}$</td>
<td>I</td>
<td>89 163</td>
</tr>
<tr>
<td>1B</td>
<td>$p \tilde{a}$</td>
<td>I</td>
<td>89 505</td>
</tr>
<tr>
<td>5</td>
<td>$p a$</td>
<td>I</td>
<td>99 433</td>
</tr>
<tr>
<td>5B</td>
<td>$p a$</td>
<td>I</td>
<td>99 791</td>
</tr>
<tr>
<td>6</td>
<td>$p a$</td>
<td>II</td>
<td>99 924</td>
</tr>
<tr>
<td>6B</td>
<td>$p a$</td>
<td>II</td>
<td>100 011</td>
</tr>
<tr>
<td>7</td>
<td>$p a$</td>
<td>III</td>
<td>100 123</td>
</tr>
<tr>
<td>7B</td>
<td>$p a$</td>
<td>III</td>
<td>100 371</td>
</tr>
</tbody>
</table>

Given the relatively small differences between the various approaches for evaluating the integral of the Bayesian decision rule, we assume that this will give a reasonable indication of the performance of the Bayesian approach in the various approximations. Table 4.3 shows that on average the Bayesian approach performs better than the corresponding ‘naive’ approximation. Table 4.4 demonstrates the relative performances of the various methods. For instance, the table shows that in 71% of the samples method 1B generates higher profits than method 1. Similarly, method 5 (the basic TPM) generates in 30% of the samples higher profits than method 7B, which is the model that generates the highest profits.

4.8 Discussion and conclusion

In order to select addresses from a list for a direct mailing campaign, an organization can build a response model and use the (consistently) estimated parameters for selection. The decision rule for selection is often defined on the basis of the estimated parameters taken as the true parameters. This chapter shows that this leads to suboptimal results. The reason for this is that the estimation uncertainty resulting from the organization’s assessment of the characteristics of the potential targets is not taken into account. Put differently, both steps of a target selection process, estimation and selection, should be considered simultaneously. We have formulated a rigorous theoretic framework, based on the organization’s profit maximizing behavior, to derive an
Table 4.4: Relative performance of methods

<table>
<thead>
<tr>
<th>Entry</th>
<th>1</th>
<th>1B</th>
<th>5</th>
<th>5B</th>
<th>6</th>
<th>6B</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1B</td>
<td>71</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>100</td>
<td>99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5B</td>
<td>100</td>
<td>99</td>
<td>74</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>100</td>
<td>99</td>
<td>63</td>
<td>55</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6B</td>
<td>100</td>
<td>99</td>
<td>65</td>
<td>59</td>
<td>60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>100</td>
<td>99</td>
<td>62</td>
<td>54</td>
<td>59</td>
<td>55</td>
<td></td>
</tr>
<tr>
<td>7B</td>
<td>100</td>
<td>99</td>
<td>70</td>
<td>61</td>
<td>66</td>
<td>63</td>
<td>72</td>
</tr>
</tbody>
</table>

Entry \((i, j)\) is the percentage of cases (in 500 bootstrap samples) where method \(i\) outperforms method \(j\).

optimal Bayesian decision rule. We have demonstrated, theoretically as well as empirically, that this approach generates higher profits.

An important aspect of our approach is the evaluation of the integral resulting from the Bayesian decision rule. We have used a normal posterior, Laplace approximation, and Monte Carlo integration to evaluate the Bayesian rule numerically. Although the normal posterior approach may be rather crude it has the advantage that a closed form expression is obtained. Moreover, it performs quite well in the empirical illustration. The advantage of having a closed form is that we do not need the computational intensive methods. Furthermore, we obtain a transparent expression for the expected profit, which explicitly shows the effect of estimation uncertainty. It must be realized, though, that the empirical results indicate that the decision rule is affected by the chosen prior density. Since the normal posterior approximation ignores the prior density, it has to be used with caution when prior information is available.

It is worth noting that Bult (1993), who used Manski’s (1975, 1985) maximum score estimator, explicitly incorporates the decision rule in the estimation procedure (see section 2.4). That is, the function he considers accounts for the fact that the cost of misclassification of an individual that responded to the mailing is much larger than the cost of misclassification of an individual that did not respond to the mailing. However, as in the naive approach, the estimation step is completed separately from the selection step. That is, the decision rule uses the point estimator as if it is the true parameter. Thus, although the
decision rule is incorporated in the estimation procedure, no attention is given to estimation uncertainty.

As to future research, it is important, for practical purposes, to analyze the possible generality of the obtained profit implications with respect to the numerical evaluation methods and the use of priors. That is, are our results generalizable to other data sets, and to the use of a larger estimation sample and/or validation set? Furthermore, we have considered uncertainty in the parameters only. There is, however, also uncertainty about the chosen model and unobserved heterogeneity. These kinds of uncertainty could be taken into account as well. It could very well be that heterogeneity has a much larger effect on the selection rule than estimation uncertainty.

In the introduction we argued that it is incorrect for a decision maker to ignore estimation uncertainty. However, the number of empirical applications that indeed incorporate estimation uncertainty is rather limited. A natural question that arises is: why is this the case? First of all, a lot of applied work does not lead to a strict decision rule. Often, it is sufficient to know whether certain variables have a significant effect on a particular phenomenon. A model is built with the intention to represent the data adequately. In that case it is reasonable to ignore the estimation uncertainty. Second, whereas it is often relatively easy to specify and estimate a model in a traditional way, it is much harder to do so in a Bayesian framework. That is, it is often difficult to specify an appropriate loss function. Moreover, since Bayesian decision problems often involve high-dimensional integrals, it may be hard to estimate the parameters. In contrast, sufficient user-friendly software packages are available to estimate the parameters in a traditional way. Third, the Bayesian ideas are not generally accepted. One of the main arguments is that it is difficult to achieve objectivity within a Bayesian framework (e.g. Efron 1986). It is, however, very difficult to incorporate estimation uncertainty, or in general, to apply statistical decision theory, in a non-Bayesian way.

In this chapter we have demonstrated how to deal with these problems. That is, by using an uninformative prior, or a crude approximations, the influence of the prior vanishes. Hence, if no prior information is available, or if the decision maker does not want to make use of any prior information, it is still possible to apply the decision theory in a Bayesian framework. Furthermore, computational difficulties are no longer an appropriate argument to ignore estimation uncertainty. In the last decade many computer-intensive algorithms have been suggested to solve high dimensional integrals. The general idea behind these methods is to replace difficult calculations by a sequence of easier
calculations, which are relatively easy to program. Recently, even standard software for Monte Carlo integration methods and convergence diagnostics has become available (e.g. BUGS of Spiegelhalter et al. (1994), and CODA of Best et al. (1995), respectively), which simplifies the implementation. Hence, given the continuing availability of high-speed computers it is possible to solve the complicated integrals resulting from the decision framework.
Chapter 5

On the value of information

5.1 Introduction

A database with customer information is an essential tool in direct marketing to maintain a structural relationship. It enables the organization to target the individuals individually. The basis for a database is a customers history file, the so-called internal list. This list includes names and addresses of individuals who have bought products or services from the organization, or who have made an inquiry, or to whom a direct marketing message has been directed. Naturally, the organization must possess sufficient information for efficient fulfilment, billing, credit screening, promotion reporting, customer inquiry and subsequent selection of promotion (Brooks 1989, Katzenstein and Sachs 1992, pp. 156-161). Furthermore, the organization would also like to have a database that contains individual-specific information on product and brand preferences, needs, wants and demands with respect to product attributes and purchase intentions. If all such data are available, an organization is able to deliver its products effectively and efficiently. In the ideal situation sketched above, perfect information is available at the individual level. However, most organizations are faced with less than perfect information.

There are several reasons why an internal list contains imperfect information. First, the internal list traditionally has no roots in the marketing department. For example, financial institutions possess financial transaction records of their clients, for reasons of business administration. These data are used as a starting point for the creation of a customer history file. It goes without saying that the records have to be adjusted in order to be useful for direct marketing. This adjustment process usually takes a lot of time and effort (Shepard 1995). Second, an internal list is a perishable commodity. Individuals
on the list show a fluctuating degree of activity (e.g. they move, marry, die, or change attitudes). According to Baier (1985), in twelve months an average customer list has as many as 25 percent address revisions. Third, most of the direct marketing organizations are involved in highly competitive markets. In order to maintain or gain market share, the organization has to expand its activities. An organization could find (Ansoff 1965) new markets for its current products (market development strategy), it could develop new products of potential interest to its current markets (product development strategy), or it could develop new products for new markets (diversification strategy). For all three strategies, the direct marketing organization needs information which is not available on the internal list. Therefore, it could, for example, rent information from external lists.

An external list contains collections of characteristics of specific individuals that can be accessed either by name and address or by some other specific identifier (Bickert 1992, Lix et al. 1995). External lists comprise lists primarily compiled for direct marketing and lists compiled for other reasons (such as telephone directories or magazine subscriber lists). External lists can be classified based according to whether the information exists at some aggregated level (usually postal code areas) or at the individual level. External lists containing aggregated information are by definition imperfect since the data are gathered at an aggregated level. In the case of postal code information systems, the value of a characteristic is not known for a specific individual but an average is known for clusters of individuals. For a specific individual this means that the value of a characteristic is known with a certain amount of ‘noise’; we call this proxy information. This proxy information on background variables can be used for selecting individuals in the same way as perfect information is used at the individual level. However, since the selection has to take place per cluster and not per individual, the maximally attainable profits will be lower than in the case where exact individual information is available.

Whether the price to be paid for employing imperfect (external) information is outweighed by the possibilities of a better focused mailing campaign depends on the ‘noise-to-signal’ ratio of the proxies and the sensitivity of the profit function to the information content in the external list. The aim of this chapter is to build a conceptual framework for this trade-off between costs and returns.

This chapter, which is based on Bult, Van der Scheer and Wansbeek (1995), is structured as follows. In section 5.2 we introduce the basic model and show the relation between the value of information, defined by $R^2$ and profits. The relation between postal code information and profits is discussed in section 5.3.
An other type of proxy information is obtained when not the variable of interest is measured but some proxy for it, the so-called measurement error. We examine the relation between this kind of proxy information and profits in section 5.4. In section 5.5 we discuss and prove some theoretical results, which are useful throughout the chapter. Section 5.6 contains a number of concluding remarks.

5.2 Model

5.2.1 Basic model

Following on the traditional approaches describing direct mail response, we assume that individual’s inclination to respond, \( \eta_i \), is linearly related to a \( k \)-vector of (observable) regressors, \( \xi_i \),

\[
\eta_i = \alpha + \xi_i' \beta + \varepsilon_i, \tag{5.1}
\]

with

\[
y_i = \begin{cases} 
1 & \text{if } \eta_i > 0 \\
0 & \text{otherwise,} \end{cases} \tag{5.2}
\]

where \( \beta \) is a \( k \)-vector of unknown parameters, \( \alpha \) is the unknown intercept, and \( \varepsilon_i \) is an (unobservable) disturbance, assumed to be distributed as \( \varepsilon_i \sim N(0, \sigma_\varepsilon^2) \); \( \sigma_\varepsilon^2 \) is also an unknown parameter. We assume that the regressors are distributed as \( \xi_i \sim N_k(0, \Sigma_\xi) \). Moreover, we assume that \( \varepsilon_i \) and \( \xi_i \) are both i.i.d. and mutually independent. We do not observe \( \eta_i \) but only whether or not an individual responded, which we call \( y_i \). We set, without loss of generality, \( y_i = 1 \) if \( \eta_i \) is positive and \( y_i = 0 \) otherwise. In other words, \( i \) will respond if \( i \)'s inclination is positive. Note that the choice of the value zero in (5.2) is not restrictive. We can set this threshold at an arbitrary constant value and subtract this value from the intercept. Furthermore, without loss of generality, we might take the mean of \( \xi_i \) equal to zero, since the mean can be absorbed in the intercept.

The probability that the random variable \( y_i \) equals one is

\[
P(y_i = 1 \mid \xi_i) = \Phi \left( \frac{\alpha + \xi_i' \beta}{\sigma_\varepsilon} \right), \tag{5.3}
\]

where \( \Phi(\cdot) \) is the standard normal distribution function. It is clear that the model is identified up to scale and thus we may impose a normalization on one of the parameters. A normalized variance of the disturbance term is the standard formulation of the probit model.
5.2.2 Optimal selection rule and profits

The first step in the selection process consists of the estimation of the parameters of model (5.1) by ML. Next, we use the resulting estimates to select targets; for the sake of simplicity we disregard the distinction between parameters and their consistent estimators. Given that \( i \) receives a mailing, the expected profits of \( i \) are equal to the response probability times the revenues of positive response (\( \equiv a \)) minus the cost of a mailing (\( \equiv c \)), i.e.

\[
a \Phi \left( \frac{\alpha + \xi'_i \beta}{\sigma_x} \right) - c.
\]

The (population) mean expected profit is the integral of (5.4) over the distribution of \( \xi_i \). The sample mean expected profit is simply the mean of (5.4) over all \( i \). Expected profits are maximized when those individuals are selected for which (5.4) is positive. We define \( n_i \equiv \xi_i / \beta \) as individual’s \( i \) index. Then, the optimal selection rule is to select \( i \) if

\[
a \Phi \left( \frac{\alpha + n_i}{\sigma_x} \right) - c \geq 0.
\]

Hence, we order the individuals by increasing values of \( n_i \), and select those with the highest values of \( n_i \). The value of \( n \) for which (5.5) holds with equality defines the optimal cutoff index, denoted by \( n_c \) and defined by

\[
n_c \equiv \tau \sigma_x - \alpha,
\]

where

\[
\tau \equiv \Phi^{-1} \left( \frac{c}{\alpha} \right).
\]

Thus, all individuals with \( n \geq n_c \) should receive a mailing. Note that \( \tau \) is defined for \( 0 < c/a < 1 \). This makes sense since \( c/a = 0 \) means that \( c = 0 \), which of course never holds in practice. Furthermore, \( c/a = 1 \) means that \( a = c \), which indicates that the expected profits are maximally zero. This maximum is only obtained when the response probability equals one, which does not hold in practice. Clearly, \( c/a \) is never negative and never larger than one (\( a \) is always larger than \( c \)), which implies that \( \Phi^{-1} (\cdot) \) is defined for practical situations.

The proportion of the variance of the dependent variable that is explained by the regressors, i.e. \( R^2 \), is an obvious choice to examine the value of information. The fact is that information is more valuable if more variation in the dependent
variable is explained. An increase in $R^2$ can either be the result of using more explanatory variables or of substituting an explanatory variable for a better one. In a binary choice model the use of $R^2$ is not straightforward because we estimate the probability of a certain outcome but this cannot be compared with the ‘true’ probability because this is unknown (see Windmeijer (1995) for a discussion on $R^2$-measures for a probit model). However, we can express the expected profits as a function of the $R^2$ of the underlying latent relation (5.1). To do so, we first decompose the variance of $\eta$,

$$\sigma_\eta^2 \equiv \text{var}(\eta) = \sigma_n^2 + \sigma_\varepsilon^2,$$

where we use the independence of $\xi_i$ and $\epsilon_i$. Hence,

$$R^2 = \frac{\sigma_n^2}{\sum_i} = \frac{\sigma_n^2}{\sigma_n^2 + \sigma_\varepsilon^2}.$$

An increase in $R^2$ means that $\sigma_n^2$ increases and that $\sigma_\varepsilon^2$ decreases by the same value since the inclination to respond and thus $\sigma_\varepsilon^2$ does not change. We use this to impose an alternative normalization on (5.1). We choose $\sigma_\varepsilon^2 = 1$, so that $R^2$ is simply $\sigma_n^2$, and $\sigma_\varepsilon^2 = 1 - R^2$. Hence, (5.3) can be written as

$$P(y_i = 1 \mid \xi_i) = \Phi \left( \frac{\alpha + n_i}{\sqrt{1 - R^2}} \right),$$

and the optimal cutoff index (5.6) becomes

$$n_c \equiv \tau \sqrt{1 - R^2} - \alpha. \quad (5.7)$$

To assess the value of information we have to express the expected profits as a function of $R^2$. For the individuals that receive the mailing, the expected profits are

$$E \left( a \Phi \left( \frac{\alpha + n_i}{\sqrt{1 - R^2}} \right) - c \mid n \geq n_c \right) = \int_{n_c}^{\infty} \Phi \left( \frac{\alpha + n_i}{\sqrt{1 - R^2}} \right) \phi \left( \frac{n_i}{R} \right) \frac{dn}{1 - \Phi \left( \frac{n_i}{R} \right)} - c, \quad (5.8)$$

where $\phi \left( \frac{n_i}{R} \right)$ is the density of $n$. The denominator is the fraction of selected individuals,

$$P(n \geq n_c) = 1 - \Phi \left( \frac{n_c}{R} \right)$$

$$= 1 - \Phi \left( \tau \sqrt{\frac{1 - R^2}{R^2}} - \frac{\alpha}{R} \right), \quad (5.9)$$
where we use (5.7). To determine the expected profits per individual on the list we have to multiply (5.8) by (5.9), which gives

$$\alpha \int_{n_c}^{\infty} \Phi \left( \frac{\alpha + n}{\sqrt{1 - R^2}} \right) \phi \left( \frac{n}{R} \right) \frac{1}{R} \, dn - c \left( 1 - \Phi \left( \frac{n_c}{R} \right) \right).$$  (5.10)

The expected total returns are obtained by multiplying this by $N$, the number of individuals on the list.

As noted before, profits are positively related with $R^2$. This results from the economic principle that better decisions will be made with better, or more informative, information. Note, however, that this relation should hold in theory but that in practice we face the problem of the optimal number of regressors. That is, statistical procedures may show a decreasing actual performance when the number of variables is increased beyond a certain bound (e.g. Breiman and Freedman 1983).

To prove algebraically that (5.8) is increasing in $R^2$ we should take the derivative of (5.8) with respect to $R^2$. Unfortunately this becomes too complicated because $R^2$ does not only appear in argument of the integral but also in $n_c$. It is, however, numerically quite easy to compute (5.8) for different values of $R^2$. Figure 5.1 shows the relation for various values of $\alpha$, with $a = 10$ and $c = 1$. It shows indeed that profits increase in $R^2$. The difference in expected profits for the three values of $\alpha$ vanishes when $R^2$ is close to one. This makes sense intuitively. The fact is that we are perfectly able to discriminate between targets and non-targets, which means that $i$ will respond if $n_i$ is above a certain threshold. Hence, all three curves go to $a - c = 9$. In contrast, if the importance of the explanatory variables vanishes, i.e. $R^2 \approx 0$, it is no longer useful to select individuals on the basis of $n$. Consequently, either all or none of the individuals should receive a mailing. The former holds if $a \Phi (\alpha) - c \geq 0$, which are then the expected profits for each individual. This is depicted in figure 5.2. It shows that if $R^2 = 0$, either all or none of the individuals are selected. The organization is indifferent between these two when $\alpha = \tau$. When $R^2 = 1$, the fraction equals $\Phi (\alpha)$, which follows directly from (5.9).

Expression (5.7) suggests that $n_c$ increases in $R^2$, because $\tau$ is generally negative ($\frac{\alpha}{n} < 0.5$). This implies that the fraction of selected individuals, i.e. $\Pr(n \geq n_c)$, decreases in $R^2$. However, according to figure 5.2, this does not always hold: it depends on the value of $\alpha$. This paradox can be explained as follows. A change in $R^2$ not only affects $n_c$ but also the shape of the distribution since $R^2$ is the variance of $n$. As illustrated in figure 5.3, an increase of $R^2$ generates a larger $n_c$ and a flatter distribution, resulting in an increase of
Figure 5.1: Relation between expected profits per selected individual and $R^2$

Figure 5.2: Fraction of selected individuals, i.e. $P(n \geq n_c)$
On the value of information

Figure 5.3: $R^2$ increases from 0.2 ($R^2_1$) to 0.6 ($R^2_2$); as a result $n_c$ increases from 0.65 ($n_{c,1}$) to 0.99 ($n_{c,2}$). Since the shape of the distribution changes, the fraction of selected individuals increases.

the fraction of selected individuals. We can also show this algebraically. The derivative of (5.9) with respect to $R$ is

$$
\frac{dP(n \geq n_c)}{dR} = -\phi \left( \frac{\sqrt{1 - R^2}}{R} - \frac{\alpha}{R} \right) \left( -\frac{\sqrt{1 - R^2}}{R} + \frac{\alpha}{R} \right) \frac{1}{R^2}.
$$

It is easy to see that the derivative is decreasing in $R$ if

$$
\alpha > \tau \sqrt{\frac{1}{1 - R^2}}.
$$

since $\phi(\cdot) > 0$ and $\frac{1}{R^2} > 0$. The maximum of the right-hand side is in $R^2 = 0$; hence, the inequality holds for all values of $R^2$ when $\alpha > \tau$, which implies that the fraction of selected individuals decreases in $R^2$. If $\alpha < \tau$, there is an interval of $R^2$ in which the derivative is positive. For example, when $\alpha = -2$, this interval for $R^2$ is $(0, 0.59)$, which becomes immediately apparent from figure 5.2.

By combining figure 5.1 and figure 5.2, i.e. expression (5.10), we obtain the expected profits for all individuals on the mailing list. This is depicted in figure 5.4. Here, we also see that the expected profits increase in $R^2$. Profits
increase, however, with a slower rate than in figure 5.1 since we have to multiply (5.8) with a value smaller than one to obtain (5.10). Note that expected profits of (5.10) equal that of (5.8) if $R^2 \approx 0$, since all or none of the individuals are selected. The curves in figure 5.4 are rather flat but it must be realized that a direct mail campaign may involve millions of individuals and this increase in profits could turn out to be a very large number for the total database. Moreover, the relation holds for a particular $x$-vector. When an additional explanatory variable is added to the equation, the value of $\alpha$ changes and so the curve changes.

In the sections below we consider (5.1) with one regressor. It is therefore useful to give the relevant expression for the optimal cutoff index and profit function here. Henceforth we also use the traditional normalization of $\sigma_x^2 = 1$. The optimal cutoff index reduces to an optimal cutoff point, which is defined by

$$\xi_c \equiv \frac{\tau - \alpha}{\beta}. \quad (5.11)$$
and (5.8) becomes

$$E \left( a \Phi(\alpha + \xi \beta) - c \mid \xi \geq \xi_c \right) = a \frac{\int_{\xi_c}^{\infty} \Phi(\alpha + \xi \beta) \phi \left( \frac{\xi - \mu}{\sigma} \right) \frac{1}{\sigma} \, d\xi}{1 - \Phi \left( \frac{\xi_c - \mu}{\sigma} \right)} - c, \quad (5.12)$$

where $\phi \left( \frac{\xi - \mu}{\sigma} \right) \frac{1}{\sigma}$ is the density of $\xi$. Analogous to (5.8), the denominator gives the fraction of selected individuals.

### 5.3 Postal code information

#### 5.3.1 What is postal code information?

The Netherlands consist of almost 400,000 postal code areas, which divide the total of approximately 6.4 million street name/house number combinations and is a standard part of the address of virtually every individual and organization. Each postal code has a unique combination of four digits and two letters. Postal code information systems (Geo-system and Mosaic) collect information on a variety of characteristics at the postal code level. In the Netherlands, information at the postal code level comprises on average 16 households or firms. The information systems can, for example, be added to an internal list, such that the organization is able to create a customer profile, which can be helpful for further direct marketing activities (e.g. Lix et al. 1995).

The information supplied by Geo-Marktprofiel is collected by specially developed, individualized single-source research on all postal code areas. This information is linked with information contained in the postal address list owned by the Dutch Postal Services. The last information is updated every four months, while every year information of about 50,000 individuals, gathered through questionnaires, is updated. The system contains over 80 neighborhood characteristics such as average income, family composition, urbanization, access to shops, school and sport facilities and house moving frequency. Based on this information, Geo-Marktprofiel classified the postal codes into 336 classes according to the dimensions 1) average income (six levels), 2) family composition and life cycle (seven levels) and 3) urbanization (eight levels).

The Mosaic system describes the postal codes by characteristics such as: types of homes, composition of the family, occupation, education, average income, social class, agricultural employment, the percentage of working woman, religion and various other characteristics. The Mosaic system is made
up of three separate information systems. First, the internal list of Wehkamp (the largest catalog retailer in the Netherlands). Second, the Wehkamp list is linked with the postal address list owned by the Dutch Postal Services. Third, individually specific information on 800 000 households is used, obtained by a written questionnaire. The three sources are updated every year.

Although individuals in the same postal code area do not possess the same characteristics, it is assumed that those individuals have a certain level of homogeneity. According to Baier (1985), people with comparable interests tend to cluster geographically. Furthermore, their purchase decisions are frequently influenced by their desire to emulate friends and neighbors. Therefore, postal codes provide the means to identify clusters of individuals that have a certain degree of similarity in purchase behavior. Postal code information systems rely on the principles of reference group theory as well on the concept of environmental influences on buying behavior (Baier 1985).

5.3.2 Profits and information

A special case of the model (5.1) is when the regressor is not observed at the individual level. For example, information about income may not be available at the individual level, so \( \xi_i \) is not observable, but a proxy \( z_i \) is available at some aggregate level, i.e. for the postal code area to which household \( i \) belongs.

The income of an individual in a specific postal code area is then the average income of that postal code area plus a deviation because obviously not all the individuals in that area have exactly the same income. We call this grouping on the basis of external information.

Consider the model with one regressor, i.e.,

\[
\eta_i = \alpha + \xi_i \beta + \epsilon_i,
\]

with \( \xi \), a scalar, which we do not observe. Please recall that we normalize the variance of \( \epsilon \), hence, \( \epsilon_i \sim N(0, 1) \). We observe \( z_i \), given by

\[
\xi_i = z_i + \omega_i,
\]

where we assume that \( z_i \sim N(0, \sigma_z^2) \) and \( \omega_i \sim N(0, \sigma_\omega^2) \); \( z_i \) and \( \omega_i \) are i.i.d. and mutually independent, and independent from \( \epsilon_i \). Note that, although suppressed by the notation, \( z_i \) is the average of a group of individuals. Substituting (5.14) into (5.13) yields

\[
\eta_i = \alpha + z_i \beta + \epsilon_i,
\]
where $e_i \equiv e_i + \omega_i \beta$. It is clear that $E(z_i e_i) = 0$. The variance of $e_i$ is:

$$\sigma_e^2 \equiv E(e_i + \omega_i \beta)^2 = 1 + \sigma_\omega^2 \beta^2. \quad (5.16)$$

Dividing (5.15) by $\sigma_e$, we obtain

$$\frac{\eta_i}{\sigma_e} = \frac{\alpha}{\sigma_e} + \frac{\beta}{\sigma_e} z_i + \frac{e_i}{\sigma_e} = \tilde{\alpha} + \tilde{z}_i \tilde{\beta} + \tilde{e}_i, \quad (5.17)$$

where $\tilde{\beta}$, $\tilde{\alpha}$, and $\tilde{e}_i$ are implicitly defined. We now have a model that is fully analogous to (5.13): on the left-hand side we have an unobservable variable of which we only observe the sign ($\eta_i/\sigma_e$ has the same sign as $\eta_i$); on the right-hand side we have a regressor (now $z_i$) and a disturbance that, by construction, has unit variance and is uncorrelated with the regressor.

Assume that a direct marketing organization neglects the noise (5.14) and just uses $z_i$ as the regressor in the analysis, obtaining consistent estimates for $\tilde{\alpha}$ and $\tilde{\beta}$ (cf. Yatchew and Griliches 1985). The consequence is that the optimal profits are reduced. This is not only due to the use of $\tilde{\alpha}$ and $\tilde{\beta}$ instead of $\alpha$ and $\beta$, but also due to the fact that the ordering of the population by $z_i$ is different from the one by $\xi_i$. Hence, two effects of imperfect information can be distinguished: the selected fraction is not optimal and the population is incorrectly ordered. These effects are shown above in figure 5.5. The upper axis gives the ordering of four individuals based on $\xi_i$ and the optimal cutoff point $\bar{z}_i$. Because $\xi_3$ and $\xi_4$ are larger than $\bar{z}_i$, individuals 3 and 4 should be selected. The middle and the lower axes give the ordering of the four individuals using the proxy $z_i$. If the same cutoff point is used only individual 2 would be selected, instead of 3 and 4. However, if the cutoff point $z_c$ is adopted, individuals 1, 2 and 4 would be selected.

### 5.3.3 Effects of using proxy information

It is straightforward to find the relevant expression for the fraction of individuals when the proxy is used. We define the optimal cutoff point, analogous to (5.11), as

$$z_c \equiv \frac{\tau - \bar{\alpha}}{\bar{\beta}}, \quad (5.18)$$

indicating that all individuals with $z_i \geq z_c$ are selected. Note that (5.18) reduces to (5.11) if there is no noise. Then $\sigma_w^2 = 0$, $\sigma_e^2 = 1$, and thus $(\bar{\alpha}, \bar{\beta}) = (\alpha, \beta)$. 


The expected profits deriving from this strategy should of course still be based on the $\xi$; this is given by

$$E \left( a\Phi(\alpha + \xi \beta) - c \mid z \geq z_c \right).$$

In section 5.5 we prove that

$$E \left( a\Phi(\alpha + \xi \beta) - c \mid z \geq z_c \right) = E \left( a\Phi(\tilde{\alpha} + z \tilde{\beta}) - c \mid z \geq z_c \right). \quad (5.19)$$

This result indicates that there is no bias in the expected profits, if they are determined on the proxy information and the estimates for $\tilde{\alpha}$ and $\tilde{\beta}$ instead of $\alpha$ and $\beta$. This is an interesting result: even though the $\xi_i$s are unobservable for the direct marketing organization, which therefore does not know $\alpha$ and $\beta$, it can still make a correct appraisal of the expected profits using the imperfect information. Thus, the organization can make an unbiased decision whether the external information is worth its costs. From (5.19) an expression for the expected profits follows directly:

$$E \left( a\Phi(\alpha + \xi \beta) - c \mid z \geq z_c \right) = a \int_{z_c}^{\infty} \Phi(\tilde{\alpha} + z \tilde{\beta}) \phi \left( \frac{\tilde{\alpha}}{\sigma_i} \right) \frac{1}{\alpha_i} \, dz = E \left( a\Phi(\tilde{\alpha} + z \tilde{\beta}) - c \mid z \geq z_c \right) - c. \quad (5.20)$$

This clearly reduces to (5.12) if $z = \xi$. We illustrate the relation between the optimally obtainable profits on the one hand and the quality of the information...
On the value of information

of the individuals on the other hand graphically. In figure 5.6 we depict (5.20), interpreted as a function of \( \sigma_z^2 \). For simplicity of presentation we use the \textit{intra-class correlation}, \( \rho \), defined by

\[
\rho \equiv \frac{\sigma_z^2}{\sigma_\xi^2},
\]

rather than \( \sigma_z^2 \) itself, since \( \rho \) is restricted to the \([0, 1]\)-interval. The figure, based on the choice \( \beta = 2, \sigma_\xi^2 = 1, a = 10 \) and \( c = 1 \), presents the expected profits \textit{per selected individual}, for four values of \( \alpha \). It shows that the expected profits are an increasing function of the intra-class correlation, as they should be.

Figure 5.7 shows how the fraction of selected individuals, i.e. \( P(z \geq z_c) \), depends on \( \rho \). On combining figure 5.6 and figure 5.7 we obtain the graphs giving expected profits for all individuals on the list (see figure 5.8). This is (5.20) multiplied by \( P(z \geq z_c) \). Also here, of course, we see that expected profits increase with the reliability of the proxy.

The interesting feature of figure 5.8, is the behavior at \( \rho = 0 \) (i.e. \( \sigma_z^2 = 0 \) or \( z_i = 0 \ \forall \ i \)). Then the importance of the proxy variable vanishes, i.e. it has no predictive power anymore. This means that it is no longer helpful to use

Figure 5.6: Expected profits per selected individual

![Expected profits per selected individual](image-url)
Figure 5.7: Fraction of selected individuals, i.e. $P(z \geq z_c)$

Figure 5.8: Expected profits per individual on the mailing list
z for the selection of individuals. Hence the direct marketing organization is confronted with the choice between mailing to all or none of the individuals; this decision depends, of course, on $\alpha$. There are two ways to look at the behavior at $\rho = 0$. First, by using (5.17). As $\sigma^2_\zeta = 0$ it follows that $\sigma^2_\zeta = \sigma^2_\epsilon$, so $\sigma^2_\zeta = 1 + \sigma^2_\zeta \beta^2$. Hence, $\tilde{\alpha} = \alpha/\sqrt{1 + \sigma^2_\zeta \beta^2}$, and the organization should send all individuals a mailing when $a \Phi(\tilde{\alpha}) \geq c$. Second, by notifying that we actually have a situation without a regressor. Hence, the organization uses the mean expected response probability, which is obtained by integrating over $\xi$ (cf. result (4.9) in chapter 4),

\[
\begin{align*}
P(\eta > 0) &= \int_{-\infty}^{\infty} \Phi(\alpha + \xi \beta) \phi \left( \frac{\xi}{\sigma_\xi} \right) \frac{1}{\sigma_\xi} \, d\xi \\
&= \Phi \left( \frac{\alpha}{\sqrt{1 + \sigma^2_\zeta \beta^2}} \right) \\
&= \Phi(\tilde{\alpha}).
\end{align*}
\]

For one particular value of $\alpha$ the organization is indifferent between mailing all or none of the individuals. This occurs if $a \Phi(\tilde{\alpha}) = c$, i.e. $\alpha = \tau \sqrt{1 + \sigma^2_\zeta \beta^2} \approx -2.87$, for the parameter setting underlying the graphs.

A glance at these figures makes clear that they are similar to those in section 5.2. This is not surprising since we can express the $R^2$ of (5.15) as a function of $\rho$ times a constant, i.e.

\[
R^2 = \frac{\sigma^2_\zeta \beta^2}{\sigma^2_\eta} = \rho \frac{\sigma^2_\zeta \beta^2}{\sigma^2_\eta}.
\]  

(5.21)

Note that the constant is the $R^2$ of model (5.13), which is 0.8 for the given parameter setting. Hence, figures 5.6 to 5.8 also give the relation between $R^2$ and expected profit, when the $x$-axes are rescaled from 0 to 0.8; the additional $x$-axes show this.

We are now in a position to discuss the use of these graphs. Assume that $z$ has already been observed. A test mailing (the cost of which we neglect for the sake of simplicity) is held to estimate consistently the values of $z_\zeta$, $\tilde{\alpha}$, $\beta$ and $\sigma^2_\zeta$. Now information is offered with intra-class correlation $\rho_\zeta$. In order to assess the value of information the organization needs to have an idea of the reliability of the information it already has; let this be denoted by $\rho_\eta$. Combining this with
the values of $\tilde{\alpha}$, $\tilde{\beta}$ and $\sigma^2_\varepsilon$ as computed from the test mailing gives us values of $\alpha$, $\beta$ and $\sigma^2_w$. It holds that

$$\tilde{\alpha} = \frac{\alpha}{\sigma_e}$$  \hspace{1cm} (5.22) \\
$$\tilde{\beta} = \frac{\beta}{\sigma_e}$$  \hspace{1cm} (5.23) \\
$$\sigma^2_\varepsilon = \sigma^2_e - \sigma^2_w$$  \hspace{1cm} (5.24) \\
$$\rho_0 = \frac{\sigma^2}{\sigma^2_\varepsilon}$$  \hspace{1cm} (5.25)

where (5.22) and (5.23) follow from (5.17), $\sigma_e$ is given by (5.16), and (5.24) implicitly follows from (5.14). The system (5.22)–(5.24) can be interpreted as a set of three equations in four unknown parameters ($\alpha$, $\beta$, $\sigma^2_e$, $\sigma^2_\varepsilon$). With an idea $\rho_0$ on $\rho$ added as in (5.25), however, the system can be solved to yield

$$\sigma^2_\varepsilon = \frac{\sigma^2}{\rho_0}$$ \\
$$\sigma^2_w = \sigma^2_\varepsilon \left( \frac{1 - \rho_0}{\rho_0} \right)$$ \\
$$\alpha = \frac{\tilde{\alpha}}{\sqrt{1 - \sigma^2_w \tilde{\beta}^2}}$$ \\
$$\beta = \frac{\tilde{\beta}}{\sqrt{1 - \sigma^2_w \tilde{\beta}^2}}$$

Given this solution the relevant graph in the sense of figure 5.8 can be drawn. Information with intra-class correlation $\rho_1$ is worthwhile if the expected profit at $\rho_1$ minus the expected profit at $\rho_0$ exceeds the purchase cost of information per individual on the mailing list.

### 5.3.4 Application

In order to illustrate the concepts introduced above, we discuss an application based on synthetic data. We consider model (5.13) with $\alpha = -1.5$, $\beta = 1$, $\xi_i \sim N(0, 0.25)$, and $\varepsilon_i \sim N(0, 1)$. Hence, the $R^2$ of the underlying model is 0.2. We choose $a = 25$ and $c = 2$. We generate the data in such a way that, apart from household information, we have two levels of grouping. At the
Table 5.1: Estimates of probit model and optimal cutoff point

<table>
<thead>
<tr>
<th>variable</th>
<th>coefficient</th>
<th>estimate</th>
<th>theoretical</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\xi$</td>
<td>$\alpha$</td>
<td>-1.526</td>
<td>-1.500</td>
</tr>
<tr>
<td></td>
<td>$\beta$</td>
<td>0.979</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>$\xi_z$</td>
<td>0.123</td>
<td></td>
</tr>
<tr>
<td>$z_1$</td>
<td>$\alpha$</td>
<td>-1.450</td>
<td>-1.420</td>
</tr>
<tr>
<td></td>
<td>$\beta$</td>
<td>0.917</td>
<td>0.947</td>
</tr>
<tr>
<td></td>
<td>$z_z$</td>
<td>0.048</td>
<td></td>
</tr>
<tr>
<td>$z_2$</td>
<td>$\alpha$</td>
<td>-1.399</td>
<td>-1.367</td>
</tr>
<tr>
<td></td>
<td>$\beta$</td>
<td>0.900</td>
<td>0.911</td>
</tr>
<tr>
<td></td>
<td>$z_z$</td>
<td>-0.006</td>
<td></td>
</tr>
</tbody>
</table>

Superscript 1 Asymptotic standard error in parentheses

first level the average is determined on the basis of 16 observations and at the second level of 256 observations. We generate 50 176 observations (to have balanced groups), of which 5 000 are used for estimation.

A certain postal code area, with 16 households, has a specific mean, which we call $\mu_1$. We have 50 176/16=3136 different values of $\mu_1$. A group of 16 postal code areas (i.e. 256 households) has mean $\mu_2$ of which we have 3136/16=196 different values. To generate the data we start with drawing 196 values of $\mu_2 \sim N(0, 0.3)$. For each $\mu_2$ we draw 16 values of $\mu_1$, $\mu_1 \sim N(\mu_2, 0.3)$. Finally, we draw 16 values, $\xi_i \sim N(\mu_1, 1)$, for each $\mu_1$; these are the household data. The variances of $\xi$, $\mu_1$ and $\mu_2$ are chosen in such a way that the intra-class correlations have the reasonable values of 0.5 and 0.15, respectively. The data have to be transformed such that $\sigma_{\xi_i}^2 = 0.25$. The values of $\xi_i$ are used to determine the sample means for the two levels of grouping: $z_{i1}$ for grouping over 16 households and $z_{i2}$ for grouping over 256 households. These sample means are used for estimation.

We follow the method described in this section. The estimated intra-class correlation coefficients are 0.536 and 0.182, respectively. First, we estimate the coefficients by ML using $\xi$, $z_1$ and $z_2$; Table 5.1 gives the results. The theoretical results are based on the asymptotic results of (5.17). The coefficients decrease as we go from $\xi$ to $z_1$ and $z_2$. Note that the theoretical values of the coefficients are within two times the standard error. The standard errors of $\alpha$ decrease and
Table 5.2: Profits and the probability of being selected

<table>
<thead>
<tr>
<th>variable</th>
<th>method I</th>
<th>method II$^1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\xi$</td>
<td>profits</td>
<td>39 471</td>
</tr>
<tr>
<td></td>
<td>$P(\xi \geq \xi_i)$</td>
<td>0.40</td>
</tr>
<tr>
<td>$z_1$</td>
<td>profits</td>
<td>29 662</td>
</tr>
<tr>
<td></td>
<td>$P(z \geq z_i)$</td>
<td>0.45</td>
</tr>
<tr>
<td>$z_2$</td>
<td>profits</td>
<td>18 876</td>
</tr>
<tr>
<td></td>
<td>$P(z \geq z_i)$</td>
<td>0.51</td>
</tr>
<tr>
<td>no info.</td>
<td>profits</td>
<td>8 648</td>
</tr>
<tr>
<td></td>
<td>$P(\text{selection})$</td>
<td>1</td>
</tr>
</tbody>
</table>

$^1$ Based on (5.20)

those of $\beta$ increase from $\xi$ to $z_1$ and $z_2$. This reflects the uncertainty of the effect of the observed variable on $\eta$.

We determine the profits (for the whole mailing list) for the three types of information in two ways. The easiest way is to use the optimal cutoff point to determine the households that would have been selected, and count the households who actually responded. That is,

$$\sum_{i=1}^{N} (ay_i - c) I(\xi_i \geq \xi),$$

where $I(\cdot)$ is an indicator function which is one if the argument is true and zero otherwise. We call this method I. Note that this gives the actual profits of the $N$ households. The other way to determine the profits, method II, is by computing (5.20) and multiplying this by the number of selected households; this gives the expected profits. Table 5.2 gives the results of both methods. The case of using no information at all is presented as a benchmark. The table shows that profits increase and the fraction of selected households decreases with the intra-class correlation (cf. figures 5.6 and 5.7).

In order to examine our approach in somewhat more detail we consider the case that the organization has observed $z_2$ and knows the intra-class correlation of this information. Then, a mailing is held to estimate the values of $(\tilde{\alpha}, \tilde{\beta}, z_c, \sigma^2_z)$ consistently. These values, combined with the intra-class correlation, give the values of $(\alpha, \beta, \sigma^2_\omega, \sigma^2_\xi)$. These values enable us to determine the
Table 5.3: Break-even price (per household) for buying better information. Entry \((i, j)\) denotes the break-even price between information on level \(i\) and information on level \(j\).

<table>
<thead>
<tr>
<th>Entry</th>
<th>(\xi)</th>
<th>(z_1)</th>
<th>(z_2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(z_1)</td>
<td>0.195</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(z_2)</td>
<td>0.410</td>
<td>0.215</td>
<td></td>
</tr>
<tr>
<td>no info.</td>
<td>0.614</td>
<td>0.419</td>
<td>0.204</td>
</tr>
</tbody>
</table>

expected profits for every value of \(\rho\). The expected profits for the intra-class correlation for \(z_1\) and \(\xi\), i.e. \(\rho = 0.536\) and \(\rho = 1\), are 27 803 and 37 139 respectively. These values are close to the expected profits given in the last column of table 5.2. Hence, if the organization knows the intra-class correlations of the observed and the new data, the proposed method is indeed able to determine the expected profits.

The bottom line is to attach a monetary value to information. In other words, what is an organization willing to pay for better information? We determine the price at which an organization is indifferent between buying and not buying better information; we called this the break-even price. Table 5.3 presents the break-even prices. For example, if the organization has information on level \(z_2\) and it can buy information on level \(z_1\), it is willing to pay 0.215 per household at most.

5.4 Measurement error

An other form of proxy information is when the regressor of interest, \(\xi\), is not measured exactly but an indicator is used instead,

\[
x_i = \xi_i + v_i,
\]

(5.26)

where \(x_i\) is the indicator and \(v_i\) is the noise when using \(x_i\) instead of \(\xi_i\), assumed to be independently normally distributed, \(v_i \sim N(0, \sigma^2)\); \(v_i\) is assumed to be independently distributed from \(v_j, j \neq i\), from \(\xi\) and from \(\epsilon\). This is the so-called measurement error model (e.g. Judge et al. 1985, chapter 17). We still consider the case with one explanatory variable. Note that the observed value of the proxy variable \(x_i\) is in principle different for each individual, unlike
the case of grouping where the observed value of the proxy variable \( z_i \) is similar for a group of individuals. Moreover, here the proxy variable \( x_i \) is not independently distributed from the error term \( \xi_i \), whereas the proxy variable \( z_i \) is independently distributed of the error term \( (\omega_i) \) in the case of grouping.

From (5.26) we have

\[
x_i \sim N(0, \sigma_x^2) = N(0, \sigma_\xi^2 + \sigma_\nu^2).
\]

Then

\[
q \equiv \frac{\sigma_\xi^2}{\sigma_x^2} = \frac{\sigma_\xi^2}{\sigma_\xi^2 + \sigma_\nu^2}, \tag{5.27}
\]

is defined as the reliability of \( x \) as a proxy for \( \xi \). We now reformulate (5.13) into a form that is based on \( x \) rather than \( \xi \):

\[
\eta_i = \alpha + \xi_i \beta + \epsilon_i
\]

\[
= \alpha + x_i \beta + \{\epsilon_i - \nu_i \beta\}
\]

\[
= \alpha + x_i \frac{\sigma_\xi^2}{\sigma_x^2} \beta + \left\{\epsilon_i - \nu_i \beta + x_i (\beta - \frac{\sigma_\xi^2}{\sigma_x^2} \beta)\right\}
\]

\[
= \alpha + x_i \frac{\sigma_\xi^2}{\sigma_x^2} \beta + \left\{\epsilon_i - \nu_i \beta + x_i \frac{\sigma_\nu^2}{\sigma_x^2} \beta\right\}
\]

\[
\equiv \alpha + x_i \frac{\sigma_\xi^2}{\sigma_x^2} \beta + u_i, \tag{5.28}
\]

where \( u_i \) is implicitly defined, and has the property

\[
E(x_i u_i) = E\left\{(\xi_i + \nu_i)(\epsilon_i - \nu_i \beta + (\xi_i + \nu_i) \frac{\sigma_\nu^2}{\sigma_x^2} \beta)\right\}
\]

\[
= \sigma_\xi^2 \frac{\sigma_\nu^2}{\sigma_x^2} \beta - \sigma_\nu^2 \beta + \frac{\sigma_\nu^4}{\sigma_x^2} \beta
\]

\[= 0.\]

For the variance of \( u_i \) there holds

\[
\sigma_u^2 = E\left\{(\epsilon_i - \nu_i \beta + x_i \frac{\sigma_\nu^2}{\sigma_x^2} \beta)^2\right\}
\]

\[
= 1 + \beta^2 \sigma_\nu^2 + \beta^2 \frac{\sigma_\nu^4}{\sigma_x^2} - 2 \beta^2 \frac{\sigma_\nu^4}{\sigma_x^2}
\]
$$= 1 + \beta^2 \left\{ \frac{\sigma^2_\epsilon}{\sigma^2_x} - \frac{\sigma^4_x}{\sigma^2_x} \right\}$$

$$= 1 + \beta^2 \frac{\sigma^2_\epsilon \sigma^2_x}{\sigma^2_x} \tag{5.29}$$

Note that $\frac{\sigma^2_\epsilon}{\sigma^2_x}$ is the probability limit of the OLS estimator of the regression of $\eta$ on $x$. Now we divide (5.28) by $\sigma_u$ to obtain

$$\frac{\eta}{\sigma_u} = \frac{\alpha}{\sigma_u} + x_i \frac{\sigma^2_\epsilon \beta}{\sigma_u} + \frac{u_i}{\sigma_u}$$

$$= \alpha^* + x_i \beta^* + u^*, \tag{5.30}$$

where $\beta^*$, $\alpha^*$ and $u^*$ are implicitly defined. Again we have a model that is fully analogous to (5.13): on the left-hand side we have a variable of which we only observe the sign; on the right-hand side we have a regressor (now $x_i$) and a disturbance that, by construction, has unit variance and is uncorrelated with the regressor. As before, the direct marketing organization neglects the noise (5.26) and just uses $x_i$ as the regressor in the analysis. Equation (5.28) shows the implication: because $u_i$ (hence $u^*$) and $x_i$ are uncorrelated, the analysis yields a consistent estimator of $\alpha^*$ and $\beta^*$, but not for $\alpha$ and $\beta$. Still neglecting the difference between parameters and their consistent estimators, we investigate the implications for expected profits when $\alpha^*$ and $\beta^*$ are used instead of $\alpha$ and $\beta$. Note that this is not a matter of choice on the organization’s behalf since $\beta$ is not identified in a linear latent variable model where all variables are normally distributed (cf. Aigner et al. 1984).

The relevant expressions are analogous to (5.11) and (5.12). The cutoff point of $x$, $x_c$, is defined as

$$x_c = \frac{T - \alpha^*}{\beta^*}, \tag{5.31}$$

which reduces to (5.11) if there is no noise, $\sigma^2_\epsilon = 0$. Hence, the organization should select individuals with $x_i \geq x_c$. Expected profits deriving from this strategy should of course still be based on the $\xi$; this is given by

$$\mathbb{E} \left( a \Phi (\alpha + \xi \beta) - c \mid x \geq x_c \right),$$

instead of the left hand side of (5.12). In section 5.5 we prove that

$$\mathbb{E} \left( a \Phi (\alpha + \xi \beta) - c \mid x \geq x_c \right) = \mathbb{E} \left( a \Phi (\alpha^* + \beta^* x) - c \mid x \geq x_c \right). \tag{5.32}$$
This result indicates that there is no bias in the expected profits, if they are determined on the proxy information and the inconsistent estimates. Hence, even though \((\xi, \beta, \alpha)\) is unknown to the direct marketing organization, it can still make a correct appraisal of the expected profits. From (5.32) an expression for the expected profits follows directly:

\[
E \left( a \Phi(\alpha + \xi \beta) - c \mid x \geq x_i \right) = a \frac{\int_{x_i}^{\infty} \Phi(\alpha^* + x \beta^*) \phi \left( \frac{x}{\sigma_\epsilon} \right) \frac{1}{\sigma_\epsilon} \, dx}{1 - \Phi \left( \frac{x_i}{\sigma_\epsilon} \right)} - c,
\]

which equals (5.12) when \(x = \xi\).

The relation between the optimally obtainable profits and the quality of the information on the individuals can be depicted in figures similar to figures 5.6, 5.7 and 5.8. Here we summarized the quality of information in terms of intra-class correlation. Now we do so using the reliability (5.27). In section 5.5 we prove that, if the underlying distributions are normal, exactly the same figures are obtained. That is, if \(\rho = q\), the expected profits and the percentage of selected individuals have the same value for both forms of imperfect information. The intuition behind this result is that the explanatory power of the two specifications is equal if \(\rho = q\). This is obvious when we express the \(R^2\) as a function of \(q\):

\[
R^2 = \frac{\sigma_\xi^2}{\sigma^2 \sigma^2} \beta^2 = q \frac{\sigma_\xi^2 \beta^2}{\sigma^2},
\]

analogous to (5.21).

Hence, the two types of imperfect information, grouping and measurement error, are mathematically largely comparable. It should be realized, however, that the assumption underlying the structure of the data is completely different. That is, in the case of measurement error, the data generating process is fully specified, in contrast with the case of grouping. Loosely speaking, it shows the difference between an econometric specification and a statistical specification.

5.5 Proofs

In this section we prove (5.19) and (5.32).

**Proof of (5.19)** Consider model (5.13) with \(\xi_i = z_i + \omega_i\). Write the response probability as

\[
P(y_i = 1 \mid z_i \geq z_c)
\]
\[ P(\eta_i > 0 \mid z_i \geq z_c) = P(\alpha + \xi_i \beta + \epsilon_i > 0 \mid z_i \geq z_c) \]
\[ = P(\alpha + (z_i + \omega_i) \beta + \epsilon_i > 0 \mid z_i \geq z_c) \]
\[ = P(\epsilon_i + \omega_i \beta > -\alpha - z_i \beta \mid z_i \geq z_c) \]
\[ = P(\tilde{\epsilon}_i > -\tilde{\alpha} - z_i \tilde{\beta} \mid z_i \geq z_c). \] (5.33)

The important aspect is the equality between (5.33) and (5.34). Evaluating these probabilities enables us to prove (5.19). That is,
\[
\begin{align*}
&\{P(\tilde{\epsilon} > -\tilde{\alpha} - z \tilde{\beta} \mid z \geq z_c)P(z \geq z_c) \\
= &\int_{z_c}^{\infty} \int_{-\tilde{\alpha} - t \tilde{\beta}}^{\infty} f_{\tilde{\epsilon},z}(s, t) \, ds \, dt \\
= &\int_{z_c}^{\infty} \int_{-\tilde{\alpha} - t \tilde{\beta}}^{\infty} f_{\tilde{\epsilon}}(s) f_z(t) \, ds \, dt \\
= &\int_{z_c}^{\infty} P(\tilde{\epsilon} \geq -\tilde{\alpha} - t \tilde{\beta}) f_z(t) \, dt \\
= &E(P(\tilde{\epsilon} > -\tilde{\alpha} - z \tilde{\beta}) \mid z \geq z_c)P(z \geq z_c) \\
= &E \left( \Phi(\tilde{\alpha} + z \tilde{\beta}) \mid z \geq z_c \right)P(z \geq z_c),
\end{align*}
\] (5.35)

where the third step uses the independence between \(\tilde{\epsilon}\) and \(z\). Moreover,
\[
\begin{align*}
&\{P(\epsilon > -\alpha - \xi \beta \mid z \geq z_c)P(z \geq z_c) \\
= &\int_{z_c}^{\infty} \int_{-\alpha - (\omega + t) \beta}^{\infty} f_{\epsilon,v,z}(s, k, t) \, ds \, dk \, dt \\
= &\int_{z_c}^{\infty} \int_{-\alpha - (\omega + t) \beta}^{\infty} f_{\epsilon,v}(s, k) f_z(t) \, ds \, dk \, dt \\
= &\int_{z_c}^{\infty} P(\epsilon > -\alpha - (\omega + t) \beta \& -\infty < \omega < \infty) f_z(t) \, dt \\
= &\int_{z_c}^{\infty} P(\epsilon > -\alpha - \xi \beta) f_z(t) \, dt \\
= &E \left( \Phi(\alpha + \xi \beta) \mid z \geq z_c \right)P(z \geq z_c),
\end{align*}
\] (5.37)
Proofs

where the fourth step uses the independence between \( \epsilon, \omega \) and \( z \). Since (5.35) equals (5.37), and (5.36) equals (5.38), it follows that

\[
E \left( a \Phi(\alpha + \xi \beta) - c \mid z \geq z_c \right) = E(a \Phi(\tilde{\alpha} + z \tilde{\beta}) - c \mid z \geq z_c),
\]

which proves (5.19).

**Proof of (5.32)** This proof is analogous to the proof of (5.19). Consider model (5.13) with \( x_i = \xi_i + \nu_i \). Write the response probability, using (5.28) and (5.30), as

\[
P(y_i = 1 \mid x_i \geq x_c) = P(\eta_i > 0 \mid x_i \geq x_c) = P(\alpha + \xi_i \beta + \epsilon_i > 0 \mid x_i \geq x_c)
\]

\[
= P \left( \frac{u_i}{\sigma_u} > -\frac{\alpha}{\sigma_u} - x_i \frac{1}{\sigma_u} \beta \mid x_i \geq x_c \right)
\]

\[
= P(u_i^* > -\alpha^* - x_i^* \beta^* \mid x_i \geq x_c)
\]

where the latter step holds since \( u_i^* \) and \( x_i \) are, by construction, mutually independent. Again, we evaluate these probabilities. Using the independence between \( u^* \) and \( x \), it clear, analogous to the equality between (5.35) and (5.36), that

\[
P(u^* + x^* \beta^* > -\alpha^* \mid x \geq x_c)P(x \geq x_c) = E(\Phi(\alpha^* + x^* \beta^*) \mid x \geq x_c) = E(\Phi(\alpha^* + x^* \beta^*) \mid x \geq x_c).
\]

Furthermore,

\[
P(\epsilon + \xi \beta > -\alpha \mid x \geq x_c)P(x \geq x_c)
\]

\[
= \int_{x_c}^{\infty} \int_{-\alpha + (k-t)\beta}^{\infty} f_{\epsilon,\nu,k}(s, k, t) \, ds \, dk \, dt
\]

\[
= \int_{x_c}^{\infty} \int_{-\alpha + (k-t)\beta}^{\infty} f_{\epsilon,\nu,k}(s, k \mid x = t) f_{\nu}(t) \, ds \, dk \, dt
\]

\[
= \int_{x_c}^{\infty} \Phi(\alpha + \xi \beta) f_{\nu}(t) \, dt
\]

\[
= E \left( \Phi(\alpha + \xi \beta) \mid x \geq x_c \right) P(x \geq x_c),
\]
which, using (5.41), and the equality between (5.39) and (5.40), implicitly proves (5.32).

In words the proofs can be sketched as follows. The probabilities of interest ((5.33) and (5.39)) are conditioned on variables that do not appear explicitly in argument $z$ and $x$ respectively. However, we can rewrite the argument in terms of either $z$ or $x$. Consequently, we have a probability that depends on two independent variables (by construction in the case of $x$) conditional on $z$ or $x$. Finally, we express the conditional probability as a conditional expectation. Since the parameters are identified up to scale, we can write the probability as a standard normal distribution function.

As indicated in section 5.3 and 5.4, the results imply that even though the direct marketing organization cannot consistently estimate the parameters and does not observe the variable of interest, it can still make a correct appraisal of the expected profits for the whole range of $p$ and $q$.

Now we prove that if $\rho = q$,

$$
E \left( \Phi(\alpha z + \beta \tilde{y}) \mid z \geq z_c \right) = E \left( \Phi(\alpha' x + \beta' \tilde{x}) \mid x \geq x_c \right). \tag{5.42}
$$

First we prove the following three auxiliary results:

$$
\frac{\sigma^2_x}{\sigma^2_z} f_{\omega|x} \left( k - \frac{\sigma^2_x}{\sigma^2_z} t \mid \xi = k \right) = f_x(t - k). \tag{5.43}
$$

$$
\frac{\sigma^2_z}{\sigma^2_x} x_c = z_c. \tag{5.44}
$$

$$
P(x \geq x_c) = P(z \geq z_c). \tag{5.45}
$$

**Proof of (5.43)**

$$
\omega \mid \xi = k \sim N \left( \frac{\sigma^2_\omega}{\sigma^2_\xi} k, \frac{\sigma^4_\omega}{\sigma^4_\xi} \right) = N \left( \frac{\sigma^2_\omega}{\sigma^2_\xi} k, \frac{\sigma^2_\omega}{\sigma^2_\xi} \frac{\sigma^2_\omega}{\sigma^2_\xi} \right),
$$

since, using $\sigma^2_\omega / \sigma^2_\xi = \sigma^2_\omega / \sigma^2_\xi$, i.e. $q = \rho$,

$$
\sigma^2_\omega - \frac{\sigma^4_\omega}{\sigma^2_\xi} = \sigma^2_\omega - \sigma^2_\omega \sigma^2_\omega = \sigma^2_\omega \frac{\sigma^2_\omega}{\sigma^2_\xi} = \frac{\sigma^2_\omega \sigma^2_\omega}{\sigma^2_\xi}.
$$

so

$$
\frac{\sigma^2_x}{\sigma^2_z} f_{\omega|x} \left( k - \frac{\sigma^2_x}{\sigma^2_z} t \mid \xi = k \right)
$$
Proofs 107

\[
\begin{align*}
&= \frac{1}{\sqrt{2\pi}} \frac{\sigma^2}{\sigma^w \sigma^z} \exp \left( -\frac{1}{2} \frac{\sigma^2}{\sigma^w \sigma^z} \left( k \frac{\sigma^2}{\sigma^w} \right)^2 \right) \\
&= \frac{1}{\sqrt{2\pi}} \frac{\sigma^z}{\sigma^w} \exp \left( -\frac{1}{2} \frac{\sigma^z}{\sigma^w} \left( t^2 - k \right)^2 \right) \\
&= \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{1}{2} \frac{\sigma^2}{\sigma^w} \left( t - k \right)^2 \right) \\
&= f_z(t - k),
\end{align*}
\]

where we use

\[1 - \frac{\sigma^2}{\sigma^z} = \frac{\sigma^2}{\sigma^w} = \frac{\sigma^2}{\sigma^z},\]

and

\[
\frac{\sigma^2}{\sigma^z \sigma^w \sigma^z} = \frac{\sigma^2}{\sigma^z \sigma^w \sigma^z} = \frac{1}{\sigma^2}, \tag{5.46}
\]

in the second and third step respectively. This proves (5.43).

**Proof of (5.44)**

\[\frac{\sigma^2}{\sigma^z} x_c = z_c \]

\[\frac{\sigma^2}{\sigma^z} (\Phi^{-1}(c/a) - \alpha^*) = \frac{\Phi^{-1}(c/a) - \alpha}{\beta} \]

\[\frac{\sigma^2}{\sigma^z} \beta^* = \frac{\alpha}{\beta} \]

using (5.31) and (5.18). Note that the numerators are equal if \(\alpha^* = \tilde{\alpha}\), i.e. \(\alpha^* = \alpha^u = \alpha^c\), and this holds if \(\alpha^* = \alpha^c\). The same argument holds for the denominators:

\[\sigma^2 \beta^* = \sigma^z \frac{\sigma^2}{\sigma^z} \beta^*/\sigma^z = \sigma^2 \beta^*/\sigma^z.\]

Note that \(\sigma^* = \sigma^u \iff 1 + \beta^2 \sigma_2^2 \sigma_1^2 / \sigma^2 = 1 + \beta^2 \sigma^2 / \sigma^2,\) i.e. \(\sigma^2 \sigma^2 / \sigma^2 = \sigma^2,\) where the latter is obtained from (5.46); this proves (5.44).

**Proof of (5.45)**

\[P(x \geq x_c) = 1 - \Phi \left( \frac{x_c}{\sigma^z} \right) = 1 - \Phi \left( \frac{\sigma^2}{\sigma^z \sigma^z} x_c \right) = 1 - \Phi \left( \frac{1}{\sigma^z} z_c \right) = P(z \geq z_c),\]
where we use (5.44) and $\sigma_x^2 = \sigma_x^2 / \sigma_z$.

**Proof of (5.42)**

$$E \left( \Phi(\alpha + \xi \beta) \mid z \geq z_c \right) P(z \geq z_c)$$

$$= \int_{z_c}^{\infty} P(\epsilon \geq -\alpha - \xi \beta) \int_{-\infty}^{\infty} f_{\xi,\omega}(k, k - s) \, dk \, ds$$

$$= \int_{z_c}^{\infty} P(\epsilon \geq -\alpha - \xi \beta) \int_{-\infty}^{\infty} f_{\omega \mid \xi}(k - s \mid \xi = k) f_{\xi}(k) \, dk \, ds$$

$$= \int_{z_c}^{\infty} P(\epsilon \geq -\alpha - \xi \beta) \int_{-\infty}^{\infty} \frac{\sigma_\xi^2}{\sigma_\omega^2} f_{\omega \mid \xi}(k - s \mid \xi = k) f_{\xi}(k) \, dk \, dt$$

$$= \int_{z_c}^{\infty} P(\epsilon \geq -\alpha - \xi \beta) \int_{-\infty}^{\infty} f_{\xi}(t - k) f_{\xi}(k) \, dk \, dt$$

$$= \int_{x_c}^{\infty} \Phi(\alpha + \xi \beta) f_x(t) \, dt$$

$$= E \left( \Phi(\alpha + \xi \beta) \mid x \geq x_c \right) P(x \geq x_c),$$

where we use the substitution $s = \sigma_\xi^2 t / \sigma_z^2$ and (5.44) in the fourth step and (5.43) in the fifth step. Using (5.45) we obtain (5.42), which completes the proof.

### 5.6 Discussion and conclusion

We have shown how the monetary value of information in direct marketing can be assessed and used in a decision process where benefits and costs of data acquisition are weighed against each other. Our approach is stylized and theoretical, but we wish to emphasize that this approach would be the core of any more extended model. Our approach provides the statistical aspects involved in assessing the value of information. There are various steps that could be taken on the road to more realism.

In order to decide whether it is profitable to buy new information, the organization should have an idea of the intra-class correlation of the new information. These could be obtained from small studies. Alternatively, a minimum intra-class correlation of new information could be specified. The result of the analysis would be a lower bound to the maximally attainable increase in
Discussion and conclusion

profits. A more refined way to assess the value of new information is by specifying a distribution of the intra-class correlation, indicating the organization’s uncertainty about $\rho$. By integrating the expected profits over $\rho$ we obtain the expected value of new information.

An important extension would be to incorporate more regressors. This extension is not trivial, because all regression coefficients are estimated inconsistently even if only one regressor is measured imperfectly (grouping or measurement error). A possible solution of this problem is to neglect the imperfect information, i.e. exclude the variable from the model. However, we then have an omitted variable problem and that also gives us inconsistent estimates (e.g. Yatchew and Griliches 1985), even if the variables are uncorrelated. A natural question that arises is whether it is preferable to include or to omit the proxy variable, which of course depends on the correlation with the other variables. If, for example, the correlation between the proxy variable and an other variable is nearly perfect, the additional information of the proxy variable is superfluous for selection and it will probably be better to omit the proxy variable.
Chapter 6

Optimizing the characteristics of the mailing

6.1 Introduction

An important aspect of the success of a direct mailing campaign are the characteristics, or the so-called communication elements, of the direct mailing package (e.g. Fraser-Robinson 1989, Roberts and Berger 1989, Throckmorton 1992 and Vögele 1992). The package often consists of an envelope, letter, brochure, response device and return device. Characteristics that are essential to the design of the mailing package relate to its form (size of the envelope, use of graphics etc.) and to aspects of the contents (style of writing, use of testimonials etc.). In order to be able to manipulate the characteristics of the mailing, the direct marketing manager needs to know to what extent the various characteristics of the mailing affect the response process.

With respect to the mailing package, two aspects of the response process can be distinguished: (1) open the envelope and (2) pay attention to the contents. The characteristics of the outside of the mailing (the envelope and other visible or otherwise noticeable elements) influence the probability of the envelope being picked up and opened. The characteristics of the mailing, e.g. the design of the letter and the brochure, influence (1) the probability of taking notice of the offer submitted in the mailing and (2) the probability of responding. Creativity, experience, and theory determine the use of specific characteristics for the envelope, letter and offer. These two aspects of the response process could be affected by the characteristics of the individuals and situational factors (e.g. season, other mailings received).

The objective of this chapter is twofold. First, we propose two methods to improve the effectiveness of direct mail campaigns by creating an optimal mailing design. We do so by analyzing several characteristics simultaneously.
A traditional way to analyze the effect of several characteristics is by studying each aspect separately (Roberts and Berger 1989, p. 198). Generally, this is inefficient since a large number of mailings is needed to achieve a certain reliability in the estimates of the effect of the characteristics on the response rate. Moreover, there is no opportunity to take interaction between the mailing characteristics into account.

The second objective is to select the targets and optimize the mailing design simultaneously. That is, we incorporate the interaction between target and mailing characteristics in a target selection model. Although the importance of both target and mailing characteristics has been recognized, hardly any attention, to the best of our knowledge, has been given to the interaction between the two, even though it may have substantial effect on the response. For example, for targets who have been customer for a long time a direct marketing organization should use other communication elements than for targets who never bought from the organization before. Therefore, the target characteristics have an effect on the relationship between the mailing characteristics and the response to the mailing. Also, the target selection process changes when, for example, the envelope and the letter contain more incentives. Thus, the mailing characteristics influence the relationship between the target characteristics and the response to the mailing. Some empirical evidence is given by Schibrowsky and Peltier (1995). They show that it is preferable to set a high decision frame, denoting the values of possible donations, to traditionally high givers and a low decision frame to traditionally low givers.

Although direct mail practitioners often apply the manipulation of the characteristics (Hoekstra and Vriens 1995), little attention has been given to the optimization of the design of the mailing in the literature. Direct marketing managerial texts generally provide ‘rules of thumb’ and some anecdotal examples to illustrate them. The research that has been done mainly focuses on characteristics that have been put forward by behavioral decision theory with respect to charity behavior. These characteristics include decision frames (Fraser et al. 1988, and Schibrowsky and Peltier 1995), and favorable descriptions of donors, e.g. as ‘helpers’ (DeJong and Oopik 1992). These studies, however, consider the effectiveness of each characteristic separately.

To our knowledge, there are three studies that consider several characteristics simultaneously. Akaah and Korgaonkar (1988) studied the relative importance of ‘risk relievers’ in a direct marketing offer, using conjoint analysis. They found that direct marketers can enhance the effectiveness of a mailing campaign by (1) offering money-back-guarantees rather than free samples, and
(2) using established manufacturer names rather than unknown manufacturer names. Furthermore, they concluded that both new and established products can be sold by means of direct marketing.

James and Li (1993) studied the importance of the design characteristics of the mailing, by interviewing both consumers and managers through a direct questioning procedure asking about the attractiveness of a number of separate design characteristics of the mailing. However, allowing respondents to explain the importance of the various design characteristics of a mailing themselves may not constitute an appropriate task for them and may produce invalid results (Green and Srinivasan 1990).

Recently, Smith and Berger (1996) analyzed the impact of direct marketing appeals on donor decisions. They considered four aspects simultaneously that are seen as important characteristics by behavioral decision theory. These are (1) anchors, denoting a suggested donation (low or high); (2) factual information (present or absent); (3) anecdotal information (present or absent) and (4) positive versus negative information, the so-called framing valence. They concluded that low anchors and positive framing have a positive effect on the response rate but not on the average donation. In contrast, the presence of factual or anecdotal information had no effect on the response rate but had an effect on the average donation. The difference with our applications is that they do not include any creative elements. As a consequence, the costs of each level of the different characteristics are equal. Thus, the preferred mailing is simply the mailing that generates the highest response rate.

This chapter is built up as follows. In the next section we briefly discuss the possibilities of conjoint analysis methodology to influence the response process. This is illustrated with three applications in section 6.3. The first two applications, which are based on Vriens, Van der Scheer, Hoekstra, and Bult (1995), aim at the development of an optimal envelope design and an optimal letter design. The third application, which is based on Van der Scheer, Hoekstra, and Vriens (1996), deals with an ‘opt out’ reply card. This is a reply card with the function to request the organization to be deleted from the mailing list. The emphasis of this application lies on this specific function of the reply card and not so much on the method, since it is the same as used in the second application. Interaction between target and mailing characteristics is discussed in section 6.4. In this section, which is based on Bult, Van der Scheer, and Wansbeek (1996), we continue with the second empirical application of section 6.3. We show that an organization is much better off by sending different mailings
to different targets for the same direct marketing campaign. We conclude in section 6.5.

6.2 Conjoint analysis experiments

Conjoint analysis methodology has received considerable attention from academics (Green and Srinivasan 1978, 1990) and commercial users (Wittink et al. 1994). The many methodological developments which have taken place enabled a broad range of marketing applications (Vriens 1994). As we will demonstrate in this chapter, conjoint analysis is also a useful methodology for identifying an optimal design of various mailing components (envelope, letter, brochure).

As a first possibility we may use traditional conjoint experiments to identify the extent to which various characteristics of e.g. an envelope contribute to its overall attractiveness. In this case we let respondents judge a set of experimentally constructed envelopes with respect to their attractiveness to open it. Using the relative importances and part-worth utilities we are able to draw conclusions about the attractiveness of many envelopes and an optimal design can be constructed. Such an experiment is performed in a laboratory setting and therefore it controls for several extraneous factors.

However, in some cases this approach cannot be implemented for practical reasons. For example, if we want to identify the extent to which various characteristics of a letter contribute to its attractiveness or its power to elicit consumer response (e.g. a donation), the traditional approach would imply that respondents would be asked to evaluate a set of letters. Such a task would make unrealistic demands on the respondents, since it is difficult for them to distinguish between the different letters; besides, they may not be able to judge the different letters due to fatigue. Moreover, it is probably too difficult for respondents to judge which letter will generate their highest response probability or which letter will generate their highest donation.

Therefore, as a second possibility to determine the optimal characteristics of the mailing, we propose the use of the so-called conjoint field-experiments. These field-experiments may be used to measure the extent to which various characteristics of a mailing component contribute to response rates and to the amount of the donation. Again, the mailing component has to be constructed by experimental design. However, instead of eliciting evaluations with respect to the constructed set, each individual in the selected sample is confronted with
only one of the experimentally varied mailings. By (randomly) sending each different mailing to a (large) group of respondents (a test-mailing), the optimal characteristics can be determined on the basis of the response figures. Hence, while attractiveness is assumed to be the underlying factor of the response, it is not judged explicitly by the respondents. It is called a field-experiment since extraneous factors are not controlled for.

In the field-experimental approach we cannot draw conclusions about the attractiveness of the mailing component. Instead, we are able to analyze (1) to what extent the characteristics affect the response (yes/no) to the mailing, and (2) to what extent the characteristics affect the amount of donation. (Cf. Smith and Berger 1996.) A straightforward way to examine these two issues is by testing whether the difference in response between two levels of a characteristic differs significantly from zero. The disadvantage of this method is that it only considers the marginal effect of a particular characteristic, i.e. unconditional on other characteristics. Since we are interested in the conditional effect of a particular mailing design on the response rate, we have to analyze the effects of the characteristics simultaneously. Therefore, we use a probit model (see section 2.4) to analyze the response (yes/no). For the amount of donation we use a tobit model (see section 3.6). This model does not filter the dependent variable down to 0 or 1 (as in probit), but to 0 if the individual does not respond to the mailing or to the amount of donation in case there is response. The estimated coefficients of these models can be used to construct an optimal mailing design and to determine the expected response rate for this design.

6.3 Applications

In this section we empirically apply the methods discussed in the previous section to a Dutch charity foundation. The foundation heavily relies upon direct marketing for the acquisition of its financial donations for health care research. It supports many types of different projects both in the Netherlands and in developing countries. Every year, the foundation sends mailings to almost 1.2 million households in the Netherlands. Because of the steadily increasing competition on the fund-raising market, the foundation has become interested in improving the design of the mailing. In sections 6.3.1 and 6.3.2 we discuss applications that consider the relative importances of a number of design characteristics of the envelope and the letter respectively. The traditional conjoint experiment is used to optimize the envelope design. A conjoint field-experiment is used to determine an optimal letter design.
Table 6.1: The characteristics included in the envelope experiment

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Format</td>
<td>A5</td>
</tr>
<tr>
<td>dummy: $d_1$</td>
<td>1</td>
</tr>
<tr>
<td>dummy: $d_2$</td>
<td>0</td>
</tr>
<tr>
<td>Extra imprint</td>
<td>Hologram</td>
</tr>
<tr>
<td>dummy: $d_3$</td>
<td>1</td>
</tr>
<tr>
<td>dummy: $d_4$</td>
<td>0</td>
</tr>
<tr>
<td>Type of paper</td>
<td>Paper without chlorine</td>
</tr>
<tr>
<td>dummy: $d_5$</td>
<td>1</td>
</tr>
<tr>
<td>dummy: $d_6$</td>
<td>1</td>
</tr>
<tr>
<td>dummy: $d_7$</td>
<td>0</td>
</tr>
<tr>
<td>Sender</td>
<td>Full name</td>
</tr>
<tr>
<td>dummy: $d_8$</td>
<td>1</td>
</tr>
<tr>
<td>Addition</td>
<td>Lottery</td>
</tr>
<tr>
<td>dummy: $d_9$</td>
<td>1</td>
</tr>
</tbody>
</table>

In section 6.3.3 we discuss an application that was executed for a charity foundation that supports activities with respect to the well-being of animals. The aim of this application is to design an optimal ‘opt out’ reply card.

6.3.1 Conjoint experiment for an envelope design

The implementation of a conjoint experiment consists of a number of steps and choices regarding both data collection and data analysis (e.g. Green and Srinivasan 1978). The steps in the data collection phase include: (1) selection of the characteristics, and determination of the relevant levels within each characteristic; (2) choice of the preference model; (3) the data collection method; (4) construction of the stimuli, i.e. the selection of experimental stimuli and the choice how to present the stimuli, and (5) the definition of the dependent
variable. These steps for our application are discussed below; next the data analysis phase will be discussed.

The characteristics and their levels (step one) were determined in consultation with the charity foundation; they are shown in table 6.1. The dummy variables mentioned in the table will be explained due course. The size of the envelope was given three formats: A5, Cabinet (1/3 A4) and Ola (giro card format). The letter can contain an extra imprint in the form of a hologram or extra window. When the paper type is without chlorine this is mentioned on the back-side of the envelope (type of paper). An extra streamer (line of text) could be printed on the envelope. If so, this is represented in type or in handwriting. The left upper corner of the envelope contains the name of the foundation, i.e. the sender. This could either be the full name or an abbreviated form. The addition denotes whether or not an additional remark with regard to a lottery is placed on the envelope.

In step two, again in consultation with the charity foundation, it was decided that no interaction effect would be included and thus we used a main-effects-only model. We discuss this model in the data analysis phase. In step three we chose the full profile method as the data collection method. Under this method, respondents are asked to evaluate a set of stimuli (envelopes) which are defined on all the characteristics that are included in the experiment. We use the basic plans of Addelman (1962) to derive an experimental design (step four). A fractional factorial design (e.g. Addelman 1962) resulted in 16 different types of envelopes used in the experiment. Although our model only contains 10 parameters (see below), the set of 16 envelopes was the smallest experimental set possible, given the number of characteristics and levels. Since we initially estimated the model at the individual level, this design allows for six degrees of freedom. Now the experimental stimuli have to be constructed. In order to confront the respondent with realistic stimuli, real envelopes were constructed. The respondents had to to evaluate the attractiveness of the envelopes with respect to opening the envelopes, using a 10-point rating scale (step 5). A ‘1’ on this scale is defined as being very unattractive (would never open this envelope) and a ‘10’ was defined as being very attractive (would always open this envelope). In addition three extra types of envelopes were constructed. Respondents had to choose the most attractive envelope from these three. These choices are used to assess predictive accuracy.

The population of interest from which our sample was drawn was defined by the foundation. It consisted of those persons in the data base who had donated at least once during the last three years. Furthermore, the population
was narrowed by considering only those persons who resided in a city or its immediate surroundings. From this population a sample of 1692 persons was drawn. From this sample, 360 persons indicated they were willing to cooperate. Subsequently, these persons were invited to a central interviewing location. Unfortunately, only 200 persons appeared at the appointed time and completed the interview; 170 questionnaires were filled out correctly and could be used for the analysis. The sample may thus not be representative for the target population. However, since the primary aim of the empirical application is illustrative, this does not constitute a major problem.

In the data analysis phase we start by specifying and estimating the model. In the data collection phase we already made one decision regarding the model specification (i.e. we assumed a main-effects-only model). In addition, we decided not to set any constraints on the relation between the levels within a characteristic. Consequently, we use the following model:

\[ y_i = \beta D + \epsilon_i, \]

where \( y_i \) is a \((1 \times 16)\) vector with individual’s \( i \) ratings of the 16 envelopes, \( D \) is a \((10 \times 16)\) matrix describing the design of the experiment, \( \alpha \) is the unknown intercept, \( \beta \) is a \((1 \times 10)\) vector of unknown parameters, and \( \epsilon_i \) is a disturbance term assumed to be independent and identically distributed. Each column of the matrix \( D \) consists of an one followed by the values of the dummy variables, \((d_1, \ldots, d_9)\), corresponding to the experimentally constructed envelopes; see table 6.1 for the specification of the dummy variables.

In the conjoint analysis literature, this is called a part-worth model. A part-worth of a level of a characteristic is the value of the parameters times the dummy variable(s) that define this level. For example, the part-worth of ‘A5 format’ is \((\beta_1, \beta_2)(d_1, d_2) = (\beta_1, \beta_2)(1, 0) = \beta_2\); the part-worth of ‘Ola format’ is \((\beta_1, \beta_2)(d_1, d_2) = (\beta_1, \beta_2)(-1, -1) = -\beta_1 - \beta_2\). A related statistic, which is often used in conjoint analysis, is the relative importance. An importance is defined as the difference between the highest and lowest part-worth of a characteristic. The relative importances are derived by normalizing the importances such that they sum to 100. As the name suggests, they are used to indicate the importance of a characteristic. They are comparable with standardized coefficients in a regression model, which are used for a direct comparison of independent variables with respect to their relative explanatory power of the dependent variable. We estimated the model with OLS since it has been found to be a robust estimation method for rating data (Wittink and Cattin 1981). The model is estimated initially at the individual level.
Table 6.2: The relative importances obtained under a three-segment representation

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Segment 1 (N=47)</th>
<th>Segment 2 (N=74)</th>
<th>Segment 3 (N=49)</th>
<th>Overall (N=170)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Format</td>
<td>17.1</td>
<td>14.3</td>
<td>33.5</td>
<td>16.5</td>
</tr>
<tr>
<td>Extra imprint</td>
<td>41.1</td>
<td>22.4</td>
<td>45.5</td>
<td>28.9</td>
</tr>
<tr>
<td>Type of paper</td>
<td>4.2</td>
<td>41.4</td>
<td>9.5</td>
<td>25.0</td>
</tr>
<tr>
<td>Streamer</td>
<td>19.2</td>
<td>8.4</td>
<td>2.7</td>
<td>14.4</td>
</tr>
<tr>
<td>Sender</td>
<td>6.5</td>
<td>2.5</td>
<td>8.1</td>
<td>0.9</td>
</tr>
<tr>
<td>Addition</td>
<td>11.8</td>
<td>11.2</td>
<td>0.7</td>
<td>14.2</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 6.3: Part-worths for the different levels per segment

<table>
<thead>
<tr>
<th>Level</th>
<th>Segment 1 (28%)</th>
<th>Segment 2 (44%)</th>
<th>Segment 3 (28%)</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>A5</td>
<td>-0.443</td>
<td>0.126</td>
<td>0.962</td>
<td>0.210</td>
</tr>
<tr>
<td>Cabinet</td>
<td>-0.014</td>
<td>0.079</td>
<td>-0.444</td>
<td>-0.097</td>
</tr>
<tr>
<td>Ola</td>
<td>0.458</td>
<td>-0.206</td>
<td>-0.518</td>
<td>-0.113</td>
</tr>
<tr>
<td>Hologram</td>
<td>-0.942</td>
<td>-0.201</td>
<td>1.293</td>
<td>0.025</td>
</tr>
<tr>
<td>Extra window</td>
<td>-0.273</td>
<td>-0.125</td>
<td>-0.569</td>
<td>-0.294</td>
</tr>
<tr>
<td>None</td>
<td>1.215</td>
<td>0.326</td>
<td>-0.723</td>
<td>0.270</td>
</tr>
<tr>
<td>Type</td>
<td>0.666</td>
<td>-0.066</td>
<td>0.045</td>
<td>0.168</td>
</tr>
<tr>
<td>Hand written</td>
<td>-0.328</td>
<td>0.133</td>
<td>-0.074</td>
<td>-0.054</td>
</tr>
<tr>
<td>None</td>
<td>-0.338</td>
<td>-0.066</td>
<td>0.029</td>
<td>-0.114</td>
</tr>
<tr>
<td>Paper without chlorine</td>
<td>-0.105</td>
<td>0.488</td>
<td>0.211</td>
<td>0.244</td>
</tr>
<tr>
<td>Recycling paper</td>
<td>0.105</td>
<td>-0.488</td>
<td>-0.211</td>
<td>-0.244</td>
</tr>
<tr>
<td>Full name</td>
<td>0.174</td>
<td>0.030</td>
<td>-0.180</td>
<td>0.009</td>
</tr>
<tr>
<td>Abbreviated form</td>
<td>-0.174</td>
<td>-0.030</td>
<td>0.180</td>
<td>-0.009</td>
</tr>
<tr>
<td>Lottery</td>
<td>-0.310</td>
<td>-0.133</td>
<td>0.017</td>
<td>-0.139</td>
</tr>
<tr>
<td>None</td>
<td>0.310</td>
<td>0.133</td>
<td>-0.017</td>
<td>0.139</td>
</tr>
</tbody>
</table>
In order to translate the individual results into an optimal mailing strategy, we constructed segments. A traditional way to construct the segments is by clustering the individuals on the basis of their estimated part-worths (e.g. Green and Krieger 1991). We employed a nonhierarchical clustering technique to identify the segments. The outcomes of this analysis were subsequently entered in a hierarchical clustering method to fine-tune these results. This procedure indicated a three-segment solution as being optimal. Since this application is illustrative, we used this segmentation procedure. It would have been better to use a more advanced segmentation procedure like latent class analyses. In order to obtain part-worth values of a segment we can either use the average value of the individual part-worths of the individuals belonging to the segment, or perform a new regression using the data of these individuals. The results are identical since the matrix $D$ does not vary across individuals. The relative importances of a segment, however, should be determined on the part-worths of that segment and are not the averages of the individual relative importances.

The relative importances of the characteristics for each of these three segments are presented in table 6.2. The corresponding part-worths for each of these three segments indicate different optimal envelopes for each of these segments (see table 6.3). Please bear in mind that the part-worths represent the importance of the levels of a characteristic, whereas the relative importances represent the importance of the characteristics.

Segment 1 prefers an envelope of the Ola format, without extra imprint, made of recycling paper with a type-written streamer, and full sender name on it; a lottery is not appreciated. Segment 2 prefers an envelope of the A5 format, no extra imprint, made of paper without chlorine, with a hand written streamer, and full sender name on it; a lottery is not appreciated either. Segment 3 prefers an envelope of the A5 format, with a hologram as extra imprint, paper without chlorine, a type-written streamer, and an abbreviated sender name. For this segment a lottery is appreciated (although it is only of minor importance).

In order to assess the predictive accuracy of each of the segment-level models we determined the first choice hits. That is, the estimated model is used to predict the value of $y$ for the three extra types of envelopes. If the envelope with the highest $y$-value is also the selected envelope, it is called a first choice hit. The percentage first choice hits for the three segment were 56%, 49% and 66%. It is simple to test whether this percentage differs significantly from a random choice, which would give a first choice hit of 33%. Hence we test $H_0 : p = 1/3$ versus $H_1 : p > 1/3$. We use the following test-statistics: $z = (\hat{p} - p) / \sqrt{p(1-p)/N}$, where $\hat{p}$ is the percentage of first choice hits.
and $N$ the sample size. This statistic is approximately normal distributed. The $p$-values are all smaller than 0.0025, and hence we conclude that the model performs well.

Segmentation results of this kind are useful only if the segment membership can be related to available characteristics of the individuals. The three segments can briefly be described as follows. Segment 1 can be characterized by respondents in the age group of 35-50 years, having children and residing outside the main city. Segment 2 can be characterized by respondents of over 65. Segment 3 can be characterized by respondents younger than 35 years, without children and residing in the city. Whether the foundation should actually send three distinct envelopes to the three segments instead of one, depends on whether the gains of such a strategy outweigh the extra costs.

For situations in which it is not possible to link segmentation results to individual characteristics, we may only compute part-worths (last column of table 6.3), and relative importances for the whole group (last column in table 6.2). The part-worths can be found by either weighing the segment part-worths (using the segment size) or by estimating the model based on all the observed data. Given these part-worths we can determine the relative importances. The optimal envelope for the whole group would be an envelope with the A5 format, without extra imprint, made of paper without chlorine, and a type-written streamer; a lottery is not necessary. The choice of the sender name is of minor importance since the mean relative importance is close to zero.

6.3.2 Conjoint field-experiment for a letter design

Again, the characteristics and levels to be studied were determined in consultation with the charity foundation; they are presented in table 6.4. The costs and the dummy variables mentioned in the table will be explained in due course. The payment device (pre-printed giro check inviting payment) is either attached at the bottom of the letter or is enclosed in the envelope. A brochure, if enclosed, gives some background information on the foundation. The letter may contain an illustration at the top left, the top right, or not at all. Amplifiers are used to stress some information given in the letter by using e.g. bold printing. There are either many, few, or no amplifiers present in the text of the letter. The Post Scriptum might contain a summary of the letter or some new information. The letter bears the signature of either the director of the foundation or a professor in health care research (in the Netherlands a professor’s title carries a lot of esteem). The address, shown through the window envelope, could either be printed on the letter or on the payment device.
Table 6.4: Characteristics included in the letter experiment

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Payment device</td>
<td>Attached, Not attached</td>
</tr>
<tr>
<td>cost</td>
<td>20 c, none</td>
</tr>
<tr>
<td>dummy: $d_{PD}$</td>
<td>1, -1</td>
</tr>
<tr>
<td>Brochure</td>
<td>Present, Absent</td>
</tr>
<tr>
<td>cost</td>
<td>40 c, none</td>
</tr>
<tr>
<td>dummy: $d_{B}$</td>
<td>1, -1</td>
</tr>
<tr>
<td>Illustration</td>
<td>Top left, Top right, None</td>
</tr>
<tr>
<td>cost</td>
<td>15 c, 15 c, no</td>
</tr>
<tr>
<td>dummy: $d_{I1}$</td>
<td>1, 0, -1</td>
</tr>
<tr>
<td>dummy: $d_{I2}$</td>
<td>0, 1, -1</td>
</tr>
<tr>
<td>Amplifier</td>
<td>Many, Few, None</td>
</tr>
<tr>
<td>cost</td>
<td>none, none, none</td>
</tr>
<tr>
<td>dummy: $d_{AM1}$</td>
<td>1, 0, -1</td>
</tr>
<tr>
<td>dummy: $d_{AM2}$</td>
<td>0, 1, -1</td>
</tr>
<tr>
<td>Post scriptum</td>
<td>Summary, New information</td>
</tr>
<tr>
<td>cost</td>
<td>no, no</td>
</tr>
<tr>
<td>dummy: $d_{PS}$</td>
<td>1, -1</td>
</tr>
<tr>
<td>Signature</td>
<td>Professor, Director</td>
</tr>
<tr>
<td>cost</td>
<td>none, none</td>
</tr>
<tr>
<td>dummy: $d_{S}$</td>
<td>1, -1</td>
</tr>
<tr>
<td>Address</td>
<td>Letter, Payment device</td>
</tr>
<tr>
<td>cost</td>
<td>none, none</td>
</tr>
<tr>
<td>dummy: $d_{AD}$</td>
<td>1, -1</td>
</tr>
</tbody>
</table>
Table 6.5: Response figures for the characteristics in the field-experiment

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Level</th>
<th>Response</th>
<th>Donations¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>Payment device</td>
<td>Attached</td>
<td>52.94%</td>
<td>9.57</td>
</tr>
<tr>
<td></td>
<td>Not attached</td>
<td>55.83%</td>
<td>9.69</td>
</tr>
<tr>
<td>Brochure</td>
<td>Present</td>
<td>54.43%</td>
<td>9.63</td>
</tr>
<tr>
<td></td>
<td>Absent</td>
<td>54.35%</td>
<td>9.63</td>
</tr>
<tr>
<td>Illustration</td>
<td>Top left</td>
<td>55.91%</td>
<td>9.77</td>
</tr>
<tr>
<td></td>
<td>Top right</td>
<td>49.51%</td>
<td>9.02</td>
</tr>
<tr>
<td></td>
<td>None</td>
<td>56.25%</td>
<td>9.97</td>
</tr>
<tr>
<td>Amplifier</td>
<td>Many</td>
<td>55.54%</td>
<td>9.73</td>
</tr>
<tr>
<td></td>
<td>Few</td>
<td>53.52%</td>
<td>9.51</td>
</tr>
<tr>
<td></td>
<td>None</td>
<td>55.46%</td>
<td>9.86</td>
</tr>
<tr>
<td>Post scriptum</td>
<td>Summary</td>
<td>56.11%</td>
<td>9.79</td>
</tr>
<tr>
<td></td>
<td>New information</td>
<td>52.17%</td>
<td>9.42</td>
</tr>
<tr>
<td>Signature</td>
<td>Professor</td>
<td>55.61%</td>
<td>9.84</td>
</tr>
<tr>
<td></td>
<td>Director</td>
<td>53.18%</td>
<td>9.42</td>
</tr>
<tr>
<td>Address</td>
<td>Letter</td>
<td>54.50%</td>
<td>9.63</td>
</tr>
<tr>
<td></td>
<td>Payment device</td>
<td>54.28%</td>
<td>9.63</td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td>54.39%</td>
<td>9.63</td>
</tr>
</tbody>
</table>

¹ Average donation (in NLG) of all the individuals

The use of a fractional factorial design resulted in 16 different types of letters. Note that 288 (3² × 2⁵) possible mailings can be constructed on the basis of the specified characteristics and levels. Every mailing was sent to 3 000 individuals who had donated at least once during the last three years. The total number of individuals in the test-mailing equals 48 000. Due to tape-transforming errors of the data the actual number of individuals that could be used was 47 635.

The overall response was 54.4 percent; the average donation amounted to NLG 9.63 (NLG 17.77 is the average of the individuals that responded). Table 6.5 presents the response figures for each of the levels of the characteristics. It shows, for example, that a signature of a professor generates 2.4% more response than the signature of the director. Moreover, it increases the average
amount of donation by NLG 0.42. We tested whether the difference between two levels of a characteristic differs significantly from zero (for amplifiers and illustration we only tested the difference between the highest two response rates). Only two characteristics, post scriptum and signature, turn out to be significantly different from zero (1% significance) with respect to response rate and donations. For the payment device there is only a significant difference for the response rate. As noted above, in this way we analyze a marginal effect, unconditional on level of the other characteristics. To obtain the conditional effect of a certain characteristic, we analyze all the characteristics simultaneously.

To do this we use a probit and tobit model. Table 6.6 gives the estimated coefficients for these models. We see that only the address is highly nonsignificant in the probit and tobit model. The brochure is significant (at the 5% level) in the probit model but not in the tobit model. The signs of the coefficients are similar between the probit and tobit model. Hence, the same level of each of the characteristics should be chosen to construct the optimal design. It is important to realize that the models are nonlinear functions. Hence, to determine the effect of a particular characteristic we need to know the values of the other characteristics, i.e. conditional on those characteristics.

Table 6.6: Coefficients\(^1\) for probit and tobit model

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Probit</th>
<th>Tobit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.103 (14.52)</td>
<td>2.747 (21.29)</td>
</tr>
<tr>
<td>Payment device</td>
<td>-0.032 (5.41)</td>
<td>-0.260 (2.56)</td>
</tr>
<tr>
<td>Brochure</td>
<td>-0.012 (2.04)</td>
<td>-0.154 (1.52)</td>
</tr>
<tr>
<td>Illustration (d_{i1})</td>
<td>0.057 (7.24)</td>
<td>0.634 (4.66)</td>
</tr>
<tr>
<td>(d_{i2})</td>
<td>-0.140 (15.03)</td>
<td>-1.831 (11.31)</td>
</tr>
<tr>
<td>Amplifiers (d_{AM1})</td>
<td>0.023 (2.43)</td>
<td>0.227 (1.41)</td>
</tr>
<tr>
<td>(d_{AM2})</td>
<td>-0.045 (5.28)</td>
<td>-0.606 (4.18)</td>
</tr>
<tr>
<td>Post scriptum</td>
<td>0.071 (11.39)</td>
<td>0.856 (7.98)</td>
</tr>
<tr>
<td>Signature</td>
<td>0.035 (5.95)</td>
<td>0.498 (4.88)</td>
</tr>
<tr>
<td>Address</td>
<td>-0.017 (0.29)</td>
<td>-0.031 (0.31)</td>
</tr>
<tr>
<td>(\sigma)</td>
<td>19.927 (204.92)</td>
<td></td>
</tr>
</tbody>
</table>

\(^1\) absolute \(t\)-values between parentheses
We use these estimates to construct an optimal mailing design. To do so we should choose the levels such that the value of the dummy times the coefficient is as large as possible. For a characteristic with three levels this means that we have to compute the three combinations, given in Table 6.4, of the two dummies. The optimal mailing design should have a payment device that is not attached, no brochure, no illustrations, many amplifiers, a post scriptum with the summary of the letter, should be signed by a professor, and have the address printed on the payment device. Since the latter is insignificant it is not really important at what level it is set. If there had been cost involved for this characteristic, it would be best to set it at the least expensive level. Whether this optimal mailing design should be used instead of the traditional mailing depends on the larger expected response rate and the cost of the optimal mailing. Since all the characteristics are set at their most inexpensive level, it is obvious that this optimal mailing is preferred. Using this optimal mailing design the expected response rate increases to 59.7%, which is a relative increase of 9.8%. The expected average amount donated, using the optimal design, is NLG 10.11, a relative increase of the average donations of 5.0%. Because direct mail may involve millions of individuals, these increases in response could turn out to be a very large figure for the total database.

6.3.3 Conjoint field-experiment for an ‘opt out’ reply card

The reply card

In order to stimulate interaction and thus build a direct relationship, it is important to make responding as easy as possible for the potential customer. The design of the reply card and the way in which the card has to be filled out influence its attractiveness and may influence the probability of the reply card being noticed and returned. The following functions of the reply card can be identified:

1. informing the organization of accepting the offer (e.g. buying a product, attending a seminar);
2. request card for (more) information about certain products, for ordering a comprehensive brochure, a catalogue or asking for a consultant to make a visit;
3. informing the organization of a change of address;
4. request to be deleted from the mailing list.

The two functions mentioned first are the commonest (Vögele 1992). In this section we focus on the last function.
Charity foundations rely heavily upon direct marketing for the acquisition of their financial donations. Although it is a fruitful way to collect money, direct marketing may be considered to be an impolite way to ask for a financial contribution. On the other hand, individuals who are aware of the charity function of the foundation may feel that they are impolite or selfish for simply refusing future charity mailings, unlike in the case of mailings from commercial organizations. Many charity foundations recognize these facts and will be highly interested in profitable solutions. One such solution may be to add the ‘opt out’ function to the reply card, giving individuals the opportunity to be deleted from the mailing list. In addition the foundation can decide to include another specific option to the reply card: the option to be deleted from the mailing list after making a once-only donation. This can be seen as ‘redemption money’. This means that the foundation will gain from individuals who otherwise would just be deleted from the mailing list.

We define the deletion of non-targets by using a reply card as self selection, as opposed to selection by the organization. We wish to emphasize that the advantage of self selection only holds if non-targets are removed from the mailing list who would not have been identified as non-targets in a target selection procedure, e.g. on the basis of their characteristics. For our empirical application self selection is especially useful as the individuals are already selected from a large population (on the basis of several characteristics), and it will be difficult to make a new (better) selection on the basis of other characteristics.

Apart from the direct advantage for the foundation, there is a possible positive image effect for the whole industry. The fact is that if the reply card creates a positive image it will not only be true for the specific organization but also for the whole direct marketing industry. The positive image is caused by the fact that the decision to receive mail in future is also the choice of the individual. A reduction of waste, as a result of improved selection, may also influence the image positively. Furthermore, it is a more decent way to get the attention of the individual; and for privacy reasons it is a polite gesture.

Two disadvantages of enclosing a reply card may be distinguished. First, there is the probability that some individuals are removed from the list although they would have responded to future mailings. Secondly, enclosing a reply card generates higher costs: the production of the card, postal cost of the return of the reply card, and the handling of reply card responses. The latter concerns the handling after the reply card is returned and the removal of these addresses from a new mailing list. Though the handling may look intensive, even without
Table 6.7: Characteristics included in the reply card experiment

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Heading</strong></td>
<td>1 Reply card for data-adjustment</td>
</tr>
<tr>
<td></td>
<td>2 I want your foundation to do the following for me:</td>
</tr>
<tr>
<td></td>
<td>3 Database-card</td>
</tr>
<tr>
<td><strong>Text lines</strong></td>
<td>1 0 I want to be deleted from your mailing list immediately</td>
</tr>
<tr>
<td></td>
<td>0 I donate once only, then I want to be deleted from the mailing list</td>
</tr>
<tr>
<td></td>
<td>2 0 I want to be deleted from your mailing list immediately</td>
</tr>
<tr>
<td></td>
<td>0 I donate once only, then I want to be deleted from the mailing list</td>
</tr>
<tr>
<td></td>
<td>0 I only want to receive a mailing from your foundation once a year</td>
</tr>
<tr>
<td><strong>Picture</strong></td>
<td>1 Present</td>
</tr>
<tr>
<td></td>
<td>2 Absent</td>
</tr>
<tr>
<td><strong>Personalization</strong></td>
<td>1 Present</td>
</tr>
<tr>
<td></td>
<td>2 Absent</td>
</tr>
<tr>
<td><strong>Font type</strong></td>
<td>1 Large</td>
</tr>
<tr>
<td></td>
<td>2 Small</td>
</tr>
</tbody>
</table>

An empirical application of the reply card was executed for a charity foundation in The Netherlands. This foundation supports activities with respect to the well-being of animals.

We used the conjoint field-experimental approach for this application. The characteristics, as well as their respective levels, were determined in consultation with the charity foundation, and are presented in table 6.7. The text lines give options for the removal from the mailing list. The picture that is shown on
the reply card is also shown in the letter. The address of the individual can be preprinted, in which case the individual only has to fill out one text line. If the reply card is personalized in this way, the back of the card can be used for informing the foundation of a change of address. The use of a fractional factorial design resulted in eight types of reply cards that we used in the experiment.

The population of interest from which the sample was drawn is selected on individual characteristics (possession of pets, interest in the environment, and interest in charity foundations) from a commercial data base with individual data on lifestyle characteristics. For each type of reply card about 800 new addresses (individuals that had never received a mailing from this foundation before) and about 800 old addresses (individuals that had received a mailing before, without reply card, but did not respond) were drawn. In total, we have 6 844 new addresses and 6 341 old addresses. There is also a group that did not receive the reply card (860 new addresses and 806 old addresses). This group is included to determine whether or not the inclusion of a reply card will generate more donations. Hence, a total of 14 851 individuals received a mailing.

**Results**

For the individuals that received a reply card there were several possible responses:
- donation without returning the reply card;
- returning the reply card without donating money;
- returning the reply card *and* donating money:
  - once-only donation;
  - request to receive one mailing per year.

The overall response with a donation, for the individuals who received a reply card, was 5%; the average donation of the individuals that responded amounts to NLG 18.25. In table 6.8 the figures for each type of the responses are presented.

Before we turn to the discussion of the optimal reply card we will formally test whether the responses differ significantly between various groups. We use a standard test for comparing population proportions (e.g. Lindgren 1993, p. 360) to examine whether the response rates differ. To examine whether there are significant differences in the donations we employ a standard test for comparing the means of two normal populations (e.g. Lindgren 1993, p. 236).

We start by testing whether the new addressees return the reply card more often than old addressees. The response rates are 6.34% (i.e. 2.12% + 4.22%)
Table 6.8: Percentages of the several types of (non)response*

<table>
<thead>
<tr>
<th></th>
<th>Mailing with reply card</th>
<th></th>
<th>Mailing without reply card</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>non-response</td>
<td>money donated</td>
<td>money donated</td>
<td>non-response</td>
</tr>
<tr>
<td></td>
<td>card not returned</td>
<td>card returned</td>
<td>returned</td>
<td></td>
</tr>
<tr>
<td>old addresses</td>
<td>92.56%</td>
<td>2.93% (17.2)</td>
<td>1.72% (20.1)</td>
<td>2.79%</td>
</tr>
<tr>
<td></td>
<td>A: 1.01% (18.1)</td>
<td>B: 0.71% (22.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>new addresses</td>
<td>90.49%</td>
<td>3.17% (18.2)</td>
<td>2.12% (19.9)</td>
<td>4.22%</td>
</tr>
<tr>
<td></td>
<td>A: 1.24% (18.3)</td>
<td>B: 0.88% (22.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>average</td>
<td>91.48%</td>
<td>3.06% (17.7)</td>
<td>1.93% (20.0)</td>
<td>3.53%</td>
</tr>
<tr>
<td></td>
<td>A: 1.13% (18.2)</td>
<td>B: 0.80% (22.6)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Average amount of donation of the individuals between parentheses (in NLG)

1“Money donated card returned” is divided in a once-only donation (A) and a request to receive one mailing per year (B)
and 4.51% (i.e. 1.72% + 2.79%), respectively. The p-value is less than 0.001, hence the new addressees return the reply card significantly more often than the old addressees. This difference is mainly caused by the response without a donation (4.22% versus 2.79%, with p-value < 0.001). A possible explanation could be that the old addressees are less involved (none of them donated before) with the charity foundation than the new addressees. No significant differences were obtained between new and old addressees with respect to (1) the percentage of returned reply cards (2.12% versus 1.72%), and (2) the response rate without returning the reply card. That latter is 3.17% versus 2.93% for the individual that received a reply card, and 4.53% versus 4.71% for the individuals that did not receive a reply card.

The average donations of the individuals that received a reply card (NLG 17.7) and of those that did not receive it (NLG 15.9), do not differ significantly (p-value of 0.11). However, the p-value indicates that there is a slight favor of the inclusion of the reply card. This may be the result of the positive image aspect. The donations within the group that returned the reply card, i.e. once-only donations (NLG 18.2) and request to receive one mailing per year (NLG 22.6), are larger for the latter group but not significantly different (p-value is 0.12).

**Profit implications**

The purpose of the conjoint field-experiment is to develop an optimal reply card. Optimal is not unambiguously defined here, since it can be based on several types of response, e.g. (1) expected response rate of returning the reply card, or (2) if the reply card is returned, the expected donations. Note that even a low expected response rate may be seen as optimal, since the risk of the removal of targets is minimal while the effect of a positive image is still there.

We employ two models to analyze the response. The first model, the probit model (see section 2.4), is used to examine the characteristics that affect the use of the reply card. The second model, the tobit model (see section 3.6), is used to examine the characteristics that affect the donation given the use of the reply card.

The dependent variable of the probit model is equal to one if an individual returned the reply card and is zero otherwise. To estimate the model we used 13185 observations, i.e. the individuals that received a reply card. Table 6.9 gives the estimation results. The characteristics that have a significant effect (at a 10% significance level) on returning the reply card are heading, personalization and font-type. We use these results to construct the optimal reply card. To do
so we choose the levels such that the value of the dummy variable times the coefficient is as large as possible. Thus, the optimal reply card with respect to the returning this card has heading 3 ("Database-card"), is personalized, has a small font-type, and has the levels of the other characteristics set on their most inexpensive level. In addition, we use the probit estimates to determine the expected response rate of the optimal reply card, which turns out to be 7.00%. Note that the response rate was 5.46% (i.e. 1.93% + 3.53%).

We used 721 observations, i.e. individuals who returned the reply card, to estimate the tobit model. The estimation results are depicted in table 6.9. Only two characteristics turn out to be significant, viz. heading and font type. Again, we can determine the optimal reply card but now with respect to the donations. In this case the card should have heading 2 ("I want your foundation to do the following for me:"")) and a large font-type.

We specified two ‘optimal’ reply cards. The objective of the first optimal reply card is to maximize the probability that an individual returns the reply card. The objective of the second optimal reply card is to maximize the expected donation for the individuals who return the reply card. Note that these may be conflicting objectives. Thus, the choice of the ultimate optimal reply card is not simply the one which generates the highest expected revenues, i.e. product of probability of returning the reply card times the expected donation conditional on returning (cf. the two-part model discussed in chapter 3). Therefore, the foundation should choose an optimal reply card on the basis of the objective which has its highest priority. Since the difference between heading 2 and 3 (with respect to donation) is quite small, the choice of the ultimate reply card actually comes down to choice of the font type here.

Since there are two possible effects of selection (the positive effect of the removal of non-targets and the negative effect of the removal of targets, respectively), it is interesting to gain insight into the profit implications. We compare profit implications on the basis of two typical strategies that can be followed by the foundation. We consider the simple case of two periods.

**Strategy 1:** The foundation decides to mail everyone on the mailing list, $N$, without a reply card, in period $t$ and $t+1$ (for instance, period $t$ is immediately after the test mailing and period $t+1$ is next year). We assume that the response probability $P_1$ is the same for both years, and the cost of a mailing is denoted by $c_1$. So, the expected profit over the two periods, $\Pi_1$, is given by

$$E(\Pi_1) = 2wNP_1 - 2Nc_1,$$

where $w$ is the (constant) revenue of a positive reply.
Optimizing the characteristics of the mailing

Table 6.9: Coefficients for probit and tobit model

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Level</th>
<th></th>
<th>Probit</th>
<th>Tobit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>1</td>
<td></td>
<td>-1.643 (31.47)</td>
<td>-24.559 (5.05)</td>
</tr>
<tr>
<td>Heading</td>
<td>2</td>
<td></td>
<td>0.032 (0.72)</td>
<td>6.566 (1.80)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td>0.106 (2.09)</td>
<td>0.220 (0.05)</td>
</tr>
<tr>
<td>Picture</td>
<td>Present</td>
<td></td>
<td>0.002 (0.05)</td>
<td>-0.706 (0.25)</td>
</tr>
<tr>
<td>Personalization</td>
<td>Present</td>
<td></td>
<td>0.064 (1.79)</td>
<td>0.035 (0.01)</td>
</tr>
<tr>
<td>Font type</td>
<td>Large</td>
<td></td>
<td>-0.062 (1.73)</td>
<td>16.242 (5.57)</td>
</tr>
<tr>
<td>Text lines</td>
<td>Two lines</td>
<td></td>
<td>-0.010 (0.29)</td>
<td>2.199 (0.77)</td>
</tr>
<tr>
<td>$\sigma$</td>
<td></td>
<td></td>
<td>31.808 (20.51)</td>
<td></td>
</tr>
</tbody>
</table>

1 absolute $t$-values between parentheses
2 Constant: heading 1, not personalized, and small font type
3 Zero-one dummy variables

Strategy 2: The foundation sends a mailing with reply card to every individual on the mailing list in period $t$, and only to the individuals that did not return the reply card in period $t + 1$. Here a distinction is made between the response probability with a reply card, $P_{rc}$ (the individuals that made a once-only donation), and the response probability without a reply card, $P_2$. We assume that $P_2$ also includes the individuals that have chosen the option to receive a mailing only once a year. Note that we do not assume $P_{rc}$ constant over time. The percentage of individuals that returned the reply card in order to be deleted from the mailing list is denoted by $f$, and the percentage that returned the reply card because they are non-targets, i.e., they will not make a donation to future mailings, by $q$. In period $t$ the expected profit is

$$E(\Pi_{2,t}) = w(N(P_2 + P_{rc}) - Nc_2),$$

where $c_2$ denotes the cost of a mailing. For period $t + 1$ there are $N(1 - f)$ individuals that receive a mailing. Like in strategy 1 we assume that $P_2 \equiv P_{2,t} = P_{2,t+1}$. Now we partition the individuals in such a way that the first part consists of the individuals that will not donate in the future and made this clear by returning the reply card; their response probability is $P_{2,t+1}^1$, and the second part are the other individuals; their response probability is $P_{2,t+1}^2$. Hence

$$P_2 = qP_{2,t+1}^1 + (1 - q)P_{2,t+1}^2$$
(1 - q) P_{2,t+1}^2;

the latter step holds since $P_{2,t+1}^1 = 0$ by definition, thus

$$p_{2,t+1}^2 = \frac{P_2}{1 - q}.$$  

Note that $P_{2,t+1}^2$ is at its maximum as $q = f$; then $N(1 - f)$ individuals that are selected generate the same absolute response as $N$ without selection. The expected profit in period $t + 1$ is

$$E(\Pi_{2,t+1}) = wN(1 - f)\left(\frac{P_2}{1 - q} + P_{r+1}^{nc}\right) - N(1 - f)c_2,$$

hence, the expected profit for strategy 2 is

$$E(\Pi_2) = E(\Pi_{2,t}) + E(\Pi_{2,t+1})$$

$$= wN(P_2 + P_{r}^{nc}) + wN(1 - f)\left(\frac{P_2}{1 - q} + P_{r+1}^{nc}\right) - N(2 - f)c_2. \quad (6.1)$$

The fraction of non-targets in the group that wants to be deleted from the mailing list is $q/f$. This value lies between zero and one; a value of one means that all the individuals that wanted to be deleted from the mailing list are non-targets, and a value of zero means that they are all targets.

The parameters $P_2$, $P_{r}^{nc}$, and $f$ of (6.1) can be consistently estimated from the data, i.e. the response figures in table 6.8. The other two parameters, $q$ and $P_{r+1}^{nc}$ cannot be estimated from the data. Note that it is unlikely that $P_{r+1}^{nc}$ is equal to $P_{r}^{nc}$, since most individuals that want to use the reply card will use it in period $t$. On the basis of additional information, for instance after the mailing in period $t + 1$, the foundation can estimate $P_{r+1}^{nc}$. Although the foundation does not have this information, it can use (6.1) to gain insight into the profit implications in the following way.

For a specific choice of the reply card, the expected response rates, except $P_{r+1}^{nc}$, can be determined. Then for two different values of $P_{r+1}^{nc}$ (without any prior information, 0 and $P_{r}^{nc}$) the relation between $q/f$ and $E(\Pi_2) - E(\Pi_1)$ can be drawn in a figure (figures 6.1 and 6.2). These figures show that the profit of strategy 2 increases with $q/f$. They also show the sign of the difference. If even in the worst case scenario ($P_{r+1}^{nc} = 0$) the difference is positive, the use of the reply card will improve the expected profits. If, in this scenario,
Optimizing the characteristics of the mailing

\[ q = f \]

Difference in expected profit (per household)

\[ P_{rc} = P_{rc}^t \]

\[ P_{t+1}^n = P_{t+1}^n = 0 \]

Figure 6.1: Difference in expected profit between a strategy with reply card (in \( t \) and \( t+1 \)) and without reply card.

Response rate as obtained from the data, \( w \) is constant, \( q/f \) is the fraction of non-targets in the group that want to be deleted from the mailing list, and \( P_{rc}^n \) is the response probability with reply card in period \( t+1 \).

the difference is positive for values of \( q/f \) and negative for other values, the foundation has to make a reasonable assumption about the value of \( q/f \) to know whether the profit-difference is positive or negative. If the difference is negative, the foundation also has to make a reasonable assumption about the value of \( P_{rc}^n \). If even in the best case scenario, i.e. \( P_{t+1}^n = P_{rc}^n \), the difference is negative for all \( q \), the inclusion of a reply card will decrease the expected profits. The importance of this analysis is that the profit implications become clear for a specific reply card.

In figure 6.1 the response rates were taken from table 6.8; in figure 6.2 the response rates are determined on the basis of the optimal reply card and the probit estimates. For both figures the revenues to a positive reply were assumed to be equal for all individuals, i.e. the average amount of donation. For the optimal reply card we divide the expected response with card returned (the three different types of response) in the same proportion as obtained from the data.

From figures 6.1 and 6.2 it is clear that the profit (of strategy 2) increases when the optimal reply card is used. In figure 6.1 the profit difference, for the worst case scenario, is negative for all values of \( q/f \). This changes when
the optimal reply card is used; then, for the worst case scenario, only for $q/f < 0.31$, the profit difference is negative. Thus, if at least 31% of the individuals who want to be deleted from the mailing list are non-targets, the inclusion of the optimal reply card leads to higher expected profits.

The above analysis is based on the optimal reply card with respect to the response rate. The same analysis can be performed using the optimal reply card with respect to amount of donation. Using the tobit analysis results we can determine the expected donation for this reply card. The response probabilities can be determined as well and similar figures can be drawn. We will not do this, for two reasons. First, the expected donation for the individuals that will respond is very sensitive to small changes in the data and to the assumption that the expected response with card returned may be divided in the same proportions for the three forms of response, as obtained from the data. Secondly, it is likely that there is a correlation between the response probability and amount of donation; this causes problems for the expected amount of donation because of the first reason mentioned.
6.4 Interaction between target and mailing characteristics

So far, we considered the optimization of the mailing design. After having chosen the design, the organization selects the targets. This section proposes a new strategy that simultaneously takes both aspects into account by using different mailings offering the same product to different targets. Using the data from the letter experiment of section 6.3.2, we show that this strategy increases the net returns of the organization.

6.4.1 Modeling the response process

We model the response process in the following way. We label individuals by the index \( i \); \( i = 1, \ldots, 47365 \). The observed reaction of individual \( i \) to the mailing is represented by \( y_i \) and can assume the values 0 (no response) or 1 (response); in the context of the present research we only analyze whether an individual responds at all and do not further consider the amount of the donation. The reaction depends on characteristics of the individual. We model this dependence by a standard probit specification. So we assume that there is a latent variable, denoted by \( \eta_i \), and interpreted as the inclination to respond, which depends on characteristics according to a linear regression model:

\[
\eta_i = \alpha_0 + \xi_i \beta_0 + \varepsilon_i, \tag{6.2}
\]

where the \((k \times 1)\) vector \( \xi_i \) contains individual’s \( i \) characteristics, \( \beta_0 \) is a \((k \times 1)\) vector of parameters, \( \alpha_0 \) is the intercept, and \( \varepsilon_i \sim \text{i.i.d.} \ N(0, 1) \) is a disturbance term. We do not observe \( \eta_i \) but observe only whether it is negative (then we set \( y_i = 0 \)) or positive (\( y_i = 1 \)). The characteristics of the individuals consists of four variables. The variables AMD90 and AMD91 measure the amount of money donated in 1990 and 1991, respectively. The variable NQSDE measures the number of quarters since the date of entry, and NQSLC measures the number of quarters since the last donation. Notice that these variables are actually RFM-variables.

If different types of mailings are tested we have to adapt the regression model. As an obvious starting point we let the intercept vary over the 16 different mailings used. So we need 16 dummies to account for the different mailings; the regression lines for the different mailing types are still assumed parallel. The intercepts are collected in the \((1 \times 16)\) vector \( \alpha \equiv (\alpha_{0,1}, \ldots, \alpha_{0,16}) \).

There are two reasons why this specification may be unsatisfactory. It may both be too strict and, in another sense, overparametrized. Starting with the last issue, as we discussed in the section 6.3, the 16 mailings used were
selected from all 288 possibilities. A reasonable hypothesis is that the \( \alpha \)'s have an additive structure. That is, each value of each characteristic has a particular contribution to the intercept, and for a particular mailing, i.e. a particular combination of values of characteristics, the various parts simply have to be added. Given the structure of the characteristics we would need \( 5 \times 2 + 2 \times 3 = 16 \) dummies, but in order to avoid multicollinearity we delete one dummy for each characteristic and add an overall intercept; so we need 10 dummies. The precise definition of the dummies is given in table 6.4.

Mathematically speaking, if the additive structure holds, the 16 elements of the \( \alpha \)-vector are restricted to lie in a 10-dimensional space. Collecting the 10 dummies in the \( (1 \times 10) \) vector \( \alpha_i \equiv (\alpha_{i,1}, \ldots, \alpha_{i,10}) \) we hence entertain the possibility that \( \alpha_i = \alpha_i D \), with \( D \) a \( (10 \times 16) \) matrix describing the design of the experiment. Since the precise form of \( D \) is not of substantive interest we do not present it. Note that we already assumed this additive structure in the conjoint field-experiment in section 6.3, which is the commonest composition rule in conjoint analysis. This rule can be extended by including interaction effects between the mailing characteristics. In that case, many more test mailings are needed.

Having thus considered a restriction on the regression model, we relax the specification in a different direction. The different mailings may interact with the target characteristics, so the regression lines may not be parallel after all. We allow for this in much the same way as we did with the intercept. In the first place we can allow for \( \beta \) varying with the 16 mailings used. Since we employ four background variables, we then have 64 regression coefficients, to be denoted by \( \beta_i \). Hence \( \beta_i \) is a \( (4 \times 16) \) matrix. As a next step, analogous to what we did with \( \alpha \), we can restrict all these \( \beta \)'s in an additive way based on the design to obtain \( \beta_i = \beta_i D \) in obvious notation. In this way we end up with \( 4 \times 10 = 40 \) \( \beta \)'s to be estimated; \( \beta_i \) is a \( (4 \times 10) \) matrix.

Hence we end up with three general specifications. One is defined by (6.2), which is the most restricted model. The unrestricted model is

\[
\eta_i = \alpha_i m_i + \xi_i \beta_i m_i + \epsilon_i, \tag{6.3}
\]

where \( m_i \) is a \( (16 \times 1) \) vector, of which the \( m \)th element is one and the other elements are zero. The \( m \)th element corresponds with the mailing that \( i \) received. Imposing the additive structure to the model we obtain

\[
\eta_i = \alpha_i d_i + \xi_i \beta_i d_i + \epsilon_i, \tag{6.4}
\]
where \( d_i \) is a column of \( D \) that corresponds with the mailing that \( i \) received. Of course, combinations of the three specifications are also possible. Thus we have three possibilities for the intercept: \( \alpha_0 \), \( \alpha_r \), and \( \alpha_u \), involving 1, 10 and 16 parameters respectively and nested from left to right. We also have three possible specifications for the regression coefficients: \( \beta_0 \), \( \beta_r \), and \( \beta_u \), with 4, 40 and 64 parameters, also nested from left to right. Combining the cases gives us nine models to be estimated.

Given the assumed normality of the disturbance term, all nine models can simply be estimated by probit analysis. Since the models are nested we can, in principle, test the various restrictions with an ML-ratio test. It is intuitively clear what the nature of the restrictions are. However, the exact form the hypotheses differ considerably between these tests. It is therefore instructive to define the hypotheses formally. We show this for tests on \( \alpha \). To test the unrestricted model (6.3) against the basis model (6.2) the hypothesis is that all the letters have the same effect on the response probability. That is,

\[
H_0 : \alpha_{u,1} = \alpha_{u,2} = \ldots = \alpha_{u,16},
\]

which then equals \( \alpha_0 \). The \( H_0 \) for testing the restricted model (6.4) against the basis model (6.2) is

\[
H_0 : \alpha_{r,1} = \alpha_{r,2} = \ldots = \alpha_{r,10} = 0.
\]

In words, the characteristics of the mailing do not affect the response probability. To test the unrestricted model (6.3) against the restricted model (6.4) we use

\[
H_0 : \alpha_u = D\alpha_r,
\]

which is the hypothesis that the additive structure holds. To select one of these nine models we employ the CAIC (Consistent Akaike’s Information Criterium, see Bozdogan (1987)), rather than the ML test since the CAIC explicitly makes a trade-off between the inaccuracy of a model and its complexity. Moreover, it takes the sample size into account. The CAIC is given by

\[
\text{CAIC} = -2l + r \times \ln(N + 1),
\]

where \( l \) denotes the log-likelihood value, \( r \) the number of parameters and \( N \) the sample size. The model with the minimum CAIC-value is preferred.
6.4.2 Empirical results

Our first interest is in model selection. In the upper panel of table 6.10 we present the log-likelihood values and the bottom panel gives the CAIC-values. We use the symbols for the variables in an obvious way to describe the models. The bottom panel of table 6.10 suggests that the model with the additive structure on intercepts and regression coefficients should be preferred.

The estimates for this model are shown in table 6.11. In the first row we tabulated the overall intercept and the main effects of the individual characteristics. We see that the amounts of money donated in 1990 and 1991 have a positive and significant influence on the response of the current mailing. The longer an individual has been present in the data base (NQSDE), the more he is likely to respond to the mailing. The coefficient is highly significant; the longer the time span since the last donation (NQSLC), the greater the likelihood of a decrease in response. The first column of table 6.11 consist of the main effects of the mailing characteristics. When looking at the main effects only the mailing should have an attached payment device, no brochure enclosed, an illustration in the top right-hand corner, many amplifiers, a post scriptum that contains a summary only, a signature from a professor, and the address printed on the payment device. Note the differences with the results of section 6.3. Three coefficients have a different sign, viz. $d_{13,1}$, $d_{12}$, which gives two differences for the optimal design. Moreover, address becomes significant and signature is no longer significant. The other cells of the table contain the interaction effects between mailing and individual characteristics. The interpretation of the first interaction effect, for example (between AMD90

---

Table 6.10: Log-likelihoods (above) and CAIC (below)

<table>
<thead>
<tr>
<th></th>
<th>$\beta_0$</th>
<th>$\beta_1$</th>
<th>$\beta_u$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_0$</td>
<td>-32 101</td>
<td>-31 788</td>
<td>-31 710</td>
</tr>
<tr>
<td>$\alpha_t$</td>
<td>-31 924</td>
<td>-31 670</td>
<td>-31 589</td>
</tr>
<tr>
<td>$\alpha_u$</td>
<td>-31 893</td>
<td>-31 638</td>
<td>-31 571</td>
</tr>
<tr>
<td>$\alpha_0$</td>
<td>64 261</td>
<td>64 055</td>
<td>64 185</td>
</tr>
<tr>
<td>$\alpha_t$</td>
<td>64 013</td>
<td>63 929</td>
<td>64 050</td>
</tr>
<tr>
<td>$\alpha_u$</td>
<td>64 021</td>
<td>63 935</td>
<td>64 084</td>
</tr>
</tbody>
</table>
Table 6.11: Estimates of the coefficients (absolute $t$-values between parentheses)

<table>
<thead>
<tr>
<th></th>
<th>intercept</th>
<th>AMD90</th>
<th>AMD91</th>
<th>NQSDE</th>
<th>NQSLC</th>
</tr>
</thead>
<tbody>
<tr>
<td>main</td>
<td>$-0.244$</td>
<td>$0.012$</td>
<td>$0.003$</td>
<td>$0.005$</td>
<td>$-0.046$</td>
</tr>
<tr>
<td></td>
<td>$(8.76)$</td>
<td>$(11.43)$</td>
<td>$(7.93)$</td>
<td>$(26.48)$</td>
<td>$(8.98)$</td>
</tr>
<tr>
<td>$d_{PD}$</td>
<td>$0.064$</td>
<td>$-0.011$</td>
<td>$0.002$</td>
<td>$0.000$</td>
<td>$0.022$</td>
</tr>
<tr>
<td></td>
<td>$(2.57)$</td>
<td>$(11.54)$</td>
<td>$(5.19)$</td>
<td>$(2.63)$</td>
<td>$(4.72)$</td>
</tr>
<tr>
<td>$d_{b}$</td>
<td>$-0.072$</td>
<td>$0.004$</td>
<td>$-0.000$</td>
<td>$0.000$</td>
<td>$-0.001$</td>
</tr>
<tr>
<td></td>
<td>$(2.86)$</td>
<td>$(4.02)$</td>
<td>$(1.16)$</td>
<td>$(1.78)$</td>
<td>$(0.21)$</td>
</tr>
<tr>
<td>$d_{I}$</td>
<td>$-0.437$</td>
<td>$0.017$</td>
<td>$0.001$</td>
<td>$-0.001$</td>
<td>$0.011$</td>
</tr>
<tr>
<td></td>
<td>$(11.22)$</td>
<td>$(11.24)$</td>
<td>$(2.40)$</td>
<td>$(4.61)$</td>
<td>$(1.54)$</td>
</tr>
<tr>
<td>$d_{I2}$</td>
<td>$0.237$</td>
<td>$-0.010$</td>
<td>$0.000$</td>
<td>$0.001$</td>
<td>$-0.004$</td>
</tr>
<tr>
<td></td>
<td>$(7.07)$</td>
<td>$(8.11)$</td>
<td>$(0.40)$</td>
<td>$(2.62)$</td>
<td>$(0.71)$</td>
</tr>
<tr>
<td>$d_{AM1}$</td>
<td>$0.186$</td>
<td>$-0.009$</td>
<td>$0.001$</td>
<td>$0.000$</td>
<td>$-0.011$</td>
</tr>
<tr>
<td></td>
<td>$(4.67)$</td>
<td>$(5.83)$</td>
<td>$(2.43)$</td>
<td>$(1.03)$</td>
<td>$(1.63)$</td>
</tr>
<tr>
<td>$d_{AM2}$</td>
<td>$-0.130$</td>
<td>$0.005$</td>
<td>$-0.001$</td>
<td>$-0.000$</td>
<td>$0.003$</td>
</tr>
<tr>
<td></td>
<td>$(3.89)$</td>
<td>$(4.35)$</td>
<td>$(2.54)$</td>
<td>$(1.01)$</td>
<td>$(0.59)$</td>
</tr>
<tr>
<td>$d_{PS}$</td>
<td>$0.179$</td>
<td>$-0.004$</td>
<td>$-0.001$</td>
<td>$0.000$</td>
<td>$-0.008$</td>
</tr>
<tr>
<td></td>
<td>$(7.12)$</td>
<td>$(4.62)$</td>
<td>$(1.68)$</td>
<td>$(0.49)$</td>
<td>$(1.75)$</td>
</tr>
<tr>
<td>$d_{S}$</td>
<td>$0.023$</td>
<td>$-0.000$</td>
<td>$-0.000$</td>
<td>$-0.001$</td>
<td>$0.007$</td>
</tr>
<tr>
<td></td>
<td>$(0.95)$</td>
<td>$(0.37)$</td>
<td>$(0.92)$</td>
<td>$(2.82)$</td>
<td>$(1.53)$</td>
</tr>
<tr>
<td>$d_{AD}$</td>
<td>$-0.065$</td>
<td>$0.004$</td>
<td>$0.000$</td>
<td>$0.001$</td>
<td>$-0.016$</td>
</tr>
<tr>
<td></td>
<td>$(2.58)$</td>
<td>$(4.71)$</td>
<td>$(0.29)$</td>
<td>$(3.93)$</td>
<td>$(3.45)$</td>
</tr>
</tbody>
</table>

AMD90: the amount of money donated in 1990;
AMD91: the amount of money donated in 1991;
NQSDE: the number of quarters since the date of entry;
NQSLC: the number of quarters since the last donation;
See table 6.4 for the definition of the dummies.
and $d_{m(d)}$, is that the more money an individual donated in 1990 the less it is influenced by the attached payment device. Since about half of the dummies are significant, there appears to be quite some interaction between mailing and individual characteristics.

### 6.4.3 Optimal target selection

Given the outcomes we now consider their implications for an optimal design of future mailings. The underlying idea is that, for each of the 47,635 individuals, we can now compute the probability of response for each of the 288 possible different mailings: from (6.2) the probability of response when individual $i$ would receive mailing $m$ is

$$P_{im} = \Phi(\eta_i > 0) = \Phi(\alpha_i d_m + \xi_i \beta_i d_m),$$

with $\Phi(\cdot)$ the standard normal distribution function, and $d_m$ the $(10 \times 1)$ vector describing the mailing $m$. Let the cost involved with a mailing of type $m$ be $c_m$ and $w$ the amount donated on average (NLG 17.77 in our sample; as stated above we disregard the individual variation in this amount). Then, by brute force calculation we can find the mailing $m$ for individual $i$ for which $wP_{im} - c_m$ is the highest. The cost of the cheapest mailings is NLG 1.65, the additional costs of the various characteristics are given in Table 6.4; only payment device, brochure, and illustration are priced.

As a first exercise, we looked for the mailings that were optimal (in this sense) for at least one individual. It turned out that 52 out of the 288 mailings survived this screening. As was to be expected, the number of individuals for whom one particular mailing was optimal varied hugely, the highest numbers being 10,585, 6,157 and 5,365, whereas thirteen mailings were optimal for fewer than 10 individuals.

Naturally this does not imply that such a fine-tuned, personalized mailing method would be optimal, since we disregard the fixed costs involved in such a complex system, which are huge. We know that for a mailing to be used in practice we must at least have a minimum number of individuals that receive this mailing, otherwise the fixed costs dominate the variable costs. In consultation with the manager of the mailing system of the organization, we put the lower bound on the number of individuals receiving a particular type of mailing at one percent of the data base. This number is rather arbitrary. A better approach would be to use the true cost equation showing the relationship between costs per mailing and the number of mailings sent. For the sake of
Optimizing the characteristics of the mailing

Table 6.12: Characteristics of the thirteen best mailings

<table>
<thead>
<tr>
<th>net returns</th>
<th>mailing characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$d_{pd}$</td>
</tr>
<tr>
<td>353 843</td>
<td>1</td>
</tr>
<tr>
<td>370 998</td>
<td>-1</td>
</tr>
<tr>
<td>401 595</td>
<td>1</td>
</tr>
<tr>
<td>403 947</td>
<td>1</td>
</tr>
<tr>
<td>409 672</td>
<td>1</td>
</tr>
<tr>
<td>417 733</td>
<td>-1</td>
</tr>
<tr>
<td>426 083</td>
<td>-1</td>
</tr>
<tr>
<td>434 794</td>
<td>-1</td>
</tr>
<tr>
<td>435 318</td>
<td>-1</td>
</tr>
<tr>
<td>441 429</td>
<td>-1</td>
</tr>
<tr>
<td>442 094</td>
<td>-1</td>
</tr>
<tr>
<td>443 197</td>
<td>-1</td>
</tr>
<tr>
<td>443 647</td>
<td>-1</td>
</tr>
</tbody>
</table>

simplicity, and because more than 95 percent of the optimal mailing per individual were contained in the mailings with more than one percent, we chose for this naive approach. Then we are left with thirteen different mailings. In table 6.12 these are presented, defined in terms of the values of the dummy variables as given in table 6.4. We ordered the mailings in the table according to the total net returns (in NLG) to be expected from the sample if all individuals were to receive that particular mailing. We see that there are large differences between these thirteen mailings, even though they were ranked as the top thirteen ones.

From table 6.12 we derive some interesting results. The only characteristic the thirteen top mailings agree on is the post scriptum. It should contain a summary of the letter, no new information. Nine out of thirteen mailings have a payment device that is not attached. Furthermore, the illustration should never be put in the top right-hand corner and the direct marketing organization should only consider the extreme levels for the amplifiers. Somewhat surprisingly, the mailing which had the highest number of individuals that should receive the mailing (10 585) is tabulated on the fifth line of table 6.12 (NLG 409 672),
scoring some ten percent below the mailing which generates the highest net returns if the single best mailing had been sent to all individuals (NLG 443 647). The single best mailing had separate payment device, no brochure, no illustrations, no amplifiers, a post scriptum containing a summary, was signed by a professor, and had the address written on the letter. Note that the fourth line from below (NLG 441 429), describes the design we obtained without taking individual characteristics and interaction effects into account (see section 6.3). Taking all thirteen mailings into account and sending the most profitable mailing to each individual, we obtain net returns of NLG 484 515, which is an increase of 9 percent over the single best mailing. If the worst mailing is used for all individuals, the net returns are expected to be NLG 247 879, nearly half the maximum, reinforcing the notion that it pays to design mailings carefully.

We summarize the findings by comparing the net returns for four mailing strategies in table 6.13. The single worst mailing is only displayed as a benchmark and is used to show that mailing characteristics do matter in terms of expected net returns. The second one is the mailing traditionally used by the organization. This mailing had an attached payment device, a brochure, no illustration, no amplifiers, a post scriptum containing a summary, a signature of the managing director, and the address on the letter. This strategy would, in net returns for the test sample, yield NLG 388 180. The third strategy uses the single best mailing out of the 288 possibilities. This would yield NLG 443 647, or 14 percent over the traditional strategy. When using the thirteen top mailings we would obtain net returns of NLG 484 515, or 25 percent over the traditional strategy and 9 percent over the single best strategy. Per individual the last strategy would generate NLG 0.86 more than the single best mailing. Given that the data base contains 1.2 million individuals, the expected net revenues increase is equal to NLG 1 029 528.

### 6.5 Discussion and conclusion

This chapter demonstrates the importance of a carefully designed direct mailing. We discussed several approaches to determine an optimal design in an efficient way. For the first approach, the traditional conjoint experiment, the respondent has to judge several mailings. For the second approach, the conjoint field-experiment, the respondent is confronted with one mailing; the judgement is unobserved, only the response to the mailing is observed. In an empirical application we determined an optimal envelope with a conjoint experiment.
We applied the conjoint field-experiment to determine the optimal design of a letter and an 'opt out' reply card. These applications show that the resulting mailing design generates a larger response rate than the design traditionally used by the organization.

We suggest that either one of these approaches should be selected on the basis of the specific requirements of the application of the mailing component. For example, a function of the envelope is to maximize the probability of the envelope being opened. Results of a conjoint field-experiment will not provide us with this information. Thus, to optimize the characteristics of an envelope, a traditional conjoint analysis experiment should be preferred. In contrast, the function of the letter is to stimulate the respondent to buy the product or to make a donation. The conjoint field-experiment will supply this information. The traditional conjoint analysis experiment, however, is difficult to apply in this case because the task for the respondent becomes too complicated.

Apart from the difference between the practical implementation of the two approaches, we distinguish two other differences that should be taken into account. The first difference concerns validity aspects. In the conjoint experiment the internal validity can be examined by the model fit and the predictive accuracy. The external validity, however, is very difficult to examine, as is often the case in an experimental setting. As a solution, a non-experimental follow-up study can be performed to examine the external validity, or the results of the optimal mailing can be compared with a former mailing (the other components should then be held equal). For the conjoint field-experiment the internal validity is difficult to examine, since the observed behavior is confounded with uncontrolled factors. Furthermore, we cannot use an estimation and validation sample to establish external validity, as the chosen design is generally not one of the mailings used in the test mailing. Again, a follow-up study is needed to determine the external validity of the results.

Table 6.13: Net returns per strategy (NLG)

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Net Return</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single worst mailing</td>
<td>247 879</td>
</tr>
<tr>
<td>Mailing traditional used by the organization</td>
<td>388 180</td>
</tr>
<tr>
<td>Single best mailing</td>
<td>443 647</td>
</tr>
<tr>
<td>Top thirteen mailings</td>
<td>484 515</td>
</tr>
</tbody>
</table>
The second difference between the two approaches are the dissimilarities in costs. In the traditional conjoint experiment all the costs are made before the mailing is sent to the individuals on the mailing list. The costs include finding individuals who are willing to cooperate, supervising them during the experiment, etc. However, all the individuals receive the optimal mailing, as opposed to the conjoint field-experiment in which a number of individuals receive a less optimal mailing (the test-mailing). The difference in revenues between the less optimal mailing and the optimal mailing, for the individuals in the sample, can be interpreted as costs in the conjoint field-experiment when comparing it with the conjoint experiment. Note that when there are no differences in the response rates between different types of mailings, the traditional conjoint experiment is much more expensive. On the other hand, if the optimal mailing generates a much higher response, the conjoint field-experiment becomes less profitable. This is especially harmful if a large number of individuals, relative to the total number of individuals on the mailing list, receives the test-mailing.

Since it is likely that there is interaction between target characteristics and mailing characteristics, we introduced a new strategy that simultaneously takes both characteristics into account by using different mailings offering the same product to different individuals. We showed that this increases the efficiency of a promotional direct marketing campaign and is more profitable than existing methods. This coincides with the well-known result that segmentation increases net profits (e.g. Frank et al. 1972). It should be realized, however, that these profit figures refer to the expected profits; implementation of this approach should prove whether this strategy does indeed generate higher profits.

Two limitations of our approach should be stressed. First, we computed the expected net revenues by assuming that the cost of a mailing does not depend on the total number that is used of that particular mailing. This may seem an inappropriate assumptions but it must be realized that we only use those mailings that will be sent to reasonably large number of individuals. However, it is simple, basically, to implement an advanced cost structure taking into account all the costs involved. Secondly, we use the average donation instead of the actual decision per individual. Hence, we implicitly assume that the mailing characteristics have an impact on the probability and much less on the amount of money donated. Of course, this approach could be extended by modeling the amount of donations as well (see chapter 3).
Chapter 7

Frequency of direct mailings

7.1 Introduction

One of the most prominent questions in direct marketing is how to define a strategy to maximize the lifetime value (LTV) of an individual. Pearson (1994) defines LTV as ‘the net present value of the stream of contributions to profit resulting from the revenues from customer transactions and allowing for the costs of delivering products, services and promised rewards to the customer’. Similar definitions are given by Jackson (1994) and Kestnbaum (1992), among others. Focusing on the LTV means that the success of a strategy is not defined in a short-term criterion such as the response rate, but in a long-term criterion. Only if current purchases are the sole consideration, as with an encyclopedia, is a short-term criterion appropriate. However, often even these purchases can be followed by purchases of related products, so-called cross-selling, and hence also the long-term aspects become of interest. One element of a strategy to maximize the LTV, which we examine in this chapter, is the frequency with which an individual should receive a mailing. Before we turn to the mailing frequency, we discuss the LTV concept in somewhat more detail.

Three aspects characterize the traditional way of calculating and employing the LTV (e.g. Courtheoux in Roberts and Berger 1989, p. 411, Hughes and Wang 1995, and Kestnbaum 1992). First, the LTV is used for decision making in the areas of customer reactivation and customer acquisition. Secondly, the calculation is solely based on variables of the organization’s database, e.g. on past purchase behavior. Thirdly, the LTV is determined for a group of individuals.

As Hoekstra en Huizingh (1996) argue, the traditional way of employing the LTV is too narrow within direct marketing. With regard to the use of the
LTV, it should be employed - apart from customer acquisition - for relationship building. That is, the LTV should be used to choose media for communication with customers, develop loyalty programs and assess the strength of the relationship. Fully exploiting the possibilities of the LTV for relationship building has several consequences for the other two aspects of the traditional way of employing the LTV.

First, when focusing on relationships rather than transactions, the calculation of the LTV should not only be based on past purchase behavior but also on data concerning future situations. That is, instead of only using the organization’s database, which is more or less the by-product of the order process, additional data should be collected, such as customer satisfaction and (repeat) purchase intention, the so-called forward-looking data. These data are also useful to capture the dynamics in the relationship between the consumer and the organization. Secondly, in order to build one-to-one relationships, the LTV should be determined at the individual level rather than at the group level. Consequently, the data should be collected and analyzed at the individual level, and the strategy should be based on these individual results.

Thus, in order to employ the LTV to its full extent, it should be used for decisions regarding creating, developing and maintaining relationships. The ultimate goal should be an individual-based normative direct marketing strategy. The strategy should depend on past purchase behavior and on forward-looking data. The normative aspect implies that the strategy indicates when an individual should receive a mailing and what kind of mailing that should be.

The first aspect, ‘when’, deals with the timing and sequencing of the mailings. The timing relates to the choice of the day of the week, month or season. Common wisdom in direct marketing suggests that, for obvious reasons, the last days of the week are preferable (Fraser-Robinson 1989, p. 106, Vögele 1992, pp. 292-293). Sequencing relates to the period between two consecutive mailings. Clearly, the optimal period differs between individuals. Furthermore, the optimal period may differ between the various mailings for one particular individual. Bitran and Mondshein (1996) derive a heuristic for a catalog retailer for when and how often to mail. Gönül and Shi (1996) extend this approach by explicitly modeling the consumers’ response behavior. They use a structural dynamic programming model to determine a strategy for a catalog retailer. Both papers conclude that it is optimal to mail individuals who have purchased a small number of times, and to mail those individuals who did not purchase anything for a long time. Note that these findings may be typical for a catalog retailer, since the response probability is not zero if an individual
does not receive a mailing. That is, the possibility remains that the individual will order from the last catalog received.

The second aspect of the strategy, ‘what mailing’, relates to the type of mailing an individual should receive. It includes the design of the mailing and the offer submitted. The importance of a carefully designed mailing is demonstrated in chapter 6. Furthermore, we showed in chapter 6 that it is useful to take the interaction between the target and mailing characteristics into account in order to choose a mailing design that maximizes the expected profits per individual. The offer submitted in the mailing includes the product itself, the price and the other elements for positioning the product. For example, the product could be submitted with an additional promotion. This promotion could be directed to reactivate an individual or to award individuals for their loyalty. In order to maximize the expected profits, the additional promotion should be determined per individual.

Note that target selection is part of such an individual-based strategy. That is, an individual is selected for a certain mailing at a particular time if that maximizes its LTV. An individual is excluded from the mailings list if his LTV is negative for all possible direct marketing activities. Furthermore, note that the decision of ‘when’ and ‘what’ are not independent. For example, the sequencing of the mailings of a catalog retailer differs between the choice of sending a reminder or a new catalog.

It goes without saying that a fine-tuned normative individual strategy is difficult to formulate and that it suffers from several practical drawbacks. The major difficulty is that, even in a rather simple setting, the structure of individual characteristics (including past purchase behavior and forward looking data), mailing and offer characteristics and situational factors, is very complex. The dynamical aspects in the analysis are enhanced by the selection effects resulting from the chosen strategy. That is, the chosen strategy is not randomly assigned to individuals but is based on the individuals’ behavior.

One of the main practical drawbacks is data availability. In order to determine individual values of the LTV on data concerning future situations, such as purchase intentions, it implies that organizations should collect additional data at the individual level. The collection of these data will be an expensive and difficult task. Not surprisingly, Hoekstra and Huizingh (1996) find that organizations with many customers have significantly less of this kind of information at the individual level than organizations with a small number of customers. Another practical problem is that a fine-tuned individual strategy may not be cost-efficient. That is, it may well be that the cost of implementing...
the strategy (data storage, computing, handling of the mailings and the like) exceeds the potential benefits.

To sum up, although a fine-tuned normative individual strategy based on the LTV should be the ultimate objective of a direct marketing organization, it will be quite hard to realize. Two main reasons are that the normative strategy should result from a complex structure and that there is a lack of specific individual data. There is, however, considerable room for improvement of various aspects of such a strategy. One of these aspects, which we examine in this chapter, is to determine the optimal frequency with which an organization should send its customers a mailing for frequently purchased consumer goods.

Roberts and Berger (1989, pp. 241-242) argue that this question has to be answered by experimentation. An experiment can for instance be performed by considering several (random) groups of the mailing list that receive the direct mailings with different frequencies. The objective of this experiment is to determine the frequency that generates the highest average profit of the addressees. Such an experiment, however, has several drawbacks:

1. The group that receives mailings with a low frequency and the group that receives mailings with a high frequency consists of “good” and “moderate” targets. This causes inefficiencies since the organization has to send some of the best (moderate) targets mailings with a much lower (higher) frequency than is preferable. This is especially harmful since the groups must be large and the experiment should be continued for some time to obtain useful and reliable results.

2. Given the length of the experiment, it takes some time before the organization has reliable results enabling it to determine the mailing frequency.

3. The results may suffer from non-stationarity. This means that one year’s optimal frequency differs from the optimal frequency a few years later.

This chapter, which is based on Van der Scheer and Leeßang (1997), proposes a method to determine the optimal frequency of direct marketing activities for frequently purchased consumer goods (e.g. books, compact discs), which does not suffer from these drawbacks. The underlying idea of this method is that consumers’ purchase behavior should be the basis for the supplier’s direct marketing behavior. The method is operationalized by specifying a conceptual framework for the purchase behavior of the individual, i.e. a model that describes the decisions an individual has to make before the purchase of the direct marketing (DM) product. The model is used to simulate the decisions of an individual for a range of frequencies of DM activities. As a result, the
optimal frequency can be determined. The model decomposes the purchase behavior in the timing of purchases, which involves interpurchase time and purchase acceleration, and the DM-product choice. Thus, the model takes all the purchases of the individuals in the product category into account, and specifies explicitly the individual's decision to buy the product through DM or in a regular store.

The order of discussion is as follows. In section 7.2 we present the maximization problem of the direct marketing organization and discuss the model describing the consumers’ purchase behavior. The specification of the components of this model is discussed in section 7.3. In section 7.4 we discuss ways to obtain data for the input parameters for the simulation. The calibration of the proposed model requires that data are collected which satisfy specific criteria. We investigated whether DM organizations collect these data on a continuous basis. We were only able to find organizations that collect the specified data on an ad-hoc basis. To calibrate our model we collected data among undergraduate students. The student population is not representative of a whole population and the outcomes of this study should therefore be interpreted with care. Thus the empirical part of this chapter is no more than an empirical illustration and application of our model. A sketch of the simulation is given in section 7.5. We present the empirical results in section 7.6, and discuss additional research issues in section 7.7.

### 7.2 The model

Under consideration is a direct marketing organization whose goal it is to maximize expected profits over a given period by deciding the number of direct mailing campaigns for a frequently purchased consumer good. We will not address the problem of target selection and assume that the DM organization selects a group of individuals for these direct mailings. We determine the frequency with which these individuals should receive a mailing by simulating the consumers’ purchase behavior. First, we formally define the maximization problem of the organization. Then, we specify the model describing the consumers’ purchase behavior.

Let \( i, \quad i = 1, \ldots, I \), be the individuals and \( m, \quad m = 1, \ldots, M \), be the mailings. The organization’s maximization problem is

\[
\max_M \sum_{m=1}^{M} \sum_{i=1}^{I} (w R_{im} - c),
\]  

(7.1)
subject to the constraint that the \( M \) mailings are sent in the specified period. \( R_{im} \) is the random variable given by

\[
R_{im} = \begin{cases} 
1 & \text{if individual } i \text{ responds to mailing } m \text{ by purchasing the DM product} \\
0 & \text{if individual } i \text{ does not respond to mailing } m;
\end{cases}
\]  

(7.2)

\( w \) are the revenues to a positive reply, and \( c \) is the cost of a mailing. Since it suffices to consider only a short period, as will become clear later on, we do not include a discount factor for future revenues. The key element in expression (7.1) is \( R_{im} \), in particular \( P(R_{im} = 1) \), i.e. the probability that \( i \) will respond to mailing \( m \).

Consider an individual that purchases products of a certain category. The individual has to decide when to purchase the product, and, in case of a purchase, where to purchase (cf. Gupta 1988). The first decision depends, among other things, on the time elapsed since the last purchase and the distribution of the interpurchase times. An interpurchase time of individual \( i \), \( T_i \), is the period between two consecutive purchases. Interpurchase times are random variables which are inversely related to the frequency with which the individual purchases the product. The distribution of interpurchase times enables us to determine the probability that the time between two purchases is at least of length \( t \): \( P(T_i > t) \). If the individual decides to purchase the product, he has to choose where to buy it. Here we assume that this can be either in a (regular) store or through a direct marketing organization.

Let \( t_m \) be the time between the last purchase and the next mailing, and let \( \theta_i \) be the interval in which an individual \( i \) takes the DM product into consideration. We allow this interval to vary across individuals. Some individuals decide immediately after they receive the mailing whether or not they want to buy the product, whereas other individuals keep the mailing to decide later on. Consequently, we can define a period \( (t_m, t_m + \theta_i) \) that the DM product is considered as an alternative. Thus,

\[
P(R_{im} = 1) = P(R_{im} = 1 \mid T_i \leq t_m \leq t_m + \theta_i) \cdot P(t_m \leq T_i \leq t_m + \theta_i) \\
+ P(R_{im} = 1 \mid T_i < t_m \lor T_i > t_m + \theta_i) \cdot P(T_i < t_m \lor T_i > t_m + \theta_i) \\
= P(R_{im} = 1 \mid t_m \leq T_i \leq t_m + \theta_i) \cdot P(t_m \leq T_i \leq t_m + \theta_i),
\]

(7.3)

where the last step holds, since a DM product can only be purchased in the period that a category purchase is made, i.e. \( P(R_{im} = 1 \mid T_i < t_m \lor T_i > t_m + \theta_i) \)
The model

\( t_m + \theta_j = 0 \). In (7.3), \( P(R_{im} = 1) \) is the product of the probability that \( i \) will purchase the DM product conditional on the purchase of a product from the category, which we call the DM-product choice, and the probability of a category purchase in \( (t_m, t_m + \theta_j) \). We now elaborate on the latter probability.

To this end, we introduce the concept of planned interpurchase time \( T^* \) (cf. Bucklin and Lattin 1991). It is an unobserved random variable that denotes the planned or intended period between two consecutive purchases. The individual purchases the product as planned, i.e. in accordance with the planned interpurchase time, except when the market behaves differently than expected. Unexpected behavior may be brought about by e.g. a price discount or a direct marketing activity. Thus, the observed interpurchase time is the planned interpurchase if the market behaves as expected; otherwise they differ.

Write \( P(t_m \leq T_i \leq t_m + \theta_j) \) conditional on the various intervals of the planned interpurchase time, i.e.

\[
P(t_m \leq T_i \leq t_m + \theta_j) = P(t_m \leq T_i \leq t_m + \theta_j, T_i^* < t_m)P(T_i^* < t_m) + P(t_m \leq T_i \leq t_m + \theta_j, T_i^* \geq t_m)P(T_i^* \geq t_m)
\]

Relation (7.4) holds because

\[
P(t_m \leq T_i \leq t_m + \theta_j, T_i^* < t_m) = 0,
\]

since individuals are not aware of the mailing at the time they plan to purchase the product, and

\[
P(t_m \leq T_i \leq t_m + \theta_j, T_m \leq T_i^* \leq t_m + \theta_j) = 1
\]

because the planned interpurchase time \( T_i^* \) and the interpurchase time \( T_i \) coincide.

If the individual receives the mailing before the planned purchase, this may influence his behavior and cause him to decide to purchase the product at an earlier time than planned. This is called purchase acceleration (or forward buying), of which the probability is given by \( P(t_m \leq T_i \leq t_m + \theta_j, T_i^* > t_m) \).

To summarize, expressions (7.3) and (7.4) define the model describing the consumers’ purchase behavior. The essential components of this model are:
1. the distribution of planned interpurchase times, which determines $P(t_m \leq T_i^* \leq t_m + \theta_i)$ and $P(T_i^* > t_m)$;
2. purchase acceleration, which determines $P(t_m \leq T_i \leq t_m + \theta_i \mid T_i^* > t_m)$;
3. DM-product choice, which determines $P(R_{im} = 1 \mid t_m \leq T_i \leq t_m + \theta_i)$.

We will specify these components in the next section.

### 7.3 Specification of the model’s components

In order to operationalize the model, we have to define the specifications underlying the (planned) interpurchase time, purchase acceleration and DM-product choice. For the time being we assume that all the information required to estimate the unknown parameters of these specifications is available.

#### Interpurchase time

To obtain a density function of interpurchase times we can use a parametric or a non-parametric approach. In the parametric approach we postulate the underlying function of the interpurchase times. The parameters of this function are estimated from the observed data. A standard approach to do so is with a hazard model (e.g. Gupta 1991, Kiefer 1988, Vilcassim and Jain 1991). However, even in frequently researched product categories there is no single function that adequately characterizes individuals’ interpurchase times (Jain and Vilcassim 1991). Moreover, there is no theory that specifies the probability function. For those reasons, Jain and Vilcassim recommend the use of a very general specification.

A non-parametric probability function is such a general specification. It has the advantage that we do not have to specify a function a priori. An obvious choice is the kernel density estimator defined by

$$f(T) = \frac{1}{hN} \sum_{n} K \left( \frac{T - t_n}{h} \right), \quad (7.5)$$

where $K(\cdot)$ is the kernel, and $h$ is the smoothing parameter; $t_n$ denotes the observed interpurchase times of all purchases of all the individuals, $n = 1, \ldots, N$. We use the Gaussian kernel with $h = 1.06\sigma N^{-1/5}$, where $\sigma$ is the standard deviation of the interpurchase times (Silverman 1986, p. 45).
Purchase acceleration

Let $t_i^*, t_i^* > t_m + \theta_i$ be a realization of $T_i^*$. We focus on purchase acceleration due to a direct marketing activity. We assume that the actual interpurchase time, $t_i$, will be $t_m + \theta_i$ if the individual decides to accelerate a purchase. Hence, the difference in time between the planned purchase and the accelerated purchase is $\delta_i = t_i^* - (t_m + \theta_i)$.

For a number of purchase situations, $j = 1, \ldots, J$, with $t_i^* > t_m + \theta_i$, we wish to determine whether or not $i$ accelerates the purchase. Let $y_{ij} = 1$ if $i$ accelerates the purchase in situation $j$, and $y_{ij} = 0$ otherwise. We denote willingness of $i$ to accelerate the purchase in situation $j$ by the latent variable $y_{ij}^*$ that satisfies a linear model,

$$y_{ij}^* = \alpha_i + \beta' x_j + \rho \delta_{ij} + \epsilon_{ij}, \quad (7.6)$$

where $x_j$ is a vector of covariates of the offer in situation $j$ (e.g. a price discount), $\delta_{ij}$ is the time between $t_i^*$ and $t_m + \theta_i$ for the $j$th purchase situation; $\beta$ and $\rho$ are unknown parameters, and $\epsilon_{ij} \sim N(0, 1)$, independently of $x_j$ and $\delta_{ij}$. The heterogeneity across individuals is represented by the $\alpha_i$, assumed to satisfy $\alpha_i \sim N(\alpha, \sigma_\alpha^2)$, with $\alpha$ and $\sigma_\alpha^2$ unknown parameters. We do not observe the willingness but the actual decision

$$y_{ij} = \begin{cases} 1 & \text{if } y_{ij}^* > 0 \\ 0 & \text{otherwise,} \end{cases}$$

which is the so-called random effects probit model (e.g. Hsiao 1986). The parameter $\rho$ gives the effect of the length of the period between the planned purchase and the accelerated purchase. It is expected that the willingness decreases in $\delta_{ij}$, i.e. $\rho$ has a negative sign.

The probability of purchase acceleration for given $x$, $t_i^*$, and $t_m + \theta_i$, so $\delta_i = t_i^* - t_m + \theta_i$, is given by

$$P(t_m \leq T_i \leq t_m + \theta_i \mid t_i^* > t_m + \theta_i) = P(y_{ij}^* > 0 \mid x, \delta_i) = \Phi(\alpha_i + \beta' x + \rho \delta_i),$$

where $\Phi(\cdot)$ is the standard normal distribution function.

Accelerated purchases can be either incremental or borrowed. It is said to be (completely) incremental if the successive planned interpurchase time does not change. It is (completely) borrowed if successive planned interpurchase time increases by $\delta_i$. Most situations will probably fall in between these two extremes. This means that the subsequent interpurchase time increases by $\kappa \delta_i$ ($0 \leq \kappa \leq 1$), where $\kappa$ is defined as the borrowing rate.
DM-product choice
Given that the individual purchases the product in \((t_m, t_m + \theta_i)\), he has to choose whether to purchase it through DM or in the store. For this decision we also use a random effects probit model, viz.

\[
R_{ij}^* = \mu_i + \gamma'X_j + u_{ij},
\]

(7.7)

where \(\gamma\) is a vector of unknown parameters, and \(u_{ij} \sim N(0, 1)\) is the disturbance term. Heterogeneity across individuals is represented by \(\mu_i\), assumed to satisfy \(\mu_i \sim N(\mu, \sigma^2_{\mu})\), with \(\mu\) and \(\sigma^2_{\mu}\) unknown parameters. For convenience of notation we assume that the same \(x_j\) occur in (7.6) and (7.7), but this is innocuous since elements of \(\beta\) and \(\gamma\) can a priori be set at zero. We do not observe \(R_{ij}^*\); we only observe whether the individual decides to purchase the DM product \((R_{ij} = 1)\) or not \((R_{ij} = 0)\), i.e.

\[
R_{ij}^* = \begin{cases} 
1 & \text{if } R_{ij}^* > 0 \\
0 & \text{otherwise},
\end{cases}
\]

given that \(t_m \leq T_i \leq t_m + \theta_i\). The latent variable, \(R_{ij}^*\), can be interpreted as the difference in utility between buying the DM product and buying the product in a store. If the utility of the former is larger, \(i\) will buy the DM product (Domencich and McFadden 1975). The probability of choosing the DM product is given by

\[
P(R_{im} = 1 \mid t_m \leq T_i \leq t_m + \theta_i) = \Phi(\mu_i + \gamma'x),
\]

where \(\Phi(\cdot)\) is the standard normal distribution function.

7.4 Data
In order to obtain estimates for the unknown parameters in the specified models and hence to employ the simulations, we need several kinds of information. The information should relate to: (1) interpurchase times, (2) purchase acceleration, and (3) DM-product choice. We consider two ways to obtain this information. First, it can be obtained from a household panel. Secondly, information can be collected by questionnaires or experiments.

Apart from demographic and geographic information, a household panel typically provides information on all the households’ purchases. Hence, it
Simulation of individuals’ decisions

provides sufficient information, in principle, to derive a distribution of inter-purchase times. In order to examine the purchase acceleration, we also need for each purchase information on the confrontations of the individuals with promotions that, at least in principle, may influence their purchase decision. Using a hazard model with covariates, the effect of promotions on the interpurchase time can be determined (e.g. Jain and Vilcassim 1991). To examine the effect of a promotion in a store versus direct marketing, we also need to know where the product is purchased, i.e. through direct marketing or in a regular store. The DM-product choice can be analyzed when information is available on all the direct mailings that the household received in the product category. Information on promotions and direct mailings is, however, not usually available through household panel data.

The other way to obtain information on the relevant variables is by a questionnaire and/or experiments. Ideally, this should be collected in addition to a household panel. A questionnaire could be used to obtain information on, for example, the period that a household takes a DM product into account. Experiments, like conjoint analysis, can be used to derive information on purchase acceleration and DM-product choice. The advantage of conjoint analysis is that the individuals are confronted with more or less real-life situations. That is, all the aspects that are expected to play a role in the individual’s decision can be incorporated in the choice sets. In particular the method of pairwise comparison is useful since it elicits choices from the individuals, in contrast with other conjoint methods. Obviously, an experiment will not give an exact picture of the individual’s behavior but it will yield helpful information to obtain the input parameters.

It is important to realize that we do not need information for all the households on the mailing list. In principle information of a small sample is sufficient. If this sample is not representative for the individuals on the mailing list, the findings could without difficulties be adjusted with the use of the geographic and demographic variables.

7.5 Simulation of individuals’ decisions

In this section we specify the set-up for the simulation of individuals’ decisions. We choose the number of mailings ($M$) in the given period. In our simulation we take these equally spread out over time, with a random start. Hence, given the time of the first mailing we know the (calendar) times of all the other mailings.
158 Frequency of direct mailings

Situation 1: The planned purchase is before the direct mailing

Situation 2: Planned purchase is in the interval \((t_m, t_m + \theta_t)\)

Situation 3: Direct mailing is before the planned purchase

Figure 7.1: Purchase situations in simulation

The \(ms\) denote the time mailings are sent, and \(p\) denotes the last purchase (indicated by the dot). The time between the last purchase and the mailing is defined by \(t_m\), and in the interval \((t_m, t_m + \theta_t)\) the DM-product is taken into consideration. Let \(t^*\) be the realization of the planned interpurchase time. In situations 1 and 2 the product will be bought in accordance with the planned interpurchase time. In situation 2 this could be the DM-product. In situation 3 the actual interpurchase time is \(t_m + \theta_t\) if the consumer accelerates the purchase by \(\delta_t\).
We describe the simulation of the purchases for one particular individual. Hence, it should be employed \( I \) times.

a. Draw \( \theta \sim \hat{g}(\theta) \), \( \alpha \sim N(\hat{\alpha}, \hat{\sigma}_\alpha^2) \), and \( \mu \sim N(\hat{\mu}, \hat{\sigma}_\mu^2) \), where \( \hat{g}(\theta) \) is the estimated (kernel) density of \( \theta \), and \( (\hat{\alpha}, \hat{\sigma}_\alpha^2) \) and \( (\hat{\mu}, \hat{\sigma}_\mu^2) \) are the estimated parameters of the normal distribution of the \( \alpha \) and \( \mu \), respectively.

b. Draw \( t^* \sim \hat{f}(T) \), a realization of \( T^* \); \( \hat{f}(T) \) is the estimated kernel density of \( T \). Determine \( t_m \): time period between the last purchase and the next mailing. Three situations can occur:

1. \( t^* < t_m \), then \( R_{im} = 0 \).
2. \( t_m \leq t^* \leq t_m + \theta_i \), then \( t_m \leq t \leq t_m + \theta_i \). Draw \( u \) from the uniform density on \((0, 1)\). \( R_{im} = 1 \) if \( u < \Phi(\mu_i + \gamma'x) \); else \( R_{im} = 0 \).
3. \( t^* > t_m + \theta_i \), then \( \delta_i = t^* - (t_m + \theta_i) \). Draw \( u \) from the uniform density on \((0, 1)\). \( R_{im} = 1 \) if \( u < \Phi(\alpha_i + \beta'x + \rho \delta_i) \Phi(\mu_i + \gamma'x) \); else \( R_{im} = 0 \).
   If \( R_{im} = 1 \), then the next interpurchase time will be increased by \( \kappa \delta_i \) (where \( \kappa \) is the borrowing rate).

c. Repeat until \( R_{im} \) is simulated.

d. Determine profits for \( i \): \( \sum_{m=1}^{M} (w_i R_{im} - c_i) \).

In step a the individual effects and the interval in which \( i \) takes the DM product into consideration are determined. These parameters do not change during the simulation. The planned interpurchase times are drawn in step b.

The three situations which are possible are depicted in figure 7.1. The \( ms \) indicate the (calendar) times of the DM activities; \( p \), indicated by the dot, is the time of the last purchase. Hence \( t_m \) is the period between \( p \) and the next \( m \). The interval \( (t_m, t_m + \theta_i) \) indicates the period in which the DM product is considered as an alternative. In situation 1 the planned purchase is before the direct mailing, so the product will be bought as planned and it is not the DM product. In situation 2 the product will also be bought as planned and this could probably be the DM product. The probability that it is the DM product is \( \Phi(\mu_i + \gamma'x) \). In situation 3 the planned interpurchase time is longer than \( t_m + \theta_i \). The DM product is bought if the individual accelerates its purchase and chooses the DM product (conditional on purchase acceleration); this probability is \( \Phi(\alpha_i + \beta'x + \rho \delta_i) \Phi(\mu_i + \gamma'x) \). If the DM product is purchased at \( t_m + \theta_i \), the successive interpurchase time will increase by \( \kappa \delta_i \) (\( \kappa \) is the borrowing rate). This is repeated (step c) until \( R_{im} \) is simulated. Employing this for \( I \) individuals we obtain the overall profits. We examine this for different values of \( M \), which gives the optimal frequency.
7.6 Empirical illustration

We applied the proposed model and simulation to compact disc purchases among students. The students filled out a questionnaire with questions on DM-product choice, purchase acceleration and compact disc purchases (in that order). Of the 146 students who received the questionnaire, 141 filled it out completely. First we briefly discuss the components of the individual response model, then we discuss the simulation results.

Data collection

Using a time line the respondents indicated the periods and prices of their CD-purchases of last year. Generally this would be difficult, but given the age and the students’ background this is not impossible. The data obtained in this way give us indications about the interpurchase times. To determine the kernel density of the interpurchase times, which is depicted in figure 7.2, we considered those respondents who bought at least four CDs at a price higher than NLG 25. We considered these respondents because they would be an interesting group for a DM organization. We chose the critical price level of NLG 25 because we want to consider CDs which have been bought at a more or less standard price, which is about NLG 40. Consequently, we used 69
Empirical illustration

Figure 7.3: Probability density function for $\theta$ (in weeks)

respondents with, in total, 321 interpurchase times to determine the kernel density. The average interpurchase time of these respondents is 8.45 weeks. The smoothing parameter, based on the formula given in section 7.3, equals 2.52.

The distribution of the $\theta$'s is obtained as follows. In the questionnaire we asked the respondents how long they kept a brochure with compact discs promotions, given that they were interested in the price discount but had not planned to purchase a compact disc. They had to choose from several digit preferences (e.g. a day, two or three days, . . . , two weeks). The average $\theta$ is 0.75 week. Figure 7.3 shows the estimated probability density function based on the Gaussian kernel. Using the smoothing parameter defined in section 7.3, which equals 0.275, we obtain a kernel that shows too much of the fixed answers in the questionnaire. Therefore we use a larger smoothing parameter (0.4), which resulted in the depicted curve.

We used a conjoint experiment, with the method of paired comparison, to obtain purchase acceleration data. The attributes we included are price (NLG 40, 35, and 30), when (direct, in 1, 2 or 3 weeks), and whether or not there was a savings plan (also called patronage awards). The latter refers to a type of promotion in which an individual saves up for a free compact disc. We kept several attribute levels fixed between the different comparisons, to keep
the task of the respondents tractable. A fractional factorial design resulted in eight comparisons. We gave the respondents one additional comparison, to be used for validation. All respondents provided information about the same nine pairs. We operationalized planned purchase in the following way: “Assume that you plan to buy a compact disc in two weeks because you receive your monthly grant (or because it is your friend’s birthday). However, you can buy the CD with a price discount now.” Then the respondents were asked to choose between two options.

Of the 141 respondents, 63 indicated that they ‘always’ preferred forward buying. This information is useful but it complicates our simulation. Moreover, we cannot use it for estimation of the random effects probit model, because these respondents would have an infinitely large positive intercept. Hence, we estimated the model on the basis of the 78 respondents with variation in their choices. This group is indicated by $G_1$. The group of respondents that do not have any variation in their answers is indicated by $G_2$. Each respondent made eight choices. Hence, $J = 8$, $I = 78$, and the covariates in $x$, are ‘price discount’ and ‘savings plan’. Price discount is the difference between the regular price (NLG 40) and the price offered. The time to a planned purchase, $\delta$, takes only three values: one, two and three weeks.

Table 7.1 gives the estimated coefficients. Since $\sigma^2$ was not significantly different from zero, we used the standard probit model. The signs of the coefficients are as expected. The price discount and the savings plan have a positive effect. Hence, an individual is more inclined to accelerate his purchases if there is a promotion. The time until the planned purchase has a negative effect on purchase acceleration. This shows that individuals are less likely to accelerate their purchase by, say, two weeks than by one week. To check the internal validity we used the additional paired comparison. For this paired comparison the predicted probability of purchase acceleration is 0.35. The actual percentage that chose purchase acceleration was 0.37.

In our simulation we need a probability of purchase acceleration for each value of $\delta$. Since a probit model defines the shape of the distribution function, we are able to do so by interpolation and extrapolation for all values of $\delta$ for respondents in $G_1$. However, we do not have a model for $G_2$. We approach this problem by making the bold assumption that the parameter values, except for the intercept, are equal for both groups. For $G_2$ we choose the intercept in such a way that the probability of purchase acceleration up to three weeks equals one. For the simulation this means that we have a mixture model with
Table 7.1: Probit model estimates, with standard errors between parentheses, for purchase acceleration (8 observations for each of the 78 respondents)

<table>
<thead>
<tr>
<th></th>
<th>Estimate (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-0.278 (0.230)</td>
</tr>
<tr>
<td>Price discount</td>
<td>0.370 (0.026)</td>
</tr>
<tr>
<td>Time to planned purchase ($\delta$)</td>
<td>-0.978 (0.087)</td>
</tr>
<tr>
<td>Savings plan DM</td>
<td>0.545 (0.127)</td>
</tr>
<tr>
<td>log-likelihood</td>
<td>-277.29</td>
</tr>
</tbody>
</table>

two segments ($G_1$ and $G_2$). This is incorporated in our simulation by assigning each individual to $G_1$ or $G_2$ with probability $\frac{78}{141}$ and $\frac{63}{141}$, respectively.

We also used a conjoint experiment with paired comparisons for the DM-product choice. In each comparison, the choice is between making a purchase in a store and through direct marketing. The price in the store equals NLG 40. The attributes in the experiment are price of DM product, delivery time, savings plan of the store, and savings plan of DM organization. Table 7.2 presents the estimates of the random effects probit model. The signs of the coefficients are as expected. The choice of a DM product is positively affected by a price discount and a savings plan of the DM organization. Delivery time and a savings plan of the store have a negative effect on DM-product choice. The DM-organization’s savings plan has a stronger effect than the store’s savings plan. This may indicate that the switching behavior between DM organizations is smaller than that between stores. Thus, a savings plan of a DM organization is more valuable to the individual. The large value of $\sigma_\epsilon^2$ indicates that there is much heterogeneity among individuals. That is, some individuals are much more inclined to buy the DM product than others. The respondents had to evaluate a particular paired comparison twice (albeit presented in different ways). Of the 141 respondents 86.5% chose the same option both times. This means that the internal validity of the approach is satisfactory. We distinguished two groups in the purchase acceleration experiment, one with variation in their choices ($G_1$) and one without ($G_2$). To see whether these groups differ with respect to DM-product choice, we estimated the random effects probit model for the two groups separately. On the basis of a likelihood ratio test we concluded that there was no significant difference.
Table 7.2: Random effect probit model estimates for DM-product choice (9 observations for each of the 141 respondents), with standard errors between parentheses

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-0.908</td>
<td>0.032</td>
</tr>
<tr>
<td>Price discount</td>
<td>0.221</td>
<td>0.030</td>
</tr>
<tr>
<td>Delivery time</td>
<td>-0.442</td>
<td>0.014</td>
</tr>
<tr>
<td>Savings plan DM</td>
<td>0.320</td>
<td>0.019</td>
</tr>
<tr>
<td>Savings plan store</td>
<td>-0.117</td>
<td>0.024</td>
</tr>
<tr>
<td>$\sigma_{\alpha}^2$</td>
<td>2.463</td>
<td>0.024</td>
</tr>
<tr>
<td>log-likelihood</td>
<td>-436.54</td>
<td></td>
</tr>
</tbody>
</table>

Simulation results

Figure 7.4 shows a typical simulation result of the integrated model ($w =$NLG 20, $c =$NLG 2.5, $I =$1000). It depicts the average expected profit per individual for different frequencies and price discounts. The frequency varies from one to ten mailings per year. As expected, profits first increase and later decrease with frequency. When the price discount increases from five to ten guilders, the optimal frequency increases from four to five mailings. If the price discount equals zero, the optimal frequency is equal to zero. However, we may not conclude from figure 7.4 that a rise in the price discount always causes an increase in the optimal frequency. The fact is that there are two effects. First, the probability of buying the DM product increases. Higher probabilities imply higher frequencies. Secondly, the probability of purchase acceleration increases, which implies a decrease in the optimal frequency. The net result of these two effects is not unambiguous.

Another promotion tool for the DM organization is the savings plan. We also performed a simulation with this variable. The results are similar to the outcomes of a simulation with a price discount. The only difference is the absolute value of expected profits. The similarity is not an unexpected result, since the coefficients of a savings plan, in the probability model for purchase acceleration and for DM-product choice, have the same sign as those of a price discount. To examine the effect of the borrowing rate we ran simulations for the two extremes, viz. borrowing rate equal to zero and one, respectively. Simulation results indicate that there is hardly any difference between these
two extremes. In other words, the lack of information is not harmful in this case.

We assumed that there is an interval in which respondent $i$ takes the DM product into consideration ($\theta_i$). There are two ways to implement this interval in the simulation: 1) the same value is taken for each individual, i.e., the average value; 2) for each individual we draw a $\theta_i$ from the estimated probability density function. The optimal frequency and the expected profit differ depending on the approach taken. Expected profits are slightly larger for a fixed $\theta$. An additional simulation showed that a rise in the average $\theta$ implies that the optimal frequency increases. Even though the optimal frequency changes with another parameter value or assumption of $\theta$, the simulations indicate that the effect on profits is small. In other words, the results are not very sensitive to this parameter value or assumption.

### 7.7 Discussion and conclusion

Traditional direct mail research focuses on optimization of aspects associated with one particular direct mail campaign. A major drawback of this research is that it focuses on a short-run criterion rather than on a long-run criterion.
such as the lifetime value. In order to employ the LTV to its full extent, it should be used for decisions regarding creating, developing and maintaining relationships. The ultimate goal should be an individual-based normative direct marketing strategy. The strategy should depend on past purchase behavior and on forward-looking data. The normative aspect implies that the resulting strategy indicates when an individual should receive a mailing and what kind of mailing that should be. Unfortunately, such a fine-tuned strategy based on the LTV will be quite hard to realize, since (1) this strategy should result from a complex structure, and (2) there is a lack of specific individual data. There is, however, considerable room for improvement of various aspects of such a strategy. One of these aspects, which we examine in this chapter, is the question with what frequency the organization should employ its direct marketing activities.

We proposed a method that seeks the optimal frequency of direct mailings for frequently purchased consumer goods. The method is based on the idea that the consumers’ purchase behavior is the basis of the supplier’s direct marketing behavior. The method is operationalized by specifying a model that describes the decisions an individual has to make before the purchase of the DM product. The proposed model decomposes the purchase behavior in the timing of purchases, which involves interpurchase time and purchase acceleration, and the DM product choice. The model takes all the purchases of the individuals in the product category into account, and explicitly specifies the individual’s decision to buy the product through DM or in a regular store. The model is used to simulate the decisions of an individual for various frequencies of DM activities. As a result, the optimal frequency can be determined. We illustrated the method with an application for which the input parameters were obtained by a questionnaire and conjoint analyses. It demonstrates that the proposed method is a relatively easy way to determine the optimal frequency of direct mailings.

There are various limitations that should be recognized. These limitations refer to: (a) specification of the maximization problem; (b) specification of the consumers’ purchase behavior, and (c) data collection.

There are several aspects related to the specification of the maximization problem. We consider the simple situation in which the DM organization solely focuses on the optimal frequency for a fixed group of individuals. Thus, target selection does not play a role, and each individual will receive the mailings with the same frequency. An obvious extension would be to identify segments and to determine the optimal frequency for each segment. Another stringent
assumption is that the individual chooses between purchasing the product either in the store or through DM. Hence, DM activities are not used to boost the incentive to purchase the product in a store (e.g. with coupons). Finally, we considered only the purchase of one product at a time. Thus, we did not pay attention to the quantity decision or to mailings that offer various products (e.g. catalogues).

A drawback of our model specification is that it does not take into account the possible long-term impact of DM activities. In other words, DM-product choice is independent of the number of mailings the individual has received. Furthermore, impulse purchases are only accounted for to a limited extent.

The structure of the proposed model can be modified in various ways. For example, it is possible to specify other distribution functions and to incorporate heterogeneity into the model in an alternative manner. In the present model, heterogeneity is included by an individual specific intercept in the probit model and in the interval in which the DM product is taken into account. However, we assumed that these forms of heterogeneity are independently distributed. Moreover, we did not assume heterogeneity in the slope parameters.

Some of the limitations of the data collection of our illustration have already been discussed. In particular the way in which we obtained the data of the past purchases is open for discussion. However, it should be realized that our objective was solely to collect data on interpurchase times.

The choices of the various aspects of these three components obviously play a crucial role in the simulation. Consequently, the reliability of the results is enhanced by the validity of the assumptions. We wish to emphasize, however, that several aspects of the maximization problem and the specification and implementation of the model are easy to adapt. This implies that it is straightforward, at least in principle, to apply the proposed method to more complex situations. Moreover, it implies that the sensitivity of the assumptions can be easily explored. We have demonstrated this by examining the effect of (1) the borrowing rate and (2) the distribution of the interval in which the individual takes the DM product into consideration.
Chapter 8

Summary, future research and management implications

8.1 Summary

An effective direct marketing campaign aims at selecting those targets, offer and communication elements - at the right time - that maximize the net profits. The list of individuals to be mailed, i.e. the targets, is considered to be the most important component. Therefore, a large amount of direct marketing research focuses on target selection techniques. In the recent literature some papers have been published on improving the communication elements. The objective of the thesis is to propose several modifications to existing methods and to introduce new approaches in order to improve the effectiveness, and hence the net profits, of addressed direct mailing campaigns.

We started in chapter 2 with an overview of the various target selection techniques, which have either been proposed in the literature or are commonly employed by direct marketers. Unfortunately, we cannot draw a general conclusion about the most appropriate technique for target selection. The reason for this is that many, sometimes conflicting, aspects play a role. However, three general conclusions about these selection techniques can be made. First, data exploratory methods like CHAID are less useful for selection than regression type models (probit, discriminant models). Secondly, regression models perform reasonably well and have the advantage of a clear interpretation. Thirdly, neural networks did not turn out to be the breakthrough in target selection, because the results are comparable with conventional statistical techniques.

Nearly all the selection techniques proposed so far deal with the case of fixed revenues to a positive reply and hence concentrate on binary choice modeling (response versus nonresponse). Thus, the quantity of response is implicitly assumed to be equal over the individuals. However, many direct mailing
campaigns do not generate simple binary response, but rather a response where the quantity differs between individuals. This quantity can be e.g. the total revenues of purchases from a catalog retailer, or the amount of money donated to a charitable foundation. In chapter 3 we specified a model that incorporates the quantity of response \(a\) as well as the probability of response \(p\). We derived an optimal selection rule on the basis of \(a\) and \(p\). We showed that this selection rule should take into account the differences between the density functions of the (estimated) response quantity of the respondents and nonrespondents. We examined three approximations to determine these densities. The first approximation simply neglects the difference. It results in an intuitively appealing curve in the \((p, a)\) space. The second approximation assumes that both densities are normal with different means and the same variance. This also results in a curve in the \((p, a)\) space and it is still very easy to apply. The third approximation is obtained by employing a nonparametric technique to estimate the densities. The results of the empirical application suggest that adding quantity modeling to probability modeling can contribute significantly to profitability. Even the first approximation generates much higher profits than the current practice of solely modeling the response probability.

Generally, the parameters of a selection model are unknown and have to be estimated. On the basis of these estimates, the individuals are ranked in order to select the targets. All of the proposed selection methods consider the estimation and selection step separately. Since by separation of these two steps the estimation uncertainty is neglected, these methods generally lead to a suboptimal decision rule and hence not to optimal profits. In chapter 4 we formulated a decision theoretic framework that integrates the estimation and decision steps. This framework provides an optimal Bayesian decision rule that follows from the organization’s profit function. That is, the estimation uncertainty resulting from parameter estimation is explicitly taken into account. One of the difficulties of such an approach is how to evaluate the high-dimensional integral resulting from the Bayesian decision rule. We discussed and applied three methods to evaluate this integral numerically. As a first approach, we assumed that the posterior density is normal. Then the integral can be expressed in a closed form and hence it is very easy to compute. The disadvantage, however, is that it may be a crude approximation since the prior density is completely ignored. A more refined approximation of the integral is by the Laplace approximation as proposed by Tierney and Kadane (1986). This approximation only requires maximization (a single Newton-Raphson step) and differentiation. Hence, there is no need for numerical integration. The third approximation
is by applying Markov chain Monte Carlo integration. These procedures have become very popular on account of the availability of high-speed computers. The advantage of these algorithms is that they are easy to implement and that they do not require the evaluation of the normalizing constant. We applied these methods with an informative and an uninformative prior. The results indicate that the Bayesian decision rule approach yields higher profits indeed; the difference between the various approximations is, however, rather small.

The quality of the mailing list is an important aspect of the list of individuals to be targeted. Roughly speaking, the degree to which the targets can be identified represents the value of information on the list. In chapter 5 we examined the relation between the value of information and profits. We first showed the relation between the $R^2$ of the underlying model and the profits. As expected, profits increase with $R^2$. Our particular interest is in the value of postal code information. This is information on the households at the postal code level, which in the Netherlands comprises 16 households on average. The selection procedures can still be employed but the maximally attainable profit will be lower. However, the profit could still be higher than in the situation where no information is available at all. The trade-off between information at the individual level versus information at the postal code level depends on the homogeneity within the postal code areas. If the degree of homogeneity is high, individual information will hardly improve the selection results. In contrast, if there is hardly any homogeneity within the postal code areas, postal code information is not useful. We showed the relation between the quality of information, expressed as the intra-class correlation, which is a measure of homogeneity, and the expected profits. This enables an organization to attach a monetary value to postal code information and assess whether better information is worth its costs. This is a relevant consideration for an organization which is confronted with the question whether it should buy information at the individual level, at the postal code level, or no information at all. Although our approach is a rather stylized and mainly theoretical, it gives a clear insight into the essential ingredients and problems that play a role when information is valued.

In chapter 6 we focused on the communication elements of a direct mailing campaign, i.e. the characteristics of a mailing. A traditional approach to evaluate the effect of many mailing characteristics is by analyzing them separately, which is obviously inefficient, generally speaking. We presented two approaches, based on conjoint analysis, to determine the optimal design in a more efficient manner. We illustrated these approaches by means of three
applications. An extension of these approaches is to combine target selection with design optimization. Although the importance of both aspects has been recognized, hardly any attention has been given to the interaction between these two. It is obvious that interaction does exist. For example, a direct marketing organization should employ different communication elements for targets who have been customers for a long time, than for targets who never bought from the organization before. We proposed a model which simultaneously optimizes the mailing design and selects the targets. The empirical application showed that interaction does indeed exist. The optimal strategy resulted in thirteen different types of mailings. Each selected individual should receive one of these thirteen mailings. The strategy results in an increase of the net (expected) returns of 25% over the traditional used mailing, and 9% over the single best mailing.

All marketing is, or should be, a continuous process. With rare exceptions, a purchase is not the end of a relationship with an individual but rather the beginning or continuation. Consequently, the success of a mailing campaign should actually be defined in long-term criteria rather than in short-term criteria. However, all the methods considered so far are concerned with a short-term criterion, i.e. maximization of one particular mailing campaign. One of the most challenging questions in direct marketing is to extend these methods to a long-term profit maximizing strategy. This strategy should focus on the maximizing lifetime value (LTV) for each individual. Ideally, the LTV should be calculated on individual data, and it should be employed for a normative strategy regarding creating, developing and maintaining relationships. It is obvious that exploiting the LTV in this way may be too ambitious. However, there is considerable room for improvement of various aspects of such a strategy. In chapter 7 we discussed a particular one, viz. with what frequency an organization should send its customers a mailing. The underlying idea of our proposed method is that the consumers’ purchase behavior should be the basis for the supplier’s direct marketing behavior. The method is operationalized by specifying a framework for the purchase behavior of the individual, i.e. a model that describes the decisions an individual has to make before the purchase of the DM product. The proposed model decomposes the purchase behavior into the timing of purchases, which involves interpurchase time and purchase acceleration and the DM product choice. The model takes all the purchases of the individuals in the product category into account and explicitly specifies the individual’s decision to buy the product either through DM or in a regular store. The model is used to simulate the decisions of an individual for a range of frequencies of DM activities. As a result, the optimal frequency can be determined. We illustrated
the method with an application for which the input parameters were obtained by a questionnaire and conjoint analyses. The illustration demonstrated that the proposed method is a relatively easy way to determine the optimal frequency of direct mailings.

8.2 Future research

In the final section of each chapter we suggested several topics for future research. Many of these are of course closely related to the subject discussed in that chapter; we will not repeat them here. In this section we discuss some general issues that should be addressed in the future.

Generalization

In chapter 2 we concluded that it is not possible to draw general conclusions about the best model for target selection. The main reason for this is that many, sometimes conflicting, aspects play a role. We gave three general conclusions for the target selection techniques; it would be interesting to extend these conclusions. Particularly, a generalization with respect to the net profits would be valuable. That is, what technique, given a certain decision rule, generates the highest net profits? Such a general conclusion could be derived by considering several different data sets. Ideally, the data sets should differ with respect to sample size of the test mailing, the product submitted in the mailing, types of selection variables, response rate, etc. The approaches proposed in chapters 3 and 4 should, of course, be examined as well. Evidently, it is quite likely that such a comprehensive analysis will not lead to a general conclusion with respect to the net profits. However, it will at least give some insight into the relative performance of the various techniques. In addition, it will indicate the factors that affect the performance. Hence, even without an unambiguous conclusion, an extensive examination of the various techniques would be valuable.

Decision making under uncertainty

Many models are designed to help marketing managers make better decisions. On the one hand, a model can be built with the intention to represent the data adequately. In that case, the model assists a manager in the decision making process. This means that the model is just a source of information which is helpful in making a decision. On the other hand, a model can be specified in
order to formulate an explicit decision to be taken. In that case, managerial judgement, feeling and intuition do not play a role once the model is defined.

Generally, a decision has to be made under uncertainty simply because an organization faces an uncertain future. Furthermore, there is uncertainty in the estimated parameters. In addition, there is uncertainty in the specific parametric form of the specified model and due to the possible heterogeneity in the parameters. In the case that the model only assists in the decision making process, the researcher should take these aspects into account. That is, he knows the assumptions and limitations of the model and should be able to employ their implications adequately. In contrast, a normative model defines the decision to be taken and hence there is no room for the researcher’s or manager’s judgement. Evidently, in order to make the most appropriate decision, the normative model should take the uncertainty into account. In chapter 4 we formulated a framework which incorporated the estimation uncertainty. As to future research, it is useful to consider the uncertainty of the specific parametric form and of the possible heterogeneity as well. We briefly sketch the implications of these types of uncertainty for the target selection techniques.

Kalyanam (1996) examines uncertainty about the specified parametric model by employing a Bayesian mixture model approach. For each potential model specification the researcher should define a prior density for the parameters and a prior probability for that specification. The former is used to derive a statistic that represents the evidence in a certain specification. Combining this statistic with the prior probability gives the posterior probability that indicates the extent to which the observed data support the specific specification among the models under consideration. The final decision is a combination of the ‘optimal’ decisions of each specified model weighed by the posterior probabilities. In the context of binary choice models it implies that several possible functional forms should be formulated. Each specification generates a response probability for each individual. Ideally, these response probabilities should be derived within the Bayesian framework to account for estimation uncertainty. A combination of these response probabilities, appropriately weighed, results in a response probability on which the selection should take place. Although it seems to be a valuable approach, two points should be raised. First, another way to incorporate uncertainty about the parametric form is by employing a semi-parametric or nonparametric approach. Bult and Wansbeek (1995) apply the semiparametric estimator developed by Cosslett (1983); their results indicate that such a flexible specification adds little to a parametric specification. Similarly, Zahavi and Levin (1995, 1997) employ neural networks, loosely speaking
Future research

a nonparametric approach, and conclude that a parametric model performs adequately. Secondly, the most widely used parametric specification of binary choice models are the probit and logit models. Because the cumulative normal distribution and the logistic distribution are very close to each other, we are not likely to get very different results using a probit or logit model. Hence, it is probably of no empirical relevance to include the uncertainty between a probit or logit model in the analysis. Thus, it is doubtful whether uncertainty about the parametric form of the binary choice model will lead to higher profits. This approach could be useful, however, for modeling the quantity of response if e.g. the two-part model and sample-selection model (see section 3.2) are the possible specifications.

A restrictive assumption in the selection models considered in this thesis is that individuals respond in a homogeneous way. That is, a single parameter vector describes the relation between the selection variables and the dependent variable. Of course, it is uncertain whether this is an appropriate assumption. To capture the uncertainty of the possible heterogeneity in the parameters a latent variable model could be employed (e.g., DeSarbo and Ramaswamy 1994). Analogous to the semiparametric method for the binary choice model, this approach is based on a very general model specification, which includes a model with and without heterogeneity in the parameters. That is, one or more segments are determined. For each segment the method provides a response probability for all the individuals. Simultaneously, the model determines a probability of membership to each of these segments. The ultimate response probability of a particular individual is obtained by weighing the response probability in each segment by the membership probabilities. Besides applying this approach in the binary context, it would be useful to employ it for modeling the response quantity, but also for the models considered in chapters 5 and 6. It should be realized, however, that the advantage of a model assuming homogeneity is its parameter parsimony. Of course, the practical relevance of these models for target selection should be addressed by the bottom line results, i.e. the net profits.

Performance

In order to reduce the individuals’ risk and compete effectively with stores that have their products displayed, many direct marketing organizations offer a very generous return policy. The result of such a return policy is that many individuals use this opportunity extensively (e.g. Hess and Mayhew 1997). In order to make an effective target selection this aspect should also be taken into
Customer receives a mailing, respond?

Yes

Purchase

R = 1

D = 1

No

R = 0

D = 0

Return

Figure 8.1: Process of fulfillment of a purchase

account. This means that the organization should focus on the performance of an individual rather than on response only. The performance is described as the fulfillment of a purchase (see figure 8.1). That is, the individual has to decide whether to respond or not, and, in the case of response, whether to purchase the ordered product or to return it. The profit function (cf. section 3.2) is now given by

$$\Pi = aRD - c - R(1 - D)c_r,$$

where

$$R = \begin{cases} 1 & \text{if the individual responds} \\ 0 & \text{if the individual does not respond} \end{cases}$$

$$D = \begin{cases} 1 & \text{if the individual purchases the product} \\ 0 & \text{if the individual returns the product} \end{cases}$$

$a$ is the (fixed) revenues of response, $c$ the cost of a mailing, and $c_r$ the costs involved if an individual returns the product; $D$ is observed only if $R = 1$. Thus, the expected profits can be written as

$$E(\Pi) = aP(R = 1)P(D = 1 | R = 1) - c - P(R = 1)P(D = 0 | R = 1)c_r.$$

Hence, in order to select individuals for the mailing campaign the organization should focus on $P(R = 1)$ and $P(D = 1 | R = 1)$, or $P(R = 1 \& D = 1)$, rather than only on $P(R = 1)$. Note that we implicitly assumed that $P(D = 1 | R = 1) = 1$, and hence $P(D = 0 | R = 1) = 0$, throughout this thesis.

Given the hierarchical or nested structure of the process, a nested logit model seems to be the appropriate specification. However, since the explanatory variables underlying the model for $R$ and $D$ are generally individual
specific characteristics, a nested logit model is not helpful. That is, the explanatory variables are equal across all outcomes and we do not have choice characteristics. An appropriate model specification is the censored bivariate probit model, since $D$ is only observed if $R = 1$ (e.g. Boyes et al. 1989). The model is defined by

$$
R^* = x' \beta_1 + \varepsilon_1 \quad R = \begin{cases} 1 & \text{if } R^* > 0 \\ 0 & \text{otherwise}, \end{cases}
$$

$$
D^* = x' \beta_2 + \varepsilon_2 \quad D = \begin{cases} 1 & \text{if } D^* > 0 \\ 0 & \text{otherwise}, \end{cases}
$$

where $R^*$ and $D^*$ are unobservable, and the errors $(\varepsilon_1, \varepsilon_2)$ are assumed to be i.i.d. as standard bivariate normal; $x$ is a vector of explanatory variables. This approach accounts for potential correlation between the two decisions and thereby corrects for the potential sample selection bias that could occur in the separate estimation of the two equations. The vectors of parameters, $\beta_1$ and $\beta_2$, can be estimated with maximum likelihood in a straightforward manner (Meng and Schmidt 1985).

When the random factors influencing the choices at the two stages are independent, i.e. $\varepsilon_1$ and $\varepsilon_2$ are independently distributed, we obtain the so-called sequential response model (Amemiya 1975). The parameters $\beta_1$ can be estimated from the entire sample by dividing it into two groups, viz. respondents and nonrespondents. The parameters $\beta_2$ can be estimated from the subsample of respondents by dividing it into two groups: “purchased the product” and “returned the product”. In both cases the binary model can be estimated by the probit or logit method. Hence, this model is very easy to apply. Note that either the censored bivariate probit model or sequential response model is also useful if the response process consists of generating leads which should successively be converted into sales. In that case $R$ denotes the leads and $D$ the conversion of these leads (cf. Hansotia 1995).

There are several aspects related to the performance question. Apart from the possibility of purchasing or returning the product ordered, there are some individuals who neither pay nor return the product. Hence, the second decision should be extended with a third option, namely ‘does not pay’. The process can then be modeled with a censored multinomial probit model. Although this model allows for potential correlation between the decisions, it is of course easier to employ a sequential model in which the second decision is specified with a multinomial logit or probit model. Obviously, the defaulters receive a reminder after some time. This is followed by more reminders if they still
Summary, future research and management implications

do not pay. Hence, the process of fulfilment of a purchase is extended with a sequence of “pay - not pay” decisions. An interesting question is how many reminders an organization should send. In order to answer this question we should determine the probability of response, i.e. the probability of paying, by an individual that received \( k \) reminders but did not respond to any of these. In principle, the whole process can be formulated in a censored multinomial probit model in which there is censoring at each stage. This provides the required response probabilities. However, such a model may be difficult to estimate. Possible solutions are to employ the sequential binary choice model or the beta-logistic model (e.g. Rao and Steckel 1995; see also section 2.4).

A closely related interesting topic for future research is an examination of the effect of a less generous return policy. On the one hand, it will result in a lower return rate, i.e. \( P(D = 0 | R = 1) \) decreases, which is of course advantageous for the organization. On the other hand, it may result in a decrease in the response probability and hence the number of individuals, which is the drawback of such a policy. It would be useful to analyze this trade-off in order to obtain an optimal return policy.

Panel data

One of the most prominent questions in direct marketing is how to specify a method to maximize the lifetime value (LTV) of an individual. In direct marketing the LTV concept should ideally be used for decisions regarding creating, developing and maintaining relationships. The ultimate goal should be to develop an individual based normative strategy. This implies that the organization should specify a strategy that indicates when an individual should receive a mailing and what kind of mailing this should be. As indicated in chapter 7 this may be too ambitious, since the structure of individual characteristics, mailing and offer characteristics and situational factors, is very complex. There are, however, several steps to be taken towards using the LTV to its full extent. An important step is by employing panel data. In a panel data model a central role is played by the individual effect. The individual response vis-à-vis direct mailing may have a strong, persistent component largely driven by unobservable variables. Put differently, the behavior of an individual may be driven by its individual effect rather than by its past behavior. Such a model will give an organization a better understanding of the consumers’ response behavior. In order to obtain the desired result, we should specify a behavioral model, in contrast with most of the (selection) models considered so far.
We introduce the various ingredients of such a model by the following dynamic panel data tobit model

\[ y_{it}^* = \alpha + \rho y_{i,t-1}^* + x_{it}' \beta + \gamma_i + u_{it} \]

with

\[ y_{it} = \begin{cases} 
  y_{it}^* & \text{if } y_{it}^* > 0 \\
  0 & \text{otherwise,} 
\end{cases} \]

where \( u_{it} \sim NID(0, \sigma_u^2) \); \( y_{it}^* \) is the latent variable which measures the inclination to respond of \( i \) at time \( t \). We cannot observe the inclination but only whether or not \( i \) responded at time \( t \), which is denoted by \( y_{it} \). The vector \( x_{it} \) includes the mailings received and the characteristics of individual \( i \). To account for the heterogeneity across individuals an individual effect, \( \gamma_i \), is included. This could be interpreted as a measure of direct marketing proneness (e.g. Blattberg and Neslin 1990, chapter 3). In the case of a charity foundation it may indicate the extent to which an individual is involved with this foundation. The term \( y_{i,t-1}^* \) reflects the learning behavior of the individual (e.g. Bush and Mosteller 1955, Lee and Boonstra 1982). The inclusion of this term is based on the assumption that a purchase increases the habit strength of performing that behavior repeatedly, the so-called habit formation (Lilien and Kotler 1983, p. 233). In the econometric literature models with a lagged latent dependent variable are called habit persistence models (Heckman 1981, Schmidt 1981). The extent to which the individual will rely on its previous purchase for current behavior is governed by \( \rho \). Habit persistence is comparable to state dependence in the discrete dynamic panel data models.

The model indicates whether the heterogeneity or the habit persistence is the main determinant for current behavior. If \( \rho \) is small the heterogeneity is the dominant component. An appropriate strategy may be to send mailings with a high frequency to individuals with a large individual effect. In contrast, this may not be the appropriate strategy if \( \rho \) is large and the individual effects vary little across individuals. In that case, the observed behavior may be closely related to the frequency of the mailings. That is, the behavior is likely to change if the frequency changes. Hence, similar behavior in the past may result in two different strategies if one type of behavior is altered by habit persistence while the other is driven by the individual effect.
8.3 Management implications

The objective of the thesis is to provide models for profit maximization in direct marketing. Throughout this thesis we have aimed at a thorough discussion of the analytical and statistical aspects involved in the proposed models. As a result, some parts of this thesis are rather stylized and theoretical. In this section we discuss the important practical implications for direct marketing management.

In chapter 2 we gave an overview of the target selection techniques which have either been proposed in the recent literature or are commonly employed by direct marketers. Although the literature comparing some of these techniques is scarce, a few conclusions can be drawn. First, data exploratory methods such as CHAID are less useful for target selection than regression type models. Secondly, regression models perform well and have the advantage of a clear interpretation. Thirdly, the performance of a neural network is comparable with that of the conventional techniques.

When selecting targets for a direct mailing campaign it is important to take into account both the probability and the quantity of response. Although this may seem an obvious statement, in the literature attention has mainly been given to the probability of response. Hereby, it is implicitly assumed that the quantity of response is equal across individuals. Evidently, the quantity of response often varies across individuals, e.g. the revenues of purchases from a catalog retailer or the donations to a charitable foundation. In chapter 3 we showed that even the simplest approach to model these two aspects jointly results in much higher profits than modeling the probability of response only. Since the incremental costs of implementing a model for the quantity of response in combination with a model for the probability of response are relatively small, we recommend to model both aspects whenever the revenues to a positive reply vary across households.

In chapter 4 we argued that an organization should define its selection rule on the basis of a strict decision theoretic framework. This is a framework that incorporates the loss of all the possible decisions. In this way, the losses, in terms of expected revenues, are minimized. Obviously, this should be the organization’s objective. The empirical application demonstrated that the gains of this approach are relatively small. It shows, however, that on average the approach will pay off indeed. Since the additional costs of implementation are relatively small, we advocate the use of this more advanced approach. Because the profit differences between the various approximations are rather small, we
suggest to use the easiest approximation. That is, to approximate the posterior density by a normal density.

In chapter 5 we examined the relation between quality of information and profits. This approach enables an organization to make a trade-off between the acquisition of information at the individual level or at the postal code level. Although the approach is mainly theoretical, it demonstrates how important it is to assess the value of information. Furthermore, it gives an insight into the essential ingredients and problems that play a role when information is valued.

The applications of chapter 6 reinforce the well-known notion that it is important to design a mailing carefully. In contrast with the traditional way to analyze the effects of various mailing characteristics, we examined several characteristics simultaneously, on the basis of the conjoint analysis methodology. This approach is preferable since it is more efficient than examining one characteristics at a time. In addition, whenever an organization has information about its targets, it is useful to take into account the interaction effects between target and mailing characteristics. Typically, this method results in different mailings being used for the same direct marketing campaign. The application of section 6.4 demonstrates that it results in higher net profits. For the management it is of course important to make a trade-off between higher profits and the costs of implementing this strategy. However, even if only one mailing design is chosen, it can be useful to take interaction effects into account. The application shows that the optimal mailing design depends on whether or not the interaction effects are taken into account.

All the methods proposed so far focus on short-run profit maximization. In chapter 7 we argued that one of the most important topics in direct marketing is to extend these methods to long-term profit maximization. Ultimately, this should result in a strategy that maximizes the lifetime value of the individuals. Although this may be too ambitious, there is considerable room for the improvement of various aspects of such a strategy. In chapter 7 we discussed a particular aspect of this, namely with what frequency an organization should send its customers a mailing. Generally speaking, in order to solve problems related to the LTV, such as the optimal mailing frequency, additional information should be collected.
Bibliography


Efron, B. (1982), The jackknife, the bootstrap and other resampling plans, Society for Industrial and Applied Mathematics, Philadelphia.


Heckman, J.J. (1976), “The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models”, *Annals of Economic and Social Measurement*, 5, 475–492.


Author index

Addelman, S., 117
Ahn, H., 33
Aigner, D.J., 102
Akaah, I.P., 112
Albert, J.H., 65
Amemiya, T., 32, 177
Ansoff, H.I., 82
Arabmazar, A., 33, 42
Baier, M., 2, 82, 91
Balestra, P., 64
Banslaben, J., 25
Bauer, C.L., 1, 15, 34
Bawa, V.S., 53
Berger, J.O., 52, 63
Berger, P.D., 2, 4, 10–12, 25–27, 111–113, 115, 147, 150
Best, N.G., 79
Beukenkamp, P.A., 1
Bickert, J., 11, 82
Bitran, G.R., 148
Blattberg, R.C., 53, 70, 179
Blundell, R., 32
Boostra, A., 179
Boyes, W.J., 177
Bozdogan, H., 138
Breiman, L., 17, 86
Brooks, F.W., 81
Bucklin, R.E., 153
Bult, J.R., 18, 20, 21, 26, 30, 77, 82, 113, 174
Bush, R.R., 179
Chopra, V.K., 53
Cossette, S.R., 18, 33, 174
Courtheoux, R.J., 30, 147
Cowles, M.K., 67
Cryer, C.R., 15
Cyert, R.M., 52
DeGroot, M.H., 52
DeJong, W., 112
DeSarbo, W.S., 21, 22, 48, 175
Domencich, T.A., 156
Duan, N., 32–34
Efron, B., 45, 78
Frank, R.E., 145
Fraser, C., 112
Fraser-Robinson, J., 111, 148
Freedman, D., 86
Gönen, F., 148
Garner, T.I., 32
Gelman, A., 67
George, E.L., 53, 70
Gilks, W.R., 65
Goldberger, A.S., 33, 42, 43
Green, P.E., 113, 114, 116, 120
Greenberg, E., 65
Greene, W.H., 21, 43
Griliches, Z., 92, 109
Gupta, S., 152, 154
Härdle, W., 14
Hansotia, B.J., 177
Hartman, R.S., 34
Haughton, D., 17
Hay, J.W., 33, 34
Heckman, J.J., 19, 32, 33, 179
<table>
<thead>
<tr>
<th>Author</th>
<th>Page Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hess, J.D.</td>
<td>175</td>
</tr>
<tr>
<td>Heyde, C.C.</td>
<td>62</td>
</tr>
<tr>
<td>Hoekstra, J.C.</td>
<td>1, 5, 112, 113, 147, 149</td>
</tr>
<tr>
<td>Hsiao, C.</td>
<td>32, 155</td>
</tr>
<tr>
<td>Hughes, A.</td>
<td>147</td>
</tr>
<tr>
<td>Huizingh, K.R.E.</td>
<td>147, 149</td>
</tr>
<tr>
<td>Hyman, M.R.</td>
<td>1</td>
</tr>
<tr>
<td>Ibrahim, J.G.</td>
<td>64</td>
</tr>
<tr>
<td>Jackson, D.R.</td>
<td>147</td>
</tr>
<tr>
<td>Jain, D.C.</td>
<td>154, 157</td>
</tr>
<tr>
<td>James, E.L.</td>
<td>113</td>
</tr>
<tr>
<td>Jeffrey, H.</td>
<td>62</td>
</tr>
<tr>
<td>Johnstone, I.M.</td>
<td>62</td>
</tr>
<tr>
<td>Jones, A.</td>
<td>32</td>
</tr>
<tr>
<td>Judge, G.G.</td>
<td>17, 100</td>
</tr>
<tr>
<td>Kadane, J.B.</td>
<td>62, 170</td>
</tr>
<tr>
<td>Kalyanam, K.</td>
<td>54, 174</td>
</tr>
<tr>
<td>Kamakura, W.A.</td>
<td>48</td>
</tr>
<tr>
<td>Kass, G.</td>
<td>16</td>
</tr>
<tr>
<td>Kass, R.E.</td>
<td>62, 63</td>
</tr>
<tr>
<td>Katzenstein, H.</td>
<td>81</td>
</tr>
<tr>
<td>Kestenbaum, R.D.</td>
<td>147</td>
</tr>
<tr>
<td>Kiefert, N.M.</td>
<td>154</td>
</tr>
<tr>
<td>Klein, R.W.</td>
<td>18, 51, 52</td>
</tr>
<tr>
<td>Kobs, J.</td>
<td>1, 5</td>
</tr>
<tr>
<td>Korgaonkar, P.K.</td>
<td>112</td>
</tr>
<tr>
<td>Kotler, P.</td>
<td>179</td>
</tr>
<tr>
<td>Krieger, A.M.</td>
<td>120</td>
</tr>
<tr>
<td>Krishnamurthi, L.</td>
<td>32</td>
</tr>
<tr>
<td>Kumar, A.</td>
<td>23, 26</td>
</tr>
<tr>
<td>Lancaster, P.</td>
<td>59</td>
</tr>
<tr>
<td>Lattin, J.M.</td>
<td>153</td>
</tr>
<tr>
<td>Laud, P.W.</td>
<td>64</td>
</tr>
<tr>
<td>Leahy, K.</td>
<td>25</td>
</tr>
<tr>
<td>Lefebvre, P.S.H.</td>
<td>1, 3, 9, 150, 179</td>
</tr>
<tr>
<td>Lehmann, E.L.</td>
<td>14, 25</td>
</tr>
<tr>
<td>Levin, N.</td>
<td>23, 24, 26, 174</td>
</tr>
<tr>
<td>Li, H.</td>
<td>113</td>
</tr>
<tr>
<td>Lilien, G.L.</td>
<td>179</td>
</tr>
<tr>
<td>Lindgren, B.W.</td>
<td>128</td>
</tr>
<tr>
<td>Lix, T.S.</td>
<td>12, 25–27, 82, 90</td>
</tr>
<tr>
<td>Maddala, G.S.</td>
<td>20</td>
</tr>
<tr>
<td>Magidson, J.</td>
<td>16, 17, 26</td>
</tr>
<tr>
<td>Magliozi, T.L.</td>
<td>25, 27</td>
</tr>
<tr>
<td>Manning, W.G.</td>
<td>33, 34</td>
</tr>
<tr>
<td>Manski, C.F.</td>
<td>20, 21, 77</td>
</tr>
<tr>
<td>Marshall, J.J.</td>
<td>2</td>
</tr>
<tr>
<td>Mayhew, G.E.</td>
<td>175</td>
</tr>
<tr>
<td>McFadden, D.</td>
<td>156</td>
</tr>
<tr>
<td>Meghir, C.</td>
<td>32</td>
</tr>
<tr>
<td>Melenberg, B.</td>
<td>33, 34</td>
</tr>
<tr>
<td>Meng, C.L.</td>
<td>177</td>
</tr>
<tr>
<td>Miglautsch, A.</td>
<td>1</td>
</tr>
<tr>
<td>Mondschein, S.V.</td>
<td>148</td>
</tr>
<tr>
<td>Morgenstern, O.</td>
<td>52</td>
</tr>
<tr>
<td>Mosteller, F.</td>
<td>179</td>
</tr>
<tr>
<td>Mroz, T.A.</td>
<td>32</td>
</tr>
<tr>
<td>Murrow, J.L.</td>
<td>1</td>
</tr>
<tr>
<td>Muus, L.</td>
<td>54</td>
</tr>
<tr>
<td>Neslin, S.A.</td>
<td>179</td>
</tr>
<tr>
<td>Oopik, A.J.</td>
<td>112</td>
</tr>
<tr>
<td>Otter, P.W.</td>
<td>30</td>
</tr>
<tr>
<td>Oulabi, S.</td>
<td>17</td>
</tr>
<tr>
<td>Pearson, S.</td>
<td>147</td>
</tr>
<tr>
<td>Peltier, J.W.</td>
<td>112</td>
</tr>
<tr>
<td>Petit, K.L.</td>
<td>2</td>
</tr>
<tr>
<td>Poirier, D.</td>
<td>64</td>
</tr>
<tr>
<td>Posnett, J.</td>
<td>32</td>
</tr>
<tr>
<td>Powell, J.L.</td>
<td>14, 33</td>
</tr>
<tr>
<td>Raaijmaakers, M.C.</td>
<td>1</td>
</tr>
<tr>
<td>Raj, S.P.</td>
<td>32</td>
</tr>
<tr>
<td>Ramaswamy, V.</td>
<td>21, 22, 48, 175</td>
</tr>
<tr>
<td>Rao, V.R.</td>
<td>18, 19, 29, 178</td>
</tr>
<tr>
<td>Ripley, B.D.</td>
<td>23</td>
</tr>
<tr>
<td>Roberts, G.O.</td>
<td>65</td>
</tr>
<tr>
<td>Roberts, M.L.</td>
<td>2, 4, 10–12, 111, 112, 147, 150</td>
</tr>
<tr>
<td>Rossi, P.E.</td>
<td>54, 62</td>
</tr>
<tr>
<td>Rubin, D.B.</td>
<td>67</td>
</tr>
<tr>
<td>Russell, G.J.</td>
<td>48</td>
</tr>
<tr>
<td>Sachs, W.S.</td>
<td>81</td>
</tr>
<tr>
<td>Sapra, S.K.</td>
<td>20</td>
</tr>
<tr>
<td>Schibrowsky, J.A.</td>
<td>112</td>
</tr>
<tr>
<td>Schmidt, P.</td>
<td>33, 42, 177, 179</td>
</tr>
</tbody>
</table>
Schofield, A., 1
Sharda, R., 22
Shepard, D., 17, 20, 81
Shi, M., 148
Silverman, B.W., 39, 154
Simon, J.L., 29
Smith, G.E., 113, 115
Sonquist, J.N., 16
Spady, R.H., 18
Spiegelhalter, D.J., 79
Srinivasan, V., 113, 114, 116
Steckel, J.H., 18, 19, 29, 45, 178
Steerneman, A.G.M., 57
Thompson, T.S., 21
Thrasher, R., 17
Throckmorton, J., 111
Tibshirani, R.J., 45
Tierney, L., 62, 65, 170
Tismenetsky, M., 59
Titterington, D.M., 23
Tobin, J., 42
Vögele, S., 111, 125, 148
Van der Scheer, H.R., 30, 54, 82, 113, 150
Van Raaij, W.F., 3
Van Soest, A., 33, 34
Vanhonacker, W.R., 45
Varian, H.R., 52
Vilcassim, N.F., 154, 157
Von Neumann, J., 52
Vredenburg, H., 2
Vriens, M., 5, 112–114
Wagner, J., 32
Wang, P., 147
Wansbeek, T.J., 18, 26, 30, 54, 82, 113, 174
Wedel, M., 9, 22
Willis, R.J., 19
Windmeijer, F.A.G., 85
Wittink, D.R., 21, 30, 114, 118
Yatchew, A., 92, 109
Zahavi, J., 23, 24, 26, 174
Zellner, A., 53, 62, 63
Ziemba, W.T., 53
Subject index

AdfoDirect, 3
AID, 16

beta-logistic model, 18, 178
binary choice model, 18, 26, 174
bootstrap, 45, 75

Calyx, 12
CHAID, 16, 26
charity foundation, 43, 115, 179
classification analysis, 19
classification and regression trees, 17
communication elements, 4, 111
Consistent Akaike’s Information Criterion (CAIC), 138
consumers’ purchase behavior, 150, 151
contingency table, 15
Cosslett’s method, 18, 26
cross-selling, 147
cross-validation, 25, 45
curse of dimensionality, 14, 16
cutoff point, 26, 43, 89, 92, 102
database, 2, 81, 147
decision making under uncertainty, 52, 173
decision rule, 51
  Bayesian, 56, 61, 74
  naive, 56
direct marketing campaign, 4, 10
discriminant analysis, 19, 26
DM-product choice, 153, 156
DMSA, 127

estimation
  sample, 25, 45
  uncertainty, 52, 55, 56, 174

follow-up mailing, 19

forward-looking data, 148
fractional factorial design, 117
frequency, 11, 150
fund-raising, 115
gains chart, 25
Gaussian kernel, 39, 154, 161
Gelman-Rubin statistics, 67
Geo-Marktprofiel, 12, 90
Geo-system, 90
Gibbs sampler, 65
habit persistence, 179
Hastings-Metropolis algorithm, 65
heterogeneity, 21, 22, 175, 179
hyperparameters, 56, 64
inclination to respond, 18, 31, 83
independence sampler, 65
information
  individual, 12, 81
  postal code, 12, 82
  proxy, 82
interaction, 17
interpurchase time, 152, 154
intra-class correlation, 94

Jeffreys’ prior, 63–65

Laplace approximation, 62
latent class analysis, 21
leads, 10, 177
lifetime value (LTV), 10, 147, 148, 150, 178
linear probability model (LPM), 17, 26
list, 2, 4
  external, 11, 82
  internal, 11, 81
logit model, 18, 19, 26, 62, 175
long-run profit maximization, 10
loss function, 20, 55
asymmetric, 68, 73
squared-error, 51, 56

mailing design, 115
mailing region, 36, 38, 40, 43, 46, 58, 59
Bayesian optimal, 56, 61
Markov chain Monte Carlo procedures (MCMC), 65, 67
maximum score estimator, 20
media of direct marketing, 2
misclassification, 20
monetary amount, 11
Mosaic, 12, 90

neural network, 22, 27, 174
nonparametric, 13, 16, 38
normal posterior approximation, 62

optimal cutoff index, 84
overfitting, 24, 25
panel data, 178
parametric, 13, 18
Pareto curve, 25
performance, 175
poisson regression model, 22
posterior
density, 55, 57
risk, 56
prior
density, 56
informative, 63
uninformative, 63
privacy, 126
probabilistic segmentation, 22
probit model, 18, 38, 55, 83, 115, 124, 130, 136, 175
censored bivariate, 177
proxy variable, 94
purchase acceleration, 153, 155

questionnaire, 91

recency, 11