Quantitative approaches for profit maximization in direct marketing

van der Scheer, H.R.

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
1998

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
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
Summary

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
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 cr the costs involved if an individual returns the product; D is observed only if R = 1. Thus, the expected profits can be written as
\[ \mathbb{E}(\Pi) = aP(R = 1)P(D = 1 | R = 1) - c - P(R = 1)P(D = 0 | R = 1)c_r. \]
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
Future research

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

$$R^* = x'\beta_1 + \varepsilon_1$$
$$D^* = x'\beta_2 + \varepsilon_2$$

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 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
do not pay. Hence, the process of fulfillment 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

\[
y_{it}^* = \alpha + \rho y_{it-1}^* + x_{it}' \beta + \gamma_i + u_{it}
\]

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 \( y_{it-1}^* \) reflects the learning behavior of the individual (e.g. Bush and Mosteller 1955, Leeflang 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.