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 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,

$$ R^* = x'y + v, $$

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

$$ P(R = 1 \mid x) = \Phi(x'y), $$

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

$$ A = x'y + u, $$

with in particular the assumption

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

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 $y$ 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

\[
P \equiv \Phi(x'\beta) \\
A \equiv x'\gamma,
\]

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; however, is only observed when \( R = 1 \). Hence, the main difference between the TPM and SSM is that the latter assumes \( \mathbb{E}(u \mid R = 1 \text{ or } R = 0, x) = 0 \) instead of \( \mathbb{E}(u \mid R = 1, x) = 0 \). 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

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

(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 distributional 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 heteroscedasticity. 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 \( x \), hence on the implied \( p \) and \( a \). 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 \( a^* (= a^*(p)) \) above which a mailing is sent. We determine \( a^* \) by maximizing the expected profit given \( p \).

So we are interested in

\[
E \equiv E(\Pi \mid p, a \geq a^*)P(a \geq a^* \mid p)
\]

\[
= E(AR - c \mid p, a \geq a^*)P(a \geq a^* \mid p)
\]
\[ E(A| R = 1, a \geq a^*) \] 

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

\[
E \approx E(A | R = 1, a \geq a^*) P(a \geq a^* | R = 1)p \\
- c \left\{ P(a \geq a^* | R = 1)p + P(a \geq a^* | R = 0)(1 - p) \right\} \\
= E(A | R = 1, a \geq a^*) P(a \geq a^* | R = 1)p \\
- c \left\{ (1 - F_1(a^*))p + (1 - F_0(a^*))(1 - p) \right\} \\
= E(a | R = 1, a \geq a^*) P(a \geq a^* | R = 1)p \\
- c \left\{ 1 - F_1(a^*)p - F_0(a^*)(1 - p) \right\} \\
= \int_{a^*}^{\infty} a f_1(a) \, da - c \left\{ 1 - F_1(a^*)p - F_0(a^*)(1 - p) \right\} 
\]

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

\[
E(a - c | R = 1, a \geq a^*) (1 - F_1(a^*)) p - c \left( 1 - F_0(a^*) \right) (1 - p).
\]
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 \( 1 - F_i(a^*) \). 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 \( 1 - F_o(a^*) \). The expected profits of these two groups are \( \mathbb{E}(a - c \mid R = 1, a \geq a^*) \) and \( -c \), respectively. By weighing the expected profits with the probability of occurrence, \( 1 - F_i(a^*) \) and \( 1 - F_o(a^*) \), 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

\[
\frac{\partial E}{\partial a^*} = -a^* f_i(a^*) p + c \left\{ f_i(a^*) p + f_o(a^*)(1 - p) \right\} = 0
\]

or

\[
a^* p = c \left\{ 1 + \left( \frac{f_o(a^*)}{f_i(a^*)} - 1 \right) (1 - p) \right\}.
\]

(3.9)

This is an implicit equation in \( a^* \), which can be solved numerically since the densities \( f_0(\cdot) \) and \( f_i(\cdot) \) 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

\[
\mathcal{M} \equiv \left\{ (p, a) \mid a \geq \frac{c}{p} \left( 1 + \left( \frac{f_o(a)}{f_i(a)} - 1 \right) (1 - p) \right) \right\},
\]

(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

\[ a = \frac{c}{p}, \]  

(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 \(E(\Pi | p, a) = ap - c \geq 0\).

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 \(\bar{\mu} \equiv (\mu_0 + \mu_1)/2\) and \(\delta \equiv \sigma^2(\mu_1 - \mu_0)\). Then

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

so

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

or

\[ a = c \frac{1 + \bar{\mu}\delta(1 - p)}{p + c\delta(1 - p)}. \]  

(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

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

(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 \(a = \tilde{\mu}\) in approximation II. Hence, the first two curves intersect in \((c/\tilde{\mu}, \tilde{\mu})\). As is obvious from (3.13), this is also a point on the third curve if \(\tilde{f}_0(\tilde{\mu}) = \tilde{f}_1(\tilde{\mu})\). 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. \(-\delta(a - \tilde{\mu}) > 0\) in approximation II and \(\tilde{f}_0(a)/\tilde{f}_1(a) > 1\) in approximation III, the mailing region can be written as \(M = \{(p, a) \mid a > \frac{c}{\lambda}\}\), 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 \(a < \tilde{\mu}\). 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 \(a > \tilde{\mu}\).

### 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, \(\tilde{a}\) as \(x'\tilde{\gamma}\) and \(\tilde{p}\) as \(\Phi(x'\tilde{\beta})\). The values of \(\tilde{a}\) 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

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

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

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 $\bar{N}_a$, 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 $\hat{p} \geq p_c = c/\bar{a}$, 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 $\hat{a} \geq a_c = c/\hat{p}$, 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>( p a )</td>
</tr>
<tr>
<td>6</td>
<td>( p a )</td>
</tr>
<tr>
<td>7</td>
<td>( 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 \( M_j, j = 1, \ldots, 7 \). In method 1, the individuals for whom \( p \geq p_c \) should be selected, hence \( M_2, M_4, M_6 \) and \( M_7 \) define the mailing region. In method 2, an individual is selected if \( a \geq a_c \), so the mailing region consists of \( M_1, M_2, M_3 \) and \( M_4 \). 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 \( p = p_c \) is the mailing region of method 1; above the
line \( a = a_c \) the mailing region of method 2.

\[ M_1 \]

\[ M_2 \]

\[ M_3 \]

\[ M_4 \]

\[ M_5 \]

\[ M_6 \]

\[ M_7 \]

\[ 0 \ 0.1 \ 0.3 \ 0.4 \ 0.5 \ 0.6 \ 0.7 \ 0.8 \]

\[ 0 \ 5 \ 10 \ 15 \ 20 \ 25 \ 30 \ 35 \]

Note that the mailing region of method 5 includes the profitable part of
\( M_1 \) and \( M_7 \) and excludes a part of \( M_4 \). Since this method also excludes the
non-profitable regions \( M_3 \) and \( M_6 \), 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

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

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

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

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

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

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

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

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

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 $E(\Pi \ | \ x)$ with respect to $\sigma$:

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

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></td>
<td>12.897</td>
<td>25.063</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(4.03)</td>
<td>(3.75)</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 $\gamma$ and $\beta$, $\mu_0$, $\mu_1$, $\sigma^2$, 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.

### 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
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
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>97</td>
<td></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 \(E(A | R = 1, p) \approx E(A | R = 1)\). 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 assumption 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
Target selection by joint modeling of the probability and quantity of response

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.