University of Groningen

Direct mail selection by joint modeling of the probability and quantity of response

Otter, P.; Scheer, H. van der; Wansbeek, T.

Publication date:
1997

Citation for published version (APA):
Otter, P., Scheer, H. V. D., & Wansbeek, T. (1997). *Direct mail selection by joint modeling of the probability and quantity of response*. s.n.

Direct mail selection by joint modeling of the probability and quantity of response

Pieter Otter   Hiek van der Scheer   Tom Wansbeek*

June 1997

Abstract

We present several methods for profit maximization when households are selected from a mailing list for a direct mail campaign. The response elicited from the campaign can vary over households, as is the case with fund raising or mail order selling. The decisions taken by the household are (a) whether to respond and, in the case of response, (b) the quantity of response, e.g. the sum donated or the monetary amount of the order. We jointly model both aspects of the response and derive a number of profit maximizing selection methods.

We empirically illustrate the methods using a data set from a charitable foundation. It appears that modeling both aspects of the response yields considerably higher profits relative to selection methods that are based on solely modeling the response probability.

1

1.   Introduction

An important topic in direct marketing is *response modeling*. It involves prediction of some measure of customer 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 revenues (given a positive reply), like 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 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 organization raising funds by mail.

The main purpose of response modeling is to rank the potential *targets*, available on a mailing list, from most to least promising, in order to make a selection. Typically, a test mailing is sent to a relatively small sample of the mailing list. Then a response model is built linking observed response behavior to households characteristics. This model is used to predict the response for the remaining households on the mailing list. By eliminating the least promising targets of the mailing list, the direct marketing company can increase their profits. The mailing list is a crucial component in this process since it has to contain sufficient information about the targets like past purchase behavior and geographic, demographic, and lifestyle variables.

In the recent literature a number of papers have been published that consider target selection; see e.g. Banslaben (1992), Bult (1993), Bult and Wansbeek (1995), DeSarbo and Ramaswamy (1994), Magidson (1988). Most of these studies deal with the case of fixed revenues to a positive reply, and hence concentrate on binary choice modeling methods like CHAID, probit, and discriminant analysis; see Bult and Wansbeek (1995) for an overview of target selection techniques. Although this literature recognizes that most direct mail campaigns do not generate simple binary response but rather household specific response, it is hard to find publications that take this aspect into account.

Simon (1987), for example, suggests taking the average amount of purchase from a random sample of the customers on the mailing list over a couple of years, and to use this as the expected value of a potential customer. 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 using an ordinary least squares (OLS) model to determine the expected revenues, and obtaining the total revenue just as

2

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 handle household specific response. Based on their past behavior, households are classified a priori into segments. For each segment a response model is estimated, which is used for the selection of households. The disadvantage of this approach, however, is that it is in general difficult to obtain a satisfactory a priori segmentation.

The purpose of this paper is to present a unified framework for modeling household specific response in order to optimally select households for a mailing campaign. Our framework specifies the relevant decisions taken by the households. 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 and that leads to a method which optimizes expected profit. For reasons of comparison we also consider simplified versions of this method that concentrate on either of the decision dimensions. An empirical application shows considerably higher profits when both decisions are modeled explicitly relative to modeling response probability only.

The paper is structured as follows. In section 2 we present a simple response model that structures both the probability of response and its quantity. This allows us to formulate a profit maximizing selection rule that takes both dimensions into account. Section 3 is devoted to this. Some simplifying approximations to this rule are presented in section 4. Section 5 adds details on practical implementation. To show how the various approaches behave in practice, we describe in section 6 the data underlying the empirical illustration. The results are presented in section 7. Section 8 concludes.

2.   The model

Consider a direct marketing firm that has to make the decision whether to send a household a mailing or not. In case a mailing is sent to a given household, the profit to the firm, Π, is given by

$$\Pi = AR - c, \tag{1}$$

where R is the random variable given by

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

3

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

$$R^* = x'\beta + v, \tag{2}$$

where x is a vector of explanatory variables and $v \sim N(0, 1)$, independently from x; x is assumed to be a random variable with unknown distribution. Whether there is a response or not is indicated by the observed dummy variable R that relates to R^* in the following way: $R = 1$ if $R^* \geq 0$ and $R = 0$ otherwise. Hence the response probability of a household is given by

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

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

$$A = x'\gamma + u, \tag{3}$$

with in particular the assumption

$$E(u \mid R = 1, x) = 0. \tag{4}$$

For convenience of notation we assume the same x in both relations but this is innocuous since elements of γ and β 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 the sequel. This way of modeling probability and corresponding quantity is called a *two-part model* (e.g. 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 large enough that we can neglect the difference between estimators and the true values, hence we in particular assume momentarily that γ and β are known. We define

$$
\begin{aligned}
p &\equiv \Phi(x'\beta) \\
a &\equiv x'\gamma,
\end{aligned}
$$

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.

4

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 list members 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

$$
\begin{aligned}
E & \equiv \mathrm{E}(\Pi \mid p, a \geq a^*) \mathrm{P}(a \geq a^* \mid p) \\
& = \mathrm{E}(AR - c \mid p, a \geq a^*) \mathrm{P}(a \geq a^* \mid p) \\
& = \mathrm{E}(AR - c \mid R = 1, p, a \geq a^*) \mathrm{P}(R = 1 \mid p, a0 \geq a^*) \mathrm{P}(a \geq a^* \mid p) \\
& \quad + \mathrm{E}(AR - c \mid R = 0, p, a \geq a^*) \mathrm{P}(R = 0 \mid p, a \geq a^*) \mathrm{P}(a \geq a^* \mid p) \\
& = \mathrm{E}(A - c \mid R = 1, p, a \geq a^*) \mathrm{P}(a \geq a^* \mid R = 1, p) \mathrm{P}(R = 1 \mid p) \\
& \quad + \mathrm{E}(-c \mid R = 0, p, a \geq a^*) \mathrm{P}(a \geq a^* \mid R = 0, p) \mathrm{P}(R = 0 \mid p) \\
& = \mathrm{E}(A - c \mid R = 1, p, a \geq a^*) \mathrm{P}(a \geq a^* \mid R = 1, p) p \\
& \quad - c \mathrm{P}(a \geq a^* \mid R = 0, p)(1 - p) \\
& = \mathrm{E}(A \mid R = 1, p, a \geq a^*) \mathrm{P}(a \geq a^* \mid R = 1, p) p \\
& \quad - c \{ \mathrm{P}(a \geq a^* \mid R = 1, p) p + \mathrm{P}(a \geq a^* \mid R = 0, p)(1 - p) \}. \quad (5)
\end{aligned}
$$

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 (5) 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 extend. Intuitively, the optimal profit to be obtained may not be affected too much if we omit the conditioning with respect to p. This yields

$$
\begin{aligned}
E & \approx \mathrm{E}(A \mid R = 1, a \geq a^*) \mathrm{P}(a \geq a^* \mid R = 1) p \\
& \quad - c \{ \mathrm{P}(a \geq a^* \mid R = 1) p + \mathrm{P}(a \geq a^* \mid R = 0)(1 - p) \} \\
& = \mathrm{E}(A \mid R = 1, a \geq a^*) \mathrm{P}(a \geq a^* \mid R = 1) p \\
& \quad - c \{ (1 - F_1(a^*)) p + (1 - F_0(a^*))(1 - p) \} \\
& = \mathrm{E}(A \mid R = 1, a \geq a^*) \mathrm{P}(a \geq a^* \mid R = 1) p \\
& \quad - c \{ 1 - F_1(a^*) p - F_0(a^*)(1 - p) \} \\
& = \mathrm{E}(a \mid R = 1, a \geq a^*) \mathrm{P}(a \geq a^* \mid R = 1) p \\
& \quad - c \{ 1 - F_1(a^*) p - F_0(a^*)(1 - p) \}
\end{aligned}
$$

5

$$= \left\{ \int_{a^*}^{\infty} af_1(a)\mathrm{d}a \right\} p - c\left\{ 1 - F_1(a^*)p - F_0(a^*)(1-p) \right\}$$

where the third strict equality is based on (4). The first-order condition 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

$$a^* p = c\left\{ 1 + \left(\frac{f_0(a^*)}{f_1(a^*)} - 1 \right)(1-p) \right\}. \tag{6}$$

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 list members. For simplicity of notation we omit the asterisk superscript to a when further discussing this curve below. The *mailing region*, denoting the households to whom a mailing should be sent, in the (p, a) space, is given by

$$\mathcal{M} \equiv \left\{ (p, a) \mid a \geq \frac{c}{p}\left(1 + \left(\frac{f_0(a)}{f_1(a)} - 1 \right)(1-p) \right) \right\}, \tag{7}$$

which follows directly from (6).

4. Approximations

In order to make (6) operational we distinguish three, increasingly precise but complex approximations to the solution of (6). The first one, further on referred to as I, is to neglect the difference between the two densities. Hence

$$a = \frac{c}{p}.$$

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 σ^2 but with different

6

means, μ_0 and μ_1 in obvious notation. Let $\bar{\mu} \equiv (\mu_0 + \mu_1)/2$ and $\delta \equiv \sigma^{-2}(\mu_1 - \mu_0)$. Then

$$
\begin{aligned}
\frac{f_0(a)}{f_1(a)} - 1 &= \exp\left\{-\frac{1}{2}\left(\frac{a - \mu_0}{\sigma}\right)^2 + \frac{1}{2}\left(\frac{a - \mu_1}{\sigma}\right)^2\right\} - 1 \\
&\approx \frac{1}{2}\left\{\left(\frac{a - \mu_1}{\sigma}\right)^2 - \left(\frac{a - \mu_0}{\sigma}\right)^2\right\} \\
&= -\frac{1}{2}\frac{\mu_1 - \mu_0}{\sigma}\frac{2a - (\mu_1 + \mu_0)}{\sigma} \\
&= -\delta(a - \bar{\mu}),
\end{aligned}
$$

so

$$
ap = c\left\{1 - \delta(a - \bar{\mu})(1 - p)\right\}
$$

or

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

Again, this is an orthogonal hyperbola; for $\delta = 0$, which holds when $\mu_0 = \mu_1$ and when $\sigma^2 \to \infty$, this reduces to the result of the first approximation. The third approximation (III) is obtained by employing a more flexible specification. We will use a nonparametric technique to approximate the two densities.

It is insightful to look at these curves and corresponding mailing regions in somewhat 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 $a = \bar{\mu}$ in approximation II. Hence, the first two curves intersect in $(c/\bar{\mu}, \bar{\mu})$. As is easily seen from (6), this is also a point on the third curve if $f_0(\bar{\mu}) = f_1(\bar{\mu})$. This is the case e.g. when the two densities are isomorphic symmetric, as in the second approximation, but will in general only hold approximately. If $f_0(a)/f_1(a) > 1$, then the mailing region becomes $\mathcal{M} = \left\{(p, a) \mid a > \frac{c}{p}\lambda\right\}$, with $\lambda > 1$. Hence, the mailing region of approximation II or III is smaller than that of approximation I. Consequently, less households should be selected. For approximation II this holds when $a < \bar{\mu}$. When $f_0(a)/f_1(a) < 1$, the opposite holds. Thus, the mailing region of approximation II or III expands with respect to approximation I, which implies that more households should be selected. For approximation II this holds when $a > \bar{\mu}$.

7

5.   Implementation

The three approximations defined in the previous section define three methods to select households from the mailing list. In order to make the methods operational we need estimates of γ and β 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 γ by OLS on (3) using the data on the respondents, and probit on (2) using the results for respondents and nonrespondents. Given these estimates, we impute, for all list members, $\hat{a}$ as $x'\hat{\gamma}$ and $\hat{p}$ as $\Phi(x'\hat{\beta})$.

In order to implement 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

$$\hat{f}_0(a) = \frac{1}{n_0 h} \sum_{i=1}^{n_0} \phi\left(\frac{\hat{a}_{0i} - a}{h}\right)$$

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

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 ω 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 implemented three methods for selection, which are straightforward to use. We consider each list member on its turn to check whether its value $(\hat{p}, \hat{a})$ falls in the mailing region. Of course, the mailing region of the three methods differs. We return to this issue when discussing the empirical example. Before we turn to this example, we want to describe four other methods that we will employ to put the results of the three methods introduced so far in perspective.

The first of these is based on substituting the average response quantity from the respondents in the test mailing, denoted by $\bar{a}$, for $\hat{a}$ for all list members. That is, we neglect the heterogeneity in the response. The selection rule is then based on the simplest hyperbola, i.e., a list member is selected solely according to its value of $\hat{p}$ and takes place if $\hat{p} \geq c/\bar{a}$. This method is interesting since it comes closest to current practice: select if the ratio of cost to (average) yield does not exceed the probability of response. The response probability is modeled but the response quantity not.

The next method has the opposite point of departure and is based on modeling the response amount but not the response probability. The response fraction $\bar{p}$ from the

test mailing is assigned to all list members. A list member is selected $\hat{a} \geq c/\bar{p}$. In other words, we confront $\hat{a}$ with the first approximating curve, the simple orthogonal hyperbola. Two variations of this method are obtained by confronting $\hat{a}$ also with the other two, more sophisticated approximating curves.

Table 7.1 summarizes the seven methods thus obtained. The first column labels methods, the second column has p if the response probability is modeled and has $\bar{p}$ 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. The last two columns of the table will be discussed below.

6. Data

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

The data sample consists of 40 000 observations. All households on the list have donated at least once to the foundation since entry on the mailing list. The dependent variable in (3) is the amount of donation in 1991, and in (2) the response/nonresponse information. The explanatory variables in both models are the amount of money donated in 1990, ditto in 1989, the interaction between these two, the date of entry on the mailing list, family size, own opinion on charitable behavior in general (four categories: donates never, donates sometimes, donates regularly, and donates always). These variables were selected from a database with 58 possible explanatory variables after a preliminary analysis.

The overall response rate $\bar{p}$ is 33.9%, which is rather high but not really surprising since charitable foundations have in general high response rates (Statistical Fact Book 1994-1995), and the mailing list only contains households 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.

7. Empirical results

In order to obtain a robust 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
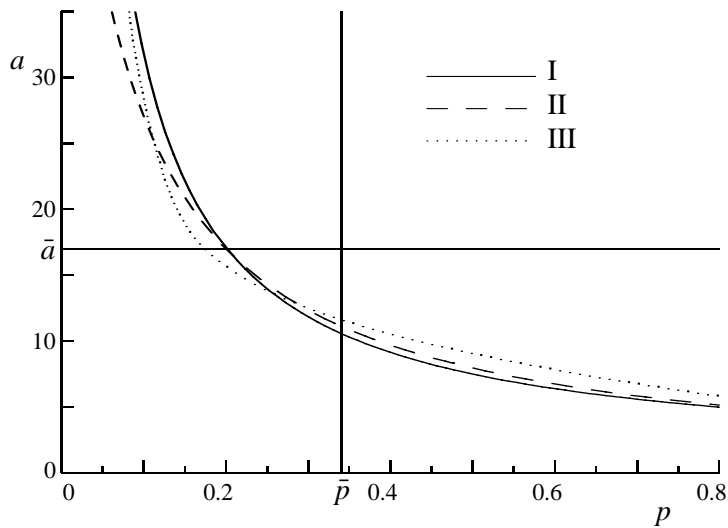
9

Figure 7.1: The mailing regions of the approximations.

observations. This sample can be interpreted as the test mailing. We then draw 39 000 observation, again with replacement, to generate a (bootstrap) validation sample. The estimation sample is used to estimate γ and β (we will not report their estimates since they are not interesting per se for our purpose), 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.

Figure 7.1 depicts the resulting selection rules. It shows the three curves, based on the three approximations, 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 7.1 contains the bottom line results. The last column shows the number of households selected when the various methods are applied. The preceding column
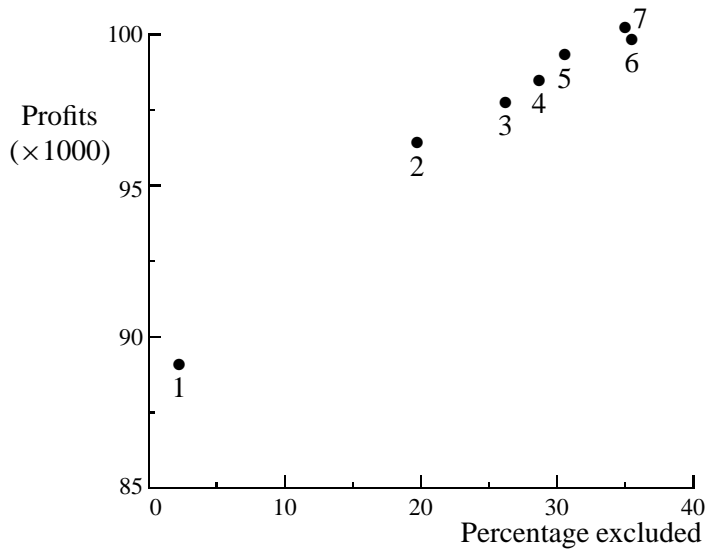
10

Figure 7.2: Profits and percentage excluded from the mailing list for the various approximations.

gives the profit obtained by this selection by considering the amounts actually donated by the selected households. 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 7.2 graphically, using the percentage of households excluded from the mailing list instead of the number selected households. A great gain results from modeling response quantity (methods 2–4), even if only the quantity but not the probability is modeled. A relatively minor but not negligible further gain results from modeling both (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.

A further analysis of the performance of the seven methods relative to each other is given in table 7.2. Figure 7.2 may be too suggestive 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 only the response probability 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.

11

Table 7.1: Performance of methods

method			appr.	profit	#selected
1	p	$\bar{a}$	I	89163	38068
2	$\bar{p}$	a	I	96672	31225
3	$\bar{p}$	a	II	97780	28699
4	$\bar{p}$	a	III	98452	27915
5	p	a	I	99433	27016
6	p	a	II	99924	25215
7	p	a	III	100123	25275

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

There are various limitations to the paper that should be addressed in future work. The results of the empirical illustration are highly evocative, especially the qualitative impression given by figure 7.2. The figure suggests that modeling only response probabilities, the focus of nearly all work in target selection, misses a dominant feature in striving for optimality; the gain to be had when quantities are taken into account is large. This may be an idiosyncratic result, and we do not claim generality.

Table 7.2: Relative performance of methods

	1	2	3	4	5	6
2	97					
3	98	70				
4	98	78	71			
5	100	82	73	63		
6	100	81	75	71	63	
7	100	85	81	76	62	59

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

The example concerns charitable donations, and the picture may be qualitatively different when the proposed methods are applied to the other leading case where response is household specific, money amounts involved in mail order buying.

Another topic for further research is the sophistication in modeling and estimating behavior. Our paper is based on the simple structure (2) and (3), which is moreover estimated by the simplest possible methods. Several more advanced estimation methods could be used as well (e.g., Melenberg and Van Soest 1996). Model (3) has no provision against negative values of A, and a possible correlation between u and v does not play a role although they are likely to correlate; incorporating such a feature could further improve selection. This suggests tobit type model, for example a type-2 tobit model (e.g. Amemiya 1985, and Blundell and Meghir 1987). However, Duan et al. (1983), Hay, Leu and Rohrer (1987), and Manning, Duan and Rogers (1987) show in extensive Monte Carlo studies that the two-part model works very well even if a type-2 tobit model is the true specification.

As a final issue, our approach is limited in the sense that the underlying model is static and does not take behavior over time into account. This issue has two aspects. In the first place, the behavioral model should be improved into a panel data model where a central role is played by the individual effect; household response vis-á-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 by us. We should note, though, that most work on response modeling focuses on static case.

References

Amemiya, T. (1985), *Advanced Econometrics*, Harvard University Press, Cambridge, Massachusetts.

Banslaben, J. (1992), "Predictive modeling", in E.L. Nash, editor, *The Direct Marketing Handbook*, McGraw-Hill, New York, 620–636.

Blundell, R. and C. Meghir (1987), "Bivariate alternatives to the tobit model", *Journal of Econometrics*, **34**, 179–200.

Bult, J.R. (1993), "Semiparametric versus parametric classification models: An application to direct marketing", *Journal of Marketing Research*, **30**, 380–390.

Bult, J.R. and T.J. Wansbeek (1995), "Optimal selection for direct mail", *Marketing Science*, **14**, 378–394.

Bult, J.R. and D.R. Wittink (1996), "Estimating and validating asymmetric heterogeneous loss functions applied to health care fund raising", *International*

Journal of Research in Marketing, **13**, 215–226.

Courtheoux, R.J. (1987), "Database modeling: Maximizing the benefits", *Direct Marketing*, **3**, 44–51.

DeSarbo, W.S. and V. Ramaswamy (1994), "CRISP: Customer response based iterative segmentation procedures for response modeling in direct marketing", *Journal of Direct Marketing*, **8 (3)**, 7–20.

Duan, N., W.G. Manning, C.N. Morris, and J.P. Newhouse (1983), "A comparison of alternative models for the demand for the medical care", *Journal of Business & Economic Statistics*, **1**, 115–126.

Efron, B. (1982), *The jackknife, the bootstrap and other resampling plans*, Society for Industrial and Applied Mathematics, Philadelphia.

Efron, B. and R.J. Tibshirani (1993), *An Introduction to the Bootstrap*, Chapman and Hall, London.

Hay, J.W., R. Leu, and P. Rohrer (1987), "Ordinary least squares and sample-selection models of health-care demand", *Journal of Business & Economic Statistics*, **5**, 499–506.

Magidson, J. (1988), "Improved statistical techniques for response modeling", *Journal of Direct Marketing*, **2 (4)**, 6–18.

Manning, W.G., N. Duan, and W.H. Rogers (1987), "Monte Carlo evidence on the choice between sample selection and two-part models", *Journal of Econometrics*, **35**, 59–82.

Melenberg, B. and A. van Soest (1996), "Parametric and semi-parametric modelling of vacation expenditures", *Journal of Applied Econometrics*, **11**, 59–76.

Rao, V.R. and J.H. Steckel (1995), "Selecting, evaluating, and updating prospects in direct mail marketing", *Journal of Direct Marketing*, **9 (2)**, 20–31.

Silverman, B.W. (1986), *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.

Simon, J.L. (1987), *How to Start and Operate a Mail-Order Business*, McGraw-Hill, New York.

Statistical Fact Book (1994-1995), Direct Marketing Association, Inc., New York, NY.