DERIVING TARGET SELECTION RULES FROM ENDOGENOUSLY SELECTED SAMPLES

BAS DONKERS, JEDID-JAH JONKER,

PHILIP HANS FRANSES AND RICHARD PAAP

ERIM REPORT SERIES *RESEARCH IN MANAGEMENT*	
ERIM Report Series reference number	ERS-2001-68-MKT
Publication	November 2001
Number of pages	38
Email address corresponding author	donkers@few.eur.nl
Address	Erasmus Research Institute of Management (ERIM) Rotterdam School of Management / Faculteit Bedrijfskunde Erasmus Universiteit Rotterdam P.O. Box 1738 3000 DR Rotterdam, The Netherlands Phone: +31 10 408 1182 Fax: +31 10 408 9640 Email: info@erim.eur.nl Internet: www.erim.eur.nl

ERASMUS **R**ESEARCH **I**NSTITUTE OF **M**ANAGEMENT

REPORT SERIES

RESEARCH IN MANAGEMENT

BIBLIOGRAPHIC DATA AND CLASSIFICATIONS		
Abstract	One of the aims of direct marketing in practice is to target the most profitable customers in the database at hand. This selection is often done based on observed behavior in the past. As a consequence, databases arising from the responses to direct mailings are not a random sample from all potential respondents. When not all heterogeneity is observed, part of the target selection rule will be based on the unobserved heterogeneity, so selection is endogenous. Treating an endogenously selected sample as a random sample results in inconsistent parameter estimates, which in general also harms the predictive performance of the model. We develop an adjustment to the likelihood of the model that corrects for the endogenous sample selection. We apply this technique to the selection of mail targets for a charitable organization. In the application we also show that, based on a model for the response rate and the amount donated simultaneously, we can create a target selection rule that maximizes expected revenues. Such a selection rule outperforms selection rules based on response rates or donated amount only. The traditional approach of maximizing response is therefore not the optimal approach to target selection.	
Library of Congress Classification (LCC)	5001-6182	Business
	5410-5417.5	Marketing
	HA 31.2	Sampling: social sciences
Journal of Economic Literature (JEL)	M	Business Administration and Business Economics
	M 31 C 44	Marketing Statistical Decision Theory
	C 19	Econometric and Statistical Methods: Other
European Business Schools Library Group (EBSLG)	85 A	Business General
	280 G 255 A	Managing the marketing function Decision theory (general)
	250 D	Statistical Analysis
Gemeenschappelijke Onderwerpsontsluiting (GOO)		
Classification GOO	85.00	Bedrijfskunde, Organisatiekunde: algemeen
	85.40 85.03	Marketing Methoden en technieken, operations research
	83.03	Methoden en technieken
Keywords GOO	Bedrijfskunde / Bedrijfseconomie	
	Marketing / Besliskunde	
	Steekproeven, direct marketing, econometrische modellen	
Free keywords	Direct marketing, econometric models, sample selection, target selection, endogeneity	

Deriving Target Selection Rules from Endogenously Selected Samples*

Bas Donkers†
Department of Marketing and Organization
Erasmus University Rotterdam

Jedid-Jah Jonker‡
Social and Cultural Planning Office

Philip Hans Franses§
Department of Marketing and Organization
Econometric Institute
Erasmus University Rotterdam

Richard Paap¶
RIBES
Erasmus University Rotterdam

*The authors gratefully acknowledge a Dutch charitable organization for providing the data used in this study. We also thank Michel Wedel and Peter Verhoef and participants at the Marketing Science 2000 Conference in Los Angeles, the Marketing Science 2001 Conference in Wiesbaden, the Econometric Society European Meeting 2001 in Lausanne and the Advanced Research Techniques Forum 2001 in Amelia Island for many helpful comments. The last author thanks the Netherlands Organization for Scientific Research (N.W.O.) for its financial support. This research has been conducted while the second author was working at the Erasmus University Rotterdam.

†Corresponding author: Erasmus University Rotterdam, Marketing and Organization, Office H15-12, P.O. Box 1738, NL-3000 DR Rotterdam, The Netherlands, Phone: (+31) 10 40 82411, Fax: (+31) 10 40 89109, E-mail: `donkers@few.eur.nl`

‡Social and Cultural Planning Office (SCP), P.O. Box 16164, NL-2500 BD The Hague, The Netherlands, E-mail: `j.jonker@scp.nl`

§Erasmus University Rotterdam, Econometric Institute, Office H11-34, P.O. Box 1738, NL-3000 DR Rotterdam, The Netherlands, E-mail: `franses@few.eur.nl`

¶Erasmus University Rotterdam, RIBES, Office H16-17, P.O. Box 1738, NL-3000 DR Rotterdam, The Netherlands, E-mail: `paap@few.eur.nl`

Modeling Target Selection

Abstract

One of the aims of direct marketing in practice is to target the most profitable customers in the database at hand. This selection is often done based on observed behavior in the past. As a consequence, databases arising from the responses to direct mailings are not a random sample from all potential respondents. When not all heterogeneity is observed, part of the target selection rule will be based on the unobserved heterogeneity, so selection is endogenous. Treating an endogenously selected sample as a random sample results in inconsistent parameter estimates, which in general also harms the predictive performance of the model. We develop an adjustment to the likelihood of the model that corrects for the endogenous sample selection. We apply this technique to the selection of mail targets for a charitable organization. In the application we also show that, based on a model for the response rate and the amount donated simultaneously, we can create a target selection rule that maximizes expected revenues. Such a selection rule outperforms selection rules based on response rates or donated amount only. The traditional approach of maximizing response is therefore not the optimal approach to target selection.

Key words: direct marketing, econometric models, sample selection, target selection, endogeneity.

1 Introduction

An important aspect of direct marketing involves the selection of target individuals or households. This is usually called target selection. Practical rules for target selection are usually derived from available databases. For example, insurance firms may keep records of the past observed behavior of their clients, which may range from requests for information, purchases of new products to actual claims. In addition, these firms can once in a while survey a subset of their clients to keep track of their attitudes. Similar databases have been (or are being) built up by investment firms, banks and also by retailers (based upon loyalty cards and bonus passes). The value of these databases for these firms is large, see Roberts and Berger (1999), among others. A common property of the databases is that firms have a substantial amount of potentially useful information about their clients. These firms know who these clients are, that is, where they live, what their income is, their family size, and so on, but often more importantly, they know what their clients have done over time. These variables usually contain what is called RFM information, that is, information on Recency, Frequency and Monetary Value of transactions. For many firms, information on past behavior is the only information they have about their customers, except for variables like the address and year of birth.

There are two important reasons to have such a database. First, it allows for target selection or segmentation to finetune marketing activities. When making a selection for a direct mailing or another type of direct marketing activity, a firm may focus on maximizing response rates, but it may also focus on profit maximization, see for example Bult and Wansbeek (1995), Gönül and Shi (1998) and Bitran and Mondschein (1996). Second, a customer database with observed behavior allows one to better understand customer behavior. Such an understanding would then help, for example, to design new products for specific groups of clients or to design better marketing strategies involving promotion and advertising. Hence, only a few customers of an insurance firm get an offer for a new type of insurance, and only a few customers of a retailer get a reduced price for specific products.

The analysis of such multivariate data is often best performed with econometric (statistical) models. Such a model may correlate observed behavior concerning, say, purchases, with covariates like past purchase behavior, the time since the last purchase, and shopping frequency, but also with household size and household income. For various reasons it would be an ideal situation when the model that is used to better understand the behavior of individuals can also be used for target selection. One reason is that such a model can be updated any time new information comes along, and hence, that one can increase knowledge over time. Another reason is that it can save effort on implementation as parts of the modeling process can perhaps be automated. Finally, when previous target selection rules did not yield success, one can try and see if the model that describes the observed behavior needs some modification. In this paper we will put forward such an econometric model, which can be used for inference as well as for target selection. To narrow focus, we build such a model for a charitable institution, which has a database containing RFM information. The extension of our model to include observed individual characteristics is straightforward.

However, the econometric analysis of databases containing information on the responses to direct mailing campaigns is not straightforward. As the aim of direct marketing is to send mailings to the most interesting customers in the database, the observations that are added to the database are not randomly selected. For example, individuals who have responded frequently in the past, might receive more mailings then others. As an extreme situation, there may be individuals, who do not receive a mailing at all, and thus are unlikely to donate by definition.

Such target selection rules are harmless when one is able to identify all systematic variation in behavior among respondents, as then the selection is based on exogenous variables. In real life situations, we are however often not able to identify all systematic variation and we are left with unobserved heterogeneity. Since selection is based on past behavior, and therefore depends on the unobserved heterogeneity, we have to deal with an endogenously selected sample of observations. Treating an endogenously selected sample as a random sample results in inconsistent parameter estimates. This leads to a wrong picture when we describe behavior, but more importantly it might also harm the target selection based on the model, resulting in a loss in the revenues that are to be generated by the direct mailing.

Previous models for target selection in the literature have addressed a number of important issues. Bult and Wansbeek (1995) developed a model that determines an optimal selection. However, their (one period) selection is only based on response. We want to examine whether the performance of selection rules can be improved by also modeling the size of the expenditure. Bitran and Mondschein (1996) use a Markov Chain model to determine the selection in a dynamic (or multi-period) environment. A drawback of their model is the rapidly expanding size of the state space, requiring a large amount of data to estimate the model. Gönül and Shi (1998) also develop a multi-period model, where they model both the decisions of the direct mailer and of the customer. In their model they have taken endogeneity into account, but they do not account for the endogenous sample selection, resulting from the partly unobserved heterogeneity. As we will show, neglecting this endogenous sample selection leads to inappropriate estimates. Furthermore, the elegant structural approach of Gönül and Shi (1998) is unfortunately not feasible in our application, as we cannot rely on a sensible specification for the utility of donating money to charity.

We extend the literature on target selection on two topics, that only partially have been dealt with before. They concern endogeneity and sample selection. Endogeneity is a problem, as the probability of receiving a mailing depends on the number of mailings received in the past through the RFM variables. Sampling selection concerns the fact that individuals who have received more mailings are also observed more often. As receiving a mailing depends on both observed and unobserved heterogeneity, the distribution of the unobserved heterogeneity in the sample is influenced by the mailing strategy. As we will show, an adjustment to the likelihood deals with both the sample selection and the endogeneity problem. However, this adjustment is based on the probability of receiving a mailing as a function of the unobserved heterogeneity. When the unobserved heterogeneity is modeled with a discrete mixture model, the correction is obtained straightforward, as the probability of receiving a mailing can be estimated for each mixture component.

In sum, in this paper we will put forward an econometric model, which can be used to obtain insights into the behavior of individuals as well as for target selection. When analyzing this model, we pay explicit attention to the notion that the data are unlikely to be random, as they are based on the same (or a similar) model used in a previous round. Hence, we focus on target selection, based on a model for response and for monetary value, taking target selection in the past into account.

The outline of our paper is as follows. In Section 2, we present a model that simultaneously describes response behavior to direct mailings and the amount donated in case of response. We then continue with a detailed discussion of the consequences of non-random sampling for the analysis of single mailshots and of sequences of mailshots. The final part of Section 2 presents the estimation strategy needed to obtain consistent estimates of the behavioral model. In Section 3, we present the details concerning the empirical specification of the econometric model, and we discuss the prediction of behavior given the estimates of the parameters. In Section 4, we apply our model to a (random) sample of 900 individuals from a database of a Dutch charitable institution. For illustrative purposes, we show how the empirical results change in case one neglects the fact that the data do not constitute a random sample. Finally, in Section 5, we conclude with a discussion on managerial implications, on limitations to our approach and on topics for further research.

2 Sampling issues and endogeneity

In this section we discuss aspects of data collection and target selection, which are important for a proper analysis of the data with an econometric model. We start with the presentation of a model for the response behavior towards a direct mailing. We then discuss the consequences of various sampling schemes for model estimation, identification and interpretation. As mailing strategies based on selection rules generally do not imply sending the same number of mailings to each individual and the number of mailings received is not independent of behavior, the usual estimation method for latent class models cannot be used. Therefore, we adapt this method, such that it provides consistent estimates for a sample where some individuals are observed more often than others, as they receive more mailings.

2.1 A model for response behavior

In this section we present a general model that describes response behavior of individuals to direct mailings. We consider a general case, even though our application focuses on a charitable organization with limited database sources. The issues we raise in this paper regarding sample selection and endogeneity are relevant for any kind of latent class model. In our empirical model we distinguish between two types of explanatory variables. The first type, denoted by x_{it}, measures observed past behavior at time t for individual i, like the number of times an individual has responded to a direct mailing in the last year, or the average amount he or she has spent in the last year, in other words, RFM variables. For further reference we label these as behavioral variables. The second type of variable, denoted by z_{it}, measures characteristics of individual i, like family size, income, or highest education level. In general, such individual-specific characteristics are approximately constant over a short range of time. Obviously, the variables in z_{it} have some explanatory value for the behavioral variables in x_{it}. It is however unlikely that individual characteristics explain behavior completely and in many cases it turns out that past behavior is a better predictor of future behavior than individual characteristics. It is therefore natural to set up an econometric model with both x_{it} and z_{it} as explanatory variables for observed behavior. As it turns out, it is crucial for a modeling strategy to know whether the explanatory variables explain all the systematic variation across the individuals, or whether they do not. Indeed, if not all systematic variation is explained, one has to incorporate unobserved heterogeneity into the model.

Our model for inference and target selection, discussed below in more detail, is based on the following general specification,

$$r_{it}^* = \beta_{1i} x_{it} + \delta_1 z_{it} + \varepsilon_{1it}, \tag{1}$$

$$r_{it} = \begin{cases} 1 & \text{if } r_{it}^* \geq 0 \\ 0 & \text{if } r_{it} < 0, \end{cases} \tag{2}$$

and

$$\ln y_{it} = \beta_{2i} x_{it} + \delta_2 z_{it} + \varepsilon_{2it}, \tag{3}$$

where equations (1) and (2) constitute a binary choice model for response ($r_{it} = 1$) or no response ($r_{it} = 0$) of individual i to a mailing at time t. Of course, parameters and variables are summarized in vectors. Equation (3) is a standard regression model for the monetary value involved. The model for the response of individual i correlates the observed response r_{it} to the observed variables x_{it} and z_{it}, through the unobserved components in the response model, β_{1i} and ε_{1it}. We refer to β_{1i} as the individual-specific parameter in the response equation, which is constant over time, and to ε_{1it} as the random error, with realizations which are mutually independent across individuals and over time. Notice that the individual-specific parameter can be modeled as either a random or a fixed parameter, so β_{1i} can also be a random variable. In our final model in Section 3.1, we assume that ε_{1it} follows a standard normal distribution. Hence, conditional on β_{1i}, the response of an individual is modeled by a probit model.

The amount of money donated by individual i, y_{it}, can be described using a traditional loglinear regression model as in equation (3). The monetary value depends on past behavior through x_{it}, on individual characteristics through z_{it}, and it also depends on the individual-specific parameter β_{2i}. The remaining random error, ε_{2it}, is uncorrelated with x_{it}, z_{it}, β_{1i}, and β_{2i}, but is allowed to be correlated with ε_{1it} in the probit model.

Similarly, the individual-specific parameters in the response model and the monetary value model can also be correlated. An example where this occurs amounts to a mailing aimed at a population that consists of two subpopulations that only differ in the intercepts in β_{1i} and β_{2i}. Individuals in the first subpopulation frequently spend a smaller amount, so the intercept in β_{1i} is high, while it is low for β_{2i}. Additionally, individuals in the second subpopulation are infrequent responders, but when they respond, the amount spent is high, so the intercept in β_{1i} is low, while it is high for β_{2i}, thereby resulting in a negative correlation between the individual-specific intercepts in the two equations.

The interpretation of the parameters in a latent class model is usually not straightforward. This is due to the fact that the parameter estimates for the effects of the observed variables x_{it} and z_{it} have to be interpreted conditional on the individual-specific parameter. This is perhaps best illustrated using an example. Consider the situation in which there is considerable unobserved heterogeneity in the amount of money spent, that is, there is substantial variation in β_{2i}. Moreover, let there be a small negative effect of the amount spent in the previous mailing on the amount spent in the current mailing for each individual. At first sight, one would expect a negative sample correlation between past and current monetary values. However, this conclusion does not necessarily hold when there is a large variation in the individual-specific effects. It may well be that some donors always give large amounts, although once a bit more and next time a bit less. The same might hold for donors with small amounts. In this case, the persistent individual effect will result in an overall positive correlation between the size of past and current donations, even though conditional on the individual effect there is a negative correlation. In general, interpretation of the individual parameter estimates is difficult and the model seems best be interpreted using simulations of the individual behavior implied by the model.

2.2 Sampling issues

The analysis of the above model depends heavily on the way in which the sample of observations is collected. For example, when information is available for a single random mailing, the parameters δ_1 and δ_2 in (1) and (3) can be identified by relying on additional assumptions about the individual effects. Obviously, without further assumptions, it is not possible to identify β_1 and β_2 as x_{it} is not observed. When the sample does not concern random individuals, and the selection rule is generally unknown but it is known that it does not depend on z_{it} only, it is in general not possible to identify the parameters. Notice that such a situation often arises when one buys a mailing list from an outside list provider.

However, when more information is available about individual behavior through observations on a series of mailings, it turns out that the parameters can be identified. As past behavior is included in the model as an explanatory variable, one should now be very careful with possible endogeneity problems. In this subsection we first discuss the situation of having information on a single mailing. We next consider multiple mailings.

A single mailing

In the situation where there is only data for a single mailing, two aspects of the general model presented above change. First, there is no information on past behavior, so one does not observe x_{it}. Second, as there is only one observation for each individual, it is impossible to disentangle the individual effects from the random errors in (1) and (3). In general, one then typically does not explicitly specify the individual effects in this situation. Individual effects, however, are relevant when analyzing non-random samples of the population of interest as we illustrate next.

Random sampling from data on a single mailing

Consider the situation where a mailing has been sent to individuals who were drawn randomly from the population. Note again that we are interested in the relationship between the observed characteristics z_{it} and observed behavior r_{it} and y_{it} for the individuals in this population. Given the randomness of the sample, this is a standard textbook situation and Maximum Likelihood [ML] estimation gives consistent estimates. If the individual effects and random errors are assumed to be independent, the parameters of the model for the monetary value can be estimated separately using Ordinary Least Squares [OLS]. To our knowledge, this situation does not often occur in practice as target selection rules are commonly used to select the individuals who receive a mailing.

Non-random sampling from data on a single mailing

In case the sample of individuals receiving a mailing is not a random draw from the population of interest, two situations can occur. The difference between these two situations is set by the potential knowledge of the variables used for target selection and also whether these are components of z_{it} or not. The simplest situation occurs when the variables used for selection are known and included in the model equations. In this situation the residuals and individual effects still constitute a random sample from their population distributions, conditional on the variables used for selection. Hence, using ML or OLS results in consistent estimates, although the standard errors of the parameter estimates have to be adjusted for the weighting scheme when sampling selectively. The main issue to understand here is that non-random sampling based on explanatory variables does not result in inconsistent parameter estimates.

This consistency result does however not hold when the sampling mechanism is unknown. In this situation it is possible that the sampling mechanism selects individuals based on the unobserved individual effect (like for example income, in case it has not been observed), resulting in a distribution of the individual effects in the sample of observations, which differs from the distribution in the population of interest. In this case, standard ML or OLS procedures do not result in consistent parameter estimates. To obtain consistent estimates of the parameters of interest, one has to make assumptions about the distribution of the individual effects and the random errors, conditional on the individual being selected. Such assumptions do not arise naturally.[1]

Multiple mailings

Interestingly, when we have information on individuals' responses to more than one mailing, identification of the individual effects and consequently of all other parameters becomes feasible. The estimation procedure depends on the target selection strategy and also on the assumptions about the individual effects.[2] We next discuss the estimation strategy in the (in practice unlikely) event of a random mailing strategy. This is then again generalized to the more relevant case of non-random sampling.

---

[1]At first sight, a Heckman-type selection model comes to mind, but, in case of a target selection model, no plausible exclusion restrictions exist and identification is then based on the assumed normality of the errors. Although it is a frequently used model, normality is however not a natural assumption to obtain identification.

[2]Modeling the individual-specific effects with a latent class model makes estimation much simpler than a model with a continuous distribution for the individual-specific effects. See Wedel *et al.* (1999) for a discussion of modeling strategies of unobserved heterogeneity.

Random sampling from data on multiple mailings

Estimation of the model with a random sample of observations seems straightforward, at least at the first sight. However, some caution has to be exercised with respect to modeling the individual effects. When a single mailing is analyzed, the observed variables are exogenous. When multiple mailings are analyzed with RFM-type variables as explanatory variables, these variables are not independent of the (possibly unobserved) individual-specific effects, as these individual effects are likely to also have influenced behavior in the past. Including the individual effects in the random error is therefore not allowed and a correct specification of the individual effects is needed. Consistent parameter estimates can be obtained from ML estimation when the individual-specific effects are modeled appropriately, where both a random-effects model and a fixed-effects model can be used. As already discussed above, a target selection rule based on exogenous individual characteristics, when these variables are also incorporated in the model equations, does not change the distribution of the individual effects or the random errors. Hence such a selected sample can be treated as if it were a random sample (of the unobserved components in the model). It should be stressed here that RFM variables are endogenous by nature and that a target selection strategy using RFM variables thus results in a non-random sample of the individual effects.

Non-random sampling from data on multiple mailings

The last situation that we treat in this section is the situation that seems to be most frequently encountered in practice. This concerns the case where one has access to response behavior on a sequence of mailings, where target selection has been based on some aspects of past behavior. We have already indicated that obtaining consistent estimates of the parameters of interest is difficult when only information on the response is available for a single mailing and selection has been based partly on the individual effects. In contrast, when information about multiple mailings is available, the individual effects are identifiable and it becomes feasible to obtain consistent estimates. However, we then still have the problem of endogeneity of the RFM variables. As for the random sample, correct modeling of the individual-specific effects circumvents such endogeneity problems. Fortunately, as we will show next, endogeneity can be seen as a sample selection issue, and hence it can again be related to a misspecification of the distribution of individual effects.

The most robust method for modeling individual-specific effects is a fixed-effects model, where for each individual a separate individual effect is modeled. However, in many practical marketing situations one does not have a large enough number of observations for each individual and the number of parameters in a fixed-effects model in general will be too large, see for example Rossi and Allenby (1993) for a discussion on this matter. Modeling the individual-specific effects as a random effect, which is distributed in the population according to some parameterized distribution, is then a frequently used solution. Two broad classes of distributions can be distinguished in the literature. The first class of distributions involves continuous distributions, such that the individual effect can take a large (infinite) number of possible values, see for example Gönül and Srinivasan (1993). The second class of distributions is discrete in nature. The model with a discrete distribution has received two interpretations in the literature. The first one is a latent class model where it is assumed that there are actually a finite number of support points, the latent classes, for the discrete distribution, see Kamakura and Russell (1989) or Wedel and Kamakura (1999). The second interpretation assumes that there is some unknown, possibly continuous, distribution of the individual effects in the population. This unknown distribution is then approximated with a finite number of discrete points. The approximation of the unknown population distribution improves with the size of the sample, as the number of support points increases with sample size. In this interpretation, the model is a semi-parametric model, see Heckman and Singer (1984).

When the sample is not a random draw from the population of interest, modeling the unobserved heterogeneity with a discrete distribution of the individual effects has large advantages over a model with a continuous distribution. The basic advantage originates from the fact that one does not necessarily obtain information on all individuals in the population.[3] And, getting information about the sample distribution of a non-random sample of a continuous distribution is known to be very difficult. When a latent class model is assumed for the individual effects, the latent classes do not change by non-random sampling from this population. The only consequence of non-random sampling from the various latent classes is that the probabilities with which each latent class occurs in the population are not identified. It is even possible that a latent class is never selected, so it gets assigned a probability mass of zero. When the discrete distribution is interpreted as an approximation to the unknown distribution, the estimated approximation will be an approximation of the sample distribution and not of the population distribution. Fortunately, this is sufficient as only information about the sample population of the individual effects is needed for parameter estimation.

2.3 Model estimation on a selected sample

Our general conclusion from the discussion in the previous subsection is that response behavior to direct mailings that are sent out using some target selection mechanism is done best with information on multiple mailings. Moreover, it is important that individual heterogeneity is taken into account and the simplest way to do this is by using a latent class model. There is now one further issue that needs to be solved. Current estimation techniques for latent class models cannot be applied to a panel with different numbers of observations for each individual, what is usually called an unbalanced panel. For standard regression models, weighting methods have been developed to correct for this, see Fitzgerald *et al.* (1998) and Ter Horst *et al.* (2001). For latent class models, the approaches are not feasible, as it is not possible to identify or to know on beforehand the latent class for each observation, when the weights still have to be determined.

A natural approach may now be to incorporate the mailing decision into the likelihood. A simple example in the Appendix shows that this leads to consistent estimates of the segment probabilities, conditional on the assumption that the other parameters

[3] When each member of the population is mailed at least once, the whole population is observed and the observed sample of individuals is again a random sample of the population.

are estimated consistently. As some segments are observed more often than others, it is likely that only the segment probabilities are influenced by this partial observability. The full likelihood to be maximized with N observations, T mailings, and S latent classes, is therefore

$$\prod_{i=1}^{N}\sum_{s=1}^{S}p_s\prod_{t=1}^{T}L(r_{it},y_{it};s)^{I_{it}}q_s^{I_{it}}(1-q_s)^{1-I_{it}}, \tag{4}$$

where I_{it} is an indicator function with value 1 if individual i received a mailing at time t and 0 otherwise, and where $L(r_{it}, y_{it}; s)$ denotes the likelihood of observing (r_{it}, y_{it}) when the individual is in latent class s. This likelihood will be discussed in more detail in the next section. Finally, q_s is the probability that an individual in latent class s receives a mailing. When $I_{it} = 0$, $L(r_{it}, y_{it}; s)^{I_{it}}$ equals 1. Notice that the probabilities to receive a mailing serve as weights in the likelihood function. As the weights depend on the latent classes, it is not feasible to estimate them a priori. It should be mentioned that the adjustment to the likelihood for the mailing strategy depends critically on the assumption of a latent class model. When a continuous mixture model is used one would have to estimate the probability of receiving a mail for each possible value of the unobserved heterogeneity. Although one could specify and estimate some parametric specification, we do not opt for this, as there is no guidance with respect to the correct specification. And, misspecification results in inconsistency of the parameters of interest. In the next section we will provide more details on the model.

3 Empirical model

To illustrate the proposed estimation strategy we apply it to the observed behavior of respondents to the mailings of a charitable organization. In Section 3.1 we specify in more detail our econometric model. Section 3.2 discusses out-of-sample prediction for donations, based on the parameter estimates. Such predictions can be used for future target selection.

3.1 Model specification

In Section 2.1 we presented the general model of interest, which deals with individuals' response behavior and the amounts being donated. Since we do not have information on individual characteristics in our sample (see Section 4), we exclude z_{it} from the model equations below. We assume that ε_{1it} and ε_{2it} are distributed with a normal distribution with covariance matrix $\Sigma = \begin{pmatrix} 1 & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}$, so $(\varepsilon_{1it}, \varepsilon_{2it}) \sim \mathrm{N}(0, \Sigma)$.

The model we propose is a type-2 Tobit model, see Amemiya (1985, p. 385–386). The likelihood contributions for this model are presented here for completeness, see Amemiya (1985), that is,

$$L(r_{it}, \ln y_{it}; \beta_i, \Sigma) = (\Phi(-\beta_{1i}x_{it}))^{1-r_{it}} \left(\Phi(e_1) \frac{1}{\sigma_{22}} \phi(e_2) \right)^{r_{it}} \tag{5}$$

with $e_1 = (\beta_{1i}x_{it} + \frac{\sigma_{12}}{\sigma_{22}^2}(\ln y_{it} - \beta_{2i}x_{it}))/\sqrt{(1 - \sigma_{12}^2\sigma_{22}^{-2})}$ and $e_2 = \frac{\ln y_{it} - \beta_{2i}x_{it}}{\sigma_{22}}$. These likelihood contributions[4] are conditional on the unobserved heterogeneity in the parameter, $\beta_i \equiv (\beta_{1i}, \beta_{2i})$. As will generally be the case, there are not enough observations for each individual to identify β_i for each individual. Therefore we have to make distributional assumptions about the unobserved heterogeneity. Wedel *et al.* (1999) argue that the choice between a discrete and a continuous distribution for the individual specific parameters is difficult and often a matter of convenience. It should be mentioned though that we have concluded in Section 2 that parameter estimation of a model based on a non-random sample is straightforward when a discrete distribution is used, while this is not the case with a continuous distribution. We therefore assume a discrete distribution for the individual specific effects in our model. More specifically, we follow the latent class interpretation, and so the distribution has a finite but unknown number of mass points.

Assume for the moment that we know the number of latent classes in our sample, say S. In this case we have to estimate S individual-specific parameter vectors, $\beta^1, \dots, \beta^S$. The sample probabilities of these classes are denoted by $p_1, \dots, p_S$, with $\sum_{s=1}^{S} p_s = 1$. These

[4] Our true interest lies in the response behavior of the individuals. The mailing strategy is not of primary concern, but it is modeled to obtain consistent estimates for the model of interest, as we have discussed in Section 2.

probabilities also need to be estimated. Notice here that there might exist more latent classes in the population and that the sample probabilities, which have to be estimated, do not necessarily correspond to the population probabilities.

The number of latent classes in the sampled population, is unknown. Unfortunately, statistical testing for the number of latent classes is not possible as under the restriction $\beta^s = \beta^l$ for $l \neq s$, one of the mixing proportions p_s is not identified. This phenomenon is known as the Davies (1977) problem. Therefore, to determine the number of latent classes S we increase the number of latent classes until the Schwarz/Bayesian Information Criterion [BIC] does not decrease anymore, see Jain *et al.* (1994) for a similar and successful approach to brand choice models.

As we have discussed in Section 2.3, estimation of latent class models is not straightforward when individuals in some latent classes receive more mailings than individuals in other classes. In our data there is a substantial amount of variation in the number of mailings received by the individuals. As the charitable organization (which has provided us with the data) sends its mailings using an RFM-based selection mechanism, it is likely that the number of mailings received also depends on the latent classes. We therefore have to model the mailing process too.

3.2 Prediction

Modeling response behavior of individuals is usually done in order to predict responses to future mailings. To generate an out-of-sample forecast for each individual, we need the explanatory variables, x_{it}, but also we have to take the unobserved heterogeneity into account. This can be done using the estimated sample distribution of the latent classes. However, this does not seem to be the best possible predictor, as we can also learn about the individual specific effects from past behavior. Therefore, we first discuss how we can update our knowledge of unobserved heterogeneity through Bayesian learning or Bayesian updating. Finally, this subsection closes with a discussion on obtaining predictions of individual behavior, given observed past behavior.

Bayesian Learning

When we estimate the model we also obtain estimates of the sample probabilities of the latent classes, $\hat{p}_s, s = 1, \ldots, S$. These probabilities provide estimates of the probabilities that an individual is member of a certain class. When we do not have additional information on an individual, these probabilities can be used for prediction. However, when we also have information on past behavior, this information can be used to update the information on class membership. This updating is done through Bayesian learning. Let $p_s(i,t)$ denote the updated probability that individual i is in class s. When no information is available, $(t = 0)$, these probabilities are the estimates for the whole sample, that is, $p_s(i,0) = \hat{p}_s$. Adding the information from observations, these probabilities can be updated according to

$$p_s(i,t) = \frac{p_s(i,t-1)L(r_{it}, \ln y_{it}; \hat{\beta}^s, \hat{\Sigma})}{\sum_{k=1}^{S} p_k(i,t-1)L(r_{it}, \ln y_{it}; \hat{\beta}^k, \hat{\Sigma})}, \tag{6}$$

for $i = 1, \ldots, I$ and $t = 1, \ldots, T_i$, where $L(r_{it}, \ln y_{it}; \hat{\beta}^k, \hat{\Sigma})$ is given in (5). In the ideal case one eventually learns to which latent class an individual belongs, that is, for large t $p_s(i,t)$ becomes 1 for the class of which the individual is a member and zero for the other classes.

Forecasting

The estimated model can be used to forecast future response behavior, and hence can be used for target selection. First we consider the probability that an individual i, who belongs to latent class s and has RFM characteristics x_{i,T_i+1}, responds to a mailing at time $T_i + 1$, that is,

$$\Pr[r_{i,T_i+1} = 1; \hat{\beta}^s, \hat{\Sigma}] = \Pr[\varepsilon_{i,T_i+1,1} \geq -\hat{\beta}_1^s x_{i,T_i+1}] = \Phi(\hat{\beta}_1^s x_{i,T_i+1}). \tag{7}$$

Given a response to this mailing, the expected donation of individual i, which belongs to class s with characteristics x_{i,T_i+1} can be shown to equal

$$\mathrm{E}[y_{i,T_i+1} | r_{i,T_i+1} = 1; \hat{\beta}^s, \hat{\Sigma}] = \exp(\hat{\beta}_2^s x_{i,T_i+1} + \frac{1}{2}\hat{\sigma}_{22}^2) \frac{(1 - \Phi(-\hat{\beta}_1^s x_{i,T_i+1} - \hat{\sigma}_{12}))}{1 - \Phi(-\hat{\beta}_1^s x_{i,T_i+1})}. \tag{8}$$

The expected donation that is to be received from an individual to which a mailing has been sent can now be obtained from multiplying the probability to response (7) with the expected donation given response (8).

The predictions of response behavior presented above all condition on knowing the latent class of the individual. As we do not know this latent class, we have to weight the predictions derived above with the probabilities with which the individual is a member of each class. Therefore, the prediction for the probability with which an individual responds to the mailing is in fact

$$\Pr[r_{i,T_i+1} = 1; \{\hat{\beta}^s, s = 1, \ldots, S\}, \hat{\Sigma}] = \sum_{s=1}^{S} p_s(i, T_i) \Pr[r_{i,T_i+1} = 1; \hat{\beta}^s, \hat{\Sigma}]. \tag{9}$$

Similarly, one can obtain predictions for the expected amount of money received, conditional on response or not. The predicted response probabilities and expected donations can now be used by the charitable organization to select which individuals from their database should receive a mailing in a next mailshot. In practice, the charitable organization faces costs that have to be taken into account in their selection rule. The charitable organization faces two types of costs, that is, the mailing cost c_m (letter and stamp) and the costs of cashing cheques c_c. The optimal selection rule will only select those individuals for whom the net unconditional expected donation (given characteristics x_{i,T_i+1}) exceeds zero, that is,

$$\mathrm{E}[y_{i,T_i+1}; \hat{\theta}] - c_m - c_c \Pr[r_{i,T_i+1} = 1; \hat{\theta}] > 0, \tag{10}$$

where $\Pr[r_{i,T_i+1} = 1; \hat{\theta}]$ is defined in (9) and $\mathrm{E}[y_{i,T_i+1}; \hat{\theta}]$ is defined similarly. The latter term corresponds to the expected costs of cashing a cheque.

4 Application to Dutch Charity Donations

We now have the complete specification of the econometric model and the tools to predict future behavior, based on the parameter estimates of the model. In this section our focus will be on estimating a model based on multiple mailings from a Dutch charitable organization, which enables us to estimate the individual-specific parameters in the model.

In Section 4.1, we describe the data. In Section 4.2, we present the estimation results and the performance of target selection rules based on our model and we compare it with the target selection rule of the charitable organization and with target selection rules based on alternative, though basically inadequate, models.

4.1 Data

Our database contains information on the donation behavior of almost 800,000 individuals for a large Dutch charitable organization for the period 1994–1999. We will use data for 1994–1998 for model specification and leave 1999 for evaluation. During the total period, the organization has sent 22 mailings (at 3 month intervals), but of course not always to everyone. Each time, the charitable organization selects targets from the full set based on RFM variables. It is known for each individual in the database whether a mailing has been sent at each of the 22 occasions, whether the individual has responded to this mailing, and the amount of the donation in Dutch guilders[5].

A random sample of 900 individuals is drawn from the total database. On average, they have received 7 mailings between 1994 and 1998. The last year in the database (1999) is used for the evaluation the model performance on target selection. The information concerning earlier mailings is used to construct RFM variables which we can use to explain response behavior and the (natural log) size of the donation. The response rate in the period 1994–1998 is about 34%. The average donation given response is 19.75 guilders with a minimum of 1 guilder and a maximum of 250 guilders.

RFM variables are used as explanatory variables in the model for response behavior and for donation. To be more precise, we use a dummy variable indicating whether an individual has responded to the last mailing (Recency), the proportion of mailings the individual has responded to (Frequency), and the natural logarithm of the average amount donated in the past (Monetary value). The variables are chosen such that they do not depend on the mailing policy used by the charitable organization. In previous research, Recency is often defined as the time since the last response (Bitran and Mondschein (1996),

[5] One Dutch guilder is about 0.45 euro.

Bult and Wansbeek (1995), Gönül and Shi (1998)). However, this variable depends on whether (and how many) mailings an individual has received since the last response. Similarly, frequency is commonly defined as the number of purchases (or reactions) in a certain time period (Bitran and Mondschein (1996), Bult and Wansbeek (1995), Gönül and Shi (1998)). Again this variable depends on the number of mailings received. To avoid dependence on the mailing strategy, we adopt the combination of these two, that is, the proportion of responded mailings. Defining Monetary Value as the average expenditure is not new, see Gönül and Shi (1998), while Bitran and Mondschein (1996) used the amount of the last expenditure. Average expenditure might have somewhat more face validity as it is a measure of long term behavior. It is of course possible to include more or other RFM-type variables in the model but some preliminary analysis indicated to us that this did not increase the target selection performance of our model, and hence we stick to the three mentioned variables.

The charitable organization itself each year decides on the number of mailings it sends to the individuals on their own list. It either does not send a mailing, or it sends 1, 2, 3, or at most 4 mailings. This decision is made using an RFM-based selection mechanism, which is not known by us.

4.2 Empirical Results

In this section we discuss the results of the model developed in Section 2. The first part of the model, equations (1) and (2), explains the response behavior of individuals. The size of the gift is modeled by the regression in equation (3). The model parameters are estimated based on the mailings sent between 1994 and 1998 to the 900 individuals in our sample.

The proposed model allows for heterogeneity in response behavior among individuals and takes into account that some individuals have received more mailings than others. The heterogeneity is modeled using a latent class formulation. The number of mixture components necessary to model the heterogeneity across the individuals is selected according to the BIC. The optimal number of mixture components turns out to be three.

We start with a discussion of the parameter estimates. Next we discuss the predicted response behavior to a mailing that follows from the parameter estimates. Based on predicted behavior we discuss various target selection rules and the resulting revenues. These revenues are compared to the revenues generated by the current selection rule of the charitable organization and to the revenues of a selection rule based on parameter estimates that are not corrected for the effect of previous target selection.

Estimation results

Table 1 shows the maximum likelihood parameter estimates for the resulting model with estimated standard errors in parentheses. The log-likelihood function is maximized using the BHHH algorithm of Berndt *et al.* (1974). To avoid identification and optimization problems that arise when latent class specific covariance matrices are allowed for, we assume that the covariance matrix, Σ, is the same for all latent classes. The estimate of σ_2, the standard error of the residual in the donation equation, equals 0.337 with a standard deviation of 0.007. The correlation between the stochastic components in the response and amount equation turns out to be reasonably high, as it is estimated to be 0.690 with a standard deviation of 0.030. The parameter estimates of the latent class specific parameters are presented in Table 1.

The general conclusion from Table 1 is that the parameters of the response equation and the donation equation differ substantially. This clearly confirms the relevance if the Tobit-2 specification, which distinguishes between the decision to respond and the amount to donate.

Although the estimation methodology uses latent classes as its terminology, marketers usually talk about segments and market segmentation. In principle, each latent class identifies a segment in the customer base of the charitable organization. In discussing the results we will talk about segments instead of classes. The parameter estimates for each segment are not straightforward to interpret, as for each segment both the parameters and the explanatory variables will be different. Consequently, we can discuss the different effects of the explanatory variables, but it is not straightforward to predict which segment

Table 1: Parameter estimates for the response equation and for the donation equation (standard errors in parentheses).

	Segment 1		Segment 2		Segment 3	
Response equation						
Intercept	−2.385	(0.136)	−0.888	(0.118)	−0.508	(0.475)
ln(Average Donation)	0.138	(0.045)	−0.130	(0.030)	−0.152	(0.136)
Prop. Resp. Mailings	2.206	(0.162)	1.940	(0.114)	0.733	(0.562)
Resp. Last Mailing	0.573	(0.098)	−0.098	(0.052)	0.147	(0.321)
Donation equation						
Intercept	−0.477	(0.079)	−0.091	(0.048)	−0.320	(0.164)
ln(Average Donation)	1.004	(0.022)	0.923	(0.011)	0.706	(0.038)
Prop. Resp. Mailings	0.389	(0.092)	0.207	(0.046)	−0.168	(0.191)
Resp. Last Mailing	0.072	(0.060)	−0.037	(0.022)	0.146	(0.124)
Segment Memb. Prob.	0.493	(0.028)	0.476	(0.026)	0.031	(0.008)
Mail Rec. Prob.	0.501	(0.010)	0.810	(0.008)	0.681	(0.030)

is most responsive or most profitable. We will discuss this issue when we look at predicted response behavior for each segment. In these predictions we take into account both the heterogeneity in explanatory variables among the segments and the different parameter estimates for each segment.

However, Table 1 does provide some useful insights. First of all, we see that there are substantial and highly significant differences between the probabilities with which each of the segments receives a mailing from the charitable organization. This indicates that there is a clear need to correct for the non-random mailing strategy of the charitable organization. The table also presents the segment membership probabilities, which indicate that there are two large segments that both represent almost 50% of the sample. The third segment is much smaller and comprises about 3% of the sample.

The estimates of the segment specific parameters also reveal an interesting fact. Comparing the two largest segments, we see that for the largest segment there is positive feedback from previous behavior. Higher donations and higher response rates in the past and responding to the last mailing all have a positive effect on the probability of responding and on the amount that is donated. This is in contrast with the estimated feedback effects for the second largest segment, where all the estimated effects are lower than for the largest segment. This is in itself not worrying, but the negative effects of average past donations and responding to the last mailing on the response probability are, at least, for the charitable organization. There is also a negative effect of responding to the last mailing on the expected amount, although this is not significant. These negative feedback effects, especially the effect of responding to the last mailing, indicate that this segment might receive too many mailings from the charitable organization. The estimates of the mail reception probabilities indicate that this segment does receive 60% more mailings than the largest segment, so an"irritation" effect for this segment seems plausible.

The information about the behavioral model contained in the smallest segment is not sufficient to obtain precise estimates. The only significant effect is the effect of average past donation on the current donation. It should also be noted that the feedback effects of the past mail response rate on the response probability and the effect of past donated amount on the current amount are significantly smaller than in the other two segments. The number of mailings this segment receives is in between that of the other two segments.

Predicted behavior and target selection

In this section we will describe the response and donation behavior of the three segments based on the predicted and actual behavior for the first mailing in 1999. This mailing was not used for estimation. The segments can be characterized by the average values of the RFM variables and by average behavior. Since we do not know which donors are in which segment, the statistics we provide in this section are based on the posterior segment membership probabilities for each donor. In Table 2 we present the average values of the RFM variables that are relevant for the response to the first mailing in 1999. From this table it is clear that Segment 2 has the best RFM characteristics, especially when it concerns the average response rate in the past and the percentage of individuals in the

Table 2: The average value of the RFM-variables.

	Segment 1	Segment 2	Segment 3
Average Donation	14.99	20.60	19.38
Prop. Reps. Mailings	41.4%	56.2%	46.7%
Resp. Last Mailing	27.6%	42.1%	31.2%

segment that responded to the last mailing. Segment 1 is characterized by low response rates and low amounts donated. Based on these observations, it might be justified that the charitable organization does not mail this segment very often.

These differences in RFM variables in combination with the segment specific parameter estimates result in different predicted responses to mailings. In Figure 1 we present the distribution function of the predicted response probabilities for the three segments. Here we see clearly that the distribution function for Segment 2 is more to the right than that of the other segments. This indicates that the individuals in Segment 2 are substantially more likely to respond than the individuals in the other segments. When we look for example at a response probability of 50%, we can see in the figure that for Segment 2, about 50% of the segment members have a lower response probability, while this is 85% and 90% for Segment 1 and Segment 3, respectively. Notice that 20% of the individuals in Segment 3 and 60% of the individuals in Segment 1 have a response rate below 10%.

Insert Figure 1 about here.

Insert Figure 2 about here.

A similar figure is presented in Figure 2, where the cumulative distributions of the amount donated, conditional upon responding to the mailing is depicted. The highest donations can be expected from the individuals in Segment 2, followed by Segment 1 and

Segment 3. The donations of individuals in Segment 3 are very low, with about 90% below 10 guilders.

The most important conclusion to be drawn from Figures 1 and 2 is that the segments differ substantially in their predicted response behavior. However, the ordering of the segments is rather stable, where Segment 2 is most promising with respect to both response probabilities and amounts donated. Whether Segment 1 is more interesting than Segment 3 depends on the dimension of response behavior one is interested in. When a rule such as proposed by Bult and Wansbeek (1995) is used (selecting the individuals with the highest response probabilities), Segment 3 is more promising, while target selection based on amount donated would focus more on Segment 1.

Insert Figure 3 about here.

The most relevant variable to use for target selection, however, might be expected revenues, which is the probability of responding times the expected amount of money that is donated, conditional on response. In principle, one might want to correct for the costs of sending a letter and the costs of collecting the money as is done in equation (10), but we do not have information on such costs, so we are not able to do so. The distribution of expected revenues for each of the segments is presented in Figure 3. This figure shows that expected revenues from Segment 3 are substantially lower than the expected revenues from the other segments. The high response rates and high donated amounts for Segment 2 result in substantial differences with the other segments.

The conclusion that one draws from the predicted behavior in the three figures is rather different from the conclusion one would draw from the average RFM variables presented in Table 2. Based on the average RFM variables, Segment 1 seems rather unattractive, even when it is compared to Segment 3. The way in which past behavior affects current behavior, however, is different, as can be concluded from Table 1. The differences in the dynamics in response behavior, captured by the differences in parameter estimates, cause the differences in RFM variables to be rather uninformative about current behavior.

When we want to develop a target selection rule, we do not have to target all the members of a particular segment at once. It is even not possible to do so, as segment membership is not observed. Using the posterior probabilities for segment membership of each of the segments, we can make predictions of the response probability and the amount donated by each individual. These predictions are weighted averages of the predictions for an individual conditional on segment membership. Using these predictions, the charitable organization could decide to send a mail to the individuals with the highest predicted probability of responding, the highest predicted amount donated, or the highest expected revenues.

To evaluate target selection based on our model, we compare it with the target selection rule currently used by the charitable organization and with a target selection rule where the parameter estimates are not corrected for previous target selection by the charitable organization. Two problems arise if one wants to compare target selection rules. First, if one considers different selection methods, one may expect that some individuals will be selected by all selection methods, while some individuals will only be selected by one of the selection strategies. Hence, to compare the different selections strategies one has to send mailings to the joint set of individuals that are selected by all target selection strategies. Fortunately, the charitable organization sends the first mailing in each year to all individuals in the database. Therefore, the response to this mailing is known for all individuals. This allows us to compare the different target selection procedures.

The second problem that arises is that we do not know the exact target selection rule implemented by the charitable organization. The organization uses some selection rule based on RFM-scores. These scores are determined at the beginning of the year and are unknown to us. We however do know a ranking of individuals into 3 groups depending on their RFM score. The individuals are ranked on their RFM scores and the best 35% end up in group A, the next 20% end up in group B, the remaining 45% end up in group C.

In the first panel of Table 3 we report the total revenue, the revenue per individual and the response rate for each group (A trough C) for the first mailing of 1999 for three selection rules. The selection rules in this panel are based on the parameter estimates

Table 3: Performance of target selection rules.

	Total Revenue	Revenue per Mail	Response Percentage
Model with latent classes and correction			
Selection on expected revenues			
A: Top 35%	4375	14.44	52%
B: Middle 20%	809	4.45	32%
C: Bottom 45%	569	1.44	12%
Selection on response rate			
A: Top 35%	3268	10.79	54%
B: Middle 20%	1445	7.94	35%
C: Bottom 45%	1040	2.64	10%
Selection on amount donated			
A: Top 35%	3812	12.58	36%
B: Middle 20%	1156	6.35	38%
C: Bottom 45%	785	1.99	22%
Selection of the charitable organization			
A: Top 35%	3168	10.46	41%
B: Middle 20%	959	5.27	28%
C: Bottom 45%	1626	4.13	23%
Model with latent classes, without correction			
Selection on expected revenues			
A: Top 35%	4307	14.21	53%
B: Middle 20%	907	4.98	32%
C: Bottom 45%	539	1.37	12%

reported in Table 1. The first case displays the performance when target selection is based on expected revenues. The individuals that have the highest predicted revenues end up in the top segment, while the individuals with the lowest predicted revenues are assigned to the bottom segment. From the response percentages and the average revenue per mail for the three segments we can conclude that our model is very well capable of differentiating between individuals with high and low potential. Revenue per mail is three times higher for the top segment than for the middle segment, while it is ten times higher than the revenue for the bottom segment.

The other two cases in the first panel of Table 3 present the performance of target selection rules when selection is based on predicted response rates and predicted amounts donated. When target selection is based on response rates, as is proposed by Bult and Wansbeek (1995), the third column of the table shows that this target selection rule is indeed better able to differentiate with respect to actual response percentages. The response rate in the top segment is 54%, which is higher than the response rate in the top segment when segmentation was based on expected revenues. However, the first two columns indicate that these higher response rates do not result in higher revenues for the top segment. When selection is based on amount donated, we see that the segments do not differ much in their observed response rates. The response rate in the top segment is even lower than the response rate in the middle segment. However, the performance of this target selection rule with respect to average revenues is better than the performance of the selection rule based on response rates.

The performance of the target selection rule currently in use by the charitable organization is presented in the second panel of Table 3. The results show that the selection rule of the charitable organization is capable of dividing the individuals according to their donating behavior. However, the performance of their target selection with respect to revenues is rather poor. It is worse than all three target selection rules we designed using the estimation results of our model. Revenues in the top segment of the charitable organization are lower than revenues in the top segment of the other selection rules, and moreover the middle and bottom segment of the charity are very similar.

The last panel of the table presents the performance of a selection rule that is based on the estimation results of the tobit-2 model without correcting for systematic differences in observation frequency. Although this model is inadequately estimated, the performance of the target selection rule based on the estimates is comparable to the performance when estimation is done properly. One reason for this is that even though we find substantial differences in mail reception probabilities, the target selection rule used by the charitable organization seems close to random, as indicated by its performance described above. When target selection is random, the correction we developed may not yield dramatic improvements and differences between the model with and without correction are not to be expected.

The comparison of the target selection rules described above results above is based on the performance on a certain (fixed) part of the sample. In practice, one also needs to determine how many individuals will be selected. Our model can explicitly determine the optimal cut-off point, using equation (10). This selection rule states that only those individuals should be selected for which expected revenue of a mailing is higher than the costs.

5 Conclusions

In this paper we proposed a target selection rule for direct mailings, based on a model that jointly estimates incidence and quantity, while accounting for historical target selection. Not everyone in the database has an equal probability of receiving a mailing, as the direct writer only mails those individuals who are considered to be most profitable. We have shown that by allowing for unobserved heterogeneity and extending the likelihood to incorporate the mailing strategy, previous target selection can be taken into account. Using data from a charitable organization, we showed how our model can lead to improved results in various respects. In this section we discuss some managerial implications of our model. Additionally, we review possible limitations and topics for further research.

5.1 Managerial Implications

In practice, companies often rely on models that maximize response and not revenues. We showed that a selection based on response does not result in optimal revenues. Extending the model to include the expected amount improves the target selection rule dramatically, but also selection on amount donated instead of response performs better. Taking the effect of previous target selection of the charitable organization on observability into account does not improve the selection rule substantially. This might be due to an ineffective selection rule that has been implemented in the past. When selection is close to random, there will be almost no sampling effects. Differences will be larger when a better target selection rule would have been applied. The main conclusion for managers is that it pays off to construct a full-blown model, while taking account of previous direct marketing strategies. The use of only RFM based selection rules is not to be indicated, as is indicated by the conclusion one would draw from Table 2 for the largest segment.

5.2 Limitations and Topics for Future Research

There are a couple of limitations to our model and empirical analysis, although they all seem to generate interesting topics for further research. A first limitation concerns the fact that we only considered individuals who did donate to the same charity. Indeed, only because of this focus we could construct RFM variables. However, for a charitable organization it is usually also of interest to acquire new donors. New prospective charity donors are often selected from mailing lists, which are provided by firms other than the charity itself. These firms provide lists of zip codes and addresses of individuals who they think might be willing to donate to charity. Hence, again there is a selection step, which should be included in the model in order to reduce sample selection effects. Even though one usually only knows some characteristics of individuals or households when they have certain zip codes, it should be possible to match this information with the RFM variables of individuals with for example the same zip codes. As this means that one has to draw inference on individuals on the basis of estimated characteristics of groups of individuals (for example, those who have the same zip codes), one has to resort to what are called ecological inference techniques.

Secondly, the selection rule provides the optimal selection rule in a one period setting. One could imagine that delaying mailings could lead to even higher revenues, see Gönül and Shi (1998). The results in this paper show that, to determine such a dynamic selection, one should better consider a model for both response and amount.

A Appendix

In this appendix we show for a simple example that maximizing the likelihood function which includes the mailing process results in consistent parameter estimates. The structure of the likelihood indicates that this correction will also work when the data have been generated with a similar mechanism.

Suppose there are two periods and two latent classes. In the first period, all individuals are observed. In the second period, individuals in latent class 1 are observed with probability q_1, while individuals in latent class 2 are observed with probability q_2. The relevant population has a fraction p_1 of individuals in latent class 1, leaving a fraction $p_2 = 1 - p_1$ of individuals in class 2. Our interest is in the estimation of model parameters based on a sample where observability depends on the latent class. Without loss of generality we assume that the likelihood of observing y for an individual in class s in period t equals $L_t(y; s)$.

When a random sample of individuals is drawn from this population, a fraction $p_1 q_1 + p_2 q_2$ of the N individuals will have two observations, while for the other individuals we have only one observation. For the individuals with two observations, the probability of observing y_1 in period 1 and y_2 in period 2 equals:

$$\frac{p_1 q_1 L_1(y_1; 1) L_2(y_2; 1) + p_2 q_2 L_1(y_1; 2) L_2(y_2; 2)}{p_1 q_1 + p_2 q_2}. \tag{11}$$

A similar expression can be obtained for the individuals who are observed only once.

The average log-likelihood function based on the likelihood in equation (4) is

$$\frac{1}{N} \sum_{i=1}^{N} \log \sum_{s=1}^{S} p_s \prod_{t=1}^{T} L_t(y_{it}; s)^{I_{it}} q_{st}^{I_{it}} (1 - q_{st})^{1 - I_{it}}. \tag{12}$$

We now analyze what happens with the estimates for the segment probabilities under the assumption that the other parameters are estimated consistently. We therefore investigate the consistency of the segment membership probabilities. Maximizing the log-likelihood in (12) with respect to p_1 results in the following first-order condition

$$0 = \frac{1}{N} \sum_{i=1}^{N} \frac{\left[\prod_{t=1}^{T} L_t(y_{it}; 1)^{I_{it}} q_{1t}^{I_{it}} (1 - q_{1t})^{1 - I_{it}} - \prod_{t=1}^{T} L_t(y_{it}; 2)^{I_{it}} q_{2t}^{I_{it}} (1 - q_{2t})^{1 - I_{it}}\right]}{\sum_{s=1}^{S} p_s \prod_{t=1}^{T} L_t(y_{it}; s)^{I_{it}} q_{st}^{I_{it}} (1 - q_{st})^{1 - I_{it}}}. \tag{13}$$

Notice here that the denominator is exactly the likelihood of an individual i, which stems from $\frac{\partial \log(x)}{\partial x} = \frac{1}{x}$. The presence of the likelihood for latent class 2 in the numerator is the result of the restriction $p_2 = 1 - p_1$. With the number of observations increasing, the average in (13) converges to its expectation, so $\frac{1}{N}\sum_{i=1}^{N}$ can be replaced by the expectation operator E. To prove consistency, we thus have to show that the condition

$$0 = \mathrm{E}\left[\frac{\prod_{t=1}^{T} L_t(y_{it};1)^{I_{it}} q_{1t}^{I_{it}}(1-q_{1t})^{1-I_{it}} - \prod_{t=1}^{T} L_t(y_{it};2)^{I_{it}} q_{2t}^{I_{it}}(1-q_{2t})^{1-I_{it}}}{\sum_{s=1}^{S} p_s \prod_{t=1}^{T} L_t(y_{it};s)^{I_{it}} q_{st}^{I_{it}}(1-q_{st})^{1-I_{it}}}\right] \tag{14}$$

holds.

This condition is satisfied by the true segment probabilities as we will show next. To distinguish the segment probabilities that have to be estimated from their true values in the data generating process, we denote the estimated segment probabilities by $\hat{p}_s$. As we assume the other parameters to be estimated consistently, this distinction is not relevant for the other parameters. Conditioning on the number of observations, we can show that

$$\begin{aligned}
&\mathrm{E}\left[\frac{\left[\prod_{t=1}^{T} L_t(y_{it};1)^{I_{it}} q_{1t}^{I_{it}}(1-q_{1t})^{1-I_{it}} - \prod_{t=1}^{T} L_t(y_{it};2)^{I_{it}} q_{2t}^{I_{it}}(1-q_{2t})^{1-I_{it}}\right]}{\sum_{s=1}^{S} \hat{p}_s \prod_{t=1}^{T} L_t(y_{it};s)^{I_{it}} q_{st}^{I_{it}}(1-q_{st})^{1-I_{it}}}\right] = \qquad (15)\\
&\quad \Pr[I_{i2}=1]\times\\
&\qquad \mathrm{E}\left[\left.\frac{\left[\prod_{t=1}^{T} L_t(y_{it};1)^{I_{it}} q_{1t}^{I_{it}}(1-q_{1t})^{1-I_{it}} - \prod_{t=1}^{T} L_t(y_{it};2)^{I_{it}} q_{2t}^{I_{it}}(1-q_{2t})^{1-I_{it}}\right]}{\sum_{s=1}^{S} \hat{p}_s \prod_{t=1}^{T} L_t(y_{it};s)^{I_{it}} q_{st}^{I_{it}}(1-q_{st})^{1-I_{it}}}\right| I_{i2}=1\right]\\
&\quad +\Pr[I_{i2}=0]\times\\
&\qquad \mathrm{E}\left[\left.\frac{\left[\prod_{t=1}^{T} L_t(y_{it};1)^{I_{it}} q_{1t}^{I_{it}}(1-q_{1t})^{1-I_{it}} - \prod_{t=1}^{T} L_t(y_{it};2)^{I_{it}} q_{2t}^{I_{it}}(1-q_{2t})^{1-I_{it}}\right]}{\sum_{s=1}^{S} \hat{p}_s \prod_{t=1}^{T} L_t(y_{it};s)^{I_{it}} q_{st}^{I_{it}}(1-q_{st})^{1-I_{it}}}\right| I_{i2}=0\right].
\end{aligned}$$

Writing out the expectation in (15) and using the fact that in the first period an individual is always observed, so $q_{s1} = 1$, $s = 1,2$, and $q_{s2} = q_s$, $s = 1,2$ we find that (15) is also equal to

$$\begin{aligned}
&\Pr[I_{i2}=1]\iint \frac{\left[q_1\prod_{t=1}^{T} L_t(y_{it};1) - q_2\prod_{t=1}^{T} L_t(y_{it};2)\right]}{\sum_{s=1}^{S}\hat{p}_s q_s \prod_{t=1}^{T} L_t(y_{it};s)} f(y_{i1},y_{i2}|I_{i2}=1)dy_{i1}dy_{i2}\\
&\qquad +\Pr[I_{i2}=0]\int \frac{[L_1(y_{i1};1)(1-q_1) - L_1(y_{i1};2)(1-q_2)]}{\sum_{s=1}^{S}\hat{p}_s L_1(y_{i1};s)(1-q_s)} f(y_{i1}|I_{i2}=0)dy_{i1}. \qquad (16)
\end{aligned}$$

We now plug in the conditional densities of the outcomes given the number of observations. Notice that the expectation concerns the real data generating process, so p_s here is not estimated, in contrast to $\hat{p}_s$. Also in the above equations $\Pr[I_{i2} = 1]$ and $\Pr[I_{i2} = 0]$ depend on the segment probabilities in the data generating process, but these are also not estimated. We therefore have that (16) can be written as

$$\begin{aligned} \Pr[I_{i2} = 1] \iint \frac{\left[q_1 \prod_{t=1}^{T} L_t(y_{it};1) - q_2 \prod_{t=1}^{T} L_t(y_{it};2)\right]}{\sum_{s=1}^{S} \hat{p}_s q_s \prod_{t=1}^{T} L_t(y_{it};s)} \frac{\sum_{s=1}^{S} p_s q_s \prod_{t=1}^{T} L_t(y_{it};s)}{\sum_{s=1}^{S} p_s q_s} dy_{i1} dy_{i2} \\ + \Pr[I_{i2} = 0] \int \frac{[L_1(y_{i1};1)(1-q_1) - L_1(y_{i1};2)(1-q_2)]}{\sum_{s=1}^{S} \hat{p}_s L_1(y_{i1};s)(1-q_s)} \frac{\sum_{s=1}^{S} p_s L_1(y_{i1};s)(1-q_s)}{\sum_{s=1}^{S} p_s (1-q_s)} dy_{i1} \end{aligned} \tag{17}$$

The denominator in the second term equals the probability in front of the integral and as the term does not depend on y_{it}, this term can be taken out of the integral and it cancels with the probability before the integral sign. When we assume that the $\hat{p}_s$ are consistently estimated, so they converge to p_s, the denominators of the first terms equal the numerator of the second term. In this situation these terms also cancel.

Notice here that because the mailing process is modeled, the q_s occur in the denominator of the first term, while they are also always present in the numerator of the second term, which stems from the data generating process. When the mailing process is not modeled, the q_s do not occur and the terms do not cancel out. In this case the maximum likelihood estimates for p_s are, in general, not consistent.

After eliminating these terms it only remains to be shown that

$$\begin{aligned} 0 = q_1 \iint \prod_{t=1}^{T} L_t(y_{it};1) dy_{i1} dy_{i2} - q_2 \iint \prod_{t=1}^{T} L_t(y_{it};2) dy_{i1} dy_{i2} \\ + (1-q_1) \int L_1(y_{i1};1) dy_{i1} - (1-q_2) \int L_1(y_{i1};2) dy_{i1} \end{aligned} \tag{18}$$

As each of the integrals above represent the integration of a density function over the complete support, these integrals are all equal to 1. Making this substitution, it is straightforward to see that the first order condition for the segment probabilities is satisfied for the true segment probabilities as

$$0 = q_1 - q_2 + (1-q_1) - (1-q_2) = 0. \tag{19}$$

References

Amemiya, T. (1985), *Advanced Econometrics*, Blackwell, Oxford.

Berndt, E. K., B. H. Hall, E. Hall, and J. A. Hausman (1974), Estimation and Inference in Non-linear Structural Models, *Annals of Economic and Social Measurement*, **3**, 653–665.

Bitran, G. R. and S. V. Mondschein (1996), Mailing Decisions in the Catalog Sales Industry, *Management Science*, **42**, 1364–1381.

Bult, J. R. and T. J. Wansbeek (1995), Optimal Selection for Direct Mail, *Marketing Science*, **14**, 378–394.

Davies, R. B. (1977), Hypothesis Testing When a Nuisance Parameter Is Present Only Under the Alternative, *Biometrika*, **64**, 247–254.

Fitzgerald, J., P. Gottschalk, and R. Moffit (1998), An Analysis of Sample Attrition in Panel Data, *Journal of Human Resources*, **33**, 251–299.

Gönül, F. and M. Z. Shi (1998), Optimal Mailing of Catalogs: A New Methodology Using Estimable Structural Dynamic Programming Models, *Management Science*, **44**, 1249–1262.

Gönül, F. and K. Srinivasan (1993), Modeling Multiple Sources of Heterogeneity in Multinomial Logit Models: Methodological and Managerial Issues, *Marketing Science*, **12**, 213–229.

Heckman, J. J. and B. Singer (1984), A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data, *Econometrica*, **52**, 271–320.

Jain, D. C., N. J. Vilcassim, and P. K. Chintagunta (1994), A Random-Coefficients Logit Brand-Choice Model Applied to Panel Data, *Journal of Business & Economic Statistics*, **12**, 317–328.

Kamakura, W. A. and G. J. Russell (1989), A Probabilistic Choice Model for Market Segmentation and Elasticity Structure, *Journal of Marketing Research*, **26**, 379–390.

Roberts, M. L. and P. D. Berger (1999), *Direct Marketing Management*, Prentice Hall, Inc., Uple Saddle River, NJ.

Rossi, P. E. and G. M. Allenby (1993), A Bayesian Approach to Estimating Household Parameters, *Journal of Marketing Research*, **30**, 171–182.

Ter Horst, J., T. Nijman, and M. Verbeek (2001), Eliminating Look Ahead Bias in Evaluating Persistence in Mutual Fund Performance, *Journal of Empirical Finance*, **8**, 345–373.

Wedel, M. and W. A. Kamakura (1999), *Market Segmentation; Conceptual and Methodological Foundations*, International Series in Quantitative Marketing, Kluwer Academic Publishers, Boston.

Wedel, M., W. A. Kamakura, A. Bemmaor, J. Chiang, T. Elrod, R. Johnson, P. Lenk, S. Neslin, and C. S. Poulsen (1999), Discrete and Continuous Representations of Unobserved Heterogeneity in Choice Modelling, *Marketing Letters*, **10**, 219–232.

Figure 1

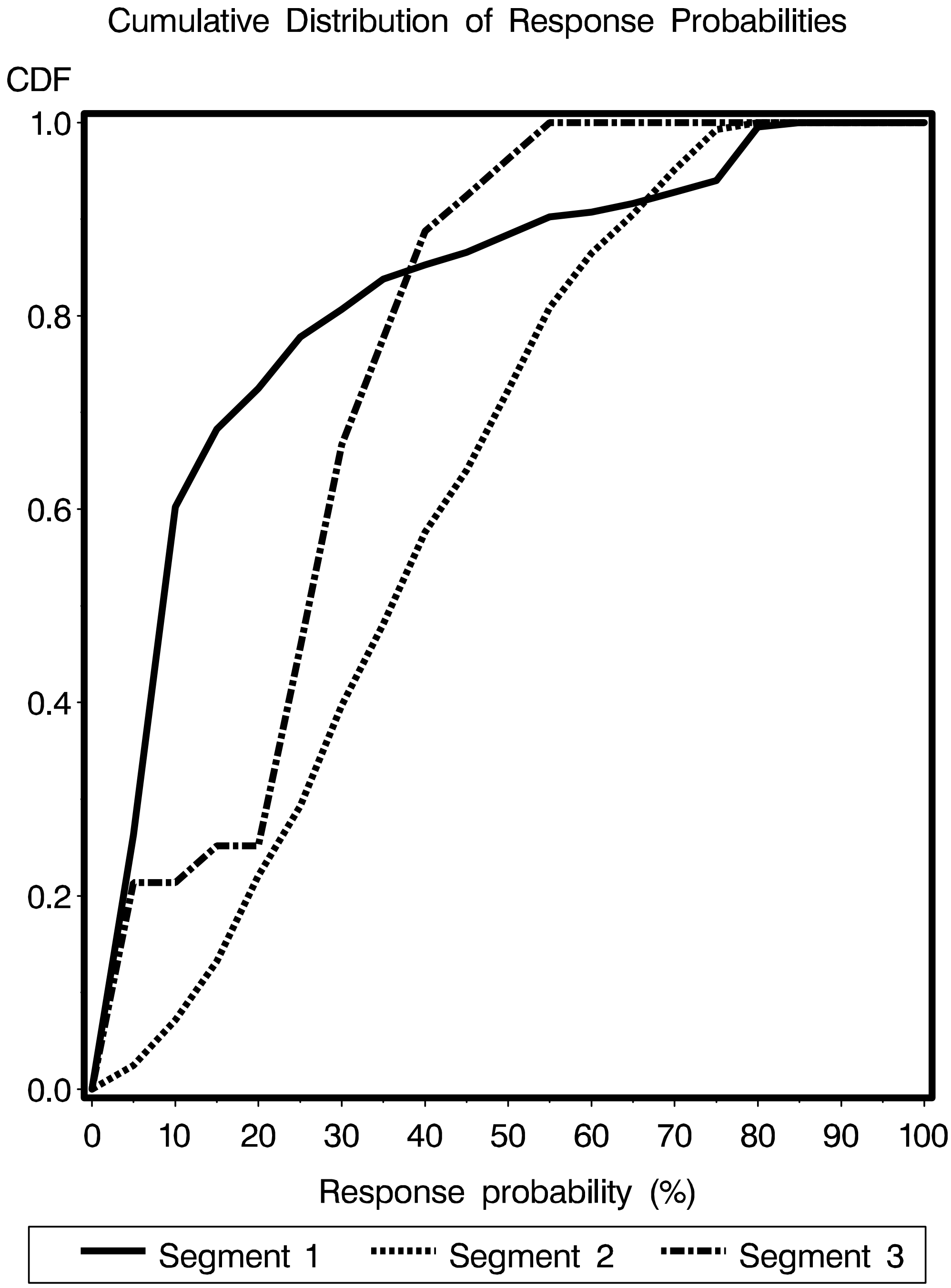

Cumulative Distribution of Response Probabilities
CDF
1.0
0.8
0.6
0.4
0.2
0.0
0
10
20
30
40
50
60
70
80
90
100
Response probability (%)
Segment 1
Segment 2
Segment 3

Figure 2

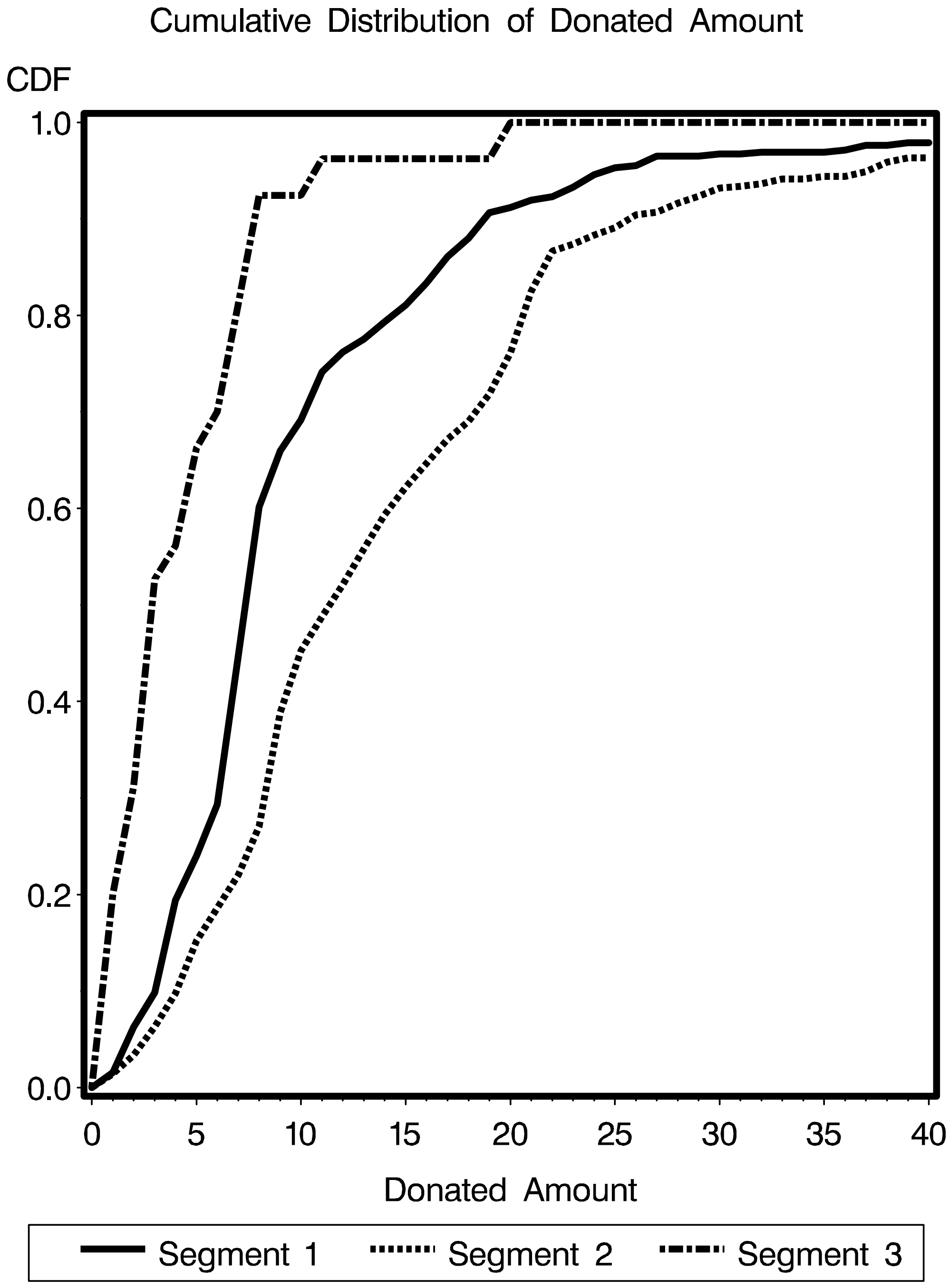
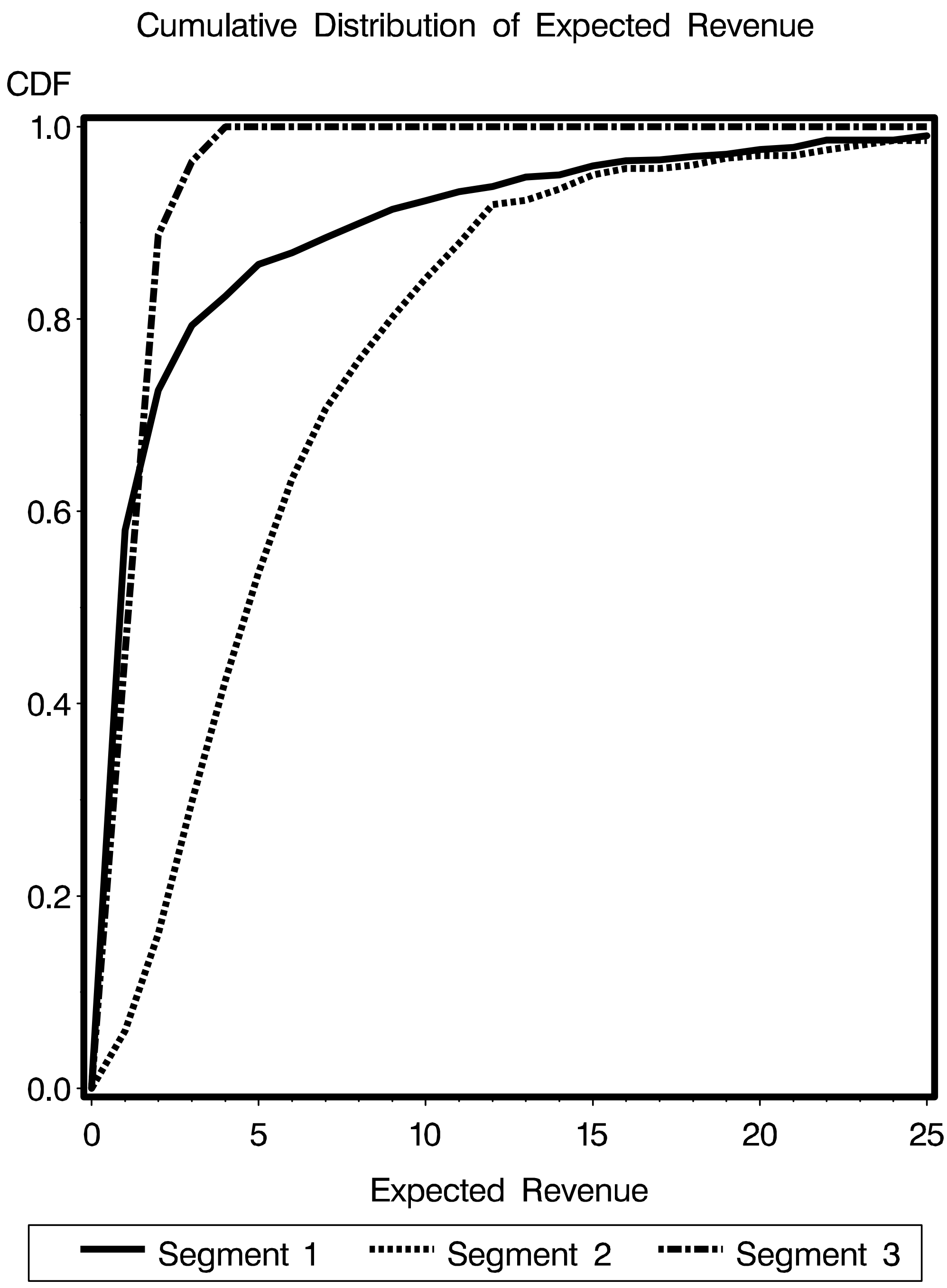

Cumulative Distribution of Donated Amount
CDF
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
25
30
35
40
Donated Amount
Segment 1
Segment 2
Segment 3

Figure 3


Cumulative Distribution of Expected Revenue
CDF
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
25
Expected Revenue
Segment 1
Segment 2
Segment 3

Publications in the Report Series Research[*] in Management

ERIM Research Program: "Marketing"

2001

Predicting Customer Potential Value. An application in the insurance industry
Peter C. Verhoef & Bas Donkers
ERS-2001-01-MKT

Modeling Potenitally Time-Varying Effects of Promotions on Sales
Philip Hans Franses, Richard Paap & Philip A. Sijthoff
ERS-2001-05-MKT

Modeling Consideration Sets and Brand Choice Using Artificial Neural Networks
Björn Vroomen, Philip Hans Franses & Erjen van Nierop
ERS-2001-10-MKT

Firm Size and Export Intensity: A Transaction Costs and Resource-Based Perspective
Ernst Verwaal & Bas Donkers
ERS-2001-12-MKT

Customs-Related Transaction Costs, Firm Size and International Trade Intensity
Ernst Verwaal & Bas Donkers
ERS-2001-13-MKT

The Effectiveness of Different Mechanisms for Integrating Marketing and R & D
Mark A.A.M. Leenders & Berend Wierenga
ERS-2001-20-MKT

Intra-Firm Adoption Decisions: Departmental Adoption of the Common European Currency
Yvonne M. van Everdingen & Berend Wierenga
ERS-2001-21-MKT

Econometric Analysis of the Market Share Attraction Model
Dennis Fok, Philip Hans Franses & Richard Paap
ERS-2001-25-MKT

Buying High Tech Products: An Embeddedness Perspective
Stefan Wuyts, Stefan Stremersch & Philip Hans Franses
ERS-2001-27-MKT

Changing Perceptions and Changing Behavior in Customer Relationships
Peter C. Verhoef, Philip Hans Franses & Bas Donkers
ERS-2001-31-MKT

How and Why Decision Models Influence Marketing Resource Allocations
Gary L. Lilien, Arvind Rangaswamy, Katrin Starke & Gerrit H. van Bruggen
ERS-2001-33-MKT

---

[*] A complete overview of the ERIM Report Series Research in Management:
http://www.ers.erim.eur.nl

ERIM Research Programs:
LIS Business Processes, Logistics and Information Systems
ORG Organizing for Performance
MKT Marketing
F&A Finance and Accounting
STR Strategy and Entrepreneurship

An Equilibrium-Correction Model for Dynamic Network Data
David Dekker, Philip Hans Franses & David Krackhardt
ERS-2001-39-MKT

Aggegration Methods in International Comparisons: What Have We Learned?
Bert M. Balk
ERS-2001-41-MKT

The Impact of Channel Function Performance on Buyer-Seller Relationships in Marketing Channels
Gerrit H. van Bruggen, Manish Kacker & Chantal Nieuwlaat
ERS-2001-44-MKT

Incorporating Responsiveness to Marketing Efforts when Modeling Brand Choice
Dennis Fok, Philip Hans Franses & Richard Paap
ERS-2001-47-MKT

Competitiveness of Family Businesses: Distinghuising Family Orientation and Business Orientation
Mark A.A.M. Leenders & Eric Waarts
ERS-2001-50-MKT

The Effectiveness of Advertising Matching Purchase Motivation: An Experimental Test
ERS-2001-65-MKT
Joost Loef, Gerrit Antonides & W. Fred van Raaij

Using Selective Sampling for Binary Choice Models to Reduce Survey Costs
ERS-2001-67-MKT
Bas Donkers, Philip Hans Franses & Peter Verhoef

Deriving Target Selction Rules from Edogenously Selected Samples
ERS-2001-68-MKT
Bas Donkers, Jedid-Jah Jonker, Philip Hans Franses & Richard Paap

2000
Impact of the Employee Communication and Perceived External Prestige on Organizational Identification
Ale Smidts, Cees B.M. van Riel & Ad Th.H. Pruyn
ERS-2000-01-MKT

Forecasting Market Shares from Models for Sales
Dennis Fok & Philip Hans Franses
ERS-2000-03-MKT

The Effect of Relational Constructs on Relationship Performance: Does Duration Matter?
Peter C. Verhoef, Philip Hans Franses & Janny C. Hoekstra
ERS-2000-08-MKT

Informants in Organizational Marketing Research: How Many, Who, and How to Aggregate Response?
Gerrit H. van Bruggen, Gary L. Lilien & Manish Kacker
ERS-2000-32-MKT

The Powerful Triangle of Marketing Data, Managerial Judgment, and Marketing Management Support Systems
Gerrit H. van Bruggen, Ale Smidts & Berend Wierenga
ERS-2000-33-MKT

Consumer Perception and Evaluation of Waiting Time: A Field Experiment
Gerrit Antonides, Peter C. Verhoef & Marcel van Aalst
ERS-2000-35-MKT

Broker Positions in Task-Specific Knowledge Networks: Effects on Perceived Performance and Role Stressors in an Account Management System
David Dekker, Frans Stokman & Philip Hans Franses
ERS-2000-37-MKT

Modeling Unobserved Consideration Sets for Household Panel Data
Erjen van Nierop, Richard Paap, Bart Bronnenberg, Philip Hans Franses & Michel Wedel
ERS-2000-42-MKT

A Managerial Perspective on the Logic of Increasing Returns
Erik den Hartigh, Fred Langerak & Harry Commandeur
ERS-2000-48-MKT

The Mediating Effect of NPD-Activities and NPD-Performance on the Relationship between Market Orientation and Organizational Performance
Fred Langerak, Erik Jan Hultink & Henry S.J. Robben
ERS-2000-50-MKT

Sensemaking from actions: Deriving organization members' means and ends from their day-to-day behavior
Johan van Rekom, Cees B.M. van Riel & Berend Wierenga
ERS-2000-52-MKT