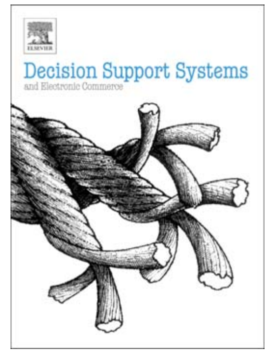
Accepted Manuscript



Comparing alternatives to account for unobserved heterogeneity in direct marketing models

Nadine Schröder, Harald Hruschka

PII: S0167-9236(17)30155-0
DOI: doi:10.1016/j.dss.2017.08.005
Reference: DECSUP 12871

To appear in: *Decision Support Systems*

Received date: 14 December 2016
Revised date: 4 August 2017
Accepted date: 30 August 2017

Please cite this article as: Nadine Schröder, Harald Hruschka, Comparing alternatives to account for unobserved heterogeneity in direct marketing models, *Decision Support Systems* (2017), doi:10.1016/j.dss.2017.08.005

Comparing alternatives to account for unobserved heterogeneity in direct marketing models

Nadine Schröder* [a], Harald Hruschka[a]

[a] *Universitätsstraße 31, 93040 Regensburg, Germany*

Abstract

We are dealing with mailing decisions of a direct marketing company and focus on assessing three alternative approaches to model unobserved heterogeneity, which are based on finite mixtures, continuous mixtures, and a mixture of Dirichlet processes (MDP), respectively. Models are estimated by Markov Chain Monte Carlo (MCMC) simulation. Based on Pseudo Bayes factors (PsBF), we find that a finite mixture model turns out to be superior both to models based on either a MDP or a continuous mixture. Whereas the MDP finds similar estimates compared to the finite mixture approach, estimates of the continuous mixture differ for some variables. According to the finite mixture, type of mailing has an effect on purchase behavior. In addition, some customers show supersaturation effects of mailings. Due to different coefficient estimates, managerial implications differ depending on which model they relate. In particular, a continuous mixture model would recommend more mailings than a finite mixture approach.

Keywords:
unobserved heterogeneity, direct mailings, hierarchical Bayesian models, mailing effects

*Corresponding author
Email addresses: `nadine.schroeder@ur.de` (Nadine Schröder*), `harald.hruschka@ur.de` (Harald Hruschka)

1. Introduction

Marketers are always interested in assessing how customers respond to marketing actions. In the area of direct marketing these actions often constitute of mailings that are sent to customers. Advertising spendings in the catalog mail order industry amount to around 8 billion U.S. dollars, which is about four times higher compared to spendings in 2010, which means that direct mailings are still an important marketing instrument despite the increase of online channels (statista, 2016). Response models can capture customer reactions and help target future mailings to these customers to induce further purchases. For a comparison of various data mining techniques see Olson and Chae (2012). However, as customers differ from one another, e.g., some customers react differently to the same marketing actions, it is important that marketers take heterogeneity of customers into account (Guelman, Guillén, Pérez-Marín, 2015). Neglecting heterogeneity may lead to biased parameter estimates (Popkowski Leszczyc & Bass, 1998). On the other hand, a survey of Verhoef, Spring, Hoekstra, and Leeflang (2002) shows that only 27% of companies employ segment specific models. Generally, different responses to mailings can be due to observed and unobserved heterogeneity. Observed heterogeneity can be assessed by customer specific variables. Unobserved heterogeneity is usually modeled with either a continuous or discrete mixture of distributions of parameters. Discrete mixtures of distributions are addressed with either a finite mixture (e.g., Gönül, Kim, & Shi, 2000) or a Mixture of Dirichlet Processes (MDP) (e.g., Ansari & Mela, 2003). A continuous mixture is usually addressed with mixtures of normal distributions (e.g., van Diepen, Donkers, & Franses, 2009). Both a finite mixture and a MDP divide customers into several segments and treat customers of the same segment equally. These two approaches are in contrast to a continuous mixture able to reproduce multimodal and skewed distributions (e.g., Hruschka, 2010). The MDP provides the number of segments as one of the estimated parameters. This property can be seen as an advantage as researchers using a MDP do not need to estimate several finite mixture models each with a different number of segments.

Our study contributes to the literature on direct marketing models in the following way. We compare three specifications of unobserved heterogeneity and investigate which of these specifications performs best in terms of model fit. To the best of our knowledge, no previous study dealing with effects of mailings has done that before. In addition, we compare how estimation results are affected by the specification of unobserved heterogeneity. We finally address different managerial implications that may result due to different specifications.

The remainder of the article is structured as follows. The following section addresses findings of selected direct marketing related studies on unobserved heterogeneity. Model specifications and our estimation method are discussed in the third section. Section 4 informs about the data set. Section 5 contains estimation results and managerial implications. In the last section we conclude the article.

2. Findings on unobserved heterogeneity in direct marketing models

In the following, we report results from past research on effects of direct mailings and unobserved heterogeneity. The first group of authors incorporates unobserved heterogeneity by finite mixture models. This group of authors tests whether a finite mixture model is superior to a homogenous model. Gönül et al. (2000) specify a Box-Cox hazard function with various predictors. Whereas they assume predictors like gender and average consumption rate to be fixed, the baseline hazard together with predictors capturing the elapsed time since the last mailing and the accumulated number of mailings since the last purchase are assumed to be heterogeneous. They estimate several models starting with a one-segment solution. Based on the Bayesian Information Criterion (BIC), they find that a two-segment solution (consisting of 62% vs. 38% of the customers) performs best. Whereas they observe a wearout effect for the customers belonging to the larger segment, recent mailings increase the response rate for customers from both segments.
Hruschka, Baumgartner, and Semmler (2003) also investigate the effect of mailings (in terms of number of mailings) and other variables on response behavior, i.e., response probability and sales. When they validate their semi-log and logistic response models, it turns out that one-segment, i.e., homogeneous, solutions are superior to two-segment solutions. They find positive relationships between number of mailings and response probability and sales, respectively.

The following two studies refer to authors who employ MDP, which automatically results in the optimal number of segments. Hruschka (2010) estimates a semi-log response model with number of catalogs as one of the predictors. In addition, he compares a policy function to an instrumental variable model. Results based on cross validation predictive densities show that an instrumental variable model with on average 14 segments of customers performs better than the policy function model with on average 31 segments of customers. The predictor number of catalogs has a positive impact on sales. The study of Ansari and Mela (2003) examines the relationship of different characteristics of online mailings, i.e., e-mails, and the probability of clicking on one of the links in the e-mails. Predictors consist of e-mail characteristics, such as, e.g., the number and order of links in an e-mail. In addition, also the editorial content is being captured. They compare a MDP with a continuous mixture approach. Based on Pseudo Bayes factors (PsBF) they find that the MDP with on average 61 segments outperforms the continuous mixture. Findings regarding their predictors include a decreasing effectiveness of links when they appear later in the e-mail. The number of links on the other hand does not affect the clicking probability.

The vast majority of studies using response models in direct marketing uses continuous mixtures, i.e., random coefficients approaches. Van Diepen et al. (2009) use a Tobit-2 model that explains response behavior towards mailings in terms of purchase probability and amount of sales. Predictors regarding mailing variables constitute of a mailing stock variable and its quadratic representation. In contrast to previous cited literature the authors differentiate between own mailings and those being sent from two of the largest competitors. The authors do not compare this model to a homogenous model. They find that mailings have a negative impact on purchase probability and amount of sales. Zantedeschi, McDonnell, and Bradlow (2016) capture purchase probability with a Tobit-1 model and differentiate between traditional mailings and online mailings, i.e., e-mails as predictors. Both types of mailings are modeled with stock variables. They compare a set of models, which differ regarding model specifications. However, all candidate models incorporate unobserved heterogeneity by random coefficients. Findings on traditional and online mailings are different. Whereas e-mails are more effective than mailings in the beginning, mailings generate a higher cumulative response after eight days. Ma, Hou, Yao, and Lee (2016) use a different response model, i.e., a hidden markov model (HMM). The authors assume that purchase probability of a mailing is affected by an unknown state at any given point in the relationship between customers and the company. They use a log-transformed mailing variable that sums the number of mailings and estimate various heterogeneous HMM, based on a continuous mixture approach, and a latent class model. It turns out that the HMM with three states is superior to other specifications, e.g., a HMM with only two states based on various predictive criteria. W.r.t. this model, they find that mailings have a positive effect on purchase probability in the short as well as in the long run.

In contrast to the previous studies, we estimate several response models considering three possible techniques of addressing unobserved heterogeneity (i.e., finite mixture, MDP, continuous mixture). We compare these models not only with respect to PsBF, which are based on cross-validation predictive densities, but we also investigate whether and how they lead to different estimation results and managerial implications.

3. Our model

3.1. Tobit-2 model

We measure the effects of (previous) mailings and previous sales on the purchase probability and sales of individual customers who receive a mailing. Any customer i who receives a mailing τ decides on whether s/he responds to this mailing by making a purchase ($R_{i\tau} = 1$) or not ($R_{i\tau} = 0$). Purchases constitute our first dependent variable. Given that a customer decides to make a purchase, s/he will also decide about the amount of money s/he spends for this purchase, which we call amount of sales $A_{i\tau}$. Of course, we observe values greater than zero for the second dependent variable, amount of sales, only if the customer makes a purchase (equivalently, if the first dependent variable, $R_{i\tau}$, equals one).

Hence, besides the binary dependent variable purchase, we have to cope with one censored dependent variable, amount of sales, whose minimum value equals zero. Dependent variables with these properties can be analyzed by Tobit models (Amemiya, 1985). In accordance with van Diepen et al. (2009), we use the Tobit-2 model. This model bears the advantage of being superior to other (e.g., logit) models (Levin & Zahavi, 1998). The Tobit-2 model consists of two equations, one for each of the two dependent variables just mentioned. In the first equation $R_{i\tau}$ equals one if the latent variable $R^*_{i\tau}$ is greater than zero, otherwise $R_{i\tau}$ equals zero:

$$R_{i\tau} = \begin{cases} 1 \text{ if } R^*_{i\tau} > 0 \\ 0 \text{ otherwise.} \end{cases} \tag{1}$$

The second equation of the Tobit-2 model refers to another latent variable, $A^*_{i\tau}$, which equals the observed amount of sales $A_{i\tau}$ if latent variable $R^*_{i\tau}$ is greater than zero:

$$A_{i\tau} = \begin{cases} A^*_{i\tau} \text{ if } R^*_{i\tau} > 0 \\ 0 \text{ otherwise.} \end{cases} \tag{2}$$

The two latent variables $R^*_{i\tau}$ and $A^*_{i\tau}$ are specified as linear combinations of explanatory variables contained in vectors $x_{Ri\tau}$ and $x_{Ai\tau}$ to which the respective error term $(\varepsilon_{Ri\tau}, \varepsilon_{Ai\tau})$ is added:

$$R^*_{i\tau} = x'_{Ri\tau}\beta_{Ri} + \varepsilon_{Ri\tau} \quad (3)$$

$$A^*_{i\tau} = x'_{Ai\tau}\beta_{Ai} + \varepsilon_{Ai\tau}. \quad (4)$$

The two coefficient vectors for customer i are captured by β_{Ri} and β_{Ai}. If unobserved heterogeneity is taken into account by either discrete or continuous mixture distributions, these vectors differ across segments or customers, respectively. We use dummy variables and z-standardized metric variables as explanatory variables.

Residuals of the Tobit-2 model follow a bivariate normal distribution with mean values equal to zero, variance σ^2_{Ai} for the amount of sales, and covariance σ_{RAi} between residuals of the first and second equation. To ensure identification, the residual variance of the purchase equation is set equal to one (Cameron & Trivedi, 2006). Thus, residuals are correlated with correlation coefficient $\rho_i = \sigma_{RAi}/\sigma_{Ai}$.

3.2. (Mailing) variables

We use several mailing variables to assess their impact on purchase behavior. This is in contrast to related publications as most studies use only one type of mailing variables (e.g, van Diepen et al., 2009; Hruschka, 2010). First of all, we distinguish between current mailings and a mailing stock resulting from past mailings. Current mailings are characterized by binary variables main_cat and many_cat. main_cat indicates whether the mailing includes the main catalog of the season (which can relate to either fall/winter or spring/summer). This mailing introduces new products of the upcoming season and thus stands apart from regular mailings. The variable many_cat refers to current mailing size and is set to one if more than two other catalogs (apart from the main catalog) are mailed on the same day. Our model also includes explanatory variables which are related to past mailing decisions of the company and past purchase behavior of customers, which both may influence current purchase behavior. We deal with such dynamic effects by creating stock variables for the size of mailings which a customer has received before. This way, we model carry over effects of mailings implicitly (Hanssens, Parsons, & Schultz, 2001). We proceed in the same manner with the amount of sales of her/his previous purchases. We compute a stock variable $Z_{i\tau}$ for each customer i and mailing τ with $z_{i\tau}$

denoting either mailing size or amount of sales in the following way:

$$Z_{i\tau} = \sum_{l=1}^{\tau-1} \lambda_i^{t_\tau - t_{\tau-l}} z_{i\tau-l}, \; 0 < \lambda < 1. \tag{5}$$

The stock variables of our model may be interpreted as time discounted mailing size and sales amount, respectively. As can be seen from Eq. (5), we only take into account previous mailings starting with mailing $\tau-1$, which immediately precedes current mailing τ. We assume that the influence of a mailing diminishes over time and weight it differently by factors $\lambda_i^{t_\tau - t_{\tau-l}}$ where $t_\tau - t_{\tau-l}$ is the time span (in years) between the current mailing τ and a previous mailing $\tau-l$. For a large time span, i.e., for a mailing that has taken place a long time ago, the respective mailing size (sales amount) is multiplied by a smaller factor, for rather recent mailings the factor is larger, as the decay parameter λ_i assumes a value between zero and one. Expression (5) also implies that, given a smaller decay parameter λ_i, customer i forgets previous mailings (previous sales) more quickly. Our stock variable for previous mailings differs from that of van Diepen et al. (2009), which is based on a binary variable only. For mailing size stock we use the total number of pages across all catalogs of previous mailings as inputs $z_{i\tau-l}$ to this expression. Sales stocks are calculated in the same manner, but with sales amounts of previous mailings as inputs. Thus, our approach has the advantage to reflect differences of sizes of previous mailings. Table 1 describes the variables of our Tobit-2 model. Please note that, depending on the content of $z_{i\tau}$, $Z_{i\tau}$ translates into size_stock or sales_stock, respectively. Similar to van Diepen et al. (2009), we allow regular expressions of mailing size and sales stocks in our model and the respective quadratic terms as well. This way, we are in line with other studies that state that mailings may not generally lead to positive effects on sales (e.g. Hruschka et al., 2003; van Diepen et al., 2009). If sizes of past mailings are too high, sales may decrease displaying an inverse u-shape. According to Lind and Mehlum (2007) an inverse u-shape and hence supersaturation is present if the following equation is satisfied

$$\xi_{1i} + 2\xi_{2i}\text{stock-variable}_{i\text{lower}} > 0 \quad \text{and} \quad \xi_{1i} + 2\xi_{2i}\text{stock-variable}_{i\text{upper}} < 0 \tag{6}$$

where ξ_{1i} and ξ_{2i} are the coefficients of the respective stock variable and its quadratic expression. Stock-variable can be either size_stock or sales_stock. Lower (upper) refers to the minimum (maximum) of that stock-variable in the respective segment.

We require that Eq. (6) holds for a relative frequency of at least 95% of the sampled coefficients, which

Table 1: Variables of the Tobit-2 model

Description	Variable name
purchase decision[a]	R
amount of sales[a], in Euro	A
age, in years	age
main catalog (binary)	main_cat
many catalogs (binary)	many_cat
(quadratic) mailing size stock, in pages	size_stock
(quadratic) sales stock, in Euro	sales_stock

a Dependent variable.

corresponds to a significance level of 0.05.

3.3. Model equations

The two model equations of the Tobit-2 model for customer i and mailing τ are as follows:

$$\begin{aligned} R^*_{i\tau} &= \beta_{R0i} + \beta_{R1i}\text{age}_{i\tau} + \beta_{R2i}\text{main\_cat}_{i\tau} + \beta_{R3i}\text{many\_cat}_{i\tau} \\ &\quad + \beta_{R4i}\text{size\_stock}_{i\tau} + \beta_{R5i}\text{size\_stock}^2_{i\tau} + \beta_{R6i}\text{sales\_stock}_{i\tau} \\ &\quad + \beta_{R7i}\text{sales\_stock}^2_{i\tau} + \varepsilon_{Ri\tau} \end{aligned} \tag{7}$$

$$\begin{aligned} A^*_{i\tau} &= \beta_{A0i} + \beta_{A1i}\text{age}_{i\tau} + \beta_{A2i}\text{many\_cat}_{i\tau} + \beta_{A3i}\text{size\_stock}_{i\tau} \\ &\quad + \beta_{A4i}\text{size\_stock}^2_{i\tau} + \beta_{A5i}\text{sales\_stock}_{i\tau} + \beta_{A6i}\text{sales\_stock}^2_{i\tau} + \varepsilon_{Ai\tau}. \end{aligned} \tag{8}$$

To ensure identification of the model, we follow Cameron and Trivedi (2006) and omit the variable main_cat from the amount equation. To account for the targeting nature of mailings, we apply a two-stage residual inclusion, which is an instrumental variable based method (Terza, Basu, & Rathouz, 2008). This way, we add additional predictors to our Tobit-2 equations, which are obtained as follows. We assume that our mailing variables $\text{main\_cat}_{i\tau}$ and $\text{many\_cat}_{i\tau}$ can be estimated via latent variables $\text{main\_cat}^*_{i\tau}$ and $\text{many\_cat}^*_{i\tau}$ defined as follows:

$$\text{main\_cat}^*_{i\tau} = x'_{\text{main}i\tau}\delta_{\text{main}} + \zeta_{\text{main}i\tau} \tag{9}$$

$$\text{many\_cat}^*_{i\tau} = x'_{\text{many}i\tau}\delta_{\text{many}} + \zeta_{\text{many}i\tau}. \tag{10}$$

For standard normally distributed error terms ($\zeta_{\text{main}i\tau}, \zeta_{\text{many}i\tau}$) with mean zero in Eq. (9) and Eq. (10) we arrive at two binary probit models. We include both error terms of Eqs. (9) and (10) as additional predictors in the purchase equation and the error term of Eq. (10) as additional predictor in the amount equation.

Both vectors, $x'_{\text{main}i\tau}$ and $x'_{\text{many}i\tau}$, consist - besides a constant term - of two sets of explanatory variables. The first set accounts for the fact that the mail order company differentiates groups of customers according to their past purchasing behavior. Group 1 (g1) contains customers who purchased in each of the previous four seasons. Customers who belong to g2 (g3) purchased in three (two) seasons. Customers belonging to g4 purchased in one of the four seasons, at most. We can use grouping variables g1, g2, and g3 as instrumental variables, though they are related to response probabilities of households in previous seasons, assuming that error terms of the two binary probit models are not autocorrelated (Griliches & Hausman, 1986). The second set consists of dummy variables to indicate the timing of a mailing. These dummy variables serve as proxies for unknown time varying cost factors, which exogenously affect mailing decisions of managers in a manner similar to costs of other non-price marketing variables (Rossi, 2014). Both sets of instrumental variables (customer groups and timing variables) may affect mailing decisions of managers, but do not influence purchase and amount of sales directly.

The timing dummy variables are set to one if a mailing has been sent to any customers in a certain time interval of an observation period. Since the mail order company sends main catalogs less frequent than regular mailings, we use five and 24 dummy variables relating to Eq. (9) and Eq.(10), respectively.

3.4. Model estimation

We use Markov Chain Monte Carlo (MCMC) simulation to estimate four types of models. Our core estimation steps, within each MCMC iteration, for the Tobit-2 model together with the two probit models can be found in steps 3 and 4 in Table 2. Dependent on how unobserved heterogeneity is being addressed, more steps have to be added. Our four types of models translate into five different models. HOM is a homogeneous model. FM2 and FM3 are finite mixture models with different numbers of latent classes. MDPM is a MDP and CM is a continuous mixture model.

The estimation process of MDPM uses algorithm 7 developed by Neal (2000). We allow coefficient estimates ($\beta_{Ri}; \beta_{Ai}$), (co-) variances ($\sigma_{RAi}; \sigma^2_{Ai}$), and decay parameters (λ_i) to vary across segments. Priors of the various model parameters and the sampling of the precision parameter are described in appendices B and C. Roughly, this algorithm works in five steps. First, it distinguishes between customer segments or

Table 2: Estimation process

steps	HOM (homogeneous)	FM2-FM3 (finite mixture)	MDPM (MDP)	CM (continuous mixture)
1. deciding on creating new segments			x	
2. allocating customers among existing segments		x	x	
3. estimating Tobit-2 coefficients	x	x	x	x
4. estimating instrumental values	x	x	x	x
5. estimating precision parameter			x	
number of iterations				
burn in	10,000	20,000	50,000	60,000
estimation	10,000	50,000	50,000	50,000

singletons (each consisting of one customer only). Based on newly sampled model parameters it decides whether new segments should be formed from existing customer segments. After that, it tries to allocate singletons to one of the other segments. In the second step, it examines allocation of customers to other segments. The algorithm decides on forming new segments and on reallocation by comparing respective likelihoods, which are calculated according to Amemiya (1985). The distribution function and density function of the standard normal distribution are given by Φ and ϕ , respectively. In calculating the log-likelihood for customer i, it is important to distinguish whether in response to a mailing τ s/he makes a purchase or not. For purchases ($R_{i\tau} = 1$) of customer i the corresponding log-likelihood is given by:

$$\begin{aligned} lnL_i \quad &= \sum_{R_{i\tau}=1} \left(ln\phi \left(\frac{x'_{Ri}\beta_{Ri} + \frac{\sigma_{RAi}}{\sigma^2_{Ai}} \left(A_{i\tau} - x'_{Ai\tau}\beta_{Ai}\right)}{\sqrt{1 - \frac{\sigma^2_{RAi}}{\sigma^2_{Ai}}}} \right) \right. \\ & \left. -ln\sigma_{Ai} + \phi \left(\frac{A_{i\tau} - x'_{Ai\tau}\beta_{Ai}}{\sigma_{Ai}} \right) \right). \end{aligned} \tag{11}$$

For non-purchases ($R_{i\tau} = 0$) of customer i the following expression applies:

$$lnL_i = \sum_{R_{i\tau}=0} ln\left(1 - \phi(x'_{Ri\tau}\beta_{Ri})\right). \tag{12}$$

In the third step of the algorithm, model parameters are estimated for each segment and singleton. Thus, this procedure results in different allocations of customers after each iteration. Parameter vectors are identical for customers of the same segment, but differ for customers of different segments. Hence, we obtain on average different vectors of model parameters for each customer making the unobserved heterogeneity evident.

The fourth step deals with estimating the first stage probit models for main_cat and many_cat. As estimation is based on iterative sampling from conditional distributions, the sequence of steps does not matter (Koop 2006) and consistent coefficient estimates result for equations 7 and 8. Please note that, e.g., Lopes and Polson (2014) in each iteration sample parameters of the structural regression model before sampling parameters of the first stage regression model.

Updates of the precision parameter are obtained in step 5. This parameter determines the number of segments estimated during MCMC iterations. Small values of α will lead to a low, large values to a high number of segments (Hruschka, 2010). HOM, FM2, and FM3 are estimated in a similar manner as MDPM, by leaving out estimation steps which are necessary for MDPM only. More details concerning the MCMC algorithm and its different steps, in which parameters are sampled, can be found in appendices B and C. The continuous mixture approach is estimated following van Diepen et al. (2009).

After MCMC iterations are complete, we employ a deterministic relabeling K-means style algorithm on parameters sampled from models FM2, FM3, and MDPM. This way, we avoid the label switching problem. According to Celeux, Hurn, and Robert (2000) averages of the resulting clusters can be used as segment specific estimates.

4. Data

We analyze a random sample based on a data set provided by a mail order company, which sells one non-durable product category (i.e., apparel). Since the company wishes to stay anonymous, we can only provide rather general information. The data result from the company's day-to-day business, i.e., the company sends mailings to their customers and stores customers' purchases from the mailings including the actual level of sales. For the time period of three years, namely from fall/winter 2006 to spring/summer 2009, a random sample of their customers has been provided for our research. To disguise the actual business performance, all sales figures have been multiplied by a factor only known to the company. We know the timing of each mailing and observe for each customer i whether s/he makes a purchase as response to a mailing τ. If s/he

Table 3: Data summary

Sample information		Mailing information	
time period	3 years	number of catalogs per mailing	1-6
number of customers	1,073	mailing size, in number of pages	80-721
number of observations	57,079	number of mailings over time period	18-105
average age of customers	69.73	proportion of main_cat, in %	8.43
		proportion of many_cat, in %	51.02

makes a purchase, we observe the amount of sales s/he spends. Our estimation is based on 57,079 $(= \sum_i T_i)$ observations where T_i denotes the number of mailings customer i receives during the three years. Our amount of observations translates into 1,073 customers. These are mainly from the demographic cohorts of the silent generation and the baby boomers with an average age of almost 70 corresponding to the company's target group. They receive mailings on average every 16 days. In more than 8% (51%) of the observations a main catalog (large mailing) has been targeted to the customers. Of these 57,079 observations 39.57% and 31.47% fall into customer groups g1 and g2, respectively. g3 contains 22.01% of the observations. Table 3 gives a short summary of the data.

5. Results

5.1. Model performance

5.1.1. Model fit

We evaluate the performance of different models by PsBF, which are based on cross-validation predictive densities (Ansari & Mela, 2003). The logarithm of a PsBF equals the difference of log pseudo marginal likelihood values (pml) across customers and mailings for two models M_A and M_B as can be seen from Eq. (13):

$$logPsBF = \sum_i \sum_{\tau=1}^{T_i} log(pml_{i\tau,M_A}) - \sum_i \sum_{\tau=1}^{T_i} log(pml_{i\tau,M_B}). \tag{13}$$

The pml corresponds to the likelihood of observed values of the dependent variables for mailing τ of customer i if model M_A is fitted to all data except this observation. It is computed as harmonic mean of conditional likelihoods across selected parameter samples.

Results can be found in Table 4. Following Kass and Raftery (1995), there is a strong evidence against the homogeneous model compared to any heterogeneous model, which shows that accounting for heterogeneity clearly enhances model fit. We also look at log(pml) values in order to choose from the three models

Table 4: Model fit

Model	log(pml)
HOM	-26,124.084
FM2; 2 classes	-25,457.645
FM3; 3 classes	-25,962.472
MDPM; on average 2 classes	-25,554.270
CM	-26,032.106

having different forms of heterogeneity. We obtain the highest log(pml) for a finite mixture model with two classes. In fact, except for three customers, this model finds the same clustering solution as the MDP. The MDP produced as well on average two classes. However, in 11.12% of our samples it estimated between 3 and 5 classes, which can be responsible for less stable estimates. Hence, the advantage of the MDP of being able to estimate the number of segments alongside the model estimation comes at a price.

The continuous mixture approach leads to the worst model fit of the heterogeneous models. This result is consistent with Ansari and Mela (2003) who also find that the continuous mixture performs worse compared to the MDP. The bad performance of the continuous mixture approach could be due to multimodal or skewed distributions of unobserved characteristics of the customers which discrete mixtures of distributions can comply with.

5.1.2. Model tests

In addition, we perform a couple of tests to examine the reliability of our results. We thank one anonymous reviewer for drawing our attention to these points. First, we examine whether estimation results are sensitive to excluding the variable main_cat from Eq. (8). Let us remind you that we proceeded this way to avoid the identification problem, which results if the same predictors appear both in the purchase and the sales equations of the Tobit-2 model (Cameron & Trivedi, 2006). We hence regress the residuals of Eq. (8) on main_cat. As the difference of the two BIC-values between Eq. (8) and this regression is close to zero (Raftery 1995), we do not lose much information by excluding the variable main_cat.

Second, we investigate the quality of our instrumental variables. We refer to the procedure of Stock, Wright, and Yogo (2002) to test for weak instruments. This way, we regress each of our endogenous variable on the remaining exogenous and instrumental variables. By means of a likelihood ratio test, we compare these models to two models that regress the respective endogenous variable on a constant term only. Since both likelihood ratio tests are significant (p-value <0.05), we can rule out the issue of weak instruments.

Besides, we also test for an over-identification issue as we use more instrumental variables than endogenous variables. We follow Bollen, Guilkey, and Mroz (1995) in testing for over-identification. To this end, we estimate another version of FM2. We include all instrumental variables except one into Eqs.(7) and (8). By means of a χ^2-test we find that these additional predictors do not improve the model significantly (p-value >0.05) and that hence over-identification is not an issue.

5.2. *Estimation differences between FM2, MDPM, and CM*

We now investigate whether models FM2, MDPM, and CM result in different parameter estimates. To test differences in parameter estimates, we calculate for each customer i average coefficient estimates that result for each of the three alternative models. Table 5 shows coefficient estimates of the three models FM2, MDPM, and CM. We use paired t-tests to examine differences between model pairs. If the ratio of mean and standard deviation of the samples of estimates exceeds the critical t-value of 1.96 in absolute terms, differences are significant at the 5% level. FM2 and MDPM do not differ much, the only exception being a significant difference for the decay parameter of size_stock.

On the other hand we obtain significant differences for a couple of coefficients when comparing FM2 to CM. Both models differ w.r.t. estimates for age (in both equations) and some estimates for size_stock and sales_stock. In addition, both decay parameters differ depending on which estimation technique is used. Average coefficients of the significantly different size_stock, size_stock^2, and sales_stock^2 variables are negative for both models. These significant differences show that CM finds a larger negative effect of size_stock and sales_stock^2 on purchase probability than FM2. According to FM2 on the other hand, size_stock^2 and sales_stock have a more negative effect on the amount equation than stated by CM. Finally, FM2 implies lower values of size_stock and sales_stock variables as can be seen from significant differences for decay parameters.

To summarize, estimation results and hence managerial implications may differ between alternative ways to account for unobserved heterogeneity. We will return to this issue in more detail in subsection 5.4. In our case however, results on model fit clearly demonstrate that FM2 is superior to the other models and should in consequence be chosen. This is why we give more detailed results on this model in the following subsection.

Table 5: Coefficient estimates of models FM2, MDPM, and CM

Variables	FM2		MDPM		CM	
	Purchase	Amount	Purchase	Amount	Purchase	Amount
Tobit-2						
constant	-1.5586	4.3498	-1.5668	4.3000	-1.5150	3.8790
age	0.0043	0.0317	0.0047	0.0320	**0.0213**	**-0.0150**
main_cat	0.8811		0.8810		0.8631	
many_cat	0.1429	-0.4642	0.1429	-0.4661	0.1376	-0.6925
size_stock	-0.0572	-0.0365	-0.0575	-0.0390	**-0.0811**	-0.1318
size_stock^2	0.0294	-0.0680	0.0270	-0.0680	0.0232	**-0.0120**
sales_stock	0.3128	-0.2815	0.3133	-0.2759	0.3179	**0.1246**
sales_stock^2	-0.0451	0.0572	-0.0478	0.0567	**-0.0776**	0.0097
$\zeta_{maini\tau}$	-0.0021		-0.0024		-0.0017	
$\zeta_{manyi\tau}$	-0.0063	0.0520	-0.0063	0.0507	-0.0055	**0.1010**
Other						
σ^2_{Ai}	4.3394		4.3164		4.1031	
λ_i (size)	0.6508		**0.6585**		**0.7233**	
λ_i (sales)	0.4011		0.4029		**0.4627**	
σ_{RAi}	-0.7775		-0.7541		-0.5208	

Significant differences at 0.05 of MDPM and CM compared to FM2 based on paired t-tests in boldface.

5.3. Empirical results based on FM2

We discuss segment specific results, which we obtain by the procedure explained in subsection 3.4. Table 6 contains averages of sampled parameters for each segment of FM2. This table also indicates whether a parameter is significant at the 5% level.

Average coefficients are different between segments. This is especially true for coefficients of the amount equation. Customers belonging to segment one have a higher estimate for the constant term. In addition, these customers react more negatively towards large current mailings as expressed by the coefficient of many_cat. The same is true for past mailings. Both size_stock coefficients are significant and negative for these customers. In contrast, only the coefficient of the quadratic size_stock is significant and negative for customers belonging to segment two. Finally, past sales have a larger negative effect on amount of sales. Besides, we also note great differences of variances and covariances between segments, which show that allowing for heterogeneity of these parameters is beneficial. Both segment specific average covariances are negative. These negative correlations reflect the fact that customers, whose purchase probabilities are higher than expected, purchase more often, but also spend less than expected on each purchase. These expected values are conditional on all the explanatory variables of the model. Decay parameters have similar average

Table 6: Segment specific estimates for FM2

Variables	Seg. 1		Seg. 2	
	Purchase	Amount	Purchase	Amount
Tobit-2				
constant	**-1.6638**	**5.3476**	**-1.4479**	**3.3004**
age	**0.0029**	**0.0235**	**0.0058**	**0.0404**
main_cat	**0.9683**		**0.7896**	
many_cat	**0.1476**	**-0.5884**	**0.1378**	**-0.3336**
size_stock	**-0.0463**	**-0.0869**	**-0.0686**	0.0164
size_stock^2	**0.0315**	**-0.0723**	**0.0271**	**-0.0635**
sales_stock	**0.2929**	**-0.4239**	**0.3338**	**-0.1317**
sales_stock^2	**-0.0404**	**0.0880**	**-0.0499**	**0.0249**
$\zeta_{maini\tau}$	**-0.0024**		**-0.0018**	
$\zeta_{manyi\tau}$	**-0.0085**	**0.0645**	**-0.0039**	**0.0388**
Other				
σ^2_{Ai}	**6.1240**		**2.4626**	
λ_i (size)	**0.6410**		**0.6612**	
λ_i (sales)	**0.4223**		**0.3789**	
σ_{RAi}	**-0.9752**		**-0.5697**	
segment size (in %)	51		49	

Significant results at 0.05 in boldface.

values for both segments.

Now we deal with assessing the impact of our mailing variables. Sending main catalogs to customers leads to an increase in purchase probability that is 6.6 and 5.7 times as high as the increase due to a large mailing. The positive effect of a main catalog can be explained with the nature of such a mailing. Main catalogs advertise all products of a mail order company within a season, that typically lasts for six months. They are usually sent at the beginning of a season and are hence the first mailing, which contains new products. In addition, they are valid throughout the whole season, which gives customers more time to make (multiple) purchases.

The effect of many_cat is mixed. For both segments, including many catalogs in a mailing might increase the probability that the customer notices the mailing, which is shown by a positive impact on purchase probability. The negative effect on amount of sales could be due to too much information consistent with research on choice overload (e.g., Iyengar & Lepper, 2000). It especially reduces amount of sales of customers that belong to segment one. Consistent with findings from van Diepen et al. (2009) we are able to show that too many and/or too large mailings may lead to frustration. W.r.t. the stock of past mailings we even observe supersaturation effects on amount of sales for customers in segment two. Here, Eq. (6) holds for 99.16% of the sampled coefficients. We proceed in the same manner when investigating the effect that a stock of past amount of sales may have on response behavior. Again, Eq. (6) holds for 100% of the sampled coefficients in segment one when the purchase equation is concerned.

5.4. Managerial implications

In this subsection we intend to show how different approaches for capturing unobserved heterogeneity may lead to different managerial implications. We therefore compare FM2, our model that has the best model fit based on PsBF (see subsection 5.1.1), and CM, which shows some different estimates based on Table 5. This way, we can show how managerial implications might deviate if another model, instead of the optimal, had been chosen. We base our managerial implications on two tables, i.e., Tables 6 and 7. Table 7 contains further information on the different segments. Besides, we use the segmentation solution based on FM2 and calculate average coefficient estimates that result under CM for this segmentation.

Table 7 contains segment specific averages of purchase probabilities and amount of sales in addition to segment specific averages of decision variables set by the mail order company. One of these averages concerns the company's timing decision. Furthermore, the company decides on the composition of the mailings as well. In this respect we give the proportion of mailings which include the main catalog. For segment one, e.g., 8.28% of all mailings contain a main catalog. As the main catalog is usually sent once per season to a customer, this proportion is lower for customers receiving many mailings compared to customers that are not exposed to such a dense mailing policy. Therefore, we also calculate the proportion of customers who receive at least one main catalog per season over the three year period. Table 7 also shows the proportion of mailings with more than two other catalogs aside from the main catalog. Our implications refer to the company's decision variables, i.e. the mailing of a main catalog (main_cat) and mailing many catalogs (many_cat). Both variables together with the timing decision influence size_stock.

As can be seen from Table 7, segments one and two differ mainly in the amount of sales per purchase and the purchase probability. Hence, managerial suggestions should mainly refer to the purchase (amount) equation of segment one (two).

5.4.1. Implications regarding the purchase equation

According to FM2 (see Table 6), sending a main catalog to customers in segment one has a larger impact on purchase probability than for customers in segment two. Hence, sending more main catalogs to customers of segment one is clearly in accordance with the goal of increasing the purchase probability of

Table 7: Segment specific statistics

	Seg. 1	Seg. 2	Average[a]
purchase probability, in %	7.18	9.84	8.49
amount of sales per purchase, in Euro	130.53	71.31	96.76
time between mailings, in days	16.53	16.30	16.42
proportion of main catalog, in %	8.28	8.6	8.43
proportion of customers receiving on average at least one main catalog per season, in %	59.82	66.92	63.19
proportion of more than two other catalogs, in %	50.44	51.61	51.02

a Average across all mailings.

these customers. CM on the other hand also finds a positive effect of main catalogs on purchase probability. However, that model does not find segment-specific effects of that variable (i.e.,0.8651 (0.8610) for segment one (two)). Since the coefficient estimate of main_cat in segment one is smaller than the one in FM2, FM2 would recommend fewer additional main catalogs than CM to reach the same effect.

In addition, FM2 (CM) finds supersaturation effects of past purchases on purchase probability for segment one (and segment two). Hence, the company should only target customers with mailings if no recent or large purchases have taken place. According to CM this holds as well for customers belonging to segment two.

5.4.2. Implications regarding the amount equation

W.r.t. the amount equation, Table 7 shows that customers belonging to segment two have a smaller amount of sales per purchase than customers belonging to segment one. Hence, increasing the amount of sales is of vital importance especially for these customers. Generally, reducing the amount of many_cat supports this goal. However, according to FM2, this holds even more for customers belonging to segment one than for segment two. Also, CM finds a larger negative impact of many_cat in segment two (i.e., -0.6952) than FM2. Hence, FM2 would require a larger reduction of many_cat than does CM in order to reach the same effect.

Furthermore, according to FM2, customers in segment two show supersaturation effects due to past mailings. If the company reduces size_stock (e.g., through a smaller amount of many_cat), this will have a positive impact on amount of sales. This implication is only based on FM2 as CM does not find a supersaturation effect.

6. Conclusion

Our results on unobserved heterogeneity show that a finite mixture approach is preferable to a MDP or continuous mixture approach. Consistent with findings on the overall model fit, a finite mixture and a MDP are more similar to each other and different from the continuous mixture model. A MDP and a finite mixture both result in the same number of segments leading to a more or less identical clustering structure. Except for estimates on the size_stock decay parameter, we do not observe any significant differences on estimates. However, the advantage of the MDP of not having to estimate several models comes at a price: the overall model fit is worse for the MDP compared to the finite mixture.

In addition, a finite mixture is superior to the continuous mixture w.r.t. PsBF. Both models value the impact of mailing variables differently and would thus lead to different managerial implications. In total, FM2 recommends fewer mailings than CM. Based on our findings, we would suggest future researchers to apply a finite mixture model in contrast to a MDP or a continuous mixture. Of course, the number of segments obtained here is not generally valid and has to be determined anew for any other data set.

We find that responses to mailings differ depending on type and sizes of mailings. In addition, we show that mailings of the sponsoring company are subject to (super-) saturation effects on the amount of sales per purchase. Firstly, in all segments, mailings with many catalogs reduce the amounts of sales. Secondly, many or frequent mailings in the past have a negative impact for 49% of the customers. These results suggest that managers of mail order companies should not ignore saturation when they decide about sending mailings to individual customers. More importantly, the company runs the risk of (super-) saturation effects.

Of course, our study is not free of limitations. First, we base our model on conventional mailings, which are in competition with online channels. However, the rising amount of spendings for regular mailings (statista, 2016) and results from previous studies show that these mailings are still important. First of all, among the various types of customers some prefer to be targeted with online mailings while others prefer mailings. Thomas and Sullivan (2005) find evidence for channel loyalty i.e., only some customers switch from offline to online channels once these are introduced. In the same line, Goldsmith and Flynn (2005) show that older customers tend to make more purchases based on mailings. Second, there is also evidence for a positive interaction between mailings and online communications. Thomas und Sullivan (2005) find that multichannel customers generate more sales than mono-channel customers. In most of their scenarios they advise that customers should also be targeted with mailings (besides another channel). According to the Direct Marketing Association (2014) response rates to mailings are much higher than to e-mails. In summary, even though online communications are becoming more and more important, it is still advisable to consider mailings as we do in this article. Although the results, and in particular our managerial implications, are based on our data set, the estimation process is generalizable to the online context. This can be seen in studies of, e.g., Ansari and Mela (2003) or Zantedeschi et al. (2016) that also need to model unobserved heterogeneity.

Another limitation of our study is due to our customer group. Findings on mailing effects might not be generalized since the average age of the analyzed customers is way above the average age of any (European) country. We thank one anonymous reviewer for pointing that out.

Acknowledgement

We thank three anonymous reviewers and the editor of the journal for their interest and feedback, which helped us to improve the article.

Appendix A - Glossary

a	parameter to sample α
$A_{\gamma i}$	parameter to sample residual covariance matrix of the Tobit-2 model
$A_{i\tau}$	amount of sales of customer i to a mailing τ, (measured in Euro); refers to the sales equation of the Tobit-2 model
α	precision parameter that determines the number of segments estimated during MCMC simulations of MDPM
b	parameter to sample α
B	parameter to sample residual covariance matrix
β_{Ai}	vector of coefficients of customer i; refers to the amount equation of the Tobit-2 model
β_{Ri}	vector of coefficients of customer i; refers to the purchase equation of the Tobit-2 model
$\overline{\beta}, \underline{\beta}$	parameter to sample coefficients of the Tobit-2 model
$Beta$	Beta distribution

C	number of segments
CM	continuous mixture model
d	normalizing constant
D_δ	parameter to sample residual covariance matrix of the probit models
δ	vector of coefficients in the probit models to describe the mailing variables main_cat or many_cat
$\overline{\delta}$	parameter to sample coefficients of the probit models
$\varepsilon_{Ri\tau}$	vector of residuals of customer i and mailing τ; refers to the purchase equation of the Tobit-2 model
$\varepsilon_{Ai\tau}$	vector of residuals of customer i and mailing τ; refers to the sales equation of the Tobit-2 model
η	parameter to sample α
F	parameter to sample residual covariance matrix
FM2	finite mixture model with 2 segments
FM3	finite mixture model with 3 segments
g1-g4	groups of customers according to past purchase behavior
Gamma	Gamma distribution
$\gamma_i, \overline{\gamma}$	parameter to sample residual covariance matrix
$H_i, \Sigma_{\varepsilon i}$	residual covariance matrix of the Tobit-2 model
i	customer index
IG	density of the inverse gamma distribution
κ	parameter to sample residual covariance matrix
lnL_i	log-likelihood for customer i
L	likelihood
λ_i	decay parameter; shows the influence of previous mailing sizes or amounts of sales when calculating the respective stock variable
MCMC	Markov Chain Monte Carlo
MDP	mixture of Dirichlet Processes
MDPM	mixture of Dirichlet Processes model
MVN	density of the multivariate normal distribution

n	number of customers
$n_{-i,c}$	the number of customers of segment c without customer i
N	density of the normal distribution
$N_{tr>0}$	normal distribution truncated to positive values
$N_{tr<0}$	normal distribution truncated to negative values
pml	pseudo marginal likelihood values
PsBF	pseudo Bayes factors
ϕ, Φ	distribution function and density function of the standard normal distribution
φ_i	auxiliary variable for customer i referring to the decay parameter
$R_{i\tau}$	response of customer i to a mailing τ, (binary variable); refers to the purchase equation of the Tobit-2 model
ρ_i	correlation of the residuals of customer i between the purchase equation and sales equation of the Tobit-2 model
S_i	parameter to sample residual covariance matrix equation of the Tobit-2 model for customer i
σ^2_{Ai}	variance of residuals of customer i; refers to the amount of sales equation of the Tobit-2 model
σ_{RAi}	covariance of the residuals of customer i between the purchase equation and sales equation of the Tobit-2 model
$t_\tau - t_{\tau-l}$	time span (in years) between the current mailing τ and a previous mailing $\tau - l$
T_i	number of mailings customer i receives during the three years
τ	mailing event index
$\overline{V}, \underline{V}$	parameter to sample coefficients of the Tobit-2 model
w	parameter to sample α
$x_{Ai\tau}, x_{Ri\tau}$	vectors of explanatory variables of customer i and mailing τ; refers to the Tobit-2 model
$x_{\text{main}i\tau}, x_{\text{main}i\tau}$	vectors of explanatory variables of customer i and mailing τ; refer to the probit models
ξ_{1i}, ξ_{2i}	coefficients of the stock variables and their quadratics

$z_{i\tau-I}$	either mailing size or amount of sales; input for calculating the stock variable
$Z_{i\tau}$	stock variable for customer i and mailing τ
$\zeta_{maini\tau}$	vector of residuals for customer i and mailing event τ; refers to probit equation w.r.t. the variable main_cat
$\zeta_{manyi\tau}$	vector of residuals for customer i and mailing event τ; refers to probit equation w.r.t. the variable many_cat

Appendix B - Overview of the MCMC Simulation

Segments are indexed as $1 \le c \le C$ with c as the number of segments currently formed. n and $n_{-i,c}$ denote the number of customers and the number of customers of segment c without customer i, respectively. One MCMC iteration for the MDP consists of the following steps (of which steps 1-3 correspond to algorithm 7 (Neal, 2000)).

1a. For each customer segment:

aa. Draw new samples of coefficients and of the residual covariance matrix from their respective prior distributions (see appendix C) and calculate the corresponding likelihood.

ab. Decide whether to form a new segment c_i^* with probability

$$\min\left[1, \frac{\alpha\, L_i^*}{(n-1)L_i}\right]. \tag{B.1}$$

1b. For each singleton:

ba. Decide on allocating the singleton to one of the segments with probability proportional to

$$\frac{n_{-i,c}}{n-1}. \tag{B.2}$$

bb. Calculate the corresponding likelihood and place the singleton into the new segment c_i^* with probability proportional to

$$\min\left[1, \frac{(n-1)L_i^*}{\alpha\, L_i}\right]. \tag{B.3}$$

2. Decide on allocating each customer belonging to a segment to one of the remaining segments with probabilities proportional to

$$d\,\frac{n_{-i,c}}{n-1}L_i \tag{B.4}$$

where d serves as normalizing constant.

3. Estimate parameters for each segment as described in appendix C.
4. Sample coefficients of the binary probit functions to correct for the targeting nature of the mailings. We unify notations of both probit models to get

$$\text{target}^*_{i\tau} = x'_{\text{target}}\delta_{\text{target}} + \zeta_{\text{target}i\tau} \tag{B.5}$$

where target might either relate to Eq. (9) or Eq. (10), respectively. We sample δ_{target} and the latent dependent variable according to Koop (2006).

5. Sample precision parameter α using data augmentation (West, 1992):

$$\begin{aligned} \eta &\sim Beta(\alpha+1, \sum_i T_i) \\ \alpha &\sim w\,Gamma(a+c, 1/(b-ln\eta)) + \\ &\quad (1-w)\,Gamma(a+c-1, 1/(b-ln\eta)) \\ &\quad \text{with} \quad w = (a+c-1)/(a+c-1+(b-ln\eta)\sum_i T_i). \end{aligned} \tag{B.6}$$

Beta and Gamma denote the $Beta$ and $Gamma$ distributions, respectively. The prior is $\alpha \sim Gamma(a,b)$ with a=0.001, b=0.001 according to Hruschka (2010).

For the continuous mixture approach we make use of the algorithm described in van Diepen et al. (2009). The finite mixture approach is obtained with the MDP approach cycling between steps 2-4.

Appendix C - Parameter Samplings for Segments

Let us remind you that in any iteration parameter values sampled are the same for all customers which belong to the same segment. The steps described here concern all customers who in the current iteration are assigned to segment c, i.e. all $i \in c$.

N, MVN, IG denote densities of the normal, the multivariate normal, and the inverse Gamma distributions, respectively. $\mathrm{N}_{tr>0}, N_{tr<0}$ symbolize densities of normal distributions truncated to positive and negative values, respectively.

1. Sampling of latent variables includes data augmentation according to Tanner and Wong (1987) and follows Koop, Poirier, and Tobias (2007) and van Diepen et al. (2009). It is necessary to distinguish

between purchases and non-purchases:

Purchases:

$$A^*_{i\tau} = A_{i\tau} \tag{C.1}$$

$$R^*_{i\tau} \sim N_{tr>0}\left(x'_{Ri\tau}\beta_{Ri} + \rho_i \frac{A^*_{i\tau} - x'_{Ai\tau}\beta_{Ai}}{\sigma_{Ai}}, 1-\rho_i^2\right) \tag{C.2}$$

Non-purchases:

$$R^*_{i\tau} \sim N_{tr<0}\left(x'_{Ri\tau}\beta_{Ri}, 1\right) \tag{C.3}$$

$$A^*_{i\tau} \sim N\left(x'_{Ai\tau}\beta_{Ai} + \sigma_{Ai}\rho_i\left(R^*_{i\tau} - x'_{Ri\tau}\beta_{Ri}\right), \left(1-\rho_i^2\right)\sigma^2_{Ai}\right) \tag{C.4}$$

with $\rho_i = \sigma_{RAi}/\sigma_{Ai}$.

2. Coefficients are sampled as follows (Koop, 2006):

$$\beta_i \mid R_{i\tau}, A_{i\tau}, H_i \sim MVN(\overline{\beta}, \overline{V}) \tag{C.5}$$

$$\overline{V} = \left(\underline{V}^{-1} + \sum_{i\in c}\sum_{\tau=1}^{T_i} X'_{i\tau}H_iX_{i\tau}\right)^{-1} \tag{C.6}$$

$$\overline{\beta} = \overline{V}\left(\underline{V}^{-1}\underline{\beta} + \sum_{i\in c}\sum_{\tau=1}^{T_i} X'_{i\tau}H_iy_{i\tau}\right) \tag{C.7}$$

with $\underline{\beta} \sim MVN(0, \underline{V})$ as prior distribution and H_i as inverse of the residual covariance matrix $\Sigma_{\varepsilon i}$. The vector $y_{i\tau} = \begin{pmatrix} R^*_{i\tau} \\ A^*_{i\tau} \end{pmatrix}$ stacks all observations of the dependent variables, $\beta_i = \begin{pmatrix} \beta_{Ri} \\ \beta_{Ai} \end{pmatrix}$ the coefficients of both equations. The data matrix is defined as: $X_{i\tau} = \begin{pmatrix} x'_{Ri\tau} & 0 \\ 0 & x'_{Ai\tau} \end{pmatrix}$.

3. Sampling of the residual covariance matrix

According to McCulloch, Polson, and Rossi (2000) and van Diepen et al. (2009) we use the reparametrization

$$\Sigma_{\varepsilon i} = \begin{pmatrix} 1 & \gamma_i \\ \gamma_i & S_i + \gamma_i^2 \end{pmatrix}. \tag{C.8}$$

These parameters are sampled as follows:

$$S_i \sim IG\left(\kappa + \sum_{i \in c} T_i, F + \sum_{i \in c} \sum_{\tau=1}^{T_i} (\varepsilon_{Ai\tau} - \varepsilon_{Ri\tau}\gamma_i)^2\right) \tag{C.9}$$

$$\gamma_i \sim N\left(A_{\gamma i}\left(\frac{\sum_{i \in c}\sum_{\tau=1}^{T_i} \varepsilon_{Ai\tau}\varepsilon_{Ri\tau}}{S_i} + B\overline{\gamma}\right), A_{\gamma i}\right) \tag{C.10}$$

$$A_{\gamma i} = \left(\frac{\sum_{i \in c}\sum_{\tau=1}^{T_i} \varepsilon_{Ri\tau}^2}{S_i} + B\right)^{-1} \tag{C.11}$$

with $S_i \sim IG(\kappa, F)$ and $\gamma_i \sim N(\overline{\gamma}, B^{-1})$ as priors. $\overline{\gamma} = 0$, B=10, $\kappa = 3$, and F=(1-B^{-1})$(\kappa - 1)$ are chosen as suggested by McCulloch et al. (2000).

4. A Metropolis-Hastings step serves to sample decay parameters $\lambda_i = (\lambda_{1i}, \lambda_{2i})$ where λ_{1i} is the decay parameter for calculating mailing size and λ_{2i} for sales stock, respectively (Ansari et al., 2008). To ensure that λ_i lies in the interval $[0,1]$, it is obtained as $\lambda_i = \exp(\varphi_i)/(1+\exp(\varphi_i))$ where $\varphi_i = (\varphi_{1i}, \varphi_{2i})$ in analogy to λ_i. Draws are accepted based on the product of likelihood and prior

$$\prod_{i \in c}\prod_{\tau=1}^{T_i} \exp\left(-\frac{1}{2}\varepsilon_{i\tau}'\Sigma_{\varepsilon i}^{-1}\varepsilon_{i\tau}\right)\exp\left(\frac{-\varphi_i^2}{2}\right) \tag{C.12}$$

with $\varphi_i \sim N(0,1)$ as prior. Each element of φ_i is drawn sequentially from the normal distribution with mean equal to its previous draw and the variance chosen in order to get acceptance probabilities of about 45%. Please note that residuals $\varepsilon_{i\tau} = (\varepsilon_{Ri\tau}, \varepsilon_{Ai\tau})$ change given another value of a decay parameter due to fact that the corresponding stock variable is modified according to Eq. (5).

References

[1] Amemiya, T. (1985). *Advanced econometrics*. Cambridge: Harvard University Press, (Chapter: 10).

[2] Ansari, A. & Mela, C.F. (2003). E-Customization. *Journal of Marketing Research*, 40, 131-145.

[3] Arellano, M. & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies*, 58, 277- 297.

[4] Bollen, K.A., Guilkey, D.K., & Mroz, T.A. (1995). Binary outcomes and endogenous explanatory variables: tests and solutions with an application to the demand for contraceptive use in Tunisia. *Demography*, 32, 111-131.

[5] Cameron, A.C. & Trivedi, P. (2006). *Microeconometrics methods and applications*. (8th ed.). Cambridge: Cambridge Univ. Press, (Chapter 16).

[6] Celeux, G., Hurn, M., & Robert, C.P. (2000). Computational and inferential difficulties with mixture posterior distributions. *Journal of the American Statistical Association*, 95, 957-970.

[7] Direct Marketing Association. Why direct mail in the 21st century? (2014). http://www.dma.org.uk/article/why-direct-mail-in-the-21st-century-1 Accessed 13.12.16.

[8] Gönül, F., Kim, B.D., & Shi, M. (2000). Mailing smarter to catalog customers. *Journal of Interactive Marketing*, 14, 2-16.

[9] Goldsmith, R.E. & Flynn, L.R. (2005). Bricks, clicks, and pix: apparel buyers' use of stores, internet, and catalogs compared. *International Journal of Retail & Distribution Management*, 33, 271-283.

[10] Griliches, Z. & Hausman, J.A. (1986). Errors in variables in panel data. *Journal of Econometrics*, 31, 93-118.

[11] Guelman, L., Guillén, M., & Pérez-Marín, A. M. (2015). A decision support framework to implement optimal personalized marketing interventions, *Decision Support Systems*, 72 , 24-32.

[12] Hanssens, D.M., Parsons, L.J., & Schultz, R.L. (2001). *Market response models, econometric and time series analysis*. (2nd ed.). Boston: Kluwer Academic Publishers, (Chapter 4).

[13] Hruschka, H., Baumgartner, B., & Semmler, M. (2003). Wirkungsmessung und Allokation von Katalogen in Versandhandel und Direktmarketing: Katalogversand an Sammelbesteller. *Zeitschrift für Betriebswirt*, 73, 7-23.

[14] Hruschka, H. (2010). Considering endogeneity for optimal catalog allocation in direct marketing. *European Journal of Operational Research*, 206, 239-247.

[15] Iyengar, S.S. & Lepper, M.R. (2000). When choice is demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology*, 79, 995-1006.

[16] Kass, R.E. & Raftery, A.E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90, 773-795.

[17] Koop, G. (2006). *Bayesian econometrics*. Chichester: Wiley, (Chapter 6).

[18] Koop. G., Poirier, D.J., & Tobias, J.L. (2007). *Bayesian econometric methods*. Cambridge: Cambridge University Press, (Chapter 14).

[19] Levin, N. & Zahavi, J. (1998). Continuous predictive modeling - a comparative analysis. *Journal of Interactive Marketing*, 12, 5-22.

[20] Lind, J.T. & Mehlum, H. (2007). With or without u? - The appropriate test for a u shaped relationship. Munich Personal RePEc Archive 4823, 1-12.

[21] Lopes, H.F. & Polson, N.G. (2014). Bayesian instrumental variables: priors and likelihoods. *Econometric Reviews*, 33, 100-121.

[22] Ma, S., Hou, L., Yao, W., & Lee, B. (2016). A nonhomegeneous hidden markov model of response dynamics and mailing optimization in direct marketing. *European Journal of Operational Research*, 253, 514-523.

[23] McCulloch, R.E., Polson, N.G., & Rossi, P.E. (2000). A bayesian analysis of the multinomial probit model with fully identified parameters. *Journal of Econometrics*, 99, 173-193.

[24] Neal, R. (2000). Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational Graphical Statistics*, 9, 249-265.

[25] Olson, D.L. & Chae,B.(K.) (2012). Direct marketing decision support through predictive customer response modeling. *Decision Support Systems*, 54, 443-451.

[26] Popkowski Leszczyc, P.T.L. & Bass, F.M. (1998). Determining the effects of observed and unobserved heterogeneity on consumer brand choice. *Applied Stochastic Models and Data Analysis*, 14, 95-115.

[27] Raftery, A.E. (1995). Bayesian model selection in social research. *Sociological Methodology*, 25, 111-163.

[28] Rossi, P. E. (2014). Even the rich can make themselves poor: a critical examination of IV methods in marketing applications. *Marketing Science*, 33, 655672.

[29] statista. Advertising spending in the catalog, mail-order houses industry in the United States from 2010 to 2016. (2016). https://www.statista.com/statistics/470620/catalog-mail-order-houses-industry-ad-spend-usa/ Accessed 13.12.16.

[30] Stock, J.H., Wright, J.H., & Yogo, M. (2002). A survey of weak instruments and weak identification in Generalized Method of Moments. *Journal of Business Economics and Statistics*, 20, 518-529.

[31] Tanner, M.A. & Wong, W.H. (1987). The calculation of porsterior distributions by data augmentation. *Journal of the American Statistical Association*, 82, 528-540.

[32] Terza, J.V., Basu, A., & Rathouz, P.J. (2008). Two-stage residual inclusion estimation: addressing endogeneity in health econometric modeling. *Journal of Health Economic*, 27, 531-543.

[33] Thomas, J.S. & Sullivan, U.Y. (2005). Managing marketing communications with multichannel customers. *Journal of Marketing*, 69, 239-251.

[34] Van Diepen, M., Donkers, B., & Franses, P.H. (2009). Dynamic and competitive effects of direct mailings: a charitable giving application. *Journal of Marketing Research*, 46, 120-133.

[35] Verhoef, P.C., Spring, P.N., Hoekstra, J.C., & Leeflang,P.S.H. (2002). The commercial use of segmentation and predictive modeling techniques for database marketing in the Netherlands. *Decision Support Systems*, 34 471-481.

[36] West, M. (1992). Hyperparameter estimation in Dirichlet process mixture models. ISDS Discussion Paper Duke University.

[37] Zantedeschi, D., McDonnell Feit, E., & Bradlow, E.T. (2016). Measuring multichannel advertising response. *Management Science*, Published online in Articles in Advance 08 Jul 2016.

Biographical Note

Nadine Schrder

Nadine Schrder holds a postdoc position at the University of Regensburg where she obtained her Ph.D. in focusing on response modeling and optimization in direct marketing. Her current research interests include direct marketing and sales response models. She holds a MSc. in statistics from the University of Georgia and a diploma in business administration from the University of Osnabrck. She has professional experience in response modeling in the mail order industry.

Harald Hruschka

Harald Hruschka holds the Marketing Chair at the University of Regensburg, Germany. His main research interests comprise brand choice and sales response models, direct marketing, semiparametric models including neural nets, and hierarchical Bayesian models. He has published in journals such as Journal of Forecasting, European Journal of Operational Research, OR Spectrum, Marketing Letters, Journal of Retailing and Consumer Services, International Journal of Research in Marketing, and others. Before joining the University of Regensburg he was associate professor at the Vienna University of Business Administration and Economics, where he also obtained his Ph.D.

Highlights

- Investigating different approaches of latent heterogeneity of customers
- The continuous mixture is inferior to Dirichlet process and finite mixture models
- Derivation of segment specific implications for mailings
- Measuring the effect of different mailing variables on purchase probability and sales