Incorporating Responsiveness to Marketing Efforts in Brand Choice Modeling

Dennis Fok*
Econometric Institute
Erasmus University Rotterdam

Richard Paap
Econometric Institute
Erasmus University Rotterdam

Philip Hans Franses
Econometric Institute
Erasmus University Rotterdam

ECONOMETRIC INSTITUTE REPORT EI 2008-15

Abstract

We put forward a brand choice model with unobserved heterogeneity that concerns responsiveness to marketing efforts. We introduce two latent segments of households. The first segment is assumed to respond to marketing efforts while households in the second segment do not do so. Whether a specific household is a member of the first or the second segment at a specific purchase occasion is described by household-specific characteristics and characteristics concerning buying behavior. Households may switch between the two responsiveness states over time.

When comparing the performance of our model with alternative choice models that account for various forms of heterogeneity for tree different datasets, we find better face validity of our parameters. Our model also forecasts better.

Key words: marketing-instrument effectiveness, heterogeneity, Multinomial Probit, mixtures

*We thank the editor and two anonymous referees for their helpful and detailed suggestions. We thank Michel Wedel and Peter Rossi for sharing useful ideas. Address for correspondence: D. Fok, Erasmus University Rotterdam, Econometric Institute, H11-2, P.O. Box 1738, NL-3000 DR Rotterdam, The Netherlands, `e-mail: dfok@few.eur.nl`.

1 Introduction

The use of brand choice models has become standard practice in marketing research (Guadagni and Little, 1983; Chintagunta *et al.*, 1991; Keane, 1997; Hansen *et al.*, 2006). In many applications of these choice models, the random utility theory framework (McFadden, 1973, 1981) is used to represent the choice process. An often made assumption used to be the homogeneity of households. That is, it was assumed that all households have similar tastes, where tastes also include features such as price elasticity and promotion sensitivity. Differences in household behavior were only allowed to the extent that they could be fully explained by observable characteristics. This corresponds with so-called *observed heterogeneity.* Taste is in this case explicitly modeled, for example, by including demographic variables (see *e.g.* Maddala, 1983), or like Horsky *et al.* (2006) who include survey data in their brand choice model to capture heterogeneity. Usually however, such survey data are not available. Also, many studies have shown that not all heterogeneity can be captured by available observed characteristics. Hence there might be so-called *unobserved heterogeneity*, see, for example, Jain *et al.* (1994) and Rossi and Allenby (1993), among others.

There are two popular techniques to deal with unobserved heterogeneity, see Allenby and Rossi (1999) and Wedel *et al.* (1999) for a discussion. These techniques are both based on the notion that when there is unobserved heterogeneity in tastes, there is a corresponding preference distribution in the population. One approach imposes a continuous distribution of a known form to capture the heterogeneity, see, for example, Rossi and Allenby (1993). The other approach tries to approximate the unknown distribution by a discrete distribution with a fixed number of probability masses. A choice model using the latter approach is an example of a finite mixture model, see, for example, Wedel and Kamakura (1999). The mixture components are usually interpreted as segments of households with similar preferences.

In the abovementioned approaches, tastes are usually assumed to be constant during the observation period for each household. This assumption is needed to identify the random heterogeneity. Additionally, the imposed unobserved heterogeneity structure has *a priori* no direct interpretation. For example, the interpretation of segments following from a mixture approach is usually done once the parameters have been estimated.

In the present paper we propose a new approach. Next to a flexible specification of possible heterogeneity in tastes, we introduce unobserved heterogeneity in a brand choice model which *a priori* has a direct and meaningful interpretation. Furthermore, we allow heterogeneity to be different across purchase occasions within the same household. Households, who choose amongst brands within a specific product category, may differ in their response to marketing efforts. For example, some households will spend more time and effort while making their choice than others do. If little time and effort is invested in the decision process, it is perhaps less likely that the household will respond to marketing instruments. For example, to be able to respond to price changes, one of course needs to recall the previous prices of all brands. To be able to respond to advertising, one has to read the newspaper in which the advertisement is printed. It may be unrealistic to assume that all households show such a strong involvement with the product category at all purchase occasions. Especially if we consider low involvement categories such as various supermarket product categories. Hence, it is likely that households will differ in the extent to which they are responsive to marketing efforts. Within a household there may also be differences in the responsiveness across purchase occasions, for example, due to different types of shopping trips.

One reason why some households are unresponsive to marketing efforts could just be a lack of interest in marketing efforts made by brand managers. On the other hand, economic motivations may also explain varying responsiveness across households and over time. For example, search costs play an important role in the decision process of a household or an individual. As mentioned before, to be responsive to price changes one needs to remember the prices of each option at each purchase occasion. Additionally, people usually face time constraints. It takes time for a household to compare the prices of all options at a specific shopping occasion at the time of purchase. Consider a household planning to buy many different items during the same shopping trip. There is obviously a limited amount of time available for the trip and therefore it may be unrealistic to assume that the household will allocate much time to each item. Following this line of thought, the more items a household purchases at a shopping trip, the less responsive this household might be to marketing efforts. Hence, the monetary value of all products purchased at a shopping trip may be inversely related to the responsiveness to marketing efforts.

As the decision process differs across households and across purchase occasions, the above implies that the observed choice of different households can unlikely be explained by the same variables. Choice behavior of *responsive* households can be explained by their base preferences, by marketing efforts, and by their purchase history. Brand choice by *unresponsive* households may only be described by base preferences and purchase history. Moreover, household characteristics are rarely seen to significantly contribute to explaining brand choice, but these might be especially informative for the type of decision process used by the household. As such, household characteristics might influence brand choice, albeit perhaps only indirectly.

In this paper we put forward a brand choice model which incorporates responsiveness to marketing efforts as an explicit form of heterogeneity. We introduce two latent segments. In the first segment the households are assumed to respond to marketing efforts, while in the second segment households are assumed not to do so. If households are not responsive their brand choice may be influenced by their previous choice or they simply purchase their most preferred brand. Whether a specific household is a member of the first or the second segment at a specific purchase occasion is described by household-specific characteristics and characteristics concerning buying behavior. Additionally, to capture differences in responsiveness over time, households are allowed to switch between the two segments across purchase occasions.

The approach in the present paper is somewhat related to structural heterogeneity, where one allows individuals to have different decision strategies. For example, Kamakura *et al.* (1996) examine brand choice within a product category where the brands carry, say, different product sizes. A household might first choose a brand and then choose the specific size to purchase. Another household might first choose a specific size and only then consider the available brands. A third household might completely ignore all this and choose directly from all available brand and product size combinations. Yang and Allenby (2000), for example, present a model in which households are allowed to differ in the reference point to which options are compared. These authors use a hierarchical Bayes model to model credit card adoption, where households are allowed to differ in their decision rule and where behavior can change over time. Wang and Fischbeck (2004) consider structural heterogeneity with respect to framing in a prospect theory setting. For a given decision, some individuals may use a gain frame, while others may adopt a loss frame. In a sense, our model is also related to the work of Bucklin and Lattin (1991). They consider a two-state model of purchase incidence and brand choice, where they distinguish between households that plan their purchases and households which act opportunistically. Bucklin and Lattin (1991) however assume homogeneous preferences, while our model also incorporates preference heterogeneity.

The outline of the paper is as follows. In Section 2, we present our responsiveness model. In Section 3, we consider parameter estimation. We opt for a Bayesian approach, see, for example, Rossi *et al.* (2005). We discuss prior specification and how to obtain posterior results using a Gibbs sampler. Furthermore, we discuss forecasting and model comparison. In Section 4 we apply our responsiveness model to three panel data sets concerning purchases of softdrinks, cereal and liquid detergent. We compare the performance of our model to two related choice models. In Section 5, we conclude with some remarks.

2 The Model

To describe our responsiveness model we first introduce some notation. We assume that household $i = 1, \ldots, I$ chooses from J brands at each purchase occasion $t = 1, \ldots, T_i$. The variable y_{ijt} denotes the chosen alternative, that is,

$$y_{ijt} = \begin{cases} 1 & \text{if household } i \text{ purchases brand } j \text{ at occasion } t \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Furthermore we will use $y_{it} \in \{1, \ldots, J\}$ to denote the index of the chosen brand at time t.

Each household is, at any point in time, either responsive or unresponsive to marketing efforts. In case a household is unresponsive to marketing efforts, the choice can only be attributed to base preference, habit, lagged choice and random influences. In the responsive state the household will also be affected by marketing efforts. We introduce a latent indicator variable Z_{it} to denote the responsiveness state of a household i at purchase occasion t, that is,

$$Z_{it} = \begin{cases} 1 & \text{if household } i \text{ is responsive} \\ & \text{to marketing efforts at purchase occasion } t \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Over time households may switch between responsiveness states. For example, the responsiveness of a household may differ according to the type of shopping trip. The type of shopping trip may be measured by the size of the shopping basket, see Bell and Lattin (1998). Of course, we do not observe the responsiveness state of a household over time, and hence these have to be inferred from the data.

To model the responsiveness, we consider a binary probit model which relates Z_{it} to an intercept and household characteristics, like, for example, family income, collected in a k-dimensional vector W_{it}. These characteristics may also include variables concerning the shopping trip itself, like recency of the last purchase and the monetary amount spent on the shopping trip. The specification of the probit model for the responsiveness state thus becomes

$$Z_{it} = \begin{cases} 1 & \text{if } Z_{it}^{*} = W_{it}'\gamma + \eta_{it} \geq 0 \\ 0 & \text{if } Z_{it}^{*} = W_{it}'\gamma + \eta_{it} < 0, \end{cases} \tag{3}$$

where γ is a k-dimensional parameter vector and $\eta_{it} \sim N(0, 1)$. Hence, the probability that household i is responsive at purchase occasion t is given by

$$\Pr[Z_{it} = 1|\gamma] = \Phi(W_{it}'\gamma), \tag{4}$$

where $\Phi(\cdot)$ is the CDF of a standard normal distribution.

In case a household is responsive to marketing efforts, then marketing instruments, as price and promotion, can have an effect on the choice made by this household. We collect the marketing instruments for brand $j = 1, \ldots, J$, as experienced by household i at purchase occasion t in the m-dimensional vector X_{ijt}. To model the choice process of a marketing-responsive household we use a Multinomial Probit [MNP] model. Conditional on responsiveness, the utility of brand j for household i at purchase occasion

$$U_{ijt} = \mu_{ij}^{(r)} + \alpha^{(r)} y_{ij,t-1} + X_{ijt}'\beta_i + \varepsilon_{ijt}, \tag{5}$$

for $j = 1, \ldots, J$, where $\varepsilon_{it} = (\varepsilon_{i1t}, \ldots, \varepsilon_{iJt}) \sim N(0, \mathbf{I}_J)$ and $\mathbf{I}_J$ denotes a J-dimensional identity matrix. The $\mu_{ij}^{(r)}$ parameters are individual-specific brand intercepts where we impose that $\mu_{iJ}^{(r)} = 0$ for identification. The $\alpha^{(r)}$ parameters measure the effect of state dependence in brand choice as $y_{ij,t-1} = 1$ if household i purchased brand j at purchase occasion $t-1$. State dependence refers to a dynamic property of the choice process, as it incorporates the household's tendency to currently buy the same brand as purchased at the previous occasion. The household-specific effects of the marketing-mix instruments are measured by the individual-specific parameters β_i. We allow for heterogeneity in these effects by assuming that

$$\beta_i \sim N(\beta, \Sigma_\beta), \tag{6}$$

such that β and Σ_β denote the population mean and covariance matrix of the effects of the marketing-mix on the brand utilities.

Of course, in case a household is unresponsive to marketing activities, the marketing instruments will not have an effect on its choice behavior. On these purchase occasions the brand choice will be mainly determined by base preferences, lagged choice, and random effects. This type of behavior can be modeled by the utility specification

$$U_{ijt} = \mu_{ij}^{(u)} + \alpha^{(u)} y_{ij,t-1} + \varepsilon_{ijt} \tag{7}$$

with $\mu_{iJ}^{(u)} = 0$, where, obviously, the X_{ijt} are not included and where we allow the brand intercepts $\mu_{ij}^{(u)}$ and the lagged choice parameter $\alpha^{(u)}$ to be different from the responsive case.

The base preferences of households are usually assumed to be constant over long periods of time. To do that here, we need to make sure that the base preference for a given household does not depend on the responsiveness state of the household on a particular purchase occasion. Note that we cannot simply restrict the utility intercepts of the two utility models (8) and (9) to be equal as the intercepts also correct for the means of the explanatory variables. Furthermore, for the unresponsive model the brand intercepts also capture differences in baseline prices across brands. To allow for constant individual-specific preferences over time independent of the responsiveness state, we therefore have to follow another strategy.

Denote the deviation of the base preference of household i from the population mean by the $J-1$-dimensional vector $\omega_i = (\omega_{i1}, \ldots, \omega_{i,J-1})'$. For the model with continuous heterogeneity, we model the population distribution of these deviations by $N(0, \Sigma_\omega)$. Furthermore, for ease of notation we define $\omega_{iJ} = 0$, $i = 1, \ldots, I$. As brand intercepts for the utilities conditional on responsiveness we now have $\mu_{ij}^{(r)} = \mu_j^{(r)} + \omega_{ij}$ and for the utilities conditional on unresponsiveness (9) we use $\mu_{ij}^{(u)} = \mu_j^{(u)} + \omega_{ij}$. Hence, the household-specific vector ω_i measures the deviation of household i's preferences from the population mean for both responsiveness states.

Household i purchases brand j at purchase occasion t when U_{ijt} is the maximum utility among all U_{ikt}, $k = 1, \ldots, J$. Strictly speaking the unresponsive specification does not correspond to a proper utility maximization problem. Under standard utility maximization, prices must enter the (reduced-form) utility model as prices are obviously part of a household's budget restriction[1]. Our implicit assumption in (7) is that households which are unresponsive to marketing efforts maximize utility without considering the *actual* price differences among the brands. Instead, they aim at an approximate utility maximization that costs less effort. In this case the average, or baseline, price for each brand is used

[1]We thank an anonymous reviewer for raising this point.

instead of the actual price. This implies that although "unresponsive" households do not take into account price promotions, they will react to permanent changes in price. The utility specification in (7) actually reads $U_{ijt} = \mu_{ij}^{(u)} + \alpha^{(u)} y_{ij,t-1} + \delta_i \bar{p}_j + \varepsilon_{ijt}$, where $\bar{p}_j$ denotes the average long-run price of brand j. However, in practically available data the long run price does not vary over time, therefore we cannot separately identify δ_i and $\mu_{ij}^{(u)}$. The utility specification we use for the unresponsive case therefore does not include prices, and the brand intercepts in (7) give a combination of base preferences and price effects.

If household i is responsive at time t then the probability of purchasing brand j is given by

$$\begin{aligned}\Pr[y_{ijt} = 1 | Z_{it} = 1; \mu_i^{(r)}, \beta_i, \alpha^{(r)}] = \\ \Pr[\varepsilon_{ikt} - \varepsilon_{ijt} < (\mu_{ik}^{(r)} - \mu_{ij}^{(r)}) + \alpha^{(r)}(y_{ik,t-1} - y_{ij,t-1}) + (X'_{ikt} - X'_{ijt})\beta_i \ \forall k \neq j]. \quad (8)\end{aligned}$$

This is the choice probability of a Multinomial Probit Model. There is no closed form expression for this probability. For small values of J one can use numerical integration methods to evaluate the probability. For large values of J one can use the GHK simulator, see Börsch-Supan and Hajivassiliou (1993). If the household is unresponsive at t, the probability of purchasing brand j is

$$\begin{aligned}\Pr[y_{ijt} = 1 | Z_{it} = 0; \mu_i^{(u)}, \alpha^{(u)}] = \\ \Pr[\varepsilon_{ikt} - \varepsilon_{ijt} < (\mu_{ik}^{(u)} - \mu_{ij}^{(u)}) + \alpha^{(u)}(y_{ik,t-1} - y_{ij,t-1}) \ \forall k \neq j]. \quad (9)\end{aligned}$$

Finally, as we do not observe whether a household at purchase occasion t belongs to the responsive segment or not, the probability that it purchases brand j at purchase occasion t is obtained by summing the conditional probabilities over the segments, that is,

$$\begin{aligned}\Pr[y_{ijt} = 1 | \theta, \theta_i] = \Pr[Z_{it} = 1 | \gamma] \Pr[y_{ijt} = 1 | Z_{it} = 1; \mu_i^{(r)}, \beta_i, \alpha^{(r)}] \\ + (1 - \Pr[Z_{it} = 1 | \gamma]) \Pr[y_{ijt} = 1 | Z_{it} = 0; \mu_i^{(u)}, \alpha^{(u)}], \quad (10)\end{aligned}$$

where $\Pr[Z_{it} = 1|\gamma]$ is given in (4), θ collects the parameters common to all households and θ_i collects the individual level parameters, that is, $\theta = (\gamma, \alpha^{(r)}, \alpha^{(u)})$ and $\theta_i = (\beta_i, \mu_i^{(r)}, \mu_i^{(u)})$.

An interesting by-product of our model concerns the possibility to calculate the conditional probability of responsiveness given the brand choice at purchase occasion t. Also conditioning on the parameters, this probability equals

$$\begin{aligned}\Pr[Z_{i,t}=1|y_{ijt}=1,\theta,\theta_i] &= \frac{\Pr[Z_{it}=1,y_{ijt}=1|\theta,\theta_i]}{\Pr[y_{ijt}=1|\theta,\theta_i]} \\ &= \frac{\Pr[y_{ijt}=1|Z_{it}=1,\mu_i^{(r)},\beta_i,\alpha^{(r)}]\Pr[Z_{it}=1|\gamma]}{\Pr[y_{ijt}=1|Z_{it}=1,\mu_i^{(r)},\beta_i,\alpha^{(r)}]\Pr[Z_{it}=1|\gamma]+\Pr[y_{ijt}=1|Z_{it}=0,\mu_i^{(u)},\alpha^{(u)}]\Pr[Z_{it}=0|\gamma]}.\end{aligned} \tag{11}$$

This expression gives the probability that household i is responsive to marketing efforts at purchase occasion t, given the parameters and the fact that brand j is purchased. In the applications we will display a histogram of the posterior means of these conditional probabilities for each purchase occasion to give an impression of the average value and the dispersion of the responsiveness in the population, below we will specify this posterior mean in more detail.

3 Inference

In this section we discuss inference within the responsiveness model. We opt for a Bayesian approach. In Section 3.1 we derive the likelihood function of the model. Section 3.2 deals with the prior specification. In Section 3.3 we discuss how to compute posterior results, while in Section 3.4 we focus on model comparison.

3.1 Likelihood function

The likelihood function of our responsiveness model is the joint density of the purchases of the I households denoted by $Y=\{Y_i\}_{i=1}^{I}$, where $Y_i=\{\{y_{ijt}\}_{t=1}^{T_i}\}_{j=1}^{J}$

$$p(Y|\Theta)=\prod_{i=1}^{I}p(Y_i|\Theta), \tag{12}$$

where $\Theta=(\theta,\beta,\mu^{(r)},\mu^{(u)},\Sigma_\beta,\Sigma_\omega)$ and where $p(Y_i|\Theta)$ equals the likelihood contribution of household i given by

$$p(Y_i|\Theta)=\iint p(Y_i|\theta,\theta_i)\phi(\omega_i;0,\Sigma_\omega)\phi(\beta_i;\beta,\Sigma_\beta)d\omega_i d\beta_i, \tag{13}$$

where $\theta_i=(\beta_i,\mu^{(r)}+w_i,\mu^{(u)}+w_i)$. The density $p(Y_i|\theta,\theta_i)$ denotes the likelihood contribution conditional on the unobserved heterogeneity parameters θ_i, that is,

$$p(Y_i|\theta,\theta_i)=\prod_{t=2}^{T_i}\prod_{j=1}^{J}\Pr[y_{ijt}=1|\theta,\theta_i]^{y_{ijt}}, \tag{14}$$

where $\Pr[y_{ijt} = 1|\theta, \theta_i]$ is given in (10). We omit the first observation in (14) as we need this observation to initialize the lagged choice dummy.

3.2 Prior specification

For our Bayesian analysis, we define independent priors for the parameters in Θ. We opt for a conjugate prior specification. For the probit parameters we take a normal prior specification, that is,

$$\gamma \sim N(\gamma_0, S_\gamma), \tag{15}$$

where γ_0 and S_γ are prior parameters. For lagged choice parameters $\alpha^{(r)}$ and $\alpha^{(u)}$, the β parameters and the $(J-1)$-dimensional vectors of brand intercept parameters $\mu^{(r)} = (\mu_1^{(r)}, \ldots, \mu_{J-1}^{(r)})$ and $\mu^{(u)} = (\mu_1^{(u)}, \ldots, \mu_{J-1}^{(u)})$ we also take a normal prior specification

$$\begin{aligned} \alpha^{(r)} &\sim N(\alpha_0^{(r)}, s^2_{\alpha^{(r)}}) \qquad & \alpha^{(u)} &\sim N(\alpha_0^{(u)}, s^2_{\alpha^{(u)}}) \\ \mu^{(r)} &\sim N(\mu_0^{(r)}, S_{\mu^{(r)}}) \qquad & \mu^{(u)} &\sim N(\mu_0^{(u)}, S_{\mu^{(u)}}) \\ \beta &\sim N(\beta_0, S_\beta). \end{aligned} \tag{16}$$

For the covariance matrices in our model we take inverted Wishart priors

$$\Sigma_\beta \sim IW(Q_\beta, \lambda_\beta) \qquad \Sigma_\omega \sim IW(Q_\omega, \lambda_\omega), \tag{17}$$

where Q_β and Q_ω are fixed scale prior parameters and λ_β, λ_ω are fixed degrees of freedom prior parameters.

The joint prior $p(\Theta)$ of the model parameters Θ follows from the product of the priors implied by (15)–(17).

3.3 Posterior Results

If we combine the prior specification $p(\Theta)$ with the likelihood function $p(Y|\Theta)$ given in (12) we obtain the posterior density

$$p(\Theta|Y) \propto p(\Theta)p(Y|\Theta). \tag{18}$$

To obtain posterior results we implement the Gibbs sampler of Geman and Geman (1984) with data augmentation (Tanner and Wong, 1987), see also Tierney (1994). The Gibbs sampler is applied to the prior times the complete data likelihood function. Hence, the latent utilities $U = \{\{\{U_{ijt}\}_{j=1}^{J}\}_{t=2}^{T_i}\}_{i=1}^{I}$, $Z^* = \{\{Z_{it}^*\}_{t=2}^{T_i}\}_{i=1}^{I}$ and the latent parameters $B = \{\beta_i\}_{i=1}^{I}$ and $\Omega = \{\omega_i\}_{i=1}^{I}$ are sampled alongside the model parameters Θ, see Albert and Chib (1993) and McCulloch and Rossi (1994) for similar approaches in choice models. The complete Gibbs sampling scheme is as follows

- $\gamma|\{U, Z^*, B, \Omega, \Theta\}\backslash\gamma$
- $B|\{U, Z^*, \Omega, \Theta\}$
- $\Sigma_\beta, \Sigma_\omega|\{U, Z^*, B, \Omega, \Theta\}\backslash\{\Sigma_\beta, \Sigma_\omega\}$
- $\mu^{(r)}, \mu^{(u)}|\{U, Z^*, B, \Omega, \Theta\}\backslash\{\mu^{(r)}, \mu^{(u)}\}$
- $Z^*|\{U, B, \Omega, \Theta\}$
- $\Omega|\{U, Z^*, B, \Theta\}$
- $\beta|\{U, Z^*, B, \Omega, \Theta\}\backslash\beta$
- $\alpha^{(r)}, \alpha^{(u)}|\{U, Z^*, B, \Omega, \Theta\}\backslash\{\alpha^{(r)}, \alpha^{(u)}\}$
- $U|\{Z^*, B, \Omega, \Theta\}$

In Appendix A we derive the full conditional posterior distributions of the model parameters in Θ and the latent variables U, Z^*, B, and Ω.

The Gibbs simulation scheme generates a Markov chain. After the chain has converged one can use the simulated values to compute posterior results. For example, the posterior probability of household i being responsive at purchase occasion t is given by

$$\Pr[Z_{i,t} = 1|y_{ijt} = 1, Y] = \iint \Pr[Z_{i,t} = 1|y_{ijt} = 1; \theta, \theta_i]p(\Theta, \theta_i|Y)d\Theta d\theta_i, \tag{19}$$

where $p(\Theta, \theta_i|Y)$ denotes the posterior density of Θ and θ_i. The posterior probability (19) is equal to $\frac{1}{M}\sum_{m=1}^{M}\mathcal{I}[Z_{it}^{*(m)} \geq 0]$ for large M, where $Z_{it}^{*(m)}$ denotes the m-th draw of Z_{it}^* in the Markov Chain and $\mathcal{I}[\cdot]$ denotes an indicator function which is one in case the argument is true and zero otherwise.

3.4 Model comparison

To judge the added value of introducing the responsiveness to marketing efforts, we compare our model with two alternative model specifications. The first specification is a standard MNP model where the utilities are given by

$$U_{ijt} = \mu_j + \omega_{ij} + \alpha y_{ij,t-1} + \beta_i' X_{ijt} + \varepsilon_{ijt} \tag{20}$$

with (6), $\omega_{iJ} = 0$, $\omega_i = (\omega_{i1}, \ldots, \omega_{i,J-1})' \sim N(0, \Sigma_\omega)$ and $\varepsilon_{it} = (\varepsilon_{i1t}, \ldots, \varepsilon_{iJt}) \sim N(0, \mathbf{I}_J)$.

The second specification is a MNP model where we relate the β_i parameters to the explanatory variables W_{ijt} in a direct way, that is, $\beta_i = \Gamma W_{ijt} + \eta_i$ with $\eta_i \sim N(0, \Sigma_\beta)$. Using the Kronecker product we compactly define the (km)-dimensional vector of cross terms of X_{ijt} and W_{ijt} by $(W_{ijt} \otimes X_{ijt})$. The model can then be written as

$$U_{ijt} = \mu_j + \omega_{ij} + \alpha y_{ij,t-1} + (W_{ijt} \otimes X_{ijt})'\beta + X_{ijt}' b_i + \varepsilon_{ijt} \tag{21}$$

with $b_i \sim N(0, \Sigma_\beta)$, $\omega_i = (\omega_{i1}, \ldots, \omega_{i,J-1})' \sim N(0, \Sigma_\omega)$ and $\varepsilon_{it} = (\varepsilon_{i1t}, \ldots, \varepsilon_{iJt}) \sim N(0, \mathbf{I}_J)$. The (km)-dimensional parameter vector β captures the cross effects between W_{ijt} and X_{ijt}, that is, $\beta = \text{vec}(\Gamma)$.

We use the same prior specification as for our proposed model, that is, a normal prior for the mean parameters and inverted Wishart priors for the covariance matrices. Posterior results of these two alternative models can be obtained using a simplified version of the Gibbs sampler in Section 3.3.

Model comparison is based on out-of-sample performance. We compare the predictive likelihoods of our model and the two alternative specifications mentioned above. The predictive likelihoods are computed for purchase $T_i + 1$ of each household collected in $Y^F = \{Y_{i,T_i+1}\}_{i=1}^{I}$. These out-of-sample purchases are not used to compute posterior results. The predictive likelihood of our model is given by

$$p(Y^F|Y) = \prod_{i=1}^{I} \iint \prod_{j=1}^{J} \Pr[y_{ij,T_i+1} = 1|\theta, \theta_i, Y]^{y_{ij,T_i+1}} p(\theta_i, \Theta|Y) d\theta_i d\Theta, \tag{22}$$

where $p(\theta_i, \Theta|Y)$ denotes the posterior density of θ_i and Θ. The predictive likelihoods of the other models are defined in a similar way.

The predictive likelihood (22) can easily be computed using the Gibbs output. Given the posterior draws $\theta_i^{(m)}$ and $\Theta^{(m)}$ we simulate $Z_{i,T_i+1}^{*(m)}$ given W_{ij,T_i+1} according to (4) for $i = 1, \ldots, I$. If $Z_{i,T_i+1}^{*(m)} \geq 0$ we simulate $U_{ij,T_i+1}^{(m)}$ according to (5) and if $Z_{i,T_i+1}^{*(m)} < 0$ we use (7) for $j = 1, \ldots, J$. The maximum utility value determines brand choice. The product of the average value of the brand choices over the simulations converges to (22).

4 Illustrations

We apply our model to three different categories of fast moving consumer goods. The same data are analyzed in Bell and Lattin (1998) for other purposes[2]. This data set contains individual scanner panel data across 24 categories. The data cover a two-year period from June 1991 to June 1993 for two separate markets in a large US city. The market we choose for our analysis concerns a suburban area. From the 24 available categories, we have randomly chosen three rather dissimilar categories, that is, softdrinks, cereal, and liquid detergent.

[2]We thank David Bell for generously sharing the data with us.

For each category we have selected households purchasing only the top brands, where the top brands are defined as having a market share of about 5% or more. In Table 1 we summarize the number of households and purchases in the three datasets and the selected brands together with their choice shares.

We perform a Bayesian analysis on the three data sets, and we consider three models. First of all, we consider our responsiveness model. The explanatory variables W_{it} in the responsiveness equation (2) are an intercept, household size, family income, amount of dollars spent on the shopping trip and weeks since last purchase in the product category. For the family income we only know the income category. The marketing mix instruments are normalized so that the coefficients can be compared across the three product categories. Table 1 shows the average values of these variables across the shopping trips. To explain brand choice we include brand intercepts μ_{ij} and a lagged choice dummy variable in both utility specification (5) and (7). We allow for unobserved heterogeneity on the brand intercepts which is the same across the responsive and non-responsive utility specification as discussed before. For the responsiveness utility specification (5) we also include price, feature and display (X_{ijt}). Again, the marketing-mix variables are normalized for the ease of comparison. We allow for continuous unobserved heterogeneity specification in the parameters explaining the effect of the marketing-mix variables.

The prior distribution for the γ parameter is given by (15) with $\gamma_0 = \mathbf{0}$ and $S_\gamma = \mathbf{I}$. For the two covariance matrices in the model Σ_β and Σ_ω we take inverted Wishart priors (17) with $Q_\beta = 10\mathbf{I}$, $Q_\omega = 10\mathbf{I}$ and $\lambda_\beta = \lambda_\omega = 10$. The prior specification of the other parameters is normal as stated in (16), where the prior means are set at $\mathbf{0}$ and the prior covariance matrices are equal to identity matrices.

The two other models are a standard MNP model (20) and a MNP model with cross effects (21). The MNP model contains brand intercepts, a lagged choice dummy, price, feature and display. The MNP model with cross effects contains, on top of that, cross effects of price, feature and display with household size, family income, amount of dollars spent on the shopping trip and weeks since last purchase in the product category as stated in (21). The prior specifications for parameters of the MNP with and without cross effects are similar to the responsiveness model. Again we allow for continuous unobserved heterogeneity on the brand intercepts and on the effect of the marketing-mix variables.

The Bayesian analysis is performed on all purchases except for the last purchase of each household. The last purchases are used for out-of-sample validation using predictive likelihoods as described in Section 3.4.

Responsiveness model

First of all we focus on the inferred probabilities of being responsive to marketing efforts for each product category. Figure 1 displays histograms of the posterior means of the responsiveness probabilities per purchase occasion (11). Across the three categories we see quite some differences. For softdrinks we find a relatively large proportion of purchase trips at which the household was responsive with a probability of almost one. We also find a cluster of observations with around a 0.2 probability of being responsive. For the cereal category the probabilities are more centered around 0.5. For the liquid detergents the distribution of the posterior probabilities is much more skewed to the right. Out of the three categories the households act most responsive here.

The differences in these graphs may be explained from the characteristics of the shopping trips. The data show that the average interpurchase times for liquid detergents are more than two times higher than for the other categories. A higher interpurchase time may imply that households are relying less on routine to make the choice. Therefore, they may not be able to remember past prices and may be more actively involved with the purchase. As a result households may be more likely to search for price information. The average interpurchase time in the softdrinks categories is the smallest. Households are more likely to rely on memory and they are less responsive. Additionally, we see that the average amount of dollars spent on the shopping trips involving softdrinks is smaller than for purchases in the other categories. This may imply that households have more time to compare prices when their shopping basket is smaller. The combination of both effects may explain the bimodality in the responsiveness distribution in the softdrinks category as shown in the first panel of Figure 1. The results for the cereal category are somewhere in between with respect to average interpurchase times and average amount of dollars spent.

To see the difference in cross-sectional variation in responsiveness and temporal variation, we compute the ratio of the average of the variance of the responsiveness probabilities per household to the variance of the average responsiveness probability per household. This ratio is 0.34 for the softdrinks category. For the cereal and detergent category the ratios equal 12.32 and 1.51, respectively. Hence, for the latter two categories the within household spread in response probabilities is larger than the spread across households.

The histograms in Figure 1 do not provide direct information on which type of household is responsive at which type of shopping trip. Such information can however be obtained from the parameter estimates related to the responsiveness equation (2). Table 2 displays the posterior results for the complete responsiveness model. We carried out a limited model selection exercise to finetune the models. Based on overwhelming support of a Bayes factor we have restricted the display variable to be zero in the responsive choice part of the model for softdrinks and cereal.

The final line of Table 2 again confirms that the average responsiveness probability is close to 50%. This again stresses the importance of the unresponsive segment, as in fact about half of the purchases can be associated with this segment. Furthermore, it shows that for liquid detergents the households tend to be most responsive.

The first panel of Table 2 shows the parameters that influence the responsiveness state. The household size is positively related to the responsiveness to marketing efforts for softdrinks and liquid detergents. For cereals, this effect is also positive but the effect is close to zero. For cereals family income is a more important driver, where here a higher family income implies less responsiveness. Note that the influence of lagged choice for cereal is small in the brand choice model for the unresponsive state. Hence, households with a higher income are more likely to go for their favorite brand without considering price. For softdrinks we find that a longer interpurchase time leads to a higher responsiveness probability. A longer interpurchase time implies a higher need to actively compare the brands, as the last purchase cannot be remembered easily. Time since the last purchase is also positive for the cereal and liquid detergent category but the posterior means are about the same as the posterior standard deviation. The average interpurchase time in these categories is in general higher than for the softdrinks category and a longer or shorter than average interpurchase time possibly does not influence the probability of being responsive anymore. For liquid detergents we find that a larger basket size leads to a smaller probability of being responsive. A large shopping basket means that less time can be devoted to each particular category, in turn this makes the household less responsive. For the other two categories we find the opposite results but the posterior means are about the same as the posterior standard deviation. In general, the amount of dollars spent on shopping trips containing cereal and softdrinks turn out to be smaller than for shopping trips containing purchases of liquid detergents which may explain the difference.

The second and third panel of Table 2 present the parameter estimates for the brand utilities. The results indicate that there can be substantial differences in the baseline preferences across the responsive and the unresponsive segment. For example, for softdrinks brand 3 has an average baseline preference within the responsiveness segment but a relatively large baseline preference within the unresponsive segment.

The influence of lagged choice also differs substantially across the two segments. For cereal and detergents we find that lagged choice is not important for unresponsive households, for softdrinks we however find the opposite. If we consider the posterior means of the brand intercepts we see that the differences in values across the brands is large for the cereal and liquid detergents category. Households seem to have a more distinct preference for a brand in these two categories. On unresponsive purchase occasions they are more likely to choose their favorite brand and lagged choice plays a less important role. For the softdrinks category the differences in posterior means in the brand intercepts is much smaller and lagged choice seems to be more important if the actual price does not matter.

Finally, the posterior means of the marketing-mix variables have the expected sign for all three product categories. The effect of price is negative and for feature and display we find a positive effect.

Unreported estimates of the variance of the brand intercepts (Σ_ω) and the variance of the marketing mix variables (Σ_β) show that there is substantial variation in base preferences and marketing mix parameters. Overall, the variation in the base preferences is largest. Comparing the different categories we find that the heterogeneity is largest for detergents and that the degree of heterogeneity is about equal for cereal and softdrinks.[3]

Standard MNP model

The results in Table 2 already indicated that models for responsive and non-responsive households can differ, and the consequences of this finding are further articulated by the estimation results for the MNP model in Table 3. Most noticeable are the differences in coefficients for price. While the posterior mean of the price parameter for the responsive households are -0.342, -0.159 and -0.523 across the three categories, the MNP model for all households would yield an underestimation of the price effect as -0.140, -0.072 and -0.306, respectively, which on average implies an underestimation of around -0.2. Note that in this MNP model we also allow for heterogeneity across households. The responsiveness model clearly allows us to additionally separate the responsive from the unresponsive purchase occasions. Of course when we do not make this split the average

[3]Detailed results are available upon request.

price elasticity will be severely affected.

The parameter concerning lagged choice is also different across the models in Tables 2 and 3. The parameter values in the MNP model are smaller than the values in the responsive part for the cereal and liquid detergent category and is larger for the softdrinks category.

MNP model with cross effects

One of the main differences between our responsiveness model and the "standard MNP model with heterogeneity" is that household variables do not interact with the marketing instruments. Of course one could extend the MNP model with these interaction effects. Table 4 reports the posterior results for such a MNP model.

Clearly the results indicate that only a very minor additional contribution can be observed from these cross effects, except for a weak effect of the weeks since last purchase and feature for liquid detergent. The fact that we allow for unobserved heterogeneity in the effects of the marketing-mix variables already seems to sufficiently describe the differences in response to marketing-mix variables.

Taking the outcomes in Tables 2, 3 and 4 together, we can conclude that the responsiveness model seems to add an important feature to the choice model. Ignoring this feature leads to substantively different results. For example, the MNP model with or without cross effect underestimates price effects relative to our model.

Next, we will focus on the fit of the different models to see whether the responsiveness model indeed fits the data better.

Model comparison

In Table 5 we present three fit measures for our three models, that is, the log predictive likelihood (22), the hit rate, and the mean squared prediction error [MSPE]. The predictive likelihood functions of the responsiveness model are clearly larger than for the MNP model and the MNP model with cross effects for all product categories. Note that because this comparison is based on observations not used during estimation we do not need to penalize for the number of parameters. Recall that we take the final purchase occasion for each household to form the test sample. Furthermore, note that the "standard MNP model" outperforms the MNP model with cross effects. Based on the results in Table 4 this was to be expected as hardly any cross effects turned out to be relevant.

Looking at the (out-of-sample) hit rate, we find that, except for the cereal category, the responsiveness model also gives the highest hit rate. A closer look at the higher hit rate of the MNP model with cross effects for the cereal category shows that the higher hit rate is accompanied by a higher prediction probabilities in the case the model produces a mishit. This explains why the log predictive likelihood of the responsiveness model of cereal is higher. However, the differences in the hit rate across all models are negligible.

As a final performance measure we consider the MSPE. In general, the MSPE of the responsiveness model is smallest. Note that the MSPE also indicates that the MNP model with cross effects performs worst.

The illustrations in this section have indicated quite convincingly that allowing for a possibly large fraction of non-responsive households leads to better fit and to a more appropriate interpretation of the effects of market efforts like price.

5 Concluding remarks

Households may not respond to marketing-mix instruments at each purchase occasion. To be able to respond to these efforts, one needs to invest time and effort in, for example, remembering price changes and reading newspapers and leaflets to notice advertisements. Households differ in the amount of effort they wish to invest in a particular purchase, and therefore they will most likely also differ in their responsiveness to marketing efforts.

The choice model we developed in this paper incorporates the responsiveness of a household at a specific purchase occasion as a form of unobserved heterogeneity. Households differ in their purchasing process. In essence, we assume there are two processes. Households either take marketing efforts into account or they base their choice on base preferences and their past experiences. The specific decision process used can differ across households and across purchase occasions. To explain and forecast the decision process, used by a specific household at a specific purchase occasion, household characteristics can be used together with information on buying behavior. To take into account this form of heterogeneity, we extended a standard brand choice model. Basically, we introduced two segments of households, one segment is unresponsive to marketing efforts whereas the other segment does respond to these efforts. The segment membership is separately modeled using a binary probit model. Household are allowed to switch over time between being responsive or not.

The illustration of our new model to three distinct categories shows that quite some different results can be obtained across our model and related MNP models. Some of the differences can be related to the circumstances and characteristics of the shopping trips in these product categories, such as interpurchase times and the size of the shopping basket. Even though there were only three cases, we can draw a generalizing conclusion and that is that the effects of market efforts will be underestimated in MNP models as in these models both responsive and non-responsive households are jointly treated as a single sample. Allowing for this specific form of heterogeneity in our model thus leads to better insights into the effects of the marketing mix. One can also identify the characteristics that result in higher probabilities of being responsive for households, and this has immediate managerial consequences. One key result of our model is that it leads to better targeting of marketing instruments, which at the same time then also yield less irritation and waste. Further research should examine if the responsiveness fraction of around 0.5 in the three studied categories is a fraction that could commonly be found for fast-moving consumer goods, or whether such a fraction could differ across different types of products.

Table 1: Characteristics of the three data sets

	softdrinks		cereal		liquid detergent	
			Selected brands with choice shares			
brand 1	Canfield	(0.14)	General Mills	(0.28)	All	(0.26)
brand 2	Schweppes	(0.12)	Kellogg's	(0.42)	Cheer	(0.10)
brand 3	Coca Cola	(0.23)	Philip Morris	(0.10)	Purex	(0.06)
brand 4	Dr. Pepper	(0.10)	Quacker	(0.09)	Surf	(0.04)
brand 5	Pepsi	(0.14)	Ralston	(0.05)	Tide	(0.27)
brand 6	Private Label	(0.13)	Nabisco	(0.05)	Wisk	(0.23)
brand 7	Royal Crown	(0.15)			Yes	(0.04)
			Number of observations			
#households	88		244		79	
#purchases	3513		6496		642	
			Average household/shopping characteristics			
household size	1.95		2.53		2.80	
family income	4.89		6.42		7.21	
dollars spent	41.70		57.65		64.24	
interp. times	2.32		3.39		8.26	

Table 2: Posterior results for the responsiveness model[a]

variable	softdrinks mean	softdrinks st. dev.	cereal mean	cereal st. dev.	liquid detergent mean	liquid detergent st. dev.
			Probit equation being responsive (3)			
intercept	0.069	0.063	0.019	0.021	0.247	0.220
household size	0.916***	0.384	0.062	0.086	0.625***	0.215
family income	0.115	0.206	−0.140*	0.076	−0.061	0.221
dollars spent	0.131	0.131	0.058	0.052	−0.310*	0.182
weeks since last purchase	0.229**	0.110	0.048	0.042	0.153	0.157
			Utility equation being responsive (5)			
brand 1	−0.211	0.201	3.110***	0.249	1.227*	0.622
brand 2	−0.205	0.233	2.759***	0.271	−0.353	0.758
brand 3	0.258	0.217	2.117***	0.262	0.804	0.560
brand 4	−0.562**	0.288	1.832***	0.314	−0.540	0.748
brand 5	0.134	0.172	0.217	0.367	1.255**	0.579
brand 6	−1.327***	0.513			0.113	0.771
lagged choice	0.089*	0.053	0.256**	0.092	1.875***	0.383
price	−0.342***	0.098	−0.159***	0.063	−0.523	0.312
feature	0.126**	0.057	0.165***	0.043	0.565**	0.223
display[b]					0.541**	0.220
			Utility equation being unresponsive (7)			
brand 1	−0.037	0.336	0.961***	0.341	−0.369	0.618
brand 2	0.552*	0.326	2.121***	0.162	0.057	0.464
brand 3	0.980***	0.379	0.824***	0.237	−1.579**	0.808
brand 4	0.593	0.364	0.714***	0.225	−1.444	1.025
brand 5	0.300	0.309	0.393**	0.168	−0.174	0.907
brand 6	0.071	0.453			1.181***	0.449
lagged choice	0.510***	0.070	0.036	0.090	0.514**	0.252
average responsiveness probability	0.475		0.507		0.582	

[a] ***, **, * denotes that 0 is not included in the 99%, 95%, 90% Highest Posterior Density interval, respectively.

[b] Bayes factors provide overwhelming posterior support for zero effect of display for softdrinks and cereal. This restriction is therefore imposed.

Table 3: Posterior results for the MNP model (20)[a]

	softdrinks		cereal		liquid detergent	
variable	mean	st. dev.	mean	st. dev.	mean	st. dev.
brand 1	−0.100	0.177	1.920***	0.168	0.647*	0.389
brand 2	0.068	0.189	2.211***	0.169	−0.116	0.493
brand 3	0.470***	0.162	1.249***	0.171	−0.270	0.462
brand 4	−0.078	0.181	0.996***	0.173	−0.750	0.496
brand 5	0.151	0.148	0.403**	0.190	1.067**	0.449
brand 6	−0.463	0.307			1.065***	0.413
lagged choice	0.298***	0.028	0.109***	0.019	0.615***	0.127
price	−0.140**	0.068	−0.072**	0.037	−0.306*	0.164
feature	0.077	0.047	0.045*	0.024	0.255**	0.116
display					0.290**	0.123

[a] ***, **, * denotes that 0 is not included in the 99%, 95%, 90% Highest Posterior Density interval, respectively.

Table 4: Posterior results for the MNP model with cross effects[a]

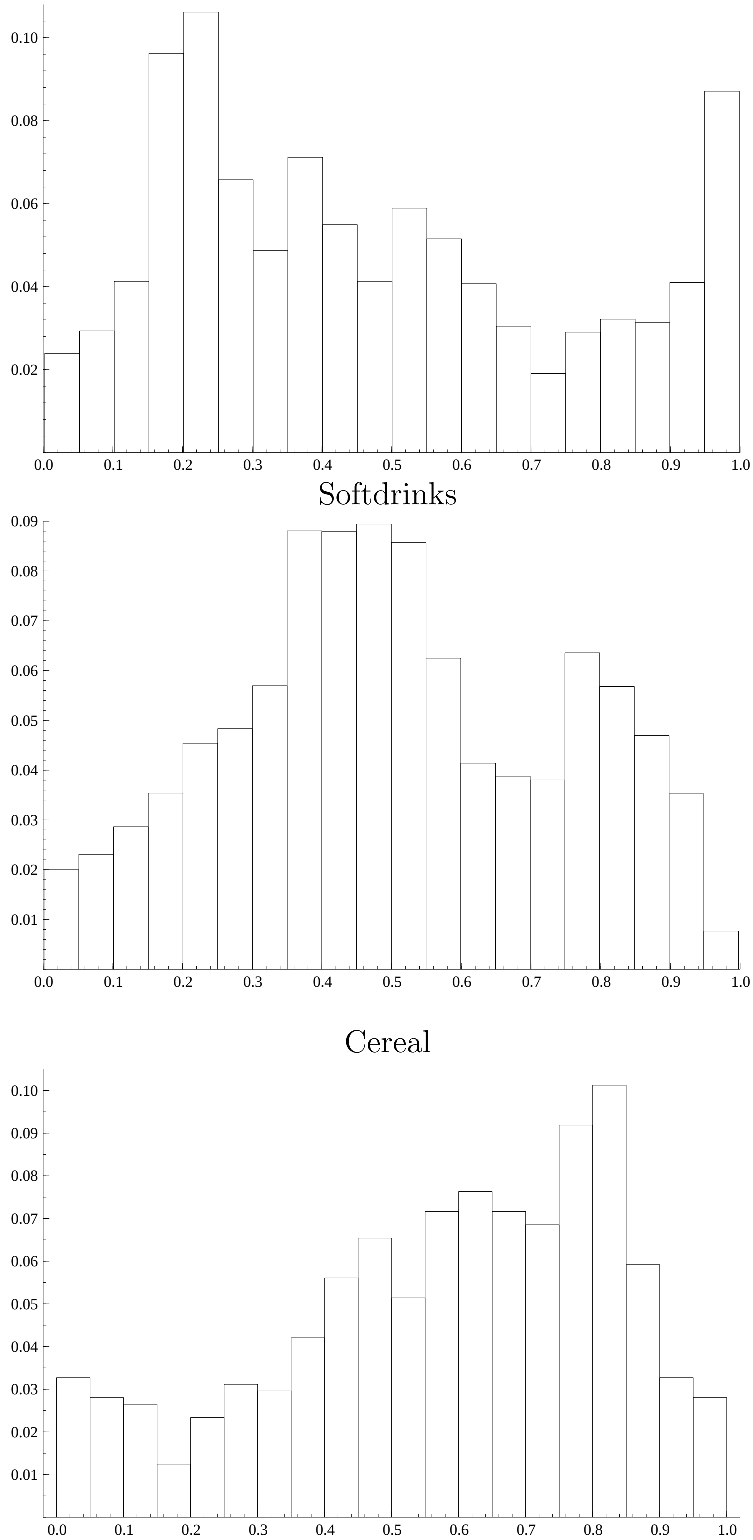
	softdrinks		cereal		liquid detergent	
variable	mean	st. dev.	mean	st. dev.	mean	st. dev.
brand 1	−0.072	0.178	1.848***	0.145	0.806	0.460
brand 2	0.106	0.181	2.130***	0.145	−0.252	0.544
brand 3	0.493***	0.158	1.203***	0.144	−0.178	0.562
brand 4	−0.084	0.187	0.962***	0.149	−0.655	0.535
brand 5	0.183	0.146	0.357**	0.156	1.004**	0.501
brand 6	−0.553*	0.320			1.153**	0.537
lagged choice	0.296***	0.028	0.114***	0.019	0.606***	0.134
price	−0.133**	0.064	−0.063*	0.037	−0.278	0.184
feature	0.081*	0.046	0.063**	0.025	0.272**	0.133
display					0.320**	0.135
			cross effect with household size			
price	0.012	0.066	−0.023	0.041	0.078	0.215
feature	−0.018	0.043	−0.026	0.026	0.101	0.159
display					0.048	0.141
			cross effect with family income			
price	0.008	0.066	0.054	0.038	0.200	0.196
feature	−0.008	0.047	0.039	0.026	−0.094	0.126
display					−0.072	0.128
			cross effect with dollars spent			
price	−0.053	0.037	0.018	0.026	0.178	0.145
feature	−0.014	0.020	−0.024	0.017	0.110	0.103
display					0.016	0.107
			cross effect with weeks since last purchase			
price	−0.017	0.027	0.009	0.016	0.029	0.143
feature	0.000	0.016	0.006	0.012	0.134*	0.080
display					−0.036	0.092

[a] ***, **, * denotes that 0 is not included in the 99%, 95%, 90% Highest Posterior Density interval, respectively.

Table 5: Model Comparison

model	softdrinks	cereal	liquid detergent
	Log predictive likelihood (22)		
responsiveness	-106.698	-279.878	-79.250
MNP (20)	-111.491	-281.031	-83.167
MNP+cross (21)	-111.920	-288.515	-85.874
	Hit rate		
responsiveness	0.580	0.533	0.709
MNP (20)	0.557	0.533	0.709
MNP+cross (21)	0.557	0.541	0.658
	Mean Squared Prediction Error[a]		
responsiveness	7.808	9.723	6.467
MNP (20)	8.203	9.966	6.573
MNP+cross (21)	8.237	10.008	6.751

[a] Mean Squared Prediction Error is defined as $100 \times \frac{1}{I}\sum_{i=1}^{I}(\mathcal{I}[y_{ij,T_i+1} = 1] - \Pr[y_{ij,T_i+1} = 1|Y])^2$.



Liquid detergent

Figure 1: Histograms of the posterior means of the responsiveness probabilities

A Full conditional posterior distributions

Sampling of γ

To simulate γ we consider

$$Z_{it}^* = W_{it}'\gamma + \eta_{it} \qquad \text{for } i = 1, \dots, I,\ t = 2, \dots, T_i \tag{23}$$

with $\eta_{it} \sim N(0,1)$. Define $Z^* = (Z_1^*, \dots, Z_I^*)'$, where $Z_i^* = (Z_{i1}, \dots, Z_{iT_i}^*)'$ and $W = (W_1', \dots, W_I')'$ with $W_i = (W_{i1}, \dots, W_{iT_i})'$. As we have a normal prior specification (15) on the regression parameter γ, the full conditional posterior distribution of γ is normal with mean $(W'W + S_\gamma^{-1})^{-1}(W'Z^* + S_\gamma^{-1}\gamma_0)$ and covariance matrix $(W'W + S_\gamma^{-1})^{-1}$, see, for example, Zellner (1971, Chapter III).

Sampling of Z^*

The full conditional distribution of Z_{it}^* is given by

$$p(Z_{it}^*|\cdot) \propto \phi(Z_{it}^*; W_{it}'\gamma, 1) \prod_{j=1}^{J} \phi(U_{ijt}; m_{ijt}, 1), \tag{24}$$

where $m_{ijt} = (\mu_{ij}^{(r)} + \alpha^{(r)} y_{ij,t-1} + X_{ijt}'\beta_i)\mathcal{I}[Z_{it}^* \geq 0] + (\mu_{ij}^{(u)} + \alpha^{(u)} y_{ij,t-1})\mathcal{I}[Z_{it}^* < 0]$. If we define $\kappa_1 = \prod_{j=1}^{J} \phi(U_{ijt}; \mu_{ij}^{(r)} + \alpha^{(r)} y_{ij,t-1} + X_{ijt}'\beta_i, 1)$ and $\kappa_0 = \prod_{j=1}^{J} \phi(U_{ijt}; \mu_{ij}^{(u)} + \alpha^{(u)} y_{ij,t-1}, 1)$ we can write

$$p(Z_{it}^*|\cdot) = \frac{1}{\kappa}(\kappa_1 \mathcal{I}[Z_{it}^* \geq 0]\phi(Z_{it}^*; W_{it}'\gamma, 1) + \kappa_0 \mathcal{I}[Z_{it}^* < 0]\phi(Z_{it}^*; W_{it}'\gamma, 1)), \tag{25}$$

where

$$\kappa = \kappa_1 \Phi(W_{it}'\gamma) + \kappa_0 \Phi(-W_{it}'\gamma). \tag{26}$$

The CDF of Z_{it}^* is given by

$$\begin{aligned} P(Z_{it}^*|\cdot) = {} & \mathcal{I}[Z_{it}^* < 0]\frac{\kappa_0}{\kappa}\Phi(Z_{it}^* - W_{it}'\gamma) \\ & + \mathcal{I}[Z_{it}^* \geq 0]\left(\frac{\kappa_1}{\kappa}(\Phi(Z_{it}^* - W_{it}'\gamma) - \Phi(-W_{it}'\gamma)) + \frac{\kappa_0}{\kappa}\Phi(-W_{it}'\gamma)\right). \end{aligned} \tag{27}$$

To sample Z_{it}^* we use the inverse CDF technique which leads to

$$Z_{it}^* = \begin{cases} \Phi^{-1}\left(\frac{\kappa u}{\kappa_0}\right) + W_{it}'\gamma & \text{if } u < \frac{\kappa_0}{\kappa}\Phi(-W_{it}'\gamma) \\ \Phi^{-1}\left(\frac{\kappa u}{\kappa_1} + \frac{\kappa_1 - \kappa_0}{\kappa_1}\Phi(-W_{it}'\gamma)\right) + W_{it}'\gamma & \text{if } u \geq \frac{\kappa_0}{\kappa}\Phi(-W_{it}'\gamma), \end{cases} \tag{28}$$

where u is a draw from a uniform distribution.

Sampling of U

To sample U_{ijt} we note that

$$U_{ijt} = (\mu_{ij}^{(r)} + \alpha^{(r)} y_{ij,t-1} + X'_{ijt}\beta_i)\mathcal{I}[Z^*_{it} \geq 0] + (\mu_{ij}^{(u)} + \alpha^{(u)} y_{ij,t-1})\mathcal{I}[Z^*_{it} < 0] + \varepsilon_{ijt} \qquad (29)$$

with $\varepsilon_{ijt} \sim N(0,1)$ for $i = 1,\ldots,I$, $t = 2,\ldots,T_i$, $j = 1,\ldots,J$. Hence, we can sample U_{ijt} from a truncated normal distribution with mean $(\mu_{ij}^{(r)} + \alpha^{(r)} y_{ij,t-1} + X'_{ijt}\beta_i)\mathcal{I}[Z^*_{it} \geq 0] + (\mu_{ij}^{(u)} + \alpha^{(u)} y_{ij,t-1})\mathcal{I}[Z^*_{it} < 0]$ and variance 1 on the region $(\max_{k\neq j} U_{ikt}, \infty)$ if $y_{ijt} = 1$ and $(-\infty, U_{ikt})$ if $y_{ikt} = 1$ with $k \neq j$, see McCulloch and Rossi (1994) for a similar approach.

Sampling of Ω

To sample ω_i $(i = 1,\ldots,I)$ we consider the system of $J-1$ equations

$$U_{ijt} - (\mu_j^{(u)} + \alpha^{(u)} y_{ij,t-1})\mathcal{I}[Z^*_{it} < 0] - (\mu_j^{(r)} + \alpha^{(r)} y_{ij,t-1} + X'_{ijt}\beta_i)\mathcal{I}[Z^*_{it} \geq 0] = \omega_{ij} + \varepsilon_{ijt}, \qquad (30)$$

for $j = 1,\ldots,J-1$ and $t = 2,\ldots,T_i$. Define the right hand side of (30) as $\tilde{U}_{ijt}$ and let $\tilde{U}_{it} = (\tilde{U}_{i1t},\ldots,\tilde{U}_{i,J-1,t})'$. If we combine this system of $J-1$ equations with the unobserved heterogeneity specification $\omega_i \sim N(0,\Sigma_\omega)$ it is easy to show that the full conditional posterior distribution of ω_i is normal with mean $(T_i\mathbf{I}_{J-1} + \Sigma_\omega^{-1})^{-1}(\sum_{t=2}^{T_i}\tilde{U}_{it})$ and covariance matrix $(T_i\mathbf{I}_{J-1} + \Sigma_\omega^{-1})^{-1}$.

Sampling of Σ_ω

The full conditional posterior of Σ_ω is given by

$$p(\Sigma_\omega|\cdot) \propto |\Sigma_\omega|^{-(I+\lambda_\omega+J)/2} \exp\Big(-\frac{1}{2}\mathrm{tr}\Big(\Sigma_\omega^{-1}\Big(\sum_{i=1}^{I}\omega_i'\omega_i + Q_\omega\Big)\Big)\Big) \qquad (31)$$

and hence we can sample Σ_ω from an inverted Wishart distribution with scale parameter $(\sum_{i=1}^{I}\omega_i'\omega_i + Q_\omega)$ and degrees of freedom $I + \lambda_\omega$.

Sampling of B

To sample β_i $(i = 1,\ldots,I)$ we collect for each household i the equations

$$U_{ijt} - \mu_{ij}^{(r)} - \alpha^{(r)} y_{ij,t-1} = X'_{ijt}\beta_i + \varepsilon_{ijt} \qquad \text{with } Z^*_{it} \geq 0 \qquad (32)$$

for $j = 1,\ldots,J$, $t = 2,\ldots,T_i$. Define $\tilde{U}_{ijt} = U_{ijt} - \mu_{ij}^{(r)} - \alpha^{(r)} y_{ij,t-1}$. If we combine the regression equations (32) with the heterogeneity specification $\beta_i \sim N(\beta,\Sigma_\beta)$ we can easily show that we have to sample β_i from a normal distribution with mean

$$\left(\sum_{t=2}^{T_i}\mathcal{I}[Z_{it}^*\geq 0]\sum_{j=1}^{J}X_{ijt}X_{ijt}'+\Sigma_\beta^{-1}\right)^{-1}\left(\sum_{t=2}^{T_i}\mathcal{I}[Z_{it}^*\geq 0]\sum_{j=1}^{J}X_{ijt}\tilde{U}_{ijt}+\Sigma_\beta^{-1}\beta\right)$$

and covariance matrix

$$\left(\sum_{t=2}^{T_i}\mathcal{I}[Z_{it}^*\geq 0]\sum_{j=1}^{J}X_{ijt}X_{ijt}'+\Sigma_\beta^{-1}\right)^{-1}.$$

Sampling of Σ_β

The full conditional posterior of Σ_β is given by

$$p(\Sigma_\beta|\cdot)\propto|\Sigma_\beta|^{-(I+\lambda_\beta+J)/2}\exp\Big(-\frac{1}{2}\mathrm{tr}\Big(\Sigma_\omega^{-1}\Big(\sum_{i=1}^{I}(\beta_i-\beta)'(\beta_i-\beta)+Q_\beta\Big)\Big)\Big) \tag{33}$$

and hence we can sample Σ_β from an inverted Wishart distribution with scale parameter $\sum_{i=1}^{I}(\beta_i-\beta)'(\beta_i-\beta)+Q_\beta)$ and degrees of freedom $I+\lambda_\beta$.

Sampling of β

The full conditional posterior density of β is given by

$$p(\beta|\cdot)\propto\exp\Big(-\frac{1}{2}\sum_{i=1}^{I}(\beta_i-\beta)'\Sigma_\beta^{-1}(\beta_i-\beta)\Big)\exp\Big(-\frac{1}{2}(\beta-\beta_0)'S_\beta^{-1}(\beta-\beta_0)\Big). \tag{34}$$

Hence, we can sample β from a normal distribution with mean $(N\Sigma_\beta^{-1}+S_\beta^{-1})^{-1}(\sum_{i=1}^{N}\Sigma_\beta^{-1}\beta_i+S_\beta^{-1}\beta_0)$ and covariance matrix $(N\Sigma_\beta^{-1}+S_\beta^{-1})^{-1}$.

Sampling of $\mu^{(u)}$ and $\mu^{(r)}$

To sample $\mu^{(u)}$ we consider the system of $J-1$ equations

$$U_{ijt}-\alpha^{(u)}y_{ij,t-1}-\omega_{ij}=\mu_j^{(u)}+\varepsilon_{ijt}\qquad\text{with } Z_{it}^*<0 \tag{35}$$

for $j=1,\ldots,J-1$, $i=1,\ldots,I$, and $t=2,\ldots,T_i$. Define $\tilde{U}_{ijt}=U_{ijt}-\alpha^{(u)}y_{ij,t-1}-\omega_{ij}$ and $\tilde{U}_{it}=(\tilde{U}_{i1t},\ldots,\tilde{U}_{i,J-1,t})'$. If we combine the system of equations (35) with the prior specification (16) it is easy to show that we have to sample $\mu^{(u)}$ from a normal distribution with mean

$$\left(\sum_{i=1}^{I}\sum_{t=2}^{T_i}\mathcal{I}[Z_{it}^*<0]+\Sigma_{\mu^{(u)}}^{-1}\right)^{-1}\left(\sum_{i=1}^{I}\sum_{t=2}^{T_i}\mathcal{I}[Z_{it}^*<0]\tilde{U}_{it}+\Sigma_{\mu^{(u)}}^{-1}\mu_0^{(u)}\right)$$

and covariance matrix

$$\left(\sum_{i=1}^{I}\sum_{t=2}^{T_i}\mathcal{I}[Z_{it}^*<0]+\Sigma_{\mu^{(u)}}^{-1}\right)^{-1},$$

see, for example, Zellner (1971, Chapter VIII).

To sample $\mu^{(r)}$ we consider the system of $J-1$ equations

$$U_{ijt} - \alpha^{(r)} y_{ij,t-1} - X'_{ijt}\beta_i - \omega_{ij} = \mu_j^{(r)} + \varepsilon_{ijt} \qquad \text{with } Z^*_{it} \geq 0, \tag{36}$$

for $j = 1, \ldots, J-1$, $i = 1, \ldots, I$, and $t = 2, \ldots, T_i$. Define now $\tilde{U}_{ijt} = U_{ijt} - \alpha^{(r)} y_{ij,t-1} - X'_{ijt}\beta_i - \omega_{ij}$ and $\tilde{U}_{it} = (\tilde{U}_{i1t}, \ldots, \tilde{U}_{i,J-1,t})'$. If we combine the system of equations (36) with the prior specification (16) it is easy to show that the full conditional distribution of $\mu^{(r)}$ is normal with mean

$$\left(\sum_{i=1}^{I} \sum_{t=2}^{T_i} \mathcal{I}[Z^*_{it} \geq 0] + \Sigma^{-1}_{\mu^{(r)}} \right)^{-1} \left(\sum_{i=1}^{I} \sum_{t=2}^{T_i} \mathcal{I}[Z^*_{it} \geq 0] \tilde{U}_{it} + \Sigma^{-1}_{\mu^{(r)}} \mu_0^{(r)} \right)$$

and covariance matrix

$$\left(\sum_{i=1}^{I} \sum_{t=2}^{T_i} \mathcal{I}[Z^*_{it} \geq 0] + \Sigma^{-1}_{\mu^{(r)}} \right)^{-1}.$$

Sampling of $\alpha^{(u)}$ and $\alpha^{(r)}$

To sample $\alpha^{(u)}$ we consider the equation

$$U_{ijt} - \mu_{ij}^{(u)} = \alpha^{(u)} y_{ij,t-1} + \varepsilon_{ijt} \qquad \text{with } Z^*_{it} < 0 \tag{37}$$

for $j = 1, \ldots, J$, $i = 1, \ldots, I$, and $t = 2, \ldots, T_i$. Define $\tilde{U}_{ijt} = U_{ijt} - \mu_{ij}^{(u)}$. If we combine the regression equation with the prior specification (16) it is easy to show that the full conditional posterior distribution of $\alpha^{(u)}$ is normal with mean

$$\left(\sum_{i=1}^{I} \sum_{t=2}^{T_i} \mathcal{I}[Z^*_{it} < 0] \sum_{j=1}^{J} y^2_{ij,t-1} + s^{-2}_{\alpha^{(u)}} \right)^{-1} \left(\sum_{i=1}^{I} \sum_{t=2}^{T_i} \mathcal{I}[Z^*_{it} < 0] \sum_{j=1}^{J} y_{ij,t-1} \tilde{U}_{ijt} + s^{-2}_{\alpha^{(u)}} \alpha_0^{(u)} \right) \tag{38}$$

and covariance matrix

$$\left(\sum_{i=1}^{I} \sum_{t=2}^{T_i} \mathcal{I}[Z^*_{it} < 0] \sum_{t=2}^{T_i} y^2_{ij,t-1} + s^{-2}_{\alpha^{(u)}} \right)^{-1}.$$

To sample $\alpha^{(r)}$ we consider the equation

$$U_{ijt} - X'_{ijt}\beta_i - \mu_{ij}^{(r)} = \alpha^{(r)} y_{ij,t-1} + \varepsilon_{ijt} \qquad \text{with } Z^*_{it} \geq 0, \tag{39}$$

for $j = 1, \ldots, J-1$, $i = 1, \ldots, I$, and $t = 2, \ldots, T_i$. Define now $\tilde{U}_{ijt} = U_{ijt} - X'_{ijt}\beta_i - \mu_{ij}^{(r)}$. If we combine the equation (39) with the prior specification (16) for $\alpha^{(r)}$ it is easy to show that the full conditional distribution of $\alpha^{(r)}$ is normal with mean

$$\left(\sum_{i=1}^{I} \sum_{t=2}^{T_i} \mathcal{I}[Z^*_{it} \geq 0] \sum_{j=1}^{J} y^2_{ij,t-1} + s^{-2}_{\alpha^{(r)}} \right)^{-1} \left(\sum_{i=1}^{I} \sum_{t=2}^{T_i} \mathcal{I}[Z^*_{it} \geq 0] \sum_{j=1}^{J} y_{ij,t-1} \tilde{U}_{ijt} + s^{-2}_{\alpha^{(r)}} \alpha_0^{(r)} \right)$$

and covariance matrix

$$\left(\sum_{i=1}^{I} \sum_{t=2}^{T_i} \mathcal{I}[Z^*_{it} \geq 0] \sum_{j=1}^{J} y^2_{ij,t-1} + s^{-2}_{\alpha^{(r)}} \right)^{-1}.$$

References

Albert, J. H. and S. Chib (1993), Bayesian Analysis of Binary and Polychotomous Response Data, *Journal of the American Statistical Association*, 88, 669–679.

Allenby, G. M. and P. E. Rossi (1999), Marketing Models of Consumer Heterogeneity, *Journal of Econometrics*, 89, 57–78.

Bell, D. R. and J. M. Lattin (1998), Shopping Behavior and Consumer Preference for Store Price Format: Why "Large-Basket Shoppers Prefer EDLP", *Marketing Science*, 17, 66–88.

Börsch-Supan, A. and V. A. Hajivassiliou (1993), Smooth Unbiased Multivariate Probability Simulators for Maximum Likelihood Estimation of Limited Dependent Variable Models, *Journal of Econometrics*, 58, 347–368.

Bucklin, R. E. and J. M. Lattin (1991), A Two-State Model of Purchase Incidence and Brand Choice, *Marketing Science*, 10, 24–39.

Chintagunta, P. K., D. C. Jain, and N. J. Vilcassim (1991), Investigating Heterogeneity in Brand Preferences in Logit Models for Panel Data, *Journal of Marketing Research*, 28, 417–428.

Geman, S. and D. Geman (1984), Stochastic Relaxations, Gibbs Distributions, and the Bayesian Restoration of Images, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 6, 721–741.

Guadagni, P. M. and J. D. C. Little (1983), A Logit Model of Brand Choice Calibrated on Scanner Data, *Marketing Science*, 2, 203–238.

Hansen, K., V. Singh, and P. Chintagunta (2006), Understanding Store-Brand Purchase Behavior Across Categories, *Marketing Science*, 25, 75–90.

Horsky, D., S. Misra, and P. Nelson (2006), Observed and Unobserved Preference Heterogeneity in Brand-Choice Models, *Marketing Science*, 25, 322–335.

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., B.-D. Kim, and J. Lee (1996), Modeling Preference and Structural Heterogeneity in Consumer Choice, *Marketing Science*, 15, 152–172.

Keane, M. P. (1997), Modeling Heterogeneity and State Dependence in Consumer Choice Behavior, *Journal of Business & Economic Statistics*, 15, 310–327.

Maddala, G. (1983), *Limited Dependent and Qualitative Variables in Econometrics*, vol. 3 of *Econometric Society Monographs*, Cambridge University Press, Cambridge.

McCulloch, R. and P. E. Rossi (1994), An Exact Likelihood Analysis of the Multinomial Probit Model, *Journal of Econometrics*, 64, 207–240.

McFadden, D. L. (1973), Conditional Logit Analysis of Qualitative Choice Behavior, in P. Zarembka (ed.), *Frontiers in Econometrics*, chap. 4, Academic Press, New York, pp. 105–142.

McFadden, D. L. (1981), Econometric Models of Probabilistic Choice, in C. Manski and D. McFadden (eds.), *Structural Analysis of Discrete Data: With Econometric Applications*, MIT Press, Cambridge, MA, pp. 197–272.

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

Rossi, P. E., G. M. Allenby, and R. E. McCulloch (2005), *Bayesian Statistics and Marketing*, John Wiley & Sons, New Jersey.

Tanner, M. A. and W. H. Wong (1987), The Calculation of Posterior Distributions by Data Augmentation, *Journal of the American Statistical Association*, 82, 528–550.

Tierney, L. (1994), Markov Chains for Exploring Posterior Distributions, *Annals of Statistics*, 22, 1701–1762.

Wang, M. and P. Fischbeck (2004), Incorporating Framing into Prospect Theory Modeling: A Mixture-Model Approach, *Journal of Risk and Uncertainty*, 29, 181–197.

Wedel, M. and W. A. Kamakura (1999), *Market Segmentation: Conceptual and Methodological Foundations*, Kluwer Academic Publishers, Dordrecht.

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

Yang, S. and G. M. Allenby (2000), A Model for Observation, Structural, and Household Heterogeneity in Panel Data, *Marketing Letters*, 11, 137–149.

Zellner, A. (1971), *An Introduction to Bayesian Inference in Econometrics*, Wiley, New York.