Retrieving Unobserved Consideration Sets from Household Panel Data

Erjen van Nierop*
Department of Marketing, Faculty of Economics, University of Groningen

Richard Paap
Econometric Institute, Erasmus University Rotterdam

Bart J. Bronnenberg
The UCLA Anderson School of Management

Philip Hans Franses
Econometric Institute, Erasmus University Rotterdam

Michel Wedel
Robert H. Smith School of Business at the University of Maryland

November 23, 2006

*Address for correspondence: University of Groningen, Faculty of Economics, Department of Marketing, POBox 800, 9700 AV Groningen, email: j.e.m.van.nierop@rug.nl.
The first author thanks the Foundation 'Vereniging Trustfonds Erasmus Universiteit Rotterdam' in the Netherlands.
We gratefully acknowledge the comments of participants of seminars in the marketing departments of the Graduate School of Industrial Administration, Carnegie Mellon University and the Wharton School, University of Pennsylvania. Also, we thank Dennis Fok, Peter Lenk, Brian Ratchford, Steven Shugan, the area editor and three anonymous referees for their many helpful suggestions.

Retrieving Unobserved Consideration Sets from Household Panel Data

Abstract

We propose, operationalize and estimate a new model to capture unobserved consideration from discrete choice data. This model consists of a multivariate binary choice model component for consideration and a multinomial probit model component for choice, given consideration. The approach allows one to analyze stated consideration set data, revealed consideration set (choice) data or both, while at the same time it allows for unobserved dependence in consideration among brands. In addition, the model can easily cope with many brands, and accommodates different effects of the marketing mix on consideration and choice, and unobserved consumer heterogeneity in both processes.

We attempt to establish the validity of existing practice to infer consideration sets from observed choices in panel data. To this end, we collect data in an on-line choice experiment involving interactive supermarket shelves and post-choice questionnaires to measure the choice protocol and stated consideration levels. We show with these experimental data that underlying consideration sets can be reliably retrieved from choice data alone and that consideration is positively affected by display and shelf space. We also find that consideration does not covary greatly across brands once we account for observed effects and unobserved heterogeneity.

Keywords: Brand choice, Consideration set, Probit models.

1 Introduction

The theory of consideration sets, developed in the seventies from the work by Bettman (1979), Howard and Sheth (1969) and Newell and Simon (1972), has led to much empirical work in marketing science (for overviews see, for example, Malhotra *et al.*, 1999; Manrai and Andrews, 1998; Roberts and Lattin, 1997) and has had important implications for marketing practice. Its basic postulate is that consumers follow a two-stage decision process of brand choice. In the first stage, they are thought to narrow down the global set of alternatives to a smaller set, the consideration set, from which a final choice is made in the second stage. Researchers in marketing have provided ample empirical evidence corroborating this two-stage process of consumer choice (Lussier and Olshavsky, 1979; Payne, 1976; Wright and Barbour, 1977).

Consideration processes are interesting for marketing purposes because they vary across households (Alba and Chattopadhyay, 1985; Belonax and Mittelstaedt, 1978; Chiang *et al.*, 1999; Roberts and Lattin, 1991) and are sensitive to marketing instruments such as promotions (Siddarth *et al.*, 1995) and advertising (Mitra, 1995). Ignoring consideration sets in models of choice may lead one to underestimate the impact of marketing control variables (Bronnenberg and Vanhonacker, 1996; Chiang *et al.*, 1999). So, with the rapid proliferation of the number of brands in the market place and the increase in cognitive demands placed on consumers choosing among them, understanding consideration set formation and how marketing affects it, has become of great relevance to marketing managers. Entering the consideration set has become an important strategic goal (see, for example, Corstjens and Corstjens, 1999).

Therefore, it is not surprising that econometric representations of choice and con-

sideration for fast moving consumer goods have received great interest from marketing researchers. Most previously used approaches are based on the random utility theory framework (see for example, McFadden, 1973; Guadagni and Little, 1983 and Zhang, 2006). Including the consideration stage into such a random utility framework is not trivial because these sets are usually neither observed nor identifiable with certainty (Ben-Akiva and Boccara, 1995).

Essentially, two approaches have been suggested to identify the sets of brands considered by consumers. One stream of research directly assesses stated consideration set membership of individual brands by modeling the marginal distribution of consideration for each brand (Roberts and Lattin, 1991, for example). This approach is usually based on an assumption of independence of consideration across brands (for example, Ben-Akiva and Boccara, 1995) that has remained untested in empirical research. Therefore, whereas this approach –which we will call the *stated consideration set approach* – works even for large choice sets, it has limitations in being based on the assumption of set-membership independence across brands.

The second stream of research identifies the distribution of consideration sets indirectly from the choice data (for example, Chiang *et al.*, 1999; Manski, 1977; Mehta *et al.*, 2003). To account for the unobserved nature of consideration, and to obtain marginal choice probabilities, it integrates over all possible consideration sets of which there are $2^J - 1$, where J is the number of choice options. This method is suited for modeling unobserved dependencies across brands, because the realization of the entire consideration set is modeled directly. This approach, which we will call the *revealed consideration set approach*, is therefore not burdened with the assumption of independence of consideration set membership across brands. But, some problems exist with empirical applications of

some of the models in question. First, the number of possible consideration sets is exponential in the number of brands contained in the global choice set (see Chiang *et al.*, 1999). With more than four brands, the method becomes rapidly unfeasible because of combinatorial complexity. Second, they offer neither a natural way to study marginal brand set-membership probabilities nor their responsiveness to marketing action. A recently proposed approach by Gilbride and Allenby (2004) addresses these problems. They model consumers' screening rules in a multinomial probit choice model, calibrated on stated preference data. Theirs is a model of consideration and choice, in which consideration is attribute-based, and implemented through a range of (compensatory, conjunctive or disjunctive) screening rules that are operationalized through thresholds on the attributes that determine whether or not an alternative is acceptable. The consideration arises from these screening rules on attributes (e.g., price, feature, display) that also occur in the choice stage. Therefore, although being a particularly powerful and versatile approach based on the Bayesian variable selection literature, it does not accommodate different attributes in the consideration and choice model stages. This is one of the issues that will be addressed in the model proposed in the present study. A second contribution of our approach is that we accommodate the truncated nature of choice data in identifying consideration sets. Data used to calibrate models of choice and consideration, are often constructed from observations of choices, conditional upon a choice being made. That is, empty consideration sets do not occur, so that estimates of the size of consideration sets at the individual level are inflated when the truncating mechanism is not taken into account. In developing our approach to modeling consideration and choice, we will accommodate the truncated nature of the data.

The two streams of consideration set research described above (the stated and re-

vealed consideration set approach) have evolved somewhat independently. There is no existing empirical evidence as to their convergence. Do the "consideration probabilities", that the models used in the revealed consideration set approach, estimated from choice data, really reflect consideration sets as stated by consumers and modeled in the stated consideration set approach? This obviously is an issue that bears directly on the validity of the interpretations of models, parameter estimates and the resulting recommendations for marketing practice. Indeed, Roberts and Lattin (1997) concluded that authors working without explicit measures of consideration "*cannot address whether the consideration stage of their model corresponds to a cognitive stage of consideration or if it is just a statistical artifact of the data. [...] Even if what is inferred is consideration, it will be estimated with substantial error.* " It may therefore be called a surprise that no research to date has addressed the issue of convergent validity of stated versus revealed consideration sets. One possible reason for this undesirable state of affairs is that in order to do so, a joint modeling framework is needed that may accommodate stated consideration data, revealed consideration data (choice) or both. This is the key intended contribution of the study in this paper: to develop a model for consideration set formation and brand choice that provides a unifying framework of the stated and revealed approaches to consideration set identification.

Such an integral approach to modeling consideration sets enables us to assess convergent validity of stated and revealed consideration sets. At the core of our approach is a multivariate binary choice [MBC] model for consideration, compounded with a multinomial probit (MNP, McCulloch and Rossi, 1994; McCulloch *et al.*, 2000) model for brand choice given consideration. In the MBC model, we directly specify the *joint* distribution of the probabilities of brands' consideration set-membership, by modeling consideration

set membership of brands as binary probits that can covary across brands like in a multivariate probit model (MVP, Edwards and Allenby, 2003). The MBC model is different from a standard MVP model as we have to account for the truncated nature of the choice data. The approach does not suffer from the curse of dimensionality providing a tractable representation of consideration set formation, the complexity of which is only *linear* in the global number of choice options. We develop our model primarily for the purpose of obtaining better substantive insights into choice processes, and the effect of marketing variables on these processes. We do not primarily aim at improving predictive validity, but at providing deeper insight into consumers' choice behavior (see also Andrews and Srinivasan, 1995). Indeed, a core contribution of this study is that we intend to validate the inference of consideration from choice data using actually measured consideration sets.

We next lay out the model and its (MCMC) estimation procedure. Then, we investigate the convergent validity of the approach to identify consideration sets from choice behavior, using data from an experimental study that was conducted for this purpose. We conclude by discussing the limitations and prospects on future research.

2 The model

2.1 Preliminaries

In this section we propose a model to describe the brand choice decision of household i ($i = 1, \ldots, I$) choosing brand j ($j = 1, \ldots, J$) at purchase occasion t ($t = 1, \ldots, T_i$). The model that we propose, consists of two stages. In the first stage, it describes which brands are considered by a household for choice. In the second stage, it describes the actual choice of the household from the brands in its consideration set.

The brand choice of household i at time t is described by the random variable D_{it}, which can take the value 1 to J. The actual brand choice is given by d_{it}. Without loss of generality we consider here the -more complex- situation where only such choice data are available and the consideration sets themselves are unobserved. Households typically do not consider all brands in their choice decision, but choose a brand from their consideration or choice set. This choice set may contain one, two or even all brands that are available to the household. For each household, there are $Q = 2^J - 1$ potential non-empty consideration sets. We model the consideration set of household i at time t by the random variable C_{it}. As we assume that households choose a brand from their unobserved consideration set, after observing the actual brand choice, the number of potential consideration sets for a household equals 2^{J-1}. We denote the collection of potential consideration sets for household i at purchase occasion t by $\mathcal{C}_{it}$. For explaining brand choice, managers are interested in the effects of marketing control variables, such as price, feature and displays. We use a subset of these variables, denoted by X_{ijt} in the consideration stage, and another, possibly overlapping subset, denoted by W_{ijt}, in the brand choice stage.

2.2 Stage 1: Consideration set

The consideration set of household i at time t, C_{it}, is described by a J-dimensional vector with binary elements

$$C_{it} = \begin{pmatrix} C_{i1t} \\ \vdots \\ C_{ijt} \end{pmatrix}, \tag{1}$$

where C_{ijt} equals 1 if brand j occurs in the consideration set of household i at time t, and 0 otherwise. In the case where household i considers buying only the first two brands

the consideration set thus equals $C_{it} = (1, 1, 0, \ldots, 0)'$. To describe if a brand is in the consideration set of household i, we consider a multivariate probit type formulation that involves

$$C^*_{ijt} = X'_{ijt}(\alpha + \alpha_i) + \varepsilon_{ijt}, \qquad j = 1, \ldots, J, \tag{2}$$

where X_{ijt} is a k_X-dimensional vector containing brand and purchase-related explanatory variables including brand-specific intercepts, where α describes the average effect and α_i the household-specific effect, and where ε_{ijt} is an unknown error process for which we will describe the distribution later on.

In a standard MVP model we would state that brand j enters the consideration set of household i at time t, that is, $C_{ijt} = 1$, if $C^*_{ijt} > 0$. Hence, if all C^*_{ijt} are negative the consideration set is empty, that is, $C_{it} = (0, \ldots, 0)'$. This occurs if at the particular purchase occasion the household does not buy from the category altogether. However, given that the choice data are observed conditional upon a choice being made, as explained above, we formally accommodate this truncation mechanism such that the empty consideration set does not occur and that at least one brand (with the highest consideration intensity C^*_{ij}) will be in the consideration set. We assume that before the actual choice, a consideration set is formed according to the rule:

Brand j enters the consideration set if

$$C^*_{ijt} > 0 \text{ or } C^*_{ikt} < C^*_{ijt} < 0 \text{ for all } k \neq j. \tag{3}$$

To illustrate this rule, we consider a simplified version of our model with only 3 brands, that is, $J = 3$. A household i chooses from these 3 brands at purchase occasion t. Hence,

there are 7 possible consideration sets which occur with probability

$$
\begin{aligned}
\Pr[C_{it} = (1,1,1)'] &= \Pr[C^*_{i1t} > 0 \wedge C^*_{i2t} > 0 \wedge C^*_{i3t} > 0] \\
\Pr[C_{it} = (1,1,0)'] &= \Pr[C^*_{i1t} > 0 \wedge C^*_{i2t} > 0 \wedge C^*_{i3t} < 0] \\
\Pr[C_{it} = (1,0,1)'] &= \Pr[C^*_{i1t} > 0 \wedge C^*_{i2t} < 0 \wedge C^*_{i3t} > 0] \\
\Pr[C_{it} = (0,1,1)'] &= \Pr[C^*_{i1t} < 0 \wedge C^*_{i2t} > 0 \wedge C^*_{i3t} > 0] \\
\Pr[C_{it} = (1,0,0)'] &= \Pr[C^*_{i1t} > C^*_{i2t} \wedge C^*_{i1t} > C^*_{i3t} \wedge C^*_{i2t} < 0 \wedge C^*_{i3t} < 0] \\
\Pr[C_{it} = (0,1,0)'] &= \Pr[C^*_{i2t} > C^*_{i1t} \wedge C^*_{i2t} > C^*_{i3t} \wedge C^*_{i1t} < 0 \wedge C^*_{i3t} < 0] \\
\Pr[C_{it} = (0,0,1)'] &= \Pr[C^*_{i3t} > C^*_{i1t} \wedge C^*_{i3t} > C^*_{i2t} \wedge C^*_{i1t} < 0 \wedge C^*_{i2t} < 0].
\end{aligned}
\tag{4}
$$

The sum of these probabilities equals 1. These probabilities can be expressed in terms of the error process ε_{it}. For example, the probability that household i only considers buying the first two brands is given by

$$
\Pr[C_{it} = (1,1,0)'] = \\
\Pr[\varepsilon_{i1t} > -X'_{i1t}(\alpha + \alpha_i), \varepsilon_{i2t} > -X'_{i2t}(\alpha + \alpha_i), \varepsilon_{i3t} \leq -X'_{i3t}(\alpha + \alpha_i)]. \tag{5}
$$

The value of the consideration set probabilities (4) depends on the distribution of the ε_{ijt}. We assume that $\varepsilon_{it} = (\varepsilon_{i1t}, \ldots, \varepsilon_{iJt})'$ is normally distributed, that is,

$$
\varepsilon_{it} \sim \mathrm{N}(\mathbf{0}, \Sigma), \tag{6}
$$

where the off-diagonal elements in the covariance matrix Σ describe the dependencies among the probabilities that the brands are contained in the consideration set. In this formulation, multiplying all utilities C^*_{ijt} by a positive constant would result in the same consideration set. Therefore, for identification purposes we set the diagonal elements of Σ all equal to 1. We will denote our model as a Multivariate Binary Choice [MBC] model.

We assume that the household-specific parameters are drawn from a population distribution, that is,

$$\alpha_i \sim \mathrm{N}(\mathbf{0}, \Sigma_\alpha), \tag{7}$$

where Σ_α is a diagonal matrix. An advantage of this approach is that it leads to a non-diagonal covariance structure in our multivariate binary choice model, that is, the unconditional covariance structure of C_{it}^* equals

$$X_{it}\Sigma_\alpha X_{it}' + \Sigma, \tag{8}$$

where $X_{it} = (X_{i1t}, \ldots, X_{iJt})'$, see also Allenby and Rossi (1999) for the same motivation and an application.

2.3 Stage 2: Brand choice

Given the consideration sets of households, we describe their brand choice by a multinomial probit model. We assume that household i perceives utility U_{ijt} from buying brand j at purchase occasion t, that is,

$$U_{ijt} = W_{ijt}'(\beta + \beta_i) + \eta_{ijt}, \qquad j = 1, \ldots, J, \tag{9}$$

where W_{ijt} is a k_W-dimensional vector containing explanatory variables including brand-specific intercepts, where β describes the average effect and β_i the household-specific effect, and where η_{ijt} is a disturbance term. The vector of the disturbances $\eta_{it} = (\eta_{i1t}, \ldots, \eta_{iJt})'$ is assumed to be normally distributed:

$$\eta_{it} \sim \mathrm{N}(\mathbf{0}, \Omega). \tag{10}$$

We also assume that the household-specific parameters β_i are drawn from a population distribution, that is,

$$\beta_i \sim \mathrm{N}(\mathbf{0}, \Sigma_\beta), \tag{11}$$

where Σ_β is a diagonal matrix. Household i purchases brand j at purchase occasion t if the perceived utility of buying brand j is the maximum over all perceived utilities for buying the other brands in the consideration set c_{it}, that is, if

$$U_{ijt} = \max(U_{ikt} \text{ for } k = 1, \ldots, J | c_{ikt} = 1). \tag{12}$$

Hence, the probability that household i chooses brand j at purchase occasion t given the consideration set c_{it} and given β_i equals

$$\begin{aligned} \Pr[D_{it} = j | c_{it}, \beta_i; \beta, \Omega] &= \Pr[U_{ijt} > U_{ikt} \ \forall k \neq j | c_{ijt} = c_{ikt} = 1] \\ &= \Pr[U_{ijt} - U_{ikt} > 0 \ \forall k \neq j | c_{ijt} = c_{ikt} = 1] \\ &= \Pr[\eta_{ikt} - \eta_{ijt} < W'_{ijt} - W'_{ikt}(\beta + \beta_i) \ \forall k \neq j | c_{ijt} = c_{ikt} = 1]. \end{aligned} \tag{13}$$

This expression shows that utility differences and not the levels of the utilities determine brand choice. Therefore, not all elements of the covariance matrix Ω are identified, see Bunch (1991) for a discussion. Additionally, Keane (1992) shows that the off-diagonal elements are often empirically non-identified, which we found to be the case as well and hence we opt for a diagonal covariance matrix. As multiplying the utilities U_{ijt} by a positive constant does not change actual brand choice, we restrict one of the diagonal elements of Ω to be 1 such that $\Omega = \text{diag}(\omega_1^2, \ldots, \omega_{J-1}^2, 1)$. This diagonal structure is generalized to a non-diagonal covariance matrix by modeling the unobserved household heterogeneity (see before in the consideration set stage and also Allenby and Rossi, 1999; Hausman and Wise, 1978, for a similar approach).

Our modeling approach has several advantages. We model the probability that a brand j is included in the consideration set, which means that we only deal with J instead of $Q = 2^J - 1$ alternatives, as would be the case when probabilities are assigned to all potential consideration sets. The covariance structure in our MBC model describes the

dependencies between the inclusion of the brands. The number of parameters in this approach therefore increases at most quadratically in J. Another contribution is that we include explanatory variables in the consideration stage of the model. These explanatory variables may be different from those in the brand choice stage.

3 Parameter estimation

We consider the case of revealed consideration data, where only choices of households have been observed. To estimate the model parameters, we consider the likelihood as a function of the brand choices of the households $D = \{\{d_{it}\}_{t=1}^{T_i}\}_{i=1}^{I}$, that is,

$$\mathcal{L}(D|\theta) = \prod_{i=1}^{I} \int_{\beta_i} \int_{\alpha_i} \prod_{t=1}^{T_i} \Bigg[\sum_{\forall c_{it} \in \mathcal{C}_{it}} \Pr[C_{it} = c_{it}|\alpha, \Sigma, \Sigma_\alpha] \\ \Pr[D_{it} = d_{it}|c_{it}, \beta_i; \beta, \Omega] \Bigg] \phi(\beta_i; \mathbf{0}, \Sigma_\beta)\phi(\alpha_i; \mathbf{0}, \Sigma_\alpha) d\alpha_i d\beta_i, \quad (14)$$

where $\theta = (\alpha, \Sigma_\alpha, \beta, \Sigma_\beta, \Sigma, \Omega)$ and $\mathcal{C}_{it}$ is the set of potential non-empty consideration sets for household i at purchase occasion t. The likelihood function contains the product of the probability of the non-empty consideration set of household i c_{it}, see, for example, (5), and the brand choice probability given c_{it}, see (13), over all purchase occasions and all households. As we do not observe the consideration sets c_{it} of the households, we have to sum over all potential consideration sets for each household. Finally, we have to integrate with respect to α_i and β_i to account for the unobserved household heterogeneity.

If we apply our model to *stated* consideration data, the situation simplifies and we observe, next to the choice indicators d_{it}, also the choice set membership indicators, c_{it}. The expression for the likelihood is similar to that shown above, but the summation across all possible consideration sets vanishes and the approach reduces to the separate estimation of the MBC and MNP components. Since that situation is more straightforward, we focus

in the further description of the estimation methodology on the more complicated case of inferring the joint process of choice and consideration from choice data alone.

The model is estimated with MCMC methods with data augmentation. In Appendix A we provide the details of the full conditional posterior distributions and sampling algorithms for the model parameters and the latent utilities C^*_{ijt} and U_{ijt}. For the estimation of the parameters of each model considered in this paper, we generate 20,000 iterations of the Gibbs sampler for burn in and 40,000 iterations for analysis, where we retain every 40th draw. The (unreported) iteration plots are inspected to see whether the sampler converges to stationary draws from the posterior distributions of the model parameters. Synthetic data analyses show that the parameters are recovered well and that the chains are stationary well before the end of the burn-in. We report the posterior means and standard deviations of the parameters in the empirical analyses below.

4 Empirical validation of identification of consideration sets from choice data

4.1 Data from an on-line experiment

We apply our model to a data set on consideration and choice collected in an on-line experiment conducted to study the effects of promotion on consideration. We use this experiment to investigate the convergent validity of stated consideration sets and the sets identified from choice data only. We demonstrate that the benefits of our model accrue in both the stated and revealed approaches to consideration set identification.

In the on-line shopping experiment, consumers interfaced with a digital image of a supermarket shelf, containing the universal set of 8 brands of laundry detergent. The choice environment was constant across individuals but varied across 10 choice occasions.

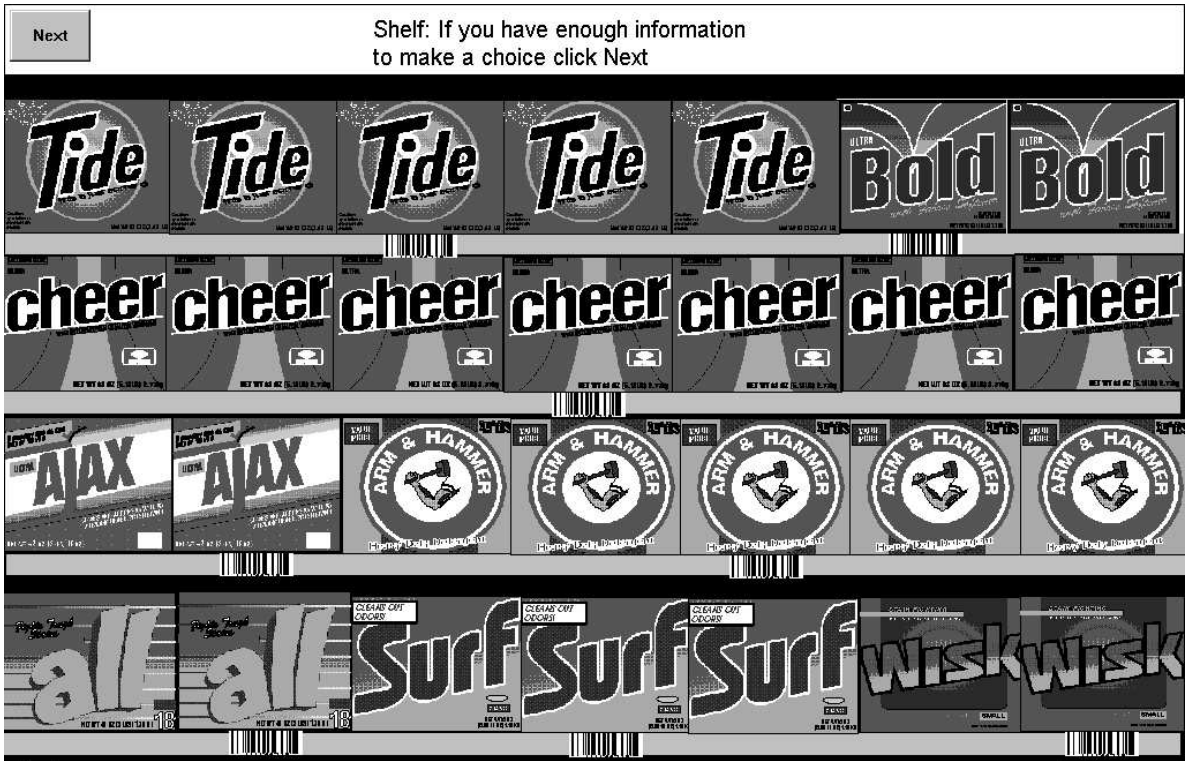

Figure 1: Screen-shot from the sixth choice occasion.

We manipulated promotion, brand position on the shelf and shelf facings as possible drivers of individuals' consideration. Figure 1 shows a screen-shot from the experiment. We also simulated a promotion environment by putting "end-of-aisle" displays into the simulation. These were created by showing the brand on promotion in isolation prior to showing the entire shelf. Subjects had the option to choose the promoted brand (and entirely bypass the shelf) or skip the "end-of-aisle" promotion and visit the regular shelf. The experiment served to measure information acquisition, stated consideration set membership and (revealed) choice. Consideration was measured through two questions using 100 point sliders: (1) did you consider brand j seriously, (2) is brand j acceptable to you? This operationalization of consideration is taken from Lehmann and Pan (1994) and Nedungadi (1990).

Table 1: Descriptive statistics for the experimental data set (432 observations)

Brand	Share	Consideration[a]	Display frequency	Average shelf
Ajax	3%	11%	0%	0.34
All	10%	28%	10%	0.33
AH	11%	21%	10%	0.40
Bold	5%	23%	10%	0.38
Cheer	26%	58%	20%	0.84
Surf	5%	17%	0%	0.40
Tide	40%	66%	10%	0.71
Wisk	2%	20%	0%	0.36

[a] This figure expresses the average consideration share, computed as the average of the two consideration questions (divided by 100) averaged across purchase occasions and individuals.

The experiment was administered to graduate subjects in a large U.S. university. Participants received a diskette with the experiment on it and were reminded once a week by e-mail to make a choice. Diskettes were collected after 10 weeks. In total, 55 subjects completed the experiment. This yielded 528 observations. Of these data, we randomly sampled 48 individuals with 432 purchases for estimation, 7 individuals were kept for hold-out validation. Table 1 shows the description of the data set.

The stated levels of consideration in this table are computed as the average of the two questions (divided by 100) averaged across purchase occasions and individuals. For estimation purposes, we need discrete consideration set memberships. These were constructed by dichotomizing the average of the two questions (divided by 100) around 0.5 for each choice occasion and each individual. Although other cutoffs could be chosen, in the absence of prior information, the scale midpoint is the optimal choice. The variable

shelf space represents the surface of the facings of the 6 brands. Display frequency is the fraction of purchase occasions that the brand was positioned at "end-of-aisle."

Table 1 shows that there is considerable variation in choice shares and consideration across brands. An interesting aspect to note from Table 1 is that the ratio between choice share and consideration is very different across brands (for a similar observation see Siddarth *et al.*, 1995). It can be inferred that, with similar unconditional shares, Arm & Hammer has a very high choice share when it is considered for choice (0.53) and that Bold, for instance, does not (0.20). Hence, whereas a single-stage choice model would treat these brands as equally large, a two-stage model would suggest that these are two very different types of brands. Arm & Hammer is more of a niche brand with high choice share but low consideration. On the other hand, Bold is a small brand with low choice share and average consideration.

4.2 Operationalizations

We assume that consideration is driven by consumer-specific preferences and by consumer-specific effects of in-store merchandizing activity aiming to make a brand more salient at point-of-purchase. Specifically, point-of-purchase merchandizing is operationalized in this study as the effect of display and shelf-space measures. In both stages, we allow for brand intercepts that serve to capture the effects of factors not depending on the marketing or choice environment as well. We do not include display and shelf in the brand choice stage as they do not increase utility for the consumer.

4.3 Estimation results from the on-line experiment

To validate the notion that it is possible to infer consideration sets from choice data, we use the experimental data with stated consideration sets. Specifically, we estimate several model specifications without using the stated consideration measures and next verify that such models in fact identify the consideration process. We consider the full multivariate binary choice/multinomial probit (MBC+MNP) model of choice and consideration estimated on *choice data alone*. Second, we consider the MBC model by itself using the stated consideration sets. Third, we consider a multinomial probit (MNP) model estimated on the choice data alone.

The estimation results for the MBCs are displayed in the top part of Table 2. On the left, the table presents the model estimates of the consideration process estimated from the choice data alone. On the right, the table lists the model estimates based on the stated consideration data. As is clear from the comparison of both columns, the parameters have very similar values. Indeed, the correlation between the parameters of the two models is very high ($r = 0.96$[1]). Interestingly, Table 2 shows that both the proposed (MBC+MNP) model (estimated on choice data) and the MBC model (estimated on consideration data) reveal that consideration is strongly determined by the displays used in the experiment. The full model suggests the presence of larger effects of marketing-mix variables such as display and shelf space (or visual salience) on consideration and ultimately choice. Especially noteworthy in Table 2 is that shelf space has no significant effects on stated consideration (ii), and also not on choices using the MNP model (iv). However, from our model (i), shelf space has positive effects on the probability that a brand is considered

[1]We note that if we omit the unobserved heterogeneity in the consideration stage, we find a correlation that is considerably lower, namely $r = 0.75$.

Table 2: Posterior means and standard deviations for parameters of the 3 models: (i+iii) On the left the full model where consideration set knowledge is ignored. On the right 2 models: (ii) a separate MBC-model on consideration set dummies and (iv) a separate MNP-model on choice dummies.

		(i) Full Model: MBC-part[a]		**(ii) MBC-only model**[b]	
		mean	stdev	mean	stdev
stage 1 consideration set	α_{Ajax}	-1.88	0.49	-2.17	0.29
	α_{All}	-1.58	0.25	-1.43	0.20
	α_{AH}	-1.29	0.26	-1.73	0.23
	α_{Bold}	-2.64	0.36	-2.17	0.31
	α_{Cheer}	-1.26	0.42	-0.68	0.37
	α_{Surf}	-0.99	0.66	-1.55	0.27
	α_{Tide}	0.40	0.61	-0.24	0.33
	α_{Wisk}	-2.13	0.32	-2.12	0.31
	display	4.14	1.40	1.75	0.18
	shelf	1.11	0.46	0.58	0.41

		(iii) Full Model: MNP-part		**(iv) Regular MNP**	
		mean	stdev	mean	stdev
stage 2 brand choice	β_{Ajax}	-1.07	0.86	-0.03	0.31
	β_{All}	0.77	0.44	0.16	0.38
	β_{AH}	-0.01	0.40	-0.46	0.52
	β_{Bold}	-0.06	0.34	-0.42	0.33
	β_{Cheer}	0.67	0.43	0.19	0.42
	β_{Surf}	-1.20	1.27	-1.41	0.64
	β_{Tide}	-0.19	0.43	1.21	0.43
	β_{Wisk}[c]	0.00	-	0.00	-
	display			2.62	0.29
	shelf			0.82	0.52

[a] Estimated on observed choice set dummies, ignoring consideration set knowledge. The associated MNP part can be read below.

[b] Estimated on observed consideration set dummies

[c] For identification purposes in the MNP models, we need to select a base brand and set its constant equal to 0. Without loss of generality, we have chosen Wisk.

for choice. In our model (i), choice is allowed to be informative about which brands are considered seriously, which is information that is absent from the consideration dummies in model (ii). Shelf space is not a significant variable in explaining choice using model (iv) because in this model it is specified to have a utility effect, rather than a consideration effect. The specification of shelf space as a utility component is functionally questionable, whereas the specification of shelf space as a component of consideration or awareness is functionally correct (see also Nedungadi 1990).

We now proceed to compare the consideration sets themselves. Using the full model, we can infer the consideration sets from which the subjects made their final choices. We call these sets the "revealed consideration sets." The self-reported measures of consideration are called "stated consideration sets". Note that both stated and revealed consideration sets comprise of numbers in-between 0 and 1, that vary across brands and purchase occasions. In order to establish the validity of inferring consideration sets from choice data, we compute for each brand, individual and choice occasion, the revealed set-membership and its correlation with stated set membership. We find that revealed and stated set membership correlate very highly for each brand. Specifically, for the eight brands these correlations are in the range of 0.44 to 0.78 with an average of 0.62. These values are lower when we use alternative consideration set models, such as the model in Bronnenberg and Vanhonacker (1996). With their model, the values range from 0.37 to 0.60, with an average of 0.50.

Table 3 shows a cross-tabulation of consideration set memberships. A zero indicates the brand is not in the consideration set, whereas a one indicates that it is. If the model predicts that the brand is not in the consideration set, this prediction is correct in 81.6% of the cases. Similarly, if the model predicts that a brand is in the consideration set, the

Table 3: Cross-tabulation of consideration set membership: stated versus estimated

		stated[a]		
		out (0)	in (1)	total
estimated[a]	out(0)	2135	481	2616
		81.6%	18.4%	
	in (1)	194	646	840
		23.1%	76.9%	
	total	2329	1127	3456 [b]

[a] For this cross-tabulation, both the revealed consideration set memberships and the stated consideration sets are rounded to 0 or 1, whichever is nearest.

[b] The data set contains 432 purchase occasions, with eight brands each. Therefore, we have 3456 observations in this cross-tabulation.

model is correct in 76.9% of the cases. These hit rates are far above the baseline hit rates that stem from random sampling off the marginal distribution of stated consideration, which would imply hit rates of 2329/3456 (67.4%) and 1127/3456 (32.6%) respectively. On a brand-by-brand basis, the hit rates of (rounded) revealed and stated consideration sets range from 62% to 94%. In total, the proposed model is correct in 80% of the purchase occasions.

The posterior mean of the covariance terms contained in Σ in the MBC model and the MBC component of the MBC+MNP model are close to 0 and all highest posterior density regions cover the value zero. Therefore, it seems that after taking into account in-store variables and unobserved heterogeneity, little covariation among consideration of brands

is left. Importantly, it thus appears that in order for a brand to enter the consideration set –at least for these data– it does not matter greatly which brands are already in it. To empirically verify that the off-diagionals are indeed indistinguishable from 0, we compute a predictive Bayes Factor to investigate whether the model with an diagonal correlation matrix performs better or worse than a model with a full correlation matrix. The predictive log-likelihood of the model with diagonal correlation matrix is -35.27. For the model with a full correlation matrix this is -36.90. These numbers show that the predictive Bayes Factor weakly favors the model with the identity matrix. This finding provides empirical support for the assumption of independence of consideration set membership across brands, which has been extensively used in the stream of research that uses the stated consideration set approach. What seems to matter is in-store merchandizing at the time of choice and the individual-specific response to this merchandizing. It is of interest that this conclusion is derived from the stated consideration set data, as well as the consideration process derived from the choice data.

We also compare the predictive loglikelihoods of the model with consideration set knowledge versus the model that estimates this from the data. As expected, the MBC model with observed consideration sets produces better predictive likelihoods, but for most brands the split model is not far behind, as can be seen from Table 4.

In sum, this application has demonstrated that (1) we can obtain very similar parameter estimates of the consideration process estimates from choice dummies vs. stated consideration dummies, and that (2) the in-sample consideration set forecasts of the full model, estimated on choice dummies alone, has a high hitrate of about 80%. We take this as strong supportive evidence for the validity of inferring consideration sets from choice data with our model. Our results support the contention that this operationalization of

Table 4: Predictive likelihoods of consideration set membership, for model where we ignore consideration set knowledge and estimate the full model on choice dummies (first row), and for the model where we do use the consideration set knowledge (second row)

	Ajax	All	AH	Bold	Cheer	Surf	Tide	Wisk
unobserved c-sets	-28.5	-27.1	-26.6	-22.5	-37.6	-43.8	-69.1	-19.8
observed c-sets	-14.2	-27.0	-20.0	-19.1	-33.5	-24.9	-47.3	-18.0

consideration, assessed from choices only, is capable of identifying consideration sets.

5 Conclusion

A major reason why estimation of consideration set formation is important to a marketing manager lies in the insights in competitive and positioning issues it provides ("Who are we competing against in the mind of the consumer?", "What is my vulnerability to competitive attacks?") and in control issues ("What will be the effect of my marketing mix variables in various stages?"). It is with these important issues that the insights derived from single-stage and two-stage models of choice really may differ. Our model may give better insight in these questions than previously possible, since it retrieves consideration sets more accurately, it can accommodate explanatory variables in the consideration stage, and because it works easily with data sets with more brands.

Entering consumers' consideration set is one of the top priorities in marketing strategy, and the implementation of those strategies is contingent upon knowledge of the consideration sets of individual consumers. Such knowledge has been obtained by either asking respondents to state their considered set of brands, or by inferring those sets from their

revealed choices. We have proposed, operationalized and estimated a new model to capture unobserved consideration from discrete choice data. This model bridges the stated and revealed approaches, enabling the analysis of either one, or both sources of data to infer sets of brands considered for purchase. Thereby, it enabled us to address the long-standing issue of whether consideration sets can be validly inferred from revealed choice data (cf. Roberts and Lattin, 1997) by studying the convergent validity of stated and revealed consideration sets. While more research in this area is needed, our first findings are promising indeed and we tentatively conclude that we are able to infer consideration from revealed choice behavior using our model.

The proposed model enables different explanatory variables to be included in the consideration and choice stages. Thus, different marketing control variables are allowed to affect the choice process in a different manner, based on theory on how they should affect that process. Our two-stage model offers a more appealing interpretation for the role of in-store merchandizing on consumer choice than a single-stage model does. In the two-stage model, in-store merchandizing has information effects. In contrast, the implication of a single-stage model is that display and shelf space are components of brand utility. But, display and shelf space do not generate the utility as vertical attributes such as price or quality of a brand. Rather, the role of these variables is to facilitate, that is, lower the cost of, consideration of brands, see Andrews and Srinivasan (1995) and Zhang (2006). Thus, we like to see our model as a useful tool in analyzing both stated and revealed consideration data and studying the role of consideration set formation in choice behavior.

A Full Conditional Posterior Distributions

In this appendix we provide the full conditional posterior distributions and sampling algorithms of the model parameters and the unobserved utilities.

Sampling of $\boldsymbol{\alpha}$

To obtain the full conditional posterior distribution of α we rewrite (2) as

$$\Sigma^{-\frac{1}{2}}(C_{it}^* - X_{it}\alpha_i) = \Sigma^{-\frac{1}{2}}X_{it}\alpha + \Sigma^{-\frac{1}{2}}\varepsilon_{it}, \tag{A.1}$$

where $X_{it} = (X_{i1t}, \ldots, X_{iJ,t})'$, for $i = 1, \ldots, I$, $t = 1, \ldots, T_i$. We can interpret (A.1) as J regression equations with regression coefficient α and uncorrelated normal distributed error terms with unit variance. In total we have $J \times \sum_{i=1}^{I} T_i$ of such regression equations. Hence, the full conditional posterior distribution of α given $\{\alpha_i\}_{i=1}^{I}$, Σ and C^* is normal. The mean and variance result from the OLS estimator of α in (A.1), see Zellner (1971, Chapter VIII).

Sampling of $\boldsymbol{\alpha_i}$

To sample α_i for $i = 1, \ldots, I$ we can follow a similar approach as for α. We rewrite (2) as

$$\begin{aligned} \Sigma^{-\frac{1}{2}}(C_{it}^* - X_{it}\alpha) &= \Sigma^{-\frac{1}{2}}X_{it}\alpha_i + \Sigma^{-\frac{1}{2}}\varepsilon_{it} \text{ for } t = 1, \ldots, T_i \\ \mathbf{0} &= \Sigma_\alpha^{-\frac{1}{2}}\alpha_i + \Sigma_\alpha^{-\frac{1}{2}}v_i. \end{aligned} \tag{A.2}$$

The last line follows from the fact that (7) can be written as $v_i = (\alpha_i - \mathbf{0}) \sim \mathrm{N}(\mathbf{0}, \Sigma_\alpha)$. This represents $k_X + JT_i$ regression equations with regression coefficient α_i and uncorrelated normal distributed error terms with unit variance. Hence, the full conditional posterior distribution of α_i given α, Σ_α, Σ and C^* is normal. The mean and variance result from the OLS estimator of α_i in (A.2).

Sampling of $\Sigma_{\boldsymbol{\alpha}}$

For the diagonal elements of Σ_α it holds that

$$p(\sigma^2_{\alpha,kk}|\cdot) \propto (\sigma^2_{\alpha,kk})^{-(I+1)} \exp\left(-\frac{1}{2\sigma^2_{\alpha,kk}} \sum_{i=1}^{I} \alpha^2_{ik}\right), \tag{A.3}$$

for $k = 1, \ldots, k_X$, where α_{ik} is the kth element of α_i. Hence, the diagonal elements of Σ_α can be sampled according to

$$\frac{\sum_{i=1}^{I} \alpha^2_{ik}}{\sigma^2_{\alpha,kk}} \sim \chi^2(I) \text{ for } k = 1, \ldots, k_X. \tag{A.4}$$

Sampling of Σ

To sample Σ we note that

$$p(\Sigma|\cdot) \propto \pi(\Sigma) = |\Sigma|^{-\frac{1}{2}\sum_{i=1}^{I} T_i} \exp(-\frac{1}{2} \sum_{i=1}^{I} \sum_{t=1}^{T_i} (C^*_{it} - X_{it}(\alpha + \alpha_i))' \Sigma^{-1} (C^*_{it} - X_{it}(\alpha + \alpha_i))), \tag{A.5}$$

for $t = 1, \ldots, T_i$ and $i = 1, \ldots, I$.

As Σ is not a free covariance matrix (the diagonal elements are 1), the full conditional distribution is not inverted Wishart. In fact the full conditional posterior distribution of Σ is not standard. To sample Σ we propose a sampler based on Besag and Green (1993) and Damien *et al.* (1999). Loosely speaking, this sampler interchanges the two steps in the Metropolis-Hastings sampler. A possible Metropolis-Hastings sampler for Σ is:

Step 1 Draw the elements of the matrix Σ from a uniform distribution on the interval $(-1, 1)$ under the restriction of positive definiteness, resulting in Σ^{new}.

Step 2 Draw u from a uniform distribution on the interval $(0, 1)$ and accept Σ^{new} if $\pi(\Sigma^{\text{new}})/\pi(\Sigma^{\text{old}}) > u$ otherwise take $\Sigma^{\text{new}} = \Sigma^{\text{old}}$.

For the sampler used in this paper we interchange these two steps. We first draw u from a uniform distribution on the interval $(0, 1)$. In the second step we keep sampling candidate draws of the elements of Σ from a uniform distribution on the interval $(-1, 1)$ until Σ^{new} is positive definite and $\pi(\Sigma^{\text{new}})/\pi(\Sigma^{\text{old}}) > u$. The advantage of the latter approach is that it always results in a new draw, which is not the case for the Metropolis-Hastings sampler, see Damien *et al.* (1999) for details. The disadvantage is that the sampler is slower as one has to draw new candidates until acceptance. Another possibility to generate Σ based on the Metropolis-Hastings sampler is given in Chib and Greenberg (1998) or the hit-and-run algorithm in Manchanda *et al.* (1999).

Sampling of $\boldsymbol{\beta}$

In the brand choice model, β is sampled in a similar way as α. We rewrite the equations (9) as

$$\omega_j^{-1}(U_{ijt} - W'_{ijt}\beta_i) = \omega_j^{-1}W'_{ijt}\beta + \omega_j^{-1}\eta_{ijt}, \tag{A.6}$$

for $j = 1, \ldots, J$, $i = 1, \ldots, I$, and $t = 1, \ldots, T_i$. This represents $J \times \sum_{i=1}^{I} T_i$ regression equations with regression coefficient β and uncorrelated normal distributed error terms with unit variance. Hence, the full conditional posterior distribution of β given $\{\beta_i\}_{i=1}^{I}$, Ω and U is normal. The mean and variance result from the OLS estimator of β in (A.6), see again Zellner (1971, Chapter VIII).

Sampling of $\boldsymbol{\beta_i}$

To sample β_i for $i = 1, \ldots, I$ we can follow a similar approach as for β. We rewrite the equation (9) as

$$\begin{aligned} \omega_j^{-1}(U_{ijt} - W'_{ijt}\beta) &= \omega_j^{-1}W'_{ijt}\beta_i + \omega_j^{-1}\eta_{ijt} \\ \mathbf{0} &= \Sigma_\beta^{-\frac{1}{2}}\beta_i + \Sigma_\beta^{-\frac{1}{2}}v_i, \end{aligned} \tag{A.7}$$

for $j = 1, \ldots, J$ and $t = 1, \ldots, T_i$. The last line follows from the fact (11) implies that $v_i = (\beta_i - \mathbf{0}) \sim \mathrm{N}(\mathbf{0}, \Sigma_\beta)$. This represents $k_W + JT_i$ regression equations with regression coefficient β_i and uncorrelated normal distributed error terms with unit variance. Hence, the full conditional posterior distribution of β_i given β, Σ_β, Ω, and U is normal. The mean and variance result from the OLS estimator of β_i in (A.7).

Sampling of $\boldsymbol{\Sigma_\beta}$

For the diagonal elements of Σ_β it holds that

$$p(\sigma^2_{\beta,kk}|\cdot) \propto (\sigma^2_{\beta,kk})^{-(I+1)} \exp\left(-\frac{1}{2\sigma^2_{\beta,kk}} \sum_{i=1}^{I} \beta^2_{ik}\right), \tag{A.8}$$

for $k = 1, \ldots, k_W$, where β_{ik} is the kth element of β_i. Hence, the diagonal elements of Σ_β can be sampled according to

$$\frac{\sum_{i=1}^{I} \beta^2_{ik}}{\sigma^2_{\beta,kk}} \sim \chi^2(I) \text{ for } k = 1, \ldots, k_W. \tag{A.9}$$

Sampling of $\boldsymbol{\Omega}$

To sample the elements of the covariance matrix Ω we use that

$$p(\omega^2_j|\cdot) \propto \omega_j^{2-(\nu+1)} \exp(-\frac{1}{2\omega^2_j} \sum_{i=1}^{I} \sum_{t=1}^{T_i} (U_{ijt} - W'_{ijt}(\beta + \beta_i))^2), \tag{A.10}$$

and hence

$$\frac{\sum_{i=1}^{I} \sum_{t=1}^{T_i} (U_{ijt} - W'_{ijt}(\beta + \beta_i))^2}{\omega^2_j} \sim \chi^2(\nu) \tag{A.11}$$

with $\nu = \sum_{i=1}^{I} T_i$ for $j = 1, \ldots, J-1$.

Sampling of $\boldsymbol{U}$

To sample U_{it} for $i = 1, \ldots, I$ and $t = 1, \ldots, T_i$, we consider

$$U_{it} = W_{it}(\beta + \beta_i) + \eta_{it}, \tag{A.12}$$

and hence U_{it} is normal distributed with mean $W_{it}(\beta+\beta_i)$ and variance Ω. The conditional distribution of U_{ijt} given $(U_{i1t}, \ldots, U_{i,j-1,t}, U_{i,j+1,t}, \ldots, U_{iJt})$ is of course also normal with, let say, mean m_j and variance s_j^2. Hence, U_{ijt} can be sampled from truncated normal distributions in the following way

$$U_{ijt}|\cdot \sim \begin{cases} \mathrm{N}(m_j, s_j^2) \times \mathbf{I}(-\infty, U_{i,d_{it},t}) & \text{if } j \neq d_{it} \wedge c_{ijt} = 1 \\ \mathrm{N}(m_j, s_j^2) \times \mathbf{I}(\max_{k|k\neq j}(U_{ikt}|c_{ikt} = 1), \infty) & \text{if } j = d_{it} \\ \mathrm{N}(m_j, s_j^2) \times \mathbf{I}(-\infty, \infty) & \text{if } c_{ijt} = 0, \end{cases} \tag{A.13}$$

for $j = 1, \ldots, J$, see Geweke (1991) for details. The value of c_{ijt} is of course determined by the values of the draws of C^*_{ijt}.

Sampling of C^*

To sample C^*_{it} for $i = 1, \ldots, I$ we have to consider (2) for period t. Rewriting this equation gives

$$-\Sigma^{-\frac{1}{2}} X_{it}(\alpha + \alpha_i) = -\Sigma^{-\frac{1}{2}} C^*_{it} + \Sigma^{-\frac{1}{2}} \varepsilon_{ijt} \tag{A.14}$$

This can again be interpreted as a regression model in the parameter C^*_{it}, which implies that the distribution of C^*_{it} is normal with mean and variance following from the OLS estimator of C^*_{it} in (A.14). The conditional distribution of C^*_{ijt} given $(C^*_{1it}, \ldots, C^*_{i,j-1,t}, C^*_{i,j+1,t}, \ldots, C^*_{iJT})$ is in this case also normal with, let say, mean m_j and variance s_j^2, for $j = 1, \ldots, J$. We need to distinguish four situations for the sampling of C^*_{ijt}:

1. The first situation corresponds to $j = d_{it}$. In this case we sample

$$C^*_{ijt}|\cdot \sim \begin{cases} \mathrm{N}(m_j, s_j^2) \times \mathbf{I}(0, \infty) & \text{if } \sum_{k=1,k\neq j}^{J} c_{ikt} > 0 \\ \mathrm{N}(m_j, s_j^2) \times \mathbf{I}(\max_{k|k\neq j}(C^*_{ikt}), \infty) & \text{if } \sum_{k=1,k\neq j}^{J} c_{ikt} = 0. \end{cases} \tag{A.15}$$

2. The second situation corresponds to $j \neq d_{it}$ and $U_{ijt} > U_{i,d_{it},t}$, in which case we sample

$$C^*_{ijt}|\cdot \sim \begin{cases} \mathrm{N}(m_j, s_j^2) \times \mathbf{I}(-\infty, 0) & \text{if } \sum_{k=1,k\neq j}^{J} c_{ikt} > 1 \\ \mathrm{N}(m_j, s_j^2) \times \mathbf{I}(-\infty, \min(0, C^*_{i,d_{it},t})) & \text{if } \sum_{j=1,k\neq j}^{J} c_{ijt} = 1. \end{cases} \tag{A.16}$$

3. The third situation corresponds to $j \neq d_{it}$ and $U_{ijt} < U_{i,d_{it},t}$ with $C^*_{i,d_{it},t} < 0$. In this case we sample

$$C^*_{ijt}|\cdot \sim \mathrm{N}(m_j, s_j^2) \times \mathbf{I}(-\infty, C^*_{i,d_{it},t}). \tag{A.17}$$

4. In all other cases we use the following approach:

$$C^*_{ijt}|\cdot \sim \mathrm{N}(m_j, s_j^2). \tag{A.18}$$

References

Alba, J. and A. Chattopadhyay (1985), The Effects of Context and Part-category Cues on the Recall of Competing Brands, *Journal of Marketing Research*, **22**, 340–349.

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

Andrews, R. L. and T. Srinivasan (1995), Studying Consideration Effects in Empirical Choice Models Using Scanner Panel Data, *Journal of Marketing Research*, **32**, 30–41.

Belonax, J. A. and R. A. Mittelstaedt (1978), Evoked Set Size as a Function of Number of Choice Criterion and Information Variability, in *Advances in Consumer Research*, vol. 5, Association for Consumer Research, Ann Arbor, MI.

Ben-Akiva, M. and B. Boccara (1995), Discrete Choice Models with Latent Choice Sets, *International Journal of Research in Marketing*, **12(1)**, 24–9.

Besag, J. and P. J. Green (1993), Spatial Statistics and Bayesian Computation, *Journal of the Royal Statistical Society B*, **55**, 25–37.

Bettman, J. R. (1979), *An Information Processing Theory of Consumer Behavior*, Addison-Wesley, Reading MA.

Bronnenberg, B. J. and W. R. Vanhonacker (1996), Limited Choice Sets, Local Price Response, and Implied Measures of Price Comptetition, *Journal of Marketing Research*, **33**, 163–174.

Bunch, D. S. (1991), Estimatibility in the Multinomial Probit Model, *Transportation Research B*, **25B**, 1–12.

Chiang, J., S. Chib, and C. Narasimhan (1999), Markov Chain Monte Carlo and Models of Consideration Set and Parameter Heterogeneity, *Journal of Econometrics*, **9**, 223–248.

Chib, S. and E. Greenberg (1998), Analysis of Multivariate Probit Models, *Biometrika*, **85**, 347–361.

Corstjens, M. and J. Corstjens (1999), *Storwars: Competing for Shelf Space and Mind Space of the Consumer*, Wiley, NY.

Damien, P., J. Wakefield, and S. Walker (1999), Gibbs Sampling for Bayesian Non-conjugate and Hierarchical Models by using Auxiliary Variables, *Journal of the Royal Statistical Society B*, **61**, 331–344.

Edwards, Y. D. and G. M. Allenby (2003), Multivariate Analysis of Multiple Response Data, *Journal of Marketing Research*, **40**, 321–334.

Geweke, J. F. (1991), Efficient Simulation from the Multivariate Normal and Student-t Distributions Subject to Linear Constraints, in *Computer Science and Statistics: Proceedings of the 23rd Symposium on Interface*, American Statistical Association, Alexandria, VA.

Gilbride, T. J. and G. M. Allenby (2004), A Choice Model with Conjunctive, Disjunctive, and Compensatory Screening Rules, *Marketing Science*, **23**, 391–406.

Guadagni, P. M. and J. D. Little (1983), A Logit Model of Brand Choice Calibrated on Scanner Data, *Marketing Science*, **2(3)**, 203–238.

Hausman, J. and D. Wise (1978), A Conditional Probit Model for Qualitative Choice:

Discrete Decsions Recognizing Interdependence and Heterogenous Preferences, *Econometrica*, **45**, 319–339.

Howard, J. A. and J. N. Sheth (1969), *The Theory of Buyer Behavior*, New York: Wiley.

Keane, M. P. (1992), A Note on Idenfication in the Multinomial Probit Model, *Journal of Business & Economic Statistics*, **10**, 193–200.

Lehmann, D. R. and Y. Pan (1994), Context Effects, New Brand Entry, and Consideration Sets, *Journal of Marketing Research*, **31**, 364–74.

Lussier, D. A. and R. W. Olshavsky (1979), Task Complexity and Contingent Processing in Brand Choice, *Journal of Consumer Research*, **6**, 154–165.

Malhotra, N. K., M. Peterson, and S. B. Kleiser (1999), Marketing Research: A State-of-the-Art Review and Directions for the Twenty-First Century, *Journal of the Academy of Marketing Science*, **27-2**, 160–183.

Manchanda, P., A. Ansari, and S. Gupta (1999), The "Shopping Basket": A Model for Multicategory Purchase Incidence Decisions, *Marketing Science*, **18:02**, 95–115.

Manrai, A. K. and R. L. Andrews (1998), Two-stage Discrete Choice Models for Scanner Panel Data: An Assessment of Process and Assumptions, *European Journal of Operational Research*, **111**, 193–215.

Manski, C. F. (1977), The Structure of Random Utility Models, *Theory and Decision*, **8**, 229–254.

McCulloch, R. E., N. G. Polson, and P. E. Rossi (2000), A Bayesian Analysis of the Multinomial Probit Model with Fully Identified Parameters, *Journal of Econometrics*, **99**, 173–193.

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

McFadden, D. (1973), Conditional Logit Analysis of Qualitative Choice Behavior, *Frontiers in econometrics*, 105–142.

Mehta, N., S. Rajiv, and K. Srinivasan (2003), Price Uncertainty and Consumer Search: A Structural Model of Consideration Set Formation, *Marketing Science*, **22**, 58–84.

Mitra, A. (1995), Advertising and the Stability of Consideration Sets Over Multiple Purchase Occasions, *International Journal of Research in Marketing*, **12**, 81–94.

Nedungadi, P. (1990), Recall and Consumer Consideration Sets: Influencing Choice Without Altering Brand Evaluations, *Journal of Consumer Research*, **17**, 263–276.

Newell, A. and H. A. Simon (1972), *Human Problem Solving*, Englewood Cliffs, NJ: Prentice-Hall.

Payne, J. W. (1976), Task Complexity and Contingent Processing in Decision-Making: An Information Search and Protocol Analysis, *Organizational Behavior and Human Performance*, **16**, 366–387.

Roberts, J. H. and J. M. Lattin (1991), Development and Testing of a Model of Consideration Set Composition, *Journal of Marketing Research*, **28**, 429–440.

Roberts, J. H. and J. M. Lattin (1997), Consideration: Review of Research and Prospects for Future Insights, *Journal of Marketing Research*, **34**, 406–410.

Siddarth, S., R. Bucklin, and D. G. Morisson (1995), Making the Cut: Modeling and Analyzing Choice Set Restriction in Scanner Panel Data, *Journal of Marketing Research*, **33**, 255–266.

Wright, P. and F. Barbour (1977), Phased Decision Strategies: Sequels to Initial Screening, in M. Starr and M. Zeleny (eds.), *Multiple Criteria Decision Making: North-Holland TIMS Studies in the Management Science*, Amsterdam: North-Holland.

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

Zhang, J. (2006), An Integrated Choice Model Incorporating Alternative Mechanisms for Consumers Reactions to In-Store Display and Feature Advertising, *Marketing Science, forthcoming*, **25**, 278–290.