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 Bronnenberg

The Anderson School of Management at UCLA

Philip Hans Franses

Econometric Institute, Erasmus University Rotterdam

Michel Wedel

Stephen M. Ross School of Business at the University of Michigan

ECONOMETRIC INSTITUTE REPORT EI 2005-49

Retrieving Unobserved Consideration Sets from Household Panel Data

Abstract

We propose a new model to describe consideration, consisting of a multivariate probit 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 accommodates different effects of the marketing mix on consideration and choice, an error process that is correlated over time, 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.

Next, we estimate the model on IRI panel data. We have two main results. First, compared with the single-stage multinomial probit model, promotion effects are larger when they are included in the consideration stage of the two-stage model. Second, we find that consideration of brands does not covary greatly across brands once we account for observed effects.

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 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 sets are interesting from a marketing perspective 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). 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 ex-

ponential 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 MNP choice model, calibrated on stated preference data. Their's 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 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.

The two streams of consideration set research described above have evolved somewhat independently. There is no existing empirical evidence as to the convergence of these two streams of consideration research. Do the "consideration probabilities", that the models used in the revealed consideration set approach, estimate from choice data, 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 one of the intended contributions of the study in this paper, that is, 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 probit model (Edwards and Allenby, 2003, MVP) for consideration, compounded with a multinomial probit (MNP, McCulloch and Rossi, 1994; McCulloch *et al.*, 2000) model for brand choice given consideration. In the MVP 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. 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 include marketing control variables and "the hand of the past" in the consideration and choice stages. We develop our model primarily for the purpose of obtaining better substantive insights into choice processes, and the effect of marketing variables 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. Subsequently we apply our model to a scanner panel data set on coffee purchases and discuss our findings. We finish 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 formulation that involves

$$C_{ijt}^* = X_{ijt}'(\alpha + \alpha_i) + e_{ijt}, \qquad j = 1, \ldots, J, \tag{2}$$

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 e_{ijt} is an unknown error process. To allow for

dynamics in the consideration set formation, we assume that the error process terms e_{ijt} follow an autoregressive process of order 1, that is,

$$e_{ijt} = \rho e_{ij,t-1} + \varepsilon_{ijt}, \tag{3}$$

where ε_{ijt} is an unobserved disturbance term. This dynamic process intends to model (habit) persistence in consideration set membership of brands. Although a similar approach was used by Allenby and Lenk (1994) in a standard MNL brand choice model, the development of a MVP with dynamics as done here is new to the literature. Note that X_{ijt} may contain lagged purchase dummies, which also allows us to model state dependence capturing possible memory effects.

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. Here we are interested primarily in characterizing consideration and not in purchase incidence. Therefore, in case the C_{ijt}^* are all negative, we assume that the highest utility brand will be (the only brand) in the consideration set.

For the household considering buying only the first two brands, the first two elements of the vector C_{it}^* are positive, while the remaining elements are all negative. To illustrate, the probability that the consideration set of household i contains only the first two brands equals

$$\Pr[C_{it} = (1, 1, 0, \ldots, 0)'] = \Pr[C_{i1t}^* > 0, C_{i2t}^* > 0, C_{i3t}^* \leq 0, \ldots, C_{iJt}^* \leq 0]$$

$$= \Pr[e_{i1t} > -X_{i1t}'(\alpha + \alpha_i), e_{i2t} > -X_{i2t}'(\alpha + \alpha_i), \tag{4}$$

$$e_{i3t} \leq -X_{i3t}'(\alpha + \alpha_i), \ldots, e_{iJt} \leq -X_{iJt}'(\alpha + \alpha_i)].$$

This probability depends on the distribution of the e_{ijt}, which follows directly from the distribution of the vector of disturbances $\varepsilon_{it} = (\varepsilon_{i1t}, \ldots, \varepsilon_{iJt})'$. We assume that ε_{it} is

normally distributed, that is,

$$\varepsilon_{it} \sim \mathrm{N}(\mathbf{0}, \Sigma), \tag{5}$$

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. Furthermore, for the first observation of each household we assume that $e_{i1} \sim \mathrm{N}(\mathbf{0}, \Sigma/(1-\rho^2))$, which is the unconditional distribution of e_{it}.

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

$$\alpha_i \sim \mathrm{N}(\mathbf{0}, \Sigma_\alpha), \tag{6}$$

where Σ_α is a diagonal matrix. An advantage of this approach is that it leads to a non-diagonal covariance structure in the multivariate probit models, that is, the unconditional covariance structure of C_{it} equals

$$X_{it}\Sigma_\alpha X_{it}' + \Sigma/(1-\rho^2), \tag{7}$$

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,

$$U_{ijt} = W_{ijt}'(\beta + \beta_i) + \eta_{ijt}, \qquad j = 1, \ldots, J, \tag{8}$$

10

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 probit disturbances $\eta_{it} = (\eta_{i1t}, \ldots, \eta_{iJt})'$ is assumed to be normally distributed:

$$\eta_{it} \sim \mathrm{N}(\mathbf{0}, \Omega). \tag{9}$$

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

$$\beta_i \sim \mathrm{N}(\mathbf{0}, \Sigma_\beta), \tag{10}$$

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

$$U_{ijt} = \max(U_{ikt} \text{ for } k = 1, \ldots, J | c_{ikt} = 1). \tag{11}$$

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

$$
\begin{aligned}
\Pr[D_{it} = j | c_{it}, \beta_i; \beta, \Omega] &= \Pr[U_{ijt} > U_{ikt} \; \forall k \neq j | c_{ijt} = c_{ikt} = 1] \\
&= \Pr[U_{ijt} - U_{ikt} > 0 \; \forall k \neq j | c_{ijt} = c_{ikt} = 1] \\
&= \Pr[\eta_{ikt} - \eta_{ijt} < W'_{ijt} - W'_{ikt}(\beta + \beta_i) \; \forall k \neq j | c_{ijt} = c_{ikt} = 1].
\end{aligned}
\tag{12}
$$

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 in several (unreported) test runs of our model and hence we opt for a diagonal covariance matrix. As

11

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 $\Omega = \mathrm{diag}(\omega_1^2, \ldots, \omega_{J-1}^2, 1)$. This diagonal structure is generalized to a non-diagonal covariance matrix by modeling the unobserved household heterogeneity (see 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 multivariate probit 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. Furthermore, we address dynamics in the consideration set formation.

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,

$$
\mathcal{L}(d|\theta) = \prod_{i=1}^{I} \int_{\beta_i} \int_{\alpha_i} \Bigg[\sum_{\forall c_{i1} \in \mathcal{C}_{i1}} \sum_{\forall c_{i2} \in \mathcal{C}_{i2}} \cdots \sum_{\forall c_{iT_i} \in \mathcal{C}_{iT_i}}
$$
$$
\Pr[C_{i1} = c_{i1}, C_{i2} = c_{i2}, \ldots, C_{iT_i} = c_{iT_i} | \alpha_i; \alpha, \rho, \Sigma, \Sigma_\alpha]
$$
$$
\prod_{t=1}^{T_i} \Pr[D_{it} = d_{it}|c_{it}, \beta_i; \beta, \Omega] \Bigg] \phi(\beta_i; \mathbf{0}, \Sigma_\beta) d\beta_i \phi(\alpha_i; \mathbf{0}, \Sigma_\alpha) d\alpha_i, \quad (13)
$$

where $\theta = (\alpha, \Sigma_\alpha, \rho, \beta, \Sigma_\beta, \Sigma, \Omega)$ and $\mathcal{C}_{it}$ is the set of potential consideration sets for household i at purchase occasion t. The likelihood function contains the joint probability that

the consideration sets of household i are equal to $c_i = (c_{i1}, \ldots, c_{iT_i})$, see, for example, (4), and the product of the brand choice probabilities given c_{it}, see (12), over 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 MVP 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. In Appendix A we provide the details of the full conditional posterior distributions and sampling algorithms for the model parameters. 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 consisting of stated choice and consideration protocol data collected in an on-line experiment. 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, subjects chose among 8 brands of laundry detergent over 10 choice occasions. In the experiment, consumers interfaced with a digital image of a supermarket shelf, containing the universal set of choice options. The choice environment was constant across individuals but varied across choice occasions. We manipulated promotion, price, brand position on the shelf and shelf facings.

Figure 1 shows a screen-shot from the sixth choice occasion. If subjects clicked on any of the brands on the shelf they received product information, that is, the brand slogan put on the front of the package by the manufacturer (for example Cheer has as its slogan "With Colorguard"). It may be noted that these slogans could not be seen by the subject by just looking at the shelf (see Figure 1). They had to make the effort to click the box. If they clicked on the corresponding bar-codes on the shelves they received price information. We simulated a promotion environment by putting "end-of-aisle" displays into the simulation. These were created by showing the brand on promotion in isolation with a price message 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

14
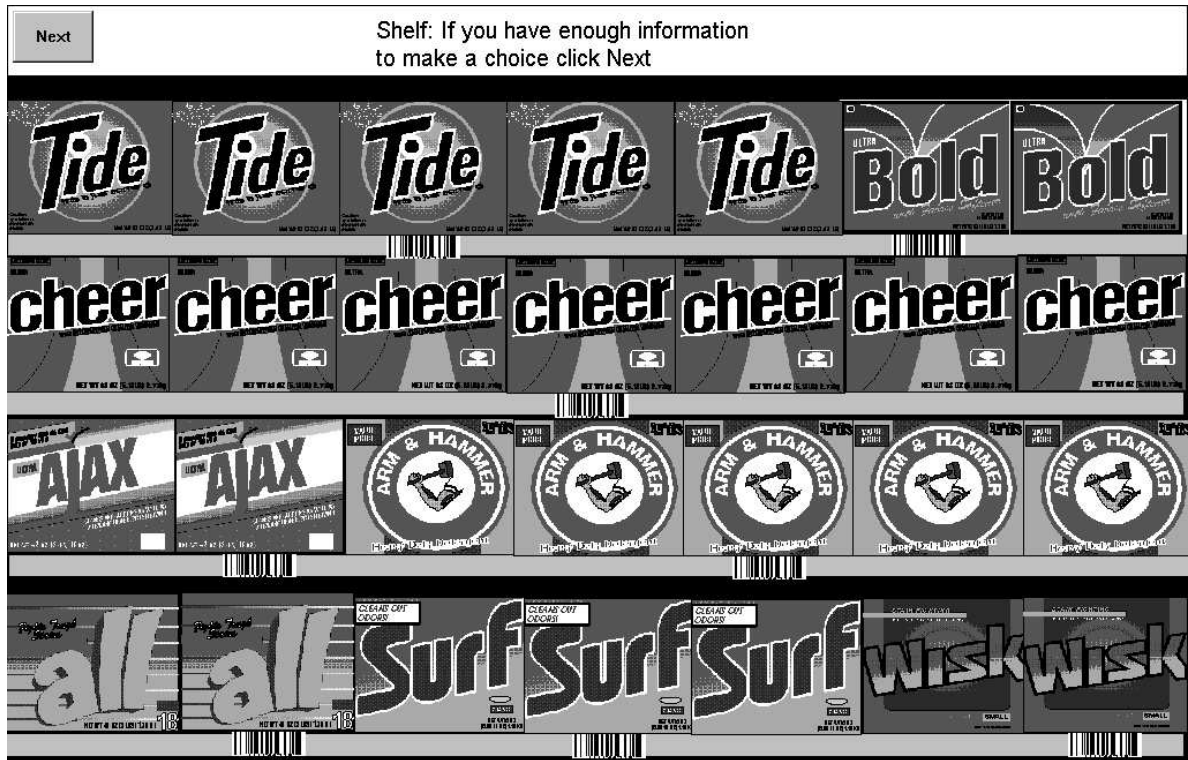
Figure 1: Screen-shot from the sixth choice occasion.

and visit the regular shelf.

The experiment served to measure the full choice protocol. This is to say, we measured (revealed) choice, information acquisition, and stated consideration set membership. The latter 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).

The experiment was administered to 65 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. Because 2 of the 8 brands were too rarely chosen, these were

Table 1: Descriptive statistics for the experimental data set ($N = 528$)

Brand	Share	Consideration[a]	Display frequency	Average shelf	Average price
All	10.6%	27.6%	10%	0.35	$3.04
Arm&Hammer	11.4%	20.3%	10%	0.39	$2.69
Bold	4.7%	23.3%	10%	0.37	$3.54
Cheer	27.3%	58.8%	20%	0.79	$3.67
Surf	4.9%	17.1%	0%	0.43	$3.59
Tide	41.1%	66.8%	20%	0.73	$3.66

[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.

dropped from the analysis. This left us with $N = 528$ observations. Of these data, we randomly sampled 48 individuals with $N = 416$ purchases for estimation. 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." The price variable is measured in US dollars.

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 (0.56) 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

To estimate the full model it is necessary to define the covariates affecting consideration and those affecting choice, respectively. In the past, some studies have simply included all variables in both stages of the model (for example, Andrews and Srinivasan, 1995), but there is increasing evidence that different marketing control variables affect the choice process in very different ways (Zhang, 2005). Therefore, in this paper, we follow a different strategy. We are explicit about which marketing actions we believe to affect consideration and choice separately and we validate our choices using the measured consideration sets from the experiment.

We assume that consideration is driven by memory for the brands considered last, and by in-store merchandizing activity to make a brand more salient at point-of-purchase. Specifically, we assume that consideration is driven by point-of-purchase merchandizing which in turn is operationalized in this study as the effect of display, feature and shelf-space measures. In addition, we allow brand-salience or consideration-utility to be correlated across purchase occasions.

With respect to brand choice, given consideration, we assume that it is determined by

17

the value of the brand to a consumer given the information that the consumer has at the time. This means that we assume that the effect of price takes hold in the choice stage. 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. Finally, as described in Section 2, we allow for unobserved household heterogeneity in the consideration stage. Including unobserved heterogeneity in the choice stage as well rendered the model unstable and gave rise to convergence problems in the MCMC chain, so that we have not included it. Apparently, unobserved heterogeneity in the consideration stage and autocorrelation in that stage are all that the data can support in this application.

So, in summary, we view consideration as a state of motivated awareness of a given choice option. In contrast, choice emanates from an evaluation of the "value proposition" (essentially benefits minus price) and this evaluation is done only for the brands for which one is sufficiently motivated at point of purchase. We attempt to capture the behavioral state of motivated awareness through a construct that is affected by (a) memory (carry-over) and (b) in-store generators of salience such as display signs and feature ads. Choice given consideration is not dependent on such aspects; there is no consumer utility attached to a feature ad, i.e., the ad is not consumed and is merely a source of information. Instead choice is made based on inherent brand benefits and price. This gives rise to the partition of the variables in the model.

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 consider three models on the data from the choice experiment with stated consideration sets. First, we consider the full multivariate probit/multinomial probit (MVP+MNP)

Table 2: Posterior means and posterior standard deviations of the full model for the experimental detergent data.

Consideration stage	mean	std. dev.	Choice stage	mean	std. dev.
α_{All}	-0.33	0.97	β_{All}	0.75	0.74
α_{AH}	0.76	1.13	β_{AH}	-1.19	1.02
α_{Bold}	0.55	1.23	β_{Bold}	-2.48	1.23
α_{Cheer}	0.19	0.72	β_{Cheer}	0.28	1.04
α_{Surf}	-1.12	1.58	β_{Surf}	-1.96	1.96
α_{Tide}	-0.42	0.67	β_{Tide}[a]		
display	-0.28	0.23			
shelf	2.85	1.17			

[a] For identification purposes, we need to select a base brand and set its intercept equal to 0. Without loss of generality, we have chosen Tide.

model of choice and consideration and we report the results in Table 2. This involves estimating the consideration effects α, the choice effects β, and the autocorrelation in consideration ρ, and its covariance of consideration, Σ, as well as the consideration heterogeneity covariance matrix Σ_α. The Bayesian analysis is based on the *choice data alone.* Second, we consider the MVP model by itself using the stated consideration sets. Third, we consider a multinomial probit (MNP) model with an autoregressive error process and unobserved heterogeneity in the effects on choice. For these models, see Table 4. For lack of variation over time, we need to drop the price variable from these analyses. When estimating the parameters, the price parameter was not separable from the brand intercepts, indicating that there is little price variation beyond the differences among brands.

The consideration set model allows for some interesting interpretations. For instance note the differences in the posterior means of the brand intercepts across the MNP and

Table 3: Posterior means and standard deviations of the unobserved heterogeneity (diagonal elements of Σ_α) and autocorrelation parameter for the experimental detergent data.

	mean	std. dev.
All	0.25	0.14
AH	0.27	0.15
Bold	0.26	0.15
Cheer	0.27	0.15
Surf	0.31	0.24
Tide	0.31	0.24
ρ	0.86	0.11

the MVP+MNP model. In the MNP model, the brand intercept is considered an overall measure of brand equity. There is a clear ordering of the brands, with Bold lowest and Tide highest. However, the MNP-component of the full model shows marked differences. For instance, while its intercept in the MNP model is low, the intercept for the brand All is high in the MNP stage of the full model. That is to say, among those who consider the brand, it is a brand that is of high value, a niche brand in other words. In effect, the full model partitions the overall equity effect into an effect that reflects the probability of consideration, and an effect that reflects brand utility (given consideration). Note that the small share brands Bold and Surf seem to suffer from double jeopardy. These brands are considered on only a few occasions. In addition, when they are considered, they have a low baseline choice probability.

Tables 2 and 4 show that the proposed (MVP+MNP) model (estimated on choice

Table 4: Posterior means and standard deviations of the separate MVP and MNP model for the experimental detergent data.

MVP[a]	mean	std. dev.	MNP[b]	mean	std. dev.
α_{All}	-1.29	0.46	β_{All}	0.32	0.70
α_{AH}	-1.63	0.49	β_{AH}	0.26	0.74
α_{Bold}	-1.13	0.20	β_{Bold}	-0.79	0.46
α_{Cheer}	0.50	0.20	β_{Cheer}	0.22	0.46
α_{Surf}	-2.15	0.30	β_{Surf}	-2.70	0.59
α_{Tide}	0.16	0.24	β_{Tide}[c]		
display	0.07	0.12	display	-0.24	0.20
shelf	1.17	0.54	shelf	1.42	1.16
ρ	0.79	0.03	ρ	0.39	0.10

[a] Estimated on consideration dummies only

[b] Estimated on choice dummies only

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

data) and the MVP model (estimated on consideration data) reveal that consideration is strongly determined by shelf space. Past consideration has strong effects as well, which is revealed by the posterior mean value of 0.86 for ρ in Table 3. This table also shows that there is much variance in the individual-specific consideration intercepts α_i. We observe that the combination of in-store activity such as shelf space, and past consideration captures a large part of the variation in consideration sets across individuals and purchase occasions. For comparison purposes, we also show the parameter estimates for the consideration set model of Bronnenberg and Vanhonacker (1996), see Table 5.

The posterior mean of the covariance terms contained in Σ in the MVP model and the MVP component of the MVP+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

Table 5: Estimation results for Bronnenberg and Vanhonacker (1996) model.

	Estimate	st. dev.
$\alpha_{\text{low price tier}}$	-0.82	0.73
threshold	14.64	0.97
α_{display}	3.06	0.31
$\alpha_{\text{shelfspace}}$	2.39	0.88
$\text{Log}(\lambda_{\text{recency}}/(1 - \lambda_{\text{recency}}))^{\text{a}}$	1.33	0.24
α_{recency}	7.47	1.01
β_{all}	0.23	0.72
β_{ArmH}	1.17	0.70
β_{bold}	-0.06	0.43
β_{cheer}	-0.81	0.21
β_{surf}	0.18	0.42

[a] This implies that the memory coefficient in the recency variable is equal to $\lambda_{\text{recency}} = 0.79$

variables and past consideration, little covariation among consideration of brands is left. This result appears even if we do not include unobserved household heterogeneity in the consideration stage of the model. 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. 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 whether a brand was considered last time and whether there is in-store merchandizing at the time of choice. It is of interest that this is conclusion is derived from the stated consideration set data, as well as the consideration process derived from the choice data.

We cannot directly compare the consideration stage parameters in Table 2 from the MVP+MNP model with those from the MVP model in Table 4, given the identification constraints imposed on these two models. Thus, rather than comparing the parameters, we will 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 six brands these correlations are in the range of 0.48 to 0.75 with an average of 0.68. These values are lower when we use alternative consideration set models, such as the model in Bronnenberg and Vanhonacker (1996). With this model, the values range from 0.37 to 0.60, with an average of 0.50.

Table 6 shows a cross-tabulation of consideration set memberships. A zero indicates the brand is not in the consideration set, whereas a one indicates the opposite. The misclassification is not symmetric, that is, if the brand is not in the consideration set according to the individual, our estimate is usually correct (in 84% of the cases, that is, 1369/1635). On the other hand, if the individual stated that the brand is contained in his or her consideration set, we estimate this correctly in 69% of the cases. On a brand-by-brand basis, the hit rates of (rounded) revealed and stated consideration sets range from 62% to 94%. In total, we are correct in 78% of the purchase occasions.

23

Table 6: Cross-tabulation of consideration set membership: stated versus estimated

| | | stated[a] | | |
		out (0)	in (1)	total
estimated[a]	out(0)	1369	272	1640
	in (1)	266	589	856
	total	1635	861	2496 [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 416 purchase occasions, with six brands each. Therefore, we have 2496 observations.

We take the above findings 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 consideration, identified from choices only, is capable of tracking the differences in choice sets both over time as well as across individuals.

5 Application to Scanner Panel Data

5.1 Data

For the illustration of the model to choice data, we also consider an optical scanner panel data set on purchases of nine brands of coffee, both ground and soluble. The data set contains information on all 2978 purchases of coffee made by 232 households in a large U.S. city, and cover a two-year period from June 1991 to June 1993. The brands and their respective number of purchase and marketing instrument statistics are given in Table 7.

Table 7: Descriptive statistics for the coffee data set ($N = 2978$)

| | Number of purchases | | Marketing instruments | | |
Brand	estimation sample	holdout sample[a]	display	feature	price/oz
Ground					
Eight O'Clock	94	12	20.4%	6.9%	$ 0.166
Folgers	378	28	5.3%	4.8%	$ 0.168
Hills Brothers	819	58	17.7%	12.7%	$ 0.152
Maxwell House	486	26	13.8%	13.1%	$ 0.188
MJB	151	17	8.4%	5.1%	$ 0.165
Papanicholas Sig	225	12	0.2%	1.1%	$ 0.275
Soluble					
Folgers	186	18	0.0%	0.6%	$ 0.518
General Foods	189	11	2.6%	2.1%	$ 0.461
Maxwell House	244	24	1.3%	5.3%	$ 0.471
Total	2772	206			

[a] This column shows the number of purchases that are in the (longitudinal) holdout sample.

The variation in choice shares of the brands is somewhat higher than for the experimental data in Table 1. The relative choice share of Hills Brothers is the highest. Among soluble coffee, Maxwell House is the brand that is purchased most. Prices are expressed in US dollars per oz. It may be observed that price variation in this data set is much larger than in the experimental data. Display and feature frequency are defined as the fraction of occasions that a brand is on display or has a feature. The variation in display frequency across brands is somewhat larger than observed in the experimental data. The data reflect

substantially different strategies in terms of promotions and pricing. The soluble items get less promotion by means of display or feature, and their prices are considerably higher than the ground items.

5.2 Estimation results from the empirical data

We consider the following three models on the coffee data, using the same operationalizations as described for the experimental data above. Again, this involves estimating the consideration effects α, the autocorrelation in consideration ρ, the covariance of consideration Σ, the choice effects β, the covariance matrix Σ_α of the random consideration set effects α_i, and the (diagonal) covariance matrix of the choice utilities Ω. Adding unobserved heterogeneity in the MNP component as well again resulted in instability of the MCMC chain, and we have not included that component in the model. We also consider a single-stage MNP choice model with again similar specifications (autoregressive error process and unobserved heterogeneity). Finally, to benchmark our model to we estimate the model of Bronnenberg and Vanhonacker (1996).

The posterior results for the coffee data are given in Table 8. Posterior results of the unobserved heterogeneity parameters are given in Table 9. From the results of the proposed (MVP+MNP) model we see that the posterior means of all marketing parameters are far away from zero (when compared to the posterior standard deviation) and that they are all of the expected sign. Consistent with the controlled choice experiment, the posterior mean of the covariance terms in Σ in the MVP model are close to 0 and all highest posterior density regions cover the value zero. This does not appear to be caused by the unobserved heterogeneity, as it also happens when we estimate the model without unobserved heterogeneity in the consideration stage. Furthermore, we have calibrated

Table 8: Posterior means and standard deviations for the full model for the coffee data.

Consideration stage	mean	std. dev.	Choice stage	mean	std. dev.
α_{Eight}	-3.79	0.28	β_{Eight}	-0.66	1.65
α_{Folgers}	-2.41	0.23	β_{Folgers}	-2.34	1.29
α_{Hills}	-1.83	0.25	β_{Hills}	-2.93	0.41
α_{Maxwell}	-2.27	0.24	β_{Maxwell}	-1.98	1.29
α_{MJB}	-3.10	0.25	β_{MJB}	-2.70	0.95
$\alpha_{\text{NAPapanicholas}}$	-4.28	0.47	$\beta_{\text{NAPapanicholas}}$	1.24	1.02
$\alpha_{\text{Folgers Soluble}}$	-3.66	0.26	$\beta_{\text{Folgers Soluble}}$	2.26	1.19
$\alpha_{\text{General Soluble}}$	-4.15	0.38	$\beta_{\text{General Soluble}}$	0.67	0.84
$\alpha_{\text{Maxwell Soluble}}$	-3.35	0.28	Price	-3.19	0.82
Display	1.16	0.13			
Feature	1.12	0.11			

[a] For identification purposes, we need to select a base brand and set its intercept equal to 0. Without loss of generality, Maxwell House Soluble is chosen as base brand.

[b] The posterior means of the covariances in the MVP model are close to 0 and are not displayed here.

versions of our model on three other data sets (cracker data with 4 brands, soft drinks with 10 brands and yoghurt, also with 10 brands), and have found little or no correlation in the consideration set stage either. Again, it appears that in order for an alternative to enter the consideration set, it does not matter greatly which alternatives are already in it.

The posterior means of the brand intercepts for the MNP-component of the full model show that Folgers (Soluble) has the highest choice share given consideration. However, the posterior mean of the MVP brand intercepts reveal that this brand has a rather low base probability of being considered, irrespective of marketing activity. The same holds for Papanicholas (Ground). These brands could therefore be considered as niche

Table 9: Posterior results for the unobserved heterogeneity (diagonal elements of Σ_α) and autocorrelation parameters for the coffee data.

Brand / parameter	mean	std. dev.
Eight	1.00	0.42
Folgers	1.31	0.30
Hills	1.64	0.34
Maxwell	0.87	0.25
MJB	0.86	0.23
NA Papanicholas	4.61	1.10
Folgers Soluble	4.22	0.74
General Soluble	4.23	0.86
Maxwell Soluble	4.35	1.01
ρ	0.20	0.04

brands. Heterogeneity in the consideration set brand intercepts is high, see Table 9. This holds in particular for the soluble brands Folgers, General and Maxwell, and for Papanicholas. Apparently, there are substantial differences among consumers in whether or not these brands are considered, irrespective of promotional efforts. The autocorrelation ρ of consideration is lower than in the experimental data, but with a posterior mean of 0.20, it still is substantial. This is an important finding for marketers, since it reveals considerable habit persistence in consideration, regardless of marketing activity. In other words, if brands enter the consideration set through promotional efforts such as display and feature, this is more than an instantaneous effect and it carries over to subsequent purchase occasions.

Both models show marketing effects with the expected sign. While the effects for feature and display are very similar for the proposed model and the MNP in Table 10, we

Table 10: Posterior means and standard deviation for the MNP model[a] for brand choice only for the coffee data.

MNP	mean	std. dev.
β_{Eight}	-0.05	0.16
β_{Folgers}	0.99	0.05
β_{Hills}	0.91	0.07
β_{Maxwell}	1.00	0.05
β_{MJB}	0.49	0.07
$\beta_{\text{NAPapanicholas}}$	0.03	0.24
$\beta_{\text{Folgers Soluble}}$	1.26	0.09
$\beta_{\text{General Soluble}}$[b]	0.37	0.11
Price	-1.57	0.21
Display	0.27	0.06
Feature	0.26	0.04
ρ	0.17	0.06

[a] To ease the comparison with the MVP+MNP model results, this single-stage MNP model has a diagonal covariance matrix, unobserved heterogeneity and autocorrelated errors.

[b] For identification purposes, we need to select a base brand and set its intercept equal to 0. Without loss of generality, Maxwell House Soluble is chosen as base brand.

would like to point to the large difference in the price coefficient. The price effect, given consideration, is three times as large. This finding, which has been previously documented in the literature, is a very important one from a strategic perspective. It shows that, once a brand has entered the consideration set through promotional efforts such as display and feature, the price instrument is very effective in increasing market share and decreasing that of competitors in the consideration set. In general, it is difficult to compare the parameters of the full model to the MNP model shown here. One reason for the smaller

price coefficient in the MNP model may be that the variables for display and feature capture part of the price variability. Indeed, when we estimate an MNP model with only intercepts and price as explanatory variables, we find a much stronger price effect. However, even then the two models are hard to compare, since the latter does not contain promotion effects at all, and the correlated error structure still appears in a different place. To facilitate comparison, we therefore used (unreported) simulation to confirm that the same price cut for a particular brand leads to a slightly higher expected market share (unconditional on consideration) in the two-stage model, compared to the single-stage MNP model presented in Table 10. The two-stage model tells us that consumers are indeed more price sensitive. The stability of the display and feature coefficients across the our proposed model and the MNP model, and their tighter distribution when included in the consideration set component of the proposed model (relative to the magnitude of the parameter estimates), may provide an indication that these two marketing control variables do primarily serve to induce consumers to consider the brands in the data set at hand.

Although the model is not purposely built to make forecasts, out-of-sample predictions show that the hit rate of the full model is 65% for hold out samples. The single-stage MNP model produces the same hit rate as we expected. The in-sample hit rate for our model equals 79%, whereas for the MNP it is somewhat lower with 77%[1]. When we apply the model used by Bronnenberg and Vanhonacker (1996) to our data set, we obtain a longitudinal hit rate of 61%. The in-sample hit rate for this model equals 64%, which is also lower than for our model. Although we see a clear difference in the forecasting hit rate

---

[1]The MNP with only intercepts and price, which was proposed before for coefficient comparison purposes, yields a longitudinal hit rate of only 61% and an in-sample hit rate of only 74%

in-sample, the posterior distributions of longitudinal forecast hit rate for the MNP and MVP + MNP models overlap. Of course one would have liked to see the added complexity of our model to result in substantial improvements in predictive performance, but as has been found previously, a simpler but theoretically less completely specified model as the MNP also predicts well. We think that the major advantage of our model accrues from its diagnostic value. We conjecture that the main reason why estimation of consideration set formation is important to a marketing manager may not be prediction, but 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.

6 Conclusion

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 applications of our model revealed substantial heterogeneity in consideration, and strong habit persistence in consideration. Thus, marketing activity directed at entering the consumers' widely differing consideration sets is likely to persist over longer periods of time. Once in the consideration set, price is an extremely effective competitive instrument – more so than predicted by previous single stage models. Looking at consideration and choice simultaneously enabled the effective identification of niche brands. These are brands that have a relatively low consideration probability, but, once considered they have a high probability of being chosen. For these brands in particular, increasing consideration through promotional effort is predicted to be highly effective.

The proposed model enables different explanatory variables to be included in the consideration and choice stages, whereby it distinguishes itself from the work by Gilbride and Allenby (2004). We included in-store merchandizing (display and feature) in the consideration stage (see also Alba and Chattopadhyay, 1985) and we included brand intercepts and prices in the choice stage. 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.

32

In the two-stage model, in-store merchandizing has information effects. In contrast, the implication of a single-stage model is that display and feature are components of brand utility. But, feature ads and display do not generate the same utility as when paying low price or receiving high 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). In-store merchandizing programs are therefore suitably seen as fulfilling the goal of lowering the mental cost of information acquisition (see Zhang, 2005, for a more detailed representation of the role of feature ads and display on consideration and choice). 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

Sampling of α

To obtain the full conditional posterior distribution of α we rewrite (2) and (3) as

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

where $X_{it} = (X_{i1t}, \ldots, X_{iJ,t})'$, for $i = 1, \ldots, I$, $t = 2, \ldots, T_i$. For the first observation of each household we have

$$\sqrt{1-\rho^2}\Sigma^{-\frac{1}{2}}(C_{i1}^* - X_{i1}\alpha_i) = \sqrt{1-\rho^2}\Sigma^{-\frac{1}{2}}X_{i1}\alpha + \sqrt{1-\rho^2}\Sigma^{-\frac{1}{2}}e_{i1}, \qquad (A.2)$$

for $i = 1, \ldots, I$. We can interpret (A.1) and (A.2) as J regression equations with regression coefficient α and uncorrelated normal distributed error terms with unit variance. In total we have $J \times \sum_{i=1}^N T_i$ of such regression equations. Hence, the full conditional posterior distribution of α given $\{\alpha_i\}_{i=1}^I$, Σ, ρ and C^* is normal. The mean and variance result from the OLS estimator of α in (A.1) and (A.2), see Zellner (1971, Chapter VIII).

Sampling of α_i

To sample α_i for $i = 1, \ldots, I$ we can follow a similar approach as for α. We rewrite (2) and (3) as

$$\Sigma^{-\frac{1}{2}}(C_{it}^* - \rho C_{i,t-1}^* - (X_{it} - \rho X_{i,t-1})\alpha) = \Sigma^{-\frac{1}{2}}(X_{it} - \rho X_{i,t-1})\alpha_i + \Sigma^{-\frac{1}{2}}\varepsilon_{it} \text{ for } t = 1, \ldots, T_i$$

$$\sqrt{1-\rho^2}\Sigma^{-\frac{1}{2}}(C_{i1}^* - X_{i1}\alpha) = \sqrt{1-\rho^2}\Sigma^{-\frac{1}{2}}X_{i1}\alpha_i + \sqrt{1-\rho^2}\Sigma^{-\frac{1}{2}}e_{i1} \qquad (A.3)$$

$$\mathbf{0} = \Sigma_\alpha^{-\frac{1}{2}}\alpha_i + \Sigma_\alpha^{-\frac{1}{2}}v_i.$$

The last line follows from the fact that (6) 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, ρ and C^* is normal. The mean and variance result from the OLS estimator of α_i in (A.3).

Sampling of Σ_α

For the diagonal elements of Σ_α it holds that

$$p(\sigma^2_{\alpha,kk}|\cdot) \propto (\sigma^2_{\alpha,kk})^{-(I+1)} \exp\left(-\frac{1}{2\sigma^2_{\alpha,kk}}\sum_{i=1}^{I}\alpha^2_{ik}\right), \tag{A.4}$$

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

$$\frac{\sum_{i=1}^{I}\alpha^2_{ik}}{\sigma^2_{\alpha,kk}} \sim \chi^2(I) \text{ for } k = 1, \ldots, k_X. \tag{A.5}$$

Sampling of ρ

To sample ρ we use the Metropolis-Hastings sampler of Metropolis *et al.* (1953) and Hastings (1970). The Metropolis-Hastings sampler amounts to sampling a candidate ρ^{new} draw from a target distribution in a first step and accept or reject this candidate in a second step based on a draw from a uniform distribution. If the draw is rejected one continues with the previous draw ρ^{old}. Given the autoregressive structure of our model we can proceed in a similar way as Chib and Greenberg (1995) in their example.

To sample the candidate we rewrite (2) and (3) as

$$\Sigma^{-\frac{1}{2}}(C^*_{it} - X_{it}(\alpha + \alpha_i)) = \Sigma^{-\frac{1}{2}}(C^*_{i,t-1} - X_{i,t-1}(\alpha + \alpha_i))\rho + \Sigma^{-\frac{1}{2}}\varepsilon_{it}, \tag{A.6}$$

for $i = 1, \ldots, I$ and $t = 2, \ldots, T_i$. Using the same arguments as above we can sample ρ from a normal distribution with mean and variance following from the OLS estimator of ρ in (A.6). This is however not the proper full conditional posterior distribution as

we have neglected the first observations of each household. The joint density of the first observations of the households as function of ρ is proportional to

$$\pi(\rho) = \left| \frac{\Sigma}{1 - \rho^2} \right|^{-\frac{1}{2}I} \prod_{i=1}^{I} \exp\left(-\frac{1}{2}(1 - \rho^2)(C_{i1}^* - X_{i1}(\alpha + \alpha_i))'\Sigma^{-1}(C_{i1}^* - X_{i1}(\alpha + \alpha_i)) \right). \tag{A.7}$$

Following Chib and Greenberg (1995) the Metropolis-Hastings sampler amounts to

Step 1 Draw ρ^{new} from a normal distribution on the interval $(-1, 1)$ using the mean and variance resulting from the OLS estimator of ρ in (A.6).

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

Sampling of Σ

To sample Σ we note that

$$p(\Sigma|\cdot) \propto \pi(\Sigma) = |\Sigma|^{-\frac{1}{2}\sum_{i=1}^{I} T_i} \exp(-\frac{1}{2}\sum_{i=1}^{I}\sum_{t=1}^{T_i} \varepsilon_{it}'\Sigma^{-1}\varepsilon_{it}), \tag{A.8}$$

where

$$\varepsilon_{it} = (C_{it}^* - X_{it}(\alpha + \alpha_i)) - \rho(C_{i,t-1}^* - X_{i,t-1}(\alpha + \alpha_i)) \qquad \text{for } t = 2, \ldots, T_i,$$
$$\varepsilon_{i1} = \sqrt{1 - \rho^2}(C_{i1}^* - X_{i1}(\alpha + \alpha_i)) \tag{A.9}$$

for $i = 1, \ldots, I$.

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 $\pi(\Sigma^{\mathrm{new}})/\pi(\Sigma^{\mathrm{old}}) > u$ otherwise take $\Sigma^{\mathrm{new}} = \Sigma^{\mathrm{old}}$.

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 $\pi(\Sigma^{\mathrm{new}})/\pi(\Sigma^{\mathrm{old}}) > u$. 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 β

In the brand choice model, β is sampled in a similar way as α. We rewrite the equations (8) for which $c_{ijt} = 1$[2] as

$$\omega_j^{-1}(U_{ijt} - W'_{ijt}\beta_i) = \omega_j^{-1}W'_{ijt}\beta + \omega_j^{-1}\eta_{ijt}, \tag{A.10}$$

for $j = 1, \ldots, J$, $i = 1, \ldots, I$, and $t = 1, \ldots, T_i$. This represents $\sum_{i=1}^{I}\sum_{t=1}^{T_i}\sum_{j=1}^{J} c_{ijt}$ regression equations with regression coefficient β and uncorrelated normal distributed error terms with unit variance. Hence, the full conditional posterior distribution of β

---

[2]The value of c_{ijt} is determined by the value of C^*_{ijt}.

given $\{\beta_i\}_{i=1}^I$, Ω, C^* and U is normal. The mean and variance result from the OLS estimator of β in (A.10), see again Zellner (1971, Chapter VIII).

Sampling of β_i

To sample β_i for $i = 1, \ldots, I$ we can follow a similar approach as for β. We rewrite the equations (8) for which $c_{ijt} = 1$ as

$$\omega_j^{-1}(U_{ijt} - W'_{ijt}\beta) = \omega_j^{-1}W'_{ijt}\beta_i + \omega_j^{-1}\eta_{ijt}$$
$$\mathbf{0} = \Sigma_\beta^{-\frac{1}{2}}\beta_i + \Sigma_\beta^{-\frac{1}{2}}v_i, \tag{A.11}$$

for $j = 1, \ldots, J$ and $t = 1, \ldots, T_i$. The last line follows from the fact (10) implies that $v_i = (\beta_i - \mathbf{0}) \sim N(\mathbf{0}, \Sigma_\beta)$. This represents $k_W + \sum_{t=1}^{T_i}\sum_{j=1}^{J} c_{ijt}$ 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, C^* and U is normal. The mean and variance result from the OLS estimator of β_i in (A.11).

Sampling of Σ_β

For the diagonal elements of Σ_β it holds that

$$p(\sigma_{\beta,kk}^2|\cdot) \propto (\sigma_{\beta,kk}^2)^{-(I+1)} \exp\left(-\frac{1}{2\sigma_{\beta,kk}^2}\sum_{i=1}^{I}\beta_{ik}^2\right), \tag{A.12}$$

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

$$\frac{\sum_{i=1}^{I}\beta_{ik}^2}{\sigma_{\beta,kk}^2} \sim \chi^2(I) \text{ for } k = 1, \ldots, k_W. \tag{A.13}$$

Sampling of Ω

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

$$p(\omega_j|\cdot) \propto \frac{1}{\omega_j^{\nu+2}} \exp\left(-\frac{1}{2\omega_j^2}\sum_{i=1}^{I}\sum_{t=1}^{T_i} I[c_{ijt} = 1](U_{ijt} - W'_{ijt}(\beta + \beta_i))^2\right), \tag{A.14}$$

and hence

$$\frac{\sum_{i=1}^{I} \sum_{t=1}^{T_i} I[c_{ijt} = 1](U_{ijt} - W'_{ijt}(\beta + \beta_i))^2}{\omega_j^2} \sim \chi^2(\nu) \qquad \text{(A.15)}$$

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

Sampling of U

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

$$U_{it} = W_{it}(\beta + \beta_i) + \eta_{it}, \qquad \text{(A.16)}$$

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

$$U_{ijt}|\cdot \sim \begin{cases} \mathrm{N}(m_j, s_j^2) \times \mathbf{I}(-\infty, U_{i,d_{it},t}) & \text{if } j \neq d_{it} \wedge c_{ijt} = 1 \\ \mathrm{N}(m_j, s_j^2) \times \mathbf{I}(\max_{k|k \neq j}(U_{ikt}|c_{ikt} = 1), \infty) & \text{if } j = d_{it} \wedge c_{ijt} = 1 \\ \mathrm{N}(m_j, s_j^2) \times \mathbf{I}(-\infty, \infty) & \text{if } c_{ijt} = 0, \end{cases} \qquad \text{(A.17)}$$

for $j = 1, \ldots, J$, see Geweke (1991) for details.

Sampling of C^*

To sample C_{it}^* for $i = 1, \ldots, I$ we have to consider (2) for period t and $t + 1$. Rewriting these equations gives

$$-\Sigma^{-\frac{1}{2}}(\rho C_{i,t-1}^* + (X_{it} - \rho X_{i,t-1})(\alpha + \alpha_i)) = -\Sigma^{-\frac{1}{2}} C_{it}^* + \Sigma^{-\frac{1}{2}} \varepsilon_{ijt}$$
$$\Sigma^{-\frac{1}{2}}(C_{i,t+1}^* - (X_{i,t+1} - \rho X_{it})(\alpha + \alpha_i)) = \rho \Sigma^{-\frac{1}{2}} C_{it}^* + \Sigma^{-\frac{1}{2}} \varepsilon_{ij,t+1}, \qquad \text{(A.18)}$$

where for $t = 1$ the first equation has to be replaced by

$$-\sqrt{1 - \rho^2} \Sigma^{-\frac{1}{2}} X_{it}(\alpha + \alpha_i) = -\sqrt{1 - \rho^2} \Sigma^{-\frac{1}{2}} C_{it}^* + \sqrt{1 - \rho^2} \Sigma^{-\frac{1}{2}} \varepsilon_{ijt}. \qquad \text{(A.19)}$$

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.18) and (A.19). 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

$$
C_{ijt}^*|\cdot \sim \begin{cases} \mathrm{N}(m_j, s_j^2) \times \mathbf{I}(0, \infty) & \text{if } \sum_{k=1, k \neq j}^{J} c_{ikt} > 0 \\ \mathrm{N}(m_j, s_j^2) \times \mathbf{I}(\max_{k|k \neq j}(C_{ikt}^*), \infty) & \text{if } \sum_{k=1, k \neq j}^{J} c_{ikt} = 0. \end{cases} \tag{A.20}
$$

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

$$
C_{ijt}^*|\cdot \sim \begin{cases} \mathrm{N}(m_j, s_j^2) \times \mathbf{I}(-\infty, 0) & \text{if } \sum_{k=1, k \neq j}^{J} c_{ikt} > 1 \\ \mathrm{N}(m_j, s_j^2) \times \mathbf{I}(-\infty, \min(0, C_{i,d_{it},t}^*)) & \text{if } \sum_{j=1, k \neq j}^{J} c_{ijt} = 1. \end{cases} \tag{A.21}
$$

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

$$
C_{ijt}^*|\cdot \sim \mathrm{N}(m_j, s_j^2) \times \mathbf{I}(-\infty, C_{i,d_{it},t}^*). \tag{A.22}
$$

4. In all other cases we use the following approach. The value of the full conditional posterior density of C_{ijt}^* changes if a brand enters the consideration set. The full conditional posterior density of C_{ijt}^* is given by

$$
f(C_{ijt}^*|\cdot) = \frac{1}{\kappa}(I[C_{ijt}^* < 0] + \frac{1}{\omega_j}\phi((U_{ijt} - X_{ijt}(\beta + \beta_i))/\omega_j)I[C_{ijt}^* > 0])\frac{1}{s_j}\phi((C_{ijt}^* - m_j)/s_j), \tag{A.23}
$$

which reflects that for $C_{ijt}^* > 0$ brand j enters the consideration set and the term $g(U_{ijt}|\cdot) = \frac{1}{\omega_j}\phi((U_{ijt} - X_{ijt}(\beta + \beta_i))/\omega_j)$ is added to the likelihood function. As the

integral of (A.23) over C^*_{ijt} has to be 1, the value of κ is given by

$$\kappa = \Phi(-m_j/s_j) + \frac{1}{\omega_j}\phi((U_{ijt} - X_{ijt}(\beta + \beta_i))/\omega_j)(1 - \Phi(-m_j/s_j)). \qquad \text{(A.24)}$$

To draw C^*_{ijt} we use the inverse method. The conditional cumulative distribution function of C^*_{ijt} is given by

$$F(C^*_{ijt}|\cdot) = \begin{cases} \frac{1}{\kappa}\Phi((C^*_{ijt} - m_j)/s_j) & \text{if } C^*_{ijt} < 0 \\ \frac{1}{\kappa}\Phi(-m_j/s_j) + \frac{1}{\kappa}g(U_{ijt}|\cdot)(\Phi((C^*_{ijt} - m_j)/s_j) - \Phi(-m_j/s_j)) & \text{if } C^*_{ijt} > 0, \end{cases}$$
$$\text{(A.25)}$$

which results in the following sampling scheme

Step 1 Draw u from a uniform distribution on the region $(0, 1)$,

Step 2 Set $C^*_{ijt} = s_j\Phi^{-1}(x) + m_j$, where

$$x = \begin{cases} \kappa u & u < \frac{1}{\kappa}\Phi(-m_j/s_j) \\ (\kappa u + (g(U_{ijt}|\cdot) - 1)\Phi(-m_j/s_j))/g(U_{ijt}|\cdot) & \text{otherwise,} \end{cases}$$

where Φ^{-1} is the inverse CDF of a standard normal distribution and κ is given in (A.24).

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. J. Lenk (1994), Modeling Household Purchase Behavior with Logistic Normal Regression, *Journal of the American Statistical Association*, **89**, 1218–1231.

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 (1995), Understanding the Metropolis-Hastings Algorithm, *American Statistician*, **49**, 327–335.

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.

Hastings, W. K. (1970), Monte Carlo Sampling using Markov Chains and their Applications, *Biometrika*, **57**, 97–10.

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.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953), Equations of State Calculations by Fast Computing Machines, *Journal of Chemical Physics*, **21**, 1087–1092.

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. (2005), An Integrated Choice Model Incorporating Alternative Mechanisms for Consumers Reactions to In-Store Display and Feature Advertising, *Marketing Science, forthcoming*.