

Retrieving Unobserved Consideration Sets from Household Panel Data

Erjen van Nierop*

Graduate School of Industrial Administration, Carnegie Mellon University

Richard Paap

Econometric Institute, Erasmus University Rotterdam

Bart Bronnenberg

The Anderson School of Management at UCLA

Philip Hans Franses

Econometric Institute and Department of Marketing and Organization
Erasmus University Rotterdam

Michel Wedel

University of Michigan Business School

September 16, 2003

*Address for correspondence: Carnegie Mellon University, Graduate School of Industrial Administration, Room 257A, 5000 Forbes Avenue Pittsburgh, PA 15213, email: nierop@andrew.cmu.edu. The first author thanks the Foundation 'Vereniging Trustfonds Erasmus Universiteit Rotterdam' in the Netherlands. The second author thanks the Netherlands Organization for Scientific Research (NWO) for its financial support.

We gratefully acknowledge the comments of participants of the Marketing Science 2001 conference and participants at 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 Brian Ratchford, the area editor and three anonymous referees for their many helpful suggestions.

Retrieving Unobserved Consideration Sets from Household Panel Data

Abstract

We propose a new model to describe consideration, consisting of a parsimonious multi-variate probit model for consideration and a multinomial probit model 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 general structures of unobserved dependence in consideration among brands. In addition, our model accommodates effects of the marketing mix on consideration, an error process that is correlated across time, and unobserved consumer heterogeneity in the choice process.

Unique to this study is that we attempt to establish the validity of existing practice to infer consideration sets from observed choices in panel data. To this end, we use data collected 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, with greater accuracy than with competing models.

Next, we estimate the model on IRI panel data. We have three main results. First, compared with the single-stage multinomial probit model, promotion effects are larger and are inferred with smaller variances when they are included in the consideration stage of the two-stage model. Second, as before, we find that consideration of brands does not covary greatly across brands once we account for observed effects. Third, we show that our model is able to analyze datasets with a larger number of brands than many other consideration set models have been able to.

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 Chattaopadhyay, 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. A frequently used approach is the random utility theory framework (see for example, McFadden, 1973 or Guadagni and Little, 1983), where one builds upon the postulate of utility maximizing consumers. 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 approaches this problem by directly assessing stated consideration set membership of individual brands. Hence, these studies model the marginal distribution of consideration for each brand (Roberts and Lattin, 1991, for example) and then transform the consideration-set inclusion probabilities into consideration sets. This usual conduit is based on an assumption of independence (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 larger global choice sets, it has limitations in being based on the assumption of set-membership independence across brands. Furthermore, in research to date, these stated consideration sets are assessed in cross sections, implicitly assuming them to be stable over time.

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) by conditioning choice on unobserved consideration. 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. Another approach is proposed by Gilbride and Allenby (2003), who model dependence in consideration set membership explicitly through similarity in product attributes. A number of 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. Third, to achieve model identification, it is often necessary to assume static consideration sets for a given household. This appears to be contrary to what consumer learning theory predicts, but this restrictive assumption is shared by both streams of consideration set research.

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 important 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 common modeling framework is needed that may accommodate stated consideration data, revealed consideration data (choice) or both. This is one of the intended core contributions of the study in this paper: to develop such a model.

We propose a model for consideration set formation and brand choice that provides a unifying framework of the stated and revealed approaches to consideration set identification. It combines their strengths and can either be estimated on revealed choice data alone or on stated consideration and choice data combined. 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 (MVP, Edwards and Allenby, 2003) 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.

This approach offers an alternative to models commonly used in the revealed approach, when estimated on actual choice data alone, for the following reasons. First, through the covariance structure of the MVP model, our approach is not based on the assumption of independence of consideration across brands and thus retains the advantage of context-dependence that is inherent in the revealed approach to consideration sets. Second, it does not suffer from the curse of dimensionality. In the worst case (that is, when we use a completely structure-free covariance matrix across brands), the number of parameters

to be estimated is quadratic in the number of brands rather than exponential. More realistically, there are many cases in which theoretical support exists for a parsimonious structure of the cross-brand consideration process. When such a structure is independent of the number of brands in the global choice set, our approach provides a fully tractable and general representation of consideration set formation, the complexity of which is only *linear* in the global number of choice options. Third, we can include marketing control variables and “the hand of the past” in the consideration stage. We develop our model primarily for the purpose of obtaining better substantive insights into choice processes, and the effect of marketing variables on it. We do not expect our model to provide improved holdout forecasts (see also Andrews and Srinivasan (1995)). The reason for that is that the consideration sets identified explain part of the variability of choices between subjects. That variability is equally well captured by a general purpose model such as the MNP with unobserved heterogeneity. Therefore, we do not primarily aim at improving predictive validity, but at providing deeper insight into consumers’ choice behavior.

Moreover, when not only choice data, but also stated consideration indicators are available, these can be included for the estimation of our model as well, without requiring any change in the model structure. Importantly, this allows us to investigate the validity of the consideration set probabilities assessed from choice data alone, by estimating the model with and without the stated consideration indicators. 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. To our knowledge this is the first attempt to do so in the consideration set literature.

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 both in a numerical and a graphical way. 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 ($i = 1, \dots, I$) choosing brand j ($j = 1, \dots, J$) at purchase occasion t ($t = 1, \dots, T_i$). 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

$$C_{it} = \begin{pmatrix} C_{i1t} \\ \vdots \\ C_{ijt} \end{pmatrix}, \quad (1)$$

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

$$C_{ijt}^* = X'_{ijt}\alpha + e_{ijt}, \quad j = 1, \dots, J, \quad (2)$$

where X_{ijt} is a vector containing brand and purchase-related explanatory variables including brand-specific intercepts, where α is a parameter vector, 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,

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

where ε_{ijt} is an unobserved disturbance term. This dynamic process intends to model persistency 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 $C_{ijt}^* > 0$. 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

$$\begin{aligned} \Pr[C_{it} = (1, 1, 0, \dots, 0)'] &= \Pr[C_{i1t}^* > 0, C_{i2t}^* > 0, C_{i3t}^* \leq 0, \dots, C_{iJt}^* \leq 0] \\ &= \Pr[e_{i1t} > -X'_{i1t}\alpha, e_{i2t} > -X'_{i2t}\alpha, e_{i3t} \leq -X'_{i3t}\alpha, \dots, e_{iJt} \leq -X'_{iJt}\alpha]. \end{aligned} \quad (4)$$

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

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

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 $e_{i1} \sim N(\mathbf{0}, \Sigma/(1 - \rho^2))$.

The multivariate probit model allows for the possibility of an empty consideration set, that is $c_{it} = (0, \dots, 0)'$. 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. Given a choice, the probability that the consideration set of a household i includes only the first two brands is then equal to probability (4) divided by 1 minus the probability of the occurrence of an empty set.

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,

$$U_{ijt} = W'_{ijt}(\beta + \beta_i) + \eta_{ijt}, \quad j = 1, \dots, J \quad (6)$$

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

$$\eta_{it} \sim N(\mathbf{0}, \Omega). \quad (7)$$

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

$$U_{ijt} = \max(U_{ikt} \text{ for all } k | c_{ikt} = 1). \quad (8)$$

Hence, the probability that household i chooses brand j at purchase occasion t given the

consideration set c_{it} equals

$$\begin{aligned}
\Pr[D_{it} = j|C_{it}] &= \Pr[U_{ijt} > U_{ikt} \text{ for all } k \neq j | c_{ijt} = c_{ikt} = 1] \\
&= \Pr[U_{ijt} - U_{ikt} > 0 \text{ for all } k \neq j | c_{ijt} = c_{ikt} = 1] \\
&= \Pr[\eta_{ikt} - \eta_{ijt} < W'_{ijt}(\beta + \beta_i) - W'_{ikt}(\beta + \beta_i) \text{ for all } k \neq j | c_{ijt} = c_{ikt} = 1].
\end{aligned} \tag{9}$$

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 was corroborated in several (unreported) test runs of our model 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 $\Omega = \text{diag}(\omega_1^2, \dots, \omega_{J-1}^2, 1)$. 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).

In general it is not possible to estimate fixed household-specific parameter vectors β_i . If a household never buys a certain brand, the household-specific base preference of this brand is not identified, see Allenby and Rossi (1991) for a discussion. Therefore, we assume that the household-specific parameters are draws from a population distribution, that is,

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

where Σ_β is a positive definite symmetric matrix. An advantage of this approach is that it leads to a non-diagonal covariance structure in the multinomial probit models, that is, the variance-covariance structure of $\eta_{it} + \beta_i$

$$W_{it}\Sigma_\beta W'_{it} + \Omega, \tag{11}$$

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

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 important contribution is that we include explanatory variables in the consideration stage of the model and that we allow for dynamics. The study in this paper is, as far as we are aware of, the first to address dynamics in the consideration set formation.

3 Estimation

3.1 Likelihood function

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 $d = \{d_{it}, i = 1, \dots, I, t = 1, \dots, T_i\}$, that is,

$$\mathcal{L}(d|\theta) = \prod_{i=1}^I \int_{\beta_i} \left[\sum_{\forall c_{i1} \in \mathcal{C}_{i1}} \sum_{\forall c_{i2} \in \mathcal{C}_{i2}} \cdots \sum_{\forall c_{iT_i} \in \mathcal{C}_{iT_i}} \Pr[C_{i1} = c_{i1}, C_{i2} = c_{i2}, \dots, C_{iT_i} = c_{iT_i} | \alpha, \rho, \Sigma] \right. \\ \left. \prod_{t=1}^{T_i} \Pr[D_{it} = d_{it} | c_{it}; \beta, \beta_i, \Sigma_\beta, \Omega] \right] \phi(\beta_i; 0, \Sigma_\beta) d\beta_i, \quad (12)$$

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

sets of household i are equal to $c_i = (c_{i1}, \dots, c_{iT_i})$, see (4), and the product of the brand choice probabilities given c_{it} , see (9), 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 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.

3.2 MCMC Approach

The likelihood function (12) is too complicated to optimize numerically over the parameter space, as the evaluation already requires the computation of many multivariate integrals. To estimate the model parameters θ we opt for a Bayesian approach, where Bayesian posterior means and posterior standard deviations are used as parameter estimates and standard errors. We assume flat priors for the model parameters such that the posterior distribution is proportional to the likelihood function (12)¹. To obtain posterior results, we use the Gibbs sampling technique of Geman and Geman (1984) with data augmentation, see Tanner and Wong (1987). The idea of Gibbs sampling is to sample iteratively from

¹In the applications described hereafter, we have experimented with different variations on the prior values, but found no difference in the resulting parameter estimates, due to the large amounts of data and the flatness of the priors. Prior means were in the order of 10^{-4} , and variances in the order 10^4

the full conditional posterior distributions of the model parameters contained in θ . This creates a Markov chain that converges under mild conditions, such that the draws can be used as draws from the joint distribution (see for example Tierney, 1994, or Casella and George, 1992 for a lucid introduction). The unobserved utilities U_{ijt} and C_{ijt}^* , and the unobserved household parameters β_i are sampled alongside with the other model parameters. The posterior means and standard deviations of the parameters of interest can be obtained by computing the sample means and variances of the draws.

The Gibbs sampling simulation algorithm to sample from the joint distribution of $(\theta, \beta_i, U, C^*)$ proceeds as follows:

Step 1 Specify starting values $(\theta^{(0)}, \{\beta_i^{(0)}\}_{i=1}^I, U^{(0)}, C^{*(0)})$ and set $g = 0$. We initialize the parameter vectors in θ at reasonable random values. We have used different random starting values and found no difference in the resulting posterior means and standard deviations.

Step 2 Draw

- $\alpha^{(g+1)} | \text{others}$
- $\rho^{(g+1)} | \text{others}$
- $\beta^{(g+1)} | \text{others}$
- $\beta_i^{(g+1)}$ for $i = 1, \dots, I$ | others
- $\Sigma_{\beta}^{(g+1)} | \text{others}$
- $\Sigma^{(g+1)} | \text{others}$
- $\Omega^{(g+1)} | \text{others}$

- $U_{it}^{(g+1)} | \text{others}$
- $C_{it}^{*(g+1)} | \text{others.}$

Step 3 Set $g = g + 1$ and go to step 2.

The described iterative scheme generates a Markov Chain. After the chain has converged, say, at G iterations (which is called the number of burn in iterations), the simulated values for $g > G$ can be used as a sample from the joint distribution of $(\theta, \beta_i, U, C^*)$ to compute posterior means, variances and marginal densities.

The derivation of the full posterior distributions of the model parameters θ , β_i , U and C^* proceeds in a similar way as in Albert and Chib (1993), McCulloch and Rossi (1994), Geweke et al. (1997), Chib and Greenberg (1998) and Paap and Franses (2000). To determine the sampling distributions of the mean (α , β_i and β) and covariance parameters (Σ_β and Ω), we rewrite the MVP and MNP models in such a way that they represent standard univariate or multivariate regression models with the parameter to be sampled acting as a regression parameter or (co-)variance parameter of the error term. For a standard regression model, we know that the full conditional posterior distribution of the regression parameter is normal with mean and variance resulting from the ordinary least squares (OLS) estimators. The full conditional posterior distribution of the variance (covariance matrix) of the error terms is an inverted χ^2 (or inverted Wishart) distribution. The sampling algorithms for ρ and Σ are both based on a Metropolis-Hastings sampler. Finally, the elements of U are sampled from truncated normal distributions, while for the elements of C^* we use the inverse CDF technique. In Appendix A we provide the appropriate full posterior distributions and sampling algorithms for the model parameters and utilities.

The model is implemented in the matrix programming language Gauss. For the estimation of the parameters of each model considered in this paper, we generate 4000 iterations of the Gibbs sampler for burn in and 8000 iterations for analysis, where we retain every eighth draw. The (unreported) iteration plots are inspected to see whether the sampler converges to stationary draws from the posterior distributions of the model parameters. Another indication of convergence is that the same posterior distributions resulted from using different sets of starting values.

3.3 Interpretation and inference

Running the Gibbs sampling scheme a large number of times results in a sample from the posterior distribution of the model parameters. All posterior inferences are based on the sample furnished by the Markov chain procedure. The analysis yields results such as posterior probabilities for each brand whether it is present in the consideration set of a household or not, influence of marketing variables on consideration and purchase probabilities and influence of marketing variables on conditional purchase probabilities, that is, purchase given consideration. Furthermore, we can use posterior results to compute posterior brand choice probabilities to predict out-of-sample brand choices of households in the data set.

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 con-

vergent 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 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 considera-

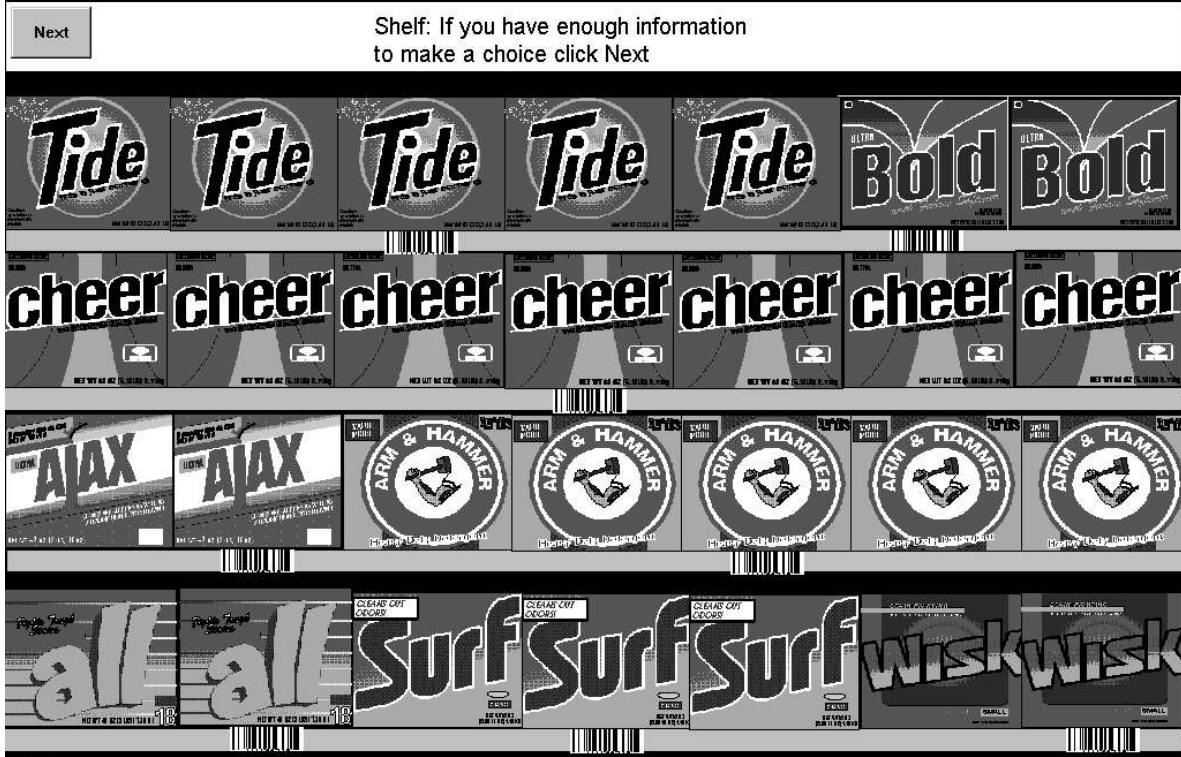


Figure 1: Screen-shot from sixth choice occasion.

tion is taken from Lehmann and Pan (1994) and Nedungadi (1990).

The experiment was administered to 65 undergraduate 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 rarely chosen, these were dropped from the analysis. This left us with $N = 528$ observations. Of these data, we randomly sampled 48 households with $N = 369$ purchases for estimation. The other 7 households are used for cross section forecasting. 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

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.

constructed by dichotomizing the average of the two questions (divided by 100) around 0.5 for each choice occasion and each individual. 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). 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 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 2, we allow for household heterogeneity in the brand choice stage.

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 estimate three models on the data from the choice experiment with stated consideration sets. First, we estimate the full multivariate probit/multinomial probit (MVP+MNP) model of choice and consideration. To recall, 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_β , of the random choice effects, β_i . For identification, we set the covariance matrix of the choice utilities, Ω to the identity matrix. This model is estimated on *choice data alone*. Second, we estimate the MVP model by itself using the reported consideration sets. Third, we estimate a multinomial probit (MNP) model with an autoregressive error process and heterogeneity in the effects on choice. For lack of variation across time, we need to drop the price variable from these analyses. When estimating the parameters, the price parameter was difficult to separate 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

Table 2: Posterior estimation results for the experimental detergent data.

Model		MVP+MNP		MVP		MNP	
Estimated on		choice dummies		consideration dummies		choice dummies	
		mean	std. dev.	mean	std. dev.	mean	std. dev.
Consideration stage	α_{All}	-1.68	0.50	-1.71	0.43		
	α_{AH}	-1.30	0.46	-1.45	0.41		
	α_{Bold}	-2.03	0.51	-2.00	0.51		
	α_{Cheer}	-0.40	0.72	-0.56	0.67		
	α_{Surf}	-1.36	0.48	-1.39	0.47		
	α_{Tide}	0.08	0.55	0.13	0.52		
	display	2.63	0.36	2.33	0.29		
	shelf	-0.20	0.62	0.02	0.61		
ρ		0.69	0.04	0.69	0.04		
Choice stage	β_{All}	0.20	0.70			-2.93	0.58
	β_{AH}	-0.37	0.57			-3.65	0.81
	β_{Bold}	-0.57	0.70			-4.82	0.77
	β_{Cheer}	-0.64	0.41			-2.65	0.64
	β_{Surf}	-0.54	0.73			-3.13	0.62
	β_{Tide}	0 ^a	—			0 ^a	—
	display					5.64	0.44
	shelf					0.00	0.01
ρ						0.95	0.01

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

note the differences in the estimates of the brand intercepts across the MNP and 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 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.

Table 2 shows that the proposed (MVP+MNP) model (estimated on choice data) and the MVP model (estimated on consideration data) reveal that consideration is strongly determined by display. Past consideration has strong effects as well, which is revealed by the value of 0.69 for ρ . We observe that the combination of in-store activity such as display, and past consideration captures a large part of the variation in consideration sets across individuals and purchase occasions.

The covariance terms in the MVP model and the MVP component of the MVP+MNP model are close to 0 and all posterior intervals cover the zero value. Therefore, it seems that after taking into account in-store variables and past consideration, little covariation among consideration of brands is left. Importantly, it 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 conclusion is derived from the stated consideration set data, as well as the consideration process derived from the choice data.

By comparing the consideration stage estimates in Table 2 from the MVP+MNP model with those from the MVP model the first result is that the consideration stage is estimated extremely well from the choice data alone. Equally interesting is that the loss of information about consideration sets going from stated consideration sets to choice data alone only results in a modest increase in the standard errors of the estimates.

Using the full model, we can infer the consideration sets from which the subjects made their final choices. We call these sets the “inferred consideration sets”. The self-reported measures of consideration are called “reported consideration sets”. Note that both reported and inferred consideration sets comprise of numbers in-between 0 and 1, that vary across brands and subjects. In order to further establish the validity of inferring consideration sets from choice data, we compute for each brand, individual and choice occasion, the inferred set-membership and its correlation with reported set membership. We find that inferred and reported set membership correlate very highly for each brand. Specifically, for the six brands these correlations are in the range of 0.57 to 0.82 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 3 shows a cross-tabulation consideration set memberships, rounded to the nearest integer. A zero indicates the brand is not in the consideration set, whereas a one indicates the opposite. The miss-classification is not symmetric, that is, if the brand is not in the consideration set according to the individual, our estimate is usually correct (in 97% of the cases). On the other hand, if the individual stated that the brand is contained in his or her consideration set, we estimate this correctly in 56% of the cases. In total, we are correct in 84% of the purchase occasions.

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

		stated^a		
		out (0)	in (1)	total
estimated^a	out(0)	1449	320	1769
	in (1)	45	400	445
total		1494	720	2214 ^b

^a For this cross-tabulation, both the inferred consideration set memberships and the reported consideration sets are rounded to 0 or 1, whichever is nearest.

^b The data set contains 369 purchase occasions, with six brands each. The product of these figures equals 2214.

We take the above findings as strong supportive evidence for the validity of inferring consideration sets from choice data with our model. Thus, our results support the contention that this operationalization of consideration, identified from choices only, is capable of tracking the differences in choice sets both across time as well as across individuals.

5 Application to Scanner Panel 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 3324 purchases of coffee made by 232 households during about 2 years. The brands and their respective number of purchase and marketing instrument statistics are given in Table 4.

Table 4: Descriptive statistics for the coffee data set ($N = 3324$)

Brand	Number of purchases			Marketing instruments		
	in sample	long	cross	display	feature	price/oz
Ground						
Eight O’Clock	94	12	14	20.4%	6.9%	\$ 0.166
Folgers	378	28	53	5.3%	4.8%	\$ 0.168
Hills Brothers	819	58	121	17.7%	12.7%	\$ 0.152
Maxwell House	486	26	39	13.8%	13.1%	\$ 0.188
MJB	151	17	44	8.4%	5.1%	\$ 0.165
Papanicholas Sig	225	12	5	0.2%	1.1%	\$ 0.275
Soluble						
Folgers	186	18	18	0.0%	0.6%	\$ 0.518
General Foods	189	11	29	2.6%	2.1%	\$ 0.461
Maxwell House	244	24	23	1.3%	5.3%	\$ 0.471
Total	2772	206	346			

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.

5.1 Estimation results from the empirical data

We estimated the following three models on the coffee data, using the same operationalizations as described for the experimental data above. The full two-stage model is estimated. 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_β , of the random choice effects, β_i , and the covariance matrix of the choice utilities, Ω . Then we estimate a single-stage MNP choice model with again similar specifications (autoregressive error process and heterogeneity). Finally, to benchmark our model to we estimate the model of Bronnenberg and Vanhonacker (1996).

The posterior estimation results for the coffee data are given in Table 5. From the results of the proposed (MVP+MNP) model we see that all marketing parameters are estimated to be far away from zero (when compared to the posterior deviation) and that they are all of the expected sign. Consistent with the controlled choice experiment, the covariance terms in the MVP model are close to 0 and all posterior intervals cover the zero value. We have calibrated our model on three other data sets (cracker data with 4 brands, softdrinks with 10 brands and yoghurt, also with 10 brands), and have found little or no covariation 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. Note that the effects of in-store merchandizing on consideration are more tightly distributed in the two-stage model than in the single-stage model. This can be seen when we compare the lower posterior standard deviations in the two-stage model, compared to its posterior means.

The brand intercepts for the MNP-component of the full model show that Eight

Table 5: Posterior estimation results for the coffee data.

Stage	Parameter	MVP+MNP		MNP	
		Mean	St. dev.	Mean	St. Dev.
Consideration stage	α_{Eight}	-4.77	0.91		
	α_{Folgers}	-1.55	0.95		
	α_{Hills}	-0.66	0.70		
	α_{Maxwell}	-1.01	0.66		
	α_{MJB}	-1.80	0.71		
	$\alpha_{\text{NAPapanicholas}}$	-1.72	0.65		
	α_{Folgers}	-1.31	0.67		
	α_{General}	-1.67	0.70		
	α_{Maxwell}	-2.18	0.77		
	α_{Display}	0.96	0.14		
	α_{Feature}	0.99	0.11		
	ρ	0.87	0.01		
Choice stage	β_{Eight}	0.34	1.01	-1.66	0.34
	β_{Folgers}	-2.05	1.16	0.60	0.27
	β_{Hills}	-2.35	0.62	1.04	0.27
	β_{Maxwell}	-2.22	0.68	0.99	0.20
	β_{MJB}	-2.65	0.80	-0.73	0.31
	$\beta_{\text{NAPapanicholas}}$	-2.27	0.91	-2.83	0.39
	β_{Folgers}	-0.73	1.01	-1.28	0.77
	β_{General}	-1.14	0.97	-2.86	0.47
	β_{Price}	-5.73	1.51	-1.86	0.55
	β_{Display}			1.03	0.20
	β_{Feature}			1.05	0.17
	ρ			0.22	0.06

^a For identification purposes, we need to select a base brand and set its constant equal to 0. Without loss of generality, Maxwell House (Soluble) is chosen as base brand.

^b The covariances in the MVP model are close to 0 and are not shown here.

^c The single-stage MNP model has an identity covariance matrix to ease the comparison with the MVP+MNP model results.

O’Clock is likely to have the highest choice share given consideration. However, the MVP brand intercepts reveal that this brand has a rather low base probability of being considered, irrespective of marketing activity. This brand could therefore be considered as a niche brand. We will confirm this when looking at purchase and revealed consideration share in the next section.

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, we 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, 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 doesn’t contain promotion effects at all, and the correlated error structure still appears in a different place. To ease comparison, we therefore used (unreported) simulation to confirm that the same price drop 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 5. The two-stage model tells us that consumers are indeed more price sensitive². The stability of the display and feature coefficients

²To check whether we would find a similarly low price coefficient in the single stage MNP model for

across the proposed and MNP models and their tighter distribution when included in the consideration set component of the proposed model may provide an indication that these two marketing control variables do primarily serve to induce consumers to consider the brands in question.

Although, the model is not purposely built to make forecasts, out-of-sample predictions show that the hit rate of the full model is 66% for hold out samples. The single-stage MNP model produces the same hit rates as we expected. The in-sample hit rate for our model equals 87%, whereas for the MNP it is lower with 77%³. When we apply the model used by Bronnenberg and Vanhonor (1996) to our data set, we obtain a longitudinal hit rate of (only) 61%. The in-sample hit rate for this model equals 64%, which is lower than for our model.

Although we see a clear difference in the forecasting hit rate in-sample, the posterior distributions of longitudinal forecast hit rate for the MNP and MVP + MNP models overlap almost completely, showing that there is no difference in longitudinal prediction between the two models. Of course one would have liked to see the added complexity of our model to result in improved predictive performance, but as has been found previously, a simpler but theoretically less completely specified model as the MNP predicts equally 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

other data sets, we also estimated the model for a yoghurt data sets (10 brands) and a softdrink data sets (10 brands). With an identity covariance matrix, estimates were similar, with price coefficients of -1.776 (posterior standard deviation 0.122) and -3.017 (posterior standard deviation 0.126) respectively.

³The 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%

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. In the next section we look at some of these added insights in more detail.

6 A Detailed Look at Consideration

In this section we describe several results that our two-stage model yields for consideration. First, at each iteration of the Gibbs sampler, we keep track of the consideration set size for each purchase occasion. The posterior average yields an average consideration set size for each purchase occasion. A histogram of these sizes is given in Figure 2. The left panel shows the distribution for all purchase occasions and the right panel provides the distribution for the household averages. These graphs show that consumers choose among 1-3 different brands of coffee at point-of-purchase, with a mode consideration set size of about 1.6. The average set size for household ranges from 1 to 2.5, with a mode of around 1.7. The graph reveals that the distribution of the average household consideration set is somewhat more symmetric and tighter than that of the consideration sets at each individual purchase occasion. That is expected, the right panel being based on the average over time. The consideration sets are fairly stable across purchase occasions, which is corroborated by the relatively high intertemporal correlation of consideration utility (0.87).

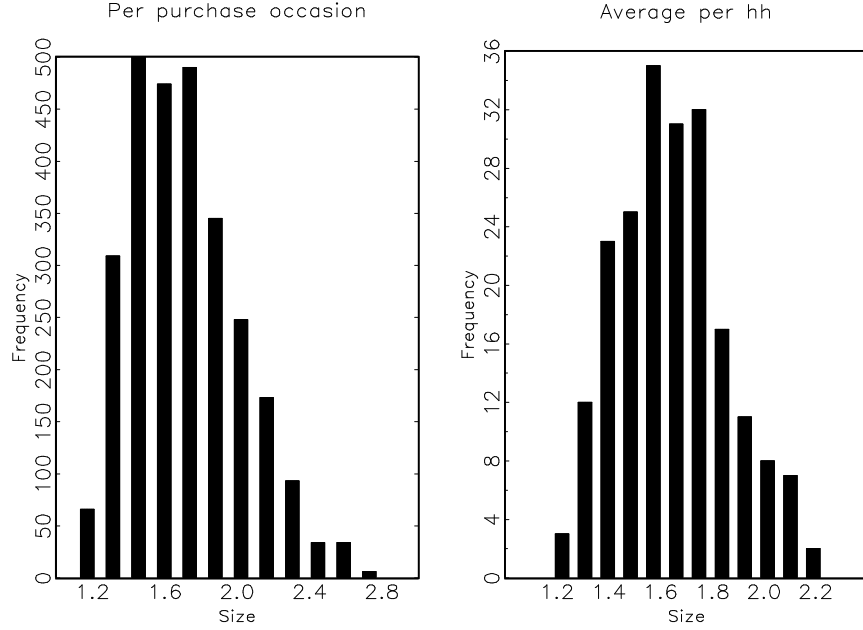


Figure 2: Histogram of consideration set size. Left panel: all purchase occasions. Right panel: household averages.

Table 6 gives additional insight in the consideration sets. The left part of the table displays purchase shares and consideration shares. Purchase shares are computed from the purchase dummies in the data sets and the consideration shares are computed by summing the posterior consideration probabilities that result from the Gibbs sampler. The fourth column of Table 6 shows the purchase share conditional on consideration, computed by dividing the purchase share by the consideration share. Brands that have a ratio close to 1 have a high probability of being purchased, once considered. The smaller brands among these may be considered niche brands. A recommendation for these brands would be to work more on entering consumers' consideration sets, although for large consideration share increases, we can't be sure that the purchase share will increase proportionally. Conversely, brands that are far below 1, are purchased not as

Table 6: Purchase and consideration shares. The left part of the table displays share of purchase (observed) and consideration (estimated). The right part of the table displays consideration share in the situation where the brand has no promotion versus the situation where the brand has a display only or feature only.

	Shares			Conditional consideration shares		
	Purchase Share %	Consid. Share %	P/C Ratio %	No Promotion %	Display Only %	Feature Only %
Ground						
Eight O Clock	3	3	98	3	7	7
Folgers	14	27	50	17	27	32
Hills Brothers	30	41	72	32	44	46
Maxwell House	18	30	58	23	32	38
MJB	5	16	35	12	22	23
Papanicholas Sig	8	14	59	13	11	24
Soluble						
Folgers	7	16	41	13	NA	17
General Foods Intl	7	13	52	10	11	17
Maxwell House	9	16	54	15	20	22

much when they are considered. This is in particular true for MJB Ground and Folgers Soluble. These brands would benefit from an increase in the consideration to purchase conversion, which may be achieved by using price promotions to convince consumers to buy the product.

To obtain further insights, Table 6 also shows the effects of display and feature on consideration. Virtually each brand is considered more on the purchase occasions where it is promoted by either a display or feature. Looking at the display and feature effects of MJB, it seems this brand is very good at entering consumer's mind-sets. Combined with the relatively low resulting purchase share, we see that the brand appears to fare poorly in the second stage of their purchase process. This doesn't seem to be due to its

price, which is among the lowest in the category, but rather due to a low intrinsic brand preference. This is confirmed by the low brand intercept estimate in the brand choice stage of -2.65 (see Table 5). As can be seen from these results, a good balance between entering consumers' consideration set through display and feature, and the conversion into a purchase in the second stage of the choice process, are necessary ingredients for achieving higher market shares.

Next, we are interested in the relationship between the average consideration set size per household and the number of different brands bought by the household. One would expect that variety seekers, who buy many different brands, have on average a larger consideration set. The left panel of Figure 3 displays this relationship. The graph shows that variety seekers have larger consideration sets⁴. Although this is not a surprising finding, it shows that our model produces intuitive results. The right panel of Figure 3 displays the relationship between the consideration set size per household and the number of purchase occasions for the household. One may expect that consumers who buy more frequently, have a larger consideration set; on the other hand, they may be more loyal and more informed about the items, resulting in a smaller consideration set. The conclusion from the graph is that consumers who buy more, have larger consideration sets⁵. Apparently, these consumers are better aware of all the brands and still consider many. A reason for this could be that they have a large family with different coffee tastes that need to be accommodated, or drink different tastes or types of coffee in the morning versus the afternoon or evening.

⁴If one would draw a (linear) regression line through these points, the parameters would be significant.

⁵Again a (linear) regression line has significant parameters.

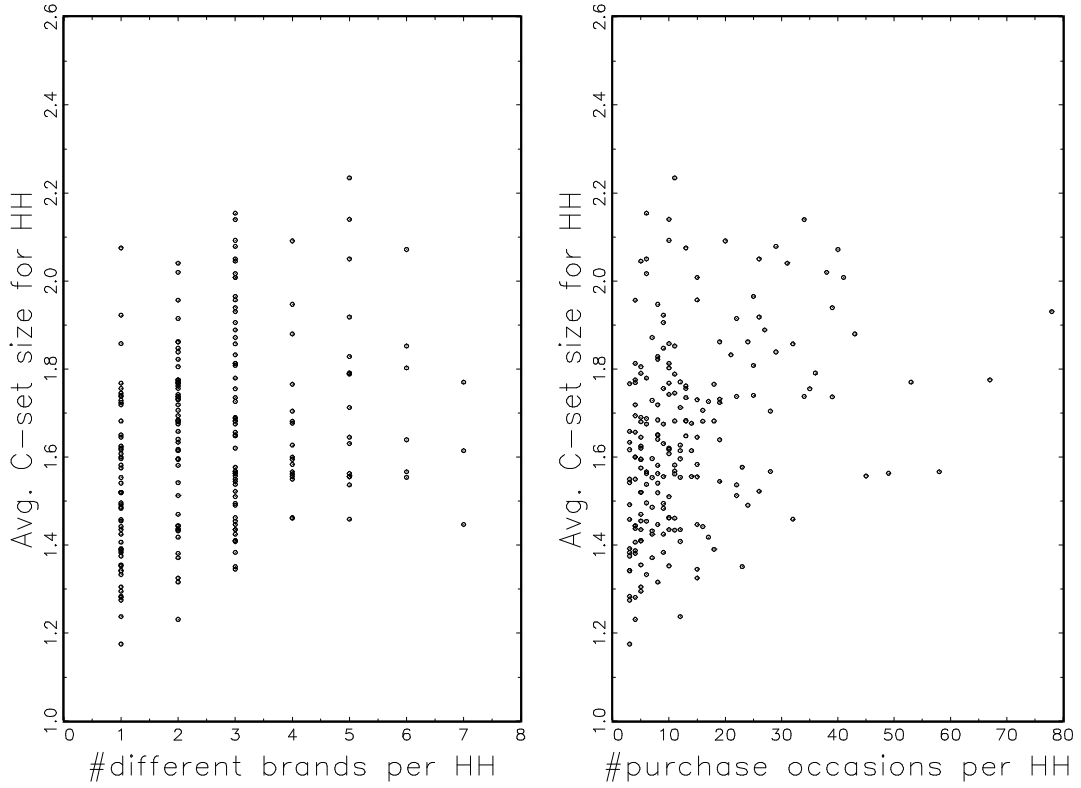


Figure 3: Left panel: relationship between average consideration set size per household and number of different brands bought by household. Right panel: relationship between average consideration set size and number of purchases by household.

7 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 a sample of respondents to state their considered set of brands, or by inferring those sets from their revealed choices. Taking the latter approach, we have proposed, operationalized and estimated a new model to capture unobserved consideration from discrete choice data. It offers important advantages of parsimony over models proposed previously and moreover

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.

The issue of whether consideration sets can be validly inferred from revealed choice data is one with a long history (cf. Roberts and Lattin, 1997). This study has begun to address this very question by studying the convergent validity of stated and revealed consideration sets in our on-line choice experiment. While more research in this area is needed, our first findings are promising indeed and we tentatively conclude that we do infer consideration from revealed choice behavior using our model.

The literature describes two classes of factors shaping the consideration set: situational and individual factors. The situational factors involve their recognition at the point of purchase, the individual factors relate to retrieval of alternatives from memory (see also Alba and Chattopadhyay, 1985). Consistent with this distinction, we included in-store merchandizing (display and feature) as an operationalization of situational factors. We included brand intercepts and, through the dynamics in the error process, consideration-set based state dependence, as an operationalization of individual memory-based factors. We therefore allow different marketing control variables to affect the choice process in a different manner based on theory on how they should affect that process: while price is assumed to affect choice directly, merchandizing is specified to affect choice through its effect on consideration. Although we found our model to reproduce consideration levels for individual brands well, our operationalization of individual and situational factors is necessarily partial and therefore has its limitations. Other situational factors may affect consideration, and memory of alternatives beyond the consideration levels of the previous purchase occasion may also have an effect. However, we think these to be empirical questions that can be addressed if sufficient data are available. We believe

that our operationalization in the two studies provides a reasonable representation of the choice processes for the products in question. This holds in particular for the choice experiment, where situational factors were almost completely under experimental control. In the analysis of the scanner panel choice data, the operationalization of individual and situational factors and the identification of their effects is limited by both the variables and the amount of information in the data set. Nevertheless, we do believe our model specification captures the main features of the choice process in that case. The fact that effects of in-store merchandizing on consideration are very strong and are more tightly distributed in the two-stage model than in the single-stage model, provides support for our model specification.

Several studies on consideration have focused on the dependencies of alternatives in the consideration set. In particular, the attractiveness of an alternative for consideration has been reported to increase if an inferior alternative is added to the set (cf. Huber and Puto, 1982). Our approach can account for such phenomena through the covariance structure of the consideration stage model component. However, our empirical analyses, both on experimental and several scanner data sets, reveal that, after accounting for in-store merchandizing and past purchase, consideration is essentially independent across brands, as evidenced by zero covariance. These effects may be absent since we study mature markets where unattractive alternatives have been eliminated from the marketplace. There may be product categories for which consideration may be dependent across brands, especially for emerging markets where unattractive alternatives may still be available. The empirical verification of that effect from revealed choice data remains an important topic for future research.

We find that the two-stage model offers a more appealing interpretation for the role

of in-store merchandizing on consumer choice than a single-stage model. 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*. This attribution is questionable on logical grounds. The goal of the consumer is to buy a (utility maximizing) brand and not to acquire brand information. Therefore, contextual information such as 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. Andrews and Srinivasan (1995) called this the direct priming effect. In-store merchandizing programs are therefore suitably seen as fulfilling the goal of lowering the mental cost of information acquisition. Economic theory suggests that the more consumers will be price-oriented, the easier it is to obtain price information (for example, Stigler, 1961). This is exactly what is implied by our model. As we have shown, single-stage choice models do not have this property. 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) as

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

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

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

for $i = 1, \dots, I$. 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. Hence, the full conditional posterior distribution of α given Σ , C^* and ρ 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 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) as

$$\Sigma^{-\frac{1}{2}}(C_{it}^* - X_{it}\alpha) = \rho\Sigma^{-\frac{1}{2}}(C_{i,t-1}^* - X_{i,t-1}\alpha) + \Sigma^{-\frac{1}{2}}\varepsilon_{it}, \quad (\text{A.3})$$

for $i = 1, \dots, I$, $t = 2, \dots, T_i$. 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.3). This is however not the proper full conditional posterior distribution as we have neglected the first observations of each household. The density of the first observations of the households as function of ρ is given by

$$\pi(\rho) = \left(\frac{1}{\sqrt{2\pi}} \right)^{JI} \left| \frac{\Sigma}{1 - \rho^2} \right|^{-\frac{1}{2}I} \prod_{i=1}^I \exp \left(-\frac{1}{2} (1 - \rho^2) (C_{i1}^* - X_{i1}\alpha)' \Sigma (C_{i1}^* - X_{i1}\alpha) \right). \quad (\text{A.4})$$

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.3).

Step 2 Draw u from a uniform distribution on the interval $(0, 1)$ and accept ρ^{new} if

$$\pi(\rho^{\text{new}})/\pi(\rho^{\text{old}}) > u, \text{ otherwise take } \rho^{\text{new}} = \rho^{\text{old}}.$$

Sampling of β

In the brand choice model, β is sampled in a similar way as α . We rewrite the equations (6) for which $c_{ijt} = 1^6$ as

$$\omega_j^{-1} U_{ijt} - \omega_j^{-1} W'_{ijt} \beta_i = \omega_j^{-1} W'_{ijt} \beta + \omega_j^{-1} \eta_{ijt}, \quad (\text{A.5})$$

for $j = 1, \dots, J$, $i = 1, \dots, I$, and $t = 1, \dots, T_i$. 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 given β_i , Ω , C^* and U is normal. The mean and variance result from the OLS estimator of β in (A.5), see again Zellner (1971, Chapter VIII).

⁶The value of c_{ijt} is determined by the value of C_{ijt}^* .

Sampling of β_i

To sample β_i we can follow a similar approach as for β . We rewrite the equations (6) for which $c_{ijt} = 1$ as

$$\begin{aligned}\omega_j^{-1}U_{ijt} - \omega_j^{-1}W'_{ijt}\beta &= \omega_j^{-1}W'_{ijt}\beta_i + \omega_j^{-1}\eta_{ijt} \\ \mathbf{0} &= \Sigma_\beta^{-\frac{1}{2}}\beta_i + \Sigma_\beta^{-\frac{1}{2}}v_i,\end{aligned}\tag{A.6}$$

for $j = 1, \dots, J$, $i = 1, \dots, I$ and $t = 1, \dots, T_i$. The last line follows from the fact that $v_i = (\beta_i - \mathbf{0}) \sim N(\mathbf{0}, \Sigma_\beta)$. This represents $1 + \sum_{t=1}^{T_i} \sum_{j=1}^J c_{ijt}$ 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.6).

Sampling of Σ_β

For Σ_β it holds that

$$p(\Sigma_\beta|\beta_i) \propto \exp\left(-\frac{1}{2}\beta_i\Sigma_\beta^{-1}\beta_i'\right),\tag{A.7}$$

and hence Σ_β can be sampled from an inverted Wishart distribution, see Zellner (1971, Chapter VIII).

Sampling of Σ

To sample Σ we note that

$$p(\Sigma|\alpha, \rho, C^*) \propto \pi(\Sigma) = |\Sigma|^{-\frac{1}{2}\sum_{i=1}^I T_i} \exp\left(-\frac{1}{2}\sum_{i=1}^I \sum_{t=2}^{T_i} \varepsilon'_{it} \Sigma^{-1} \varepsilon_{it}\right).\tag{A.8}$$

where

$$\begin{aligned}\varepsilon_{it} &= (C_{it}^* - X_{it}\alpha) - \rho(C_{i,t-1}^* - X_{i,t-1}\alpha) & \text{for } t = 2, \dots, T_i, \\ \varepsilon_{i1} &= \sqrt{1 - \rho^2}(C_{i1}^* - X_{i1}\alpha)\end{aligned}\tag{A.9}$$

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

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 Basag 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 Ω

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

$$p(\omega_j | \beta, \beta_i, U, C^*) \propto \frac{1}{\omega_j^\nu} \exp\left(-\frac{1}{2\omega_j^2} \sum_{i=1}^I \sum_{t=1}^{T_i} I[c_{ijt} = 1](U_{ijt} - W'_{ijt}(\beta + \beta_i))^2\right), \quad (\text{A.10})$$

and hence

$$\frac{\sum_{i=1}^I \sum_{t=1}^{T_i} I[c_{ijt} = 1] (U_{ijt} - W'_{ijt}(\beta + \beta_i))^2}{\omega_j^2} \sim \chi^2(\nu) \quad (\text{A.11})$$

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

Sampling of U

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

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

and hence U_{it} is normal distributed with mean $W_{it}(\beta + \beta_i)$ and variance Ω . The full conditional posterior distributions of the elements of U_{it} are of course also normal. Hence, U_{ijt} for $c_{ijt} = 1$ can be sampled from truncated normal distributions in the following way

$$U_{ijt}|U_{i,-j,t} \sim \begin{cases} \text{normal on } (-\infty, U_{i,d_{it},t}) & \text{if } d_{it} \neq j \\ \text{normal on } (\max(U_{ikt} \text{ for all } k \neq j | c_{ikt} = 1), \infty) & \text{if } d_{it} = j \end{cases} \quad (\text{A.13})$$

where $U_{i,-j,t} = (U_{ikt} \text{ for all } k \neq j | c_{ikt} = 1)$, see Geweke (1991) for details.

Sampling of C^*

To sample C_{it}^* we have to consider (2) for time t and $t + 1$. Rewriting these equations gives

$$\begin{aligned} -\Sigma^{-\frac{1}{2}}(\rho C_{i,t-1}^* + (X_{it} - \rho X_{i,t-1})\alpha) &= -\Sigma^{-\frac{1}{2}}C_{it}^* + \Sigma^{-\frac{1}{2}}\varepsilon_{ijt} \\ \Sigma^{-\frac{1}{2}}(C_{i,t+1}^* - (X_{i,t+1} - \rho X_{it})\alpha) &= \rho \Sigma^{-\frac{1}{2}}C_{it}^* + \Sigma^{-\frac{1}{2}}\varepsilon_{ij,t+1}, \end{aligned} \quad (\text{A.14})$$

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

$$-\sqrt{1 - \rho^2} \Sigma^{-\frac{1}{2}} X_{it} \alpha = -\sqrt{1 - \rho^2} \Sigma^{-\frac{1}{2}} C_{it}^* + \sqrt{1 - \rho^2} \Sigma^{-\frac{1}{2}} \varepsilon_{ijt}. \quad (\text{A.15})$$

This can again be interpreted as a regression model in the parameter C_{it}^* , which implies that the conditional distribution of C_{it}^* is normal with mean and variance following from

the OLS estimator of C_{it}^* in (A.14) and (A.15). The conditional distribution of C_{ijt}^* in the MVP model is in this case also normal with, let say, mean μ_j and variance σ_j^2 . If brand j is chosen by household i at time t , we have to sample C_{ijt}^* from a normal distribution with mean μ_j and variance σ_j^2 subject to $C_{ijt}^* > 0$. In other cases we have to follow a different approach.

The value of the full posterior density of C_{ijt}^* changes if a brand enters the consideration set. The full conditional posterior density of C_{ijt}^* is therefore

$$f(C_{ijt}^*) = \frac{1}{\kappa} (I[C_{ijt}^* < 0] + \phi(U_{ijt}|X_{ijt}(\beta + \beta_i), \omega_j^2) I[C_{ijt}^* > 0]) \frac{1}{\sigma_j} \phi((C_{ijt}^* - \mu_j)/\sigma_j). \quad (\text{A.16})$$

If $C_{ijt}^* > 0$ brand j enters the consideration set and the term $\phi(U_{ijt}|\cdot) = \phi(U_{ijt}|X_{ijt}(\beta + \beta_i), \omega_j^2)$ which is the pdf of a normal distribution is added to the likelihood function. As the integral of (A.16) over C_{ijt}^* has to be 1, the value of κ is given by

$$\kappa = \Phi(-\mu_j/\sigma_j) + \phi(U_{ijt}|X_{ijt}(\beta + \beta_i), \omega_j^2)(1 - \Phi(-\mu_j/\sigma_j)). \quad (\text{A.17})$$

To draw C_{ijt}^* we use the inverse method. The CDF of C_{ijt}^* is given by

$$F(C_{ijt}^*) = \begin{cases} \frac{1}{\kappa} \Phi((C_{ijt}^* - \mu_j)/\sigma_j) & \text{if } C_{ijt}^* < 0 \\ \frac{1}{\kappa} \Phi(-\mu_j/\sigma_j) + \frac{1}{\kappa} \phi(U_{ijt}|\cdot) (\Phi((C_{ijt}^* - \mu_j)/\sigma_j) - \Phi(-\mu_j/\sigma_j)) & \text{if } C_{ijt}^* > 0 \end{cases} \quad (\text{A.18})$$

Sample of C_{ijt}^* using the inverse CDF technique proceeds in the following way

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

Step 2 Set $C_{ijt}^* = \sigma_j \Phi^{-1}(x) + \mu_j$, where

$$x = \begin{cases} \kappa u & u < \frac{1}{\kappa} \Phi(-\mu_j/\sigma_j) \\ (\kappa u + (\phi(U_{ijt}|\cdot) - 1) \Phi(-\mu_j/\sigma_j)) / \phi(U_{ijt}|\cdot) & \text{otherwise,} \end{cases} \quad (\text{A.19})$$

where Φ^{-1} is the inverse CDF of a standard normal distribution and κ is given in (A.17).

References

- Alba, J.W., A. Chattopadhyay, 1985. The effects of context and part-category cues on the recall of competing brands. *Journal of Marketing Research* **22** 340–349.
- Albert, James H., Siddhartha Chib, 1993. Bayesian Analysis of Binary and Polychotomous Response Data. *Journal of the American Statistical Association* .
- Allenby, Greg M., Peter J. Lenk, 1994. Modeling Household Purchase Behavior with Logistic Normal Regression. *Journal of the American Statistical Association* .
- Allenby, Greg M., Peter E. Rossi, 1991. Quality Perceptions and Asymmetric Switching Between Brands. *Marketing Science* .
- , 1999. Marketing Models of Consumer Heterogeneity. *Journal of Econometrics* .
- Andrews, Rick L., T.C. Srinivasan, 1995. Studying Consideration Effects in Empirical Choice Models Using Scanner Panel Data. *Journal of Marketing Research* **32** 30–41.
- Basag, J., P. J. Green, 1993. Spatial Statistics and Bayesian Computation. *Journal of the Royal Statistical Society B* **55** 25–37.
- Belonax, Joseph A., Robert A. Mittelstaedt, 1978. Evoked set size as a function of number of choice criterion and information variability. In *Advances in Consumer Research*. Association for Consumer Research, Ann Arbor, MI, volume 5, 48–51.
- Ben-Akiva, Moshe, Bruno Boccara, 1995. Discrete Choice Models with Latent Choice Sets. *International Journal of Research in Marketing* **12(1)** 24–Sep.

- Bettman, James R., 1979. *An Information Processing Theory of Consumer Behavior*. Addison-Wesley, Reading MA.
- Bronnenberg, Bart J., Wilfried R. Vanhonnacker, 1996. Limited Choice Sets, Local Price Response, and Implied Measures of Price Competition. *Journal of Marketing Research* **33** 163–174.
- Bunch, David S., 1991. Estimability in the Multinomial Probit Model. *Transportation Research B* **25B**(1) 1–12.
- Casella, G., E. I. George, 1992. Explaining the Gibbs Sampler. *American Statistician* **46** 167–174.
- Chiang, J., S. Chib, C. Narasimhan, 1999. Markov chain Monte Carlo and models of consideration set and parameter heterogeneity. *Journal of Econometrics* **9** 223–248.
- Chib, Siddhartha, Edward Greenberg, 1995. Understanding the Metropolis-Hastings Algorithm. *American Statistician* **49** 327–335.
- , 1998. Analysis of Multivariate Probit Models. *Biometrika* .
- Corstjens, Marcel, Judy Corstjens, 1999. *Storwars: Competing for Shelf Space and Mind Space of the Consumer*. Wiley, NY.
- Damien, Paul, Jon Wakefield, Stephen 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, Yancy D., Greg M. Allenby, 2003. Multivariate Analysis of Multiple Response Data. *Journal of Marketing Research*, forthcoming .

- Geman, S., D. Geman, 1984. Stochastic Relaxations, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transaction on Pattern Analysis and Machine Intelligence* **6** 721–741.
- Geweke, John 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, 571–578.
- Geweke, John F., Michael P. Keane, David E. Runkle, 1997. Statistical Inference in the Multinomial Multiperiod Probit Model. *Journal of Econometrics* .
- Gilbride, Timothy J., Greg M. Allenby, 2003. A Choice Model with Conjunctive, Disjunctive, and Compensatory Screening Rules. *Marketing Science, forthcoming* .
- Guadagni, Peter M., John D.C. 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* .
- Hausman, J., D. Wise, 1978. A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogenous Preferences. *Econometrica* **45** 319–339.
- Howard, J.A., J.N. Sheth, 1969. *The Theory of Buyer Behavior*. New York: Wiley.
- Huber, J., C. Puto, 1982. Market Boundaries and Product Choice: Illustrating Attraction and Substitution Effects. *Journal of Consumer Research* **10 (June)** 31–44.

- Keane, Michael P., 1992. A Note on Identification in the Multinomial Probit Model. *Journal of Business & Economic Statistics* **10**(2) 193–200.
- Lehmann, D.R., Y. Pan, 1994. Context Effects, New Brand Entry, and Consideration Sets. *Journal of Marketing Research* **31** pp. 364–74.
- Lussier, J.G., R.W. Olshavsky, 1979. Task Complexity and Contingent Processing in Brand Choice. *Journal of Consumer Research* **6** 154–165.
- Malhotra, Naresh K., Mark Peterson, Susan Bardi 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, S. Gupta, 1999. The “Shopping Basket”: A Model for Multi-category Purchase Incidence Decisions. *Marketing Science* **18:02** 95–115.
- Manrai, Ajay K., Rick 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, Robert E., N. G. Polson, P. E. Rossi, 2000. A Bayesian Analysis of the Multinomial Probit Model with Fully Identified Parameters. *Journal of Econometrics* **99** 173–193.
- McCulloch, Robert E., Peter E. Rossi, 1994. An exact likelihood analysis of the multinomial probit model. *Journal of Econometrics* **64** 207–240.

- McFadden, 1973. Conditional Logit Analysis of Qualitative Choice Behavior. *Frontiers in econometrics* 105–142.
- Mehta, Nitin, Surendra Rajiv, Kannan Srinivasan, 2003. Price Uncertainty and Consumer Search: A Structural Model of Consideration Set Formation. *Marketing Science* **22**(1) pp. 58–84.
- Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller, 1953. Equations of State Calculations by Fast Computing Machines. *Journal of Chemical Physics* **21** 1087–1092.
- Mitra, Anusree, 1995. Advertising and the Stability of Consideration Sets Over Multiple Purchase Occasions. *International Journal of Research in Marketing* **12** 81–94.
- Nedungadi, Prakash, 1990. Recall and Consumer Consideration Sets: Influencing Choice Without Altering Brand Evaluations. *Journal of Consumer Research* **17** 263–276.
- Newell, A., H. Simon, 1972. *Human Problem Solving*. Englewood Cliffs, NJ: Prentice-Hall.
- Paap, Richard, Philip Hans Franses, 2000. A Dynamic Multinomial Probit Model for Brand Choice with Different Long-run and Short-run Effects of Marketing Mix Variables. *Journal of Applied Econometrics* **15** 717–744.
- 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, John H., James M. Lattin, 1991. Development and Testing of a Model of Consideration Set Composition. *Journal of Marketing Research* **28** 429–440.

- , 1997. Consideration: Review of Research and Prospects for Future Insights. *Journal of Marketing Research* **34** 406–410.
- Siddarth, S., R.E. Bucklin, D. G. Morisson, 1995. Making the Cut: Modeling and Analyzing Choice Set Restriction in Scanner Panel Data. *Journal of Marketing Research* **33** 255–266.
- Stigler, G. J., 1961. The Economics of Information. *Journal of Political Economy* **69** 213–25.
- Tanner, M.A., W.H. Wong, 1987. The Calculation of Posterior Distributions by Data Augmentation. *Journal of the American Statistical Association* .
- Tierney, L., 1994. Markov Chains for Exploring Posterior Distributions. *Annals of Statistics* .
- Wright, P., F. Barbour, 1977. Phased Decision Strategies: Sequels to Initial Screening. In *Multiple Criteria Decision Making: North-Holland TIMS Studies in the Management Science*, edited by M. Starr, M. Zeleny. Amsterdam: North-Holland.
- Zellner, Arnold, 1971. *An Introduction to Bayesian Inference in Econometrics*. Wiley, New York.