ELSEVIER Journal of Econometrics 89 (1999) 223–248

JOURNAL OF Econometrics

Markov chain Monte Carlo and models of consideration set and parameter heterogeneity

Jeongwen Chiang[a,*], Siddhartha Chib[b], Chakravarthi Narasimhan[c]

[a] *Hong Kong University of Science and Technology, Clearwater Bay, Kowloon, Hong Kong*
[b] *Department of Business and Public Administration, Washington University, St. Louis, MO 63130, USA*
[c] *John M. Olin School of Business, Washington University, St. Louis, MO 63130, USA*

Abstract

In this paper the authors propose an integrated consideration set-brand choice model that is capable of accounting for the heterogeneity in consideration set and in the parameters of the brand choice model. The model is estimated by an approximation free Markov chain Monte Carlo sampling procedure and is applied to a scanner panel data. The main findings are: ignoring consideration set heterogeneity understates the impact of marketing mix and overstates the impact of preferences and past purchase feedback even when heterogeneity in parameters is modeled; the estimate of consideration set heterogeneity is robust to the inclusion of parameter heterogeneity; when consideration set heterogeneity is included the parameter heterogeneity takes on considerably less importance; the promotional response of households depends on their consideration set even if the underlying choice parameters are identical.

JEL classification: C1; C5; D1; M3

Keywords: Heterogeneity; Consideration set; Random effect; Brand choice models; Metropolis–Hasting algorithm

1. Introduction

Econometric brand choice models have played an important role in marketing problems starting with the work of Guadagni and Little (1983). A vast

* Corresponding author. E-mail: mkchiang@usthk.ust.hk

0304-4076/99/$ – see front matter
PII: S0304-4076(98)00062-1

literature has now emerged that is concerned with the modeling of brand choice at the household level using a random utility framework. Various extensions of the basic model have been considered including models with unobserved heterogeneity (see for example Kamakura and Russell, 1989; Chintagunta et al., 1991; Gonul and Srinivasan, 1993) and models in which the choice-set is household specific (see for example Andrews and Srinivasan, 1995; Swait and Ben-Akiva, 1987; Ben-Akiva and Boccara, 1995; Siddarth et al., 1995; Bronnenberg and Vanhonacker, 1995). These extensions have been motivated by theoretical and empirical arguments. The parameter heterogeneity approach is driven mainly by statistical considerations (a la the random effects models in econometrics) and the need to model intra-individual correlation in choices. Justification for the assumption that consideration sets are household specific (and perhaps smaller than the set of available brands) has been provided by Alba and Chattaopadhyay, 1985; Belonax and Mittelstaedt, 1978; Hauser and Wernerfelt, 1990; Gensch and Soofi, 1995; Parkinson and Reilly, 1979; Roberts and Lattin, 1991; Shocker et al., 1991 and in the special issue of International Journal of Research in Marketing, 1995.

Absent from this vast literature is a unified brand-choice model that permits both parameter and consideration set heterogeneity to coexist in the same model. Since both types of heterogeneity exist, to argue that consideration-set heterogeneity is important in a brand-choice context one has to allow the model to exhibit parameter heterogeneity. Likewise, if one wishes to make a case for parameter heterogeneity, one must allow the consideration sets to vary. Thus, the purpose of this paper is to advance an econometric methodology to estimate a new brand choice model that includes heterogeneity in *both* the consideration set and parameters. In this model, parameter heterogeneity is modeled through a Gaussian random effects formulation, similar to many other papers in the literature. Household choice set formation is modeled by taking the power set (the set of all possible subsets) of the available brands and assigning a household-specific probability mass on each subset. The model is completed by assuming that, conditioned on the random effects, the consideration set and the fixed effects parameters, choice is determined by a multinomial logit distribution. This model is fitted to a scanner panel data consisting of four brands in the ketchup category. These high-dimensional models are estimated by an approximation free Markov chain Monte Carlo procedure. It does not seem to be feasible to estimate these models by any other means.

The proposed model and the estimation procedure should be of interest to researchers in both marketing and econometrics. Our model for consideration set formation is non-parametric and could be used as a benchmark against which other consideration set models (all of which rely on a parametric formulation) may be evaluated in future work. The estimation results indicate that ignoring consideration set heterogeneity understates the impact of marketing mix and overstates the impact of preferences and past purchase feedback *even* when heterogeneity in parameters is modeled; the estimate of consideration set heterogeneity is robust to the inclusion of parameter heterogeneity; when consideration set heterogeneity is included the parameter heterogeneity takes on considerably less importance; the promotional response of households depends on their consideration set even if the underlying choice parameters are identical.

The rest of the paper is organized as follows. In the next section we develop the model and discuss our econometric procedure. In Section 3 we describe the scanner panel data obtained from Nielsen and discuss the results of our model and a few benchmark models when applied to this data. Implications and concluding remarks are made in the final sections.

2. Proposed hierarchical model and estimation procedure

In this section, we begin with a discussion of the model for consideration set formation and parameter heterogeneity in the context of scanner panel data. We show that the likelihood function is quite complex and intractable. We then describe the alternative Bayesian simulation-based estimation approach which by-passes the calculation of the likelihood function.

2.1. Modeling consideration sets

Suppose that the ith household in the panel ($i = 1,2,\dots,n$) is faced with a choice of J brands at the tth purchase occasion ($t = 1,2,\dots,T_i$). We let C denote the possible subsets of the J brands (the power set of J minus the empty set) and let

$$C = \{1,2,\dots,S\},$$

where $S = 2^{J-1}$. Each element of C identifies a particular subset. For example, if the market consists of three bands denoted $\{A, B, C\}$, then C consists of the subsets $\{A\}$,$\{B\}$, $\{C\}$, $\{AB\}$, $\{AC\}$, $\{BC\}$ and $\{ABC\}$ and $C = 1$ identifies the subset $\{A\}$, and $C = 7$ identifies the subset $\{ABC\}$. We model consideration set heterogeneity as follows. Let C_i denote the consideration set of the ith household in the sample, $i = 1,\dots,n$. Of course, none of the consideration sets are known. Hence, we assume there is a (discrete) probability distribution for each C_i. The distribution consists of mass points on each possible value of C_i. Let p_i denote a $S \times 1$ vector of probabilities (indexed by i because these probabilities are household-specific) such that the probability that $C_i = s$ (denoted $\Pr(C_{is}|p_i)$) is given by the sth element of p_i:

$$\Pr(C_{is}|p_i) = p_{is}.$$

We complete the specification of the consideration set model by assuming that p_i are distributed across the population according to a Dirichlet distribution with (known) parameter vector $\alpha = (\alpha_1, \ldots, \alpha_S)$:

$$p_i | \alpha \sim \mathscr{D}(\alpha_1, \ldots, \alpha_S).$$

In this example, we take $\alpha_s = 1/S$, $s = 1, \ldots, S$ to reflect weak prior information about p_i. Several important features of this specification should be noted. First, the discrete distribution on the possible subsets is completely arbitrary (except for the constraint that the probabilities sum to one). This feature is extremely attractive since it avoids any specification error at this stage of the model. This departs from prior work where the probabilities p_i are represented parametrically by a two-step process and one first specifies a parametric model for the probability of brand consideration (for each brand) and second a model for the probability of each subset in terms of the probabilities of brand consideration. Notable papers along these lines include, for example, Andrews and Srinivasan, 1995 and Bronnenberg and Vanhonacker, 1995. Although this approach has its virtues, it should be clear that the model we propose avoids the specification problems that arise at each stage of the latter process, particularly since there is still no consensus in the literature as to how and why consumers narrow their options, and since in virtually all instances researchers do not have direct observations of choice process itself.

Second, our model assumes the consideration set of each household does not change across the T_i purchase occasions. For example, in the 3 brand example, if $T_i = 5$ and $C_i = 4$, then the brand choice on each of the 5 purchase occasions is made conditional on the subset $\{AB\}$. This assumption can be relaxed provided one is willing to specify a process for the evolution and revision of consideration sets across time. We refrain from doing this in the interest of simplicity. Furthermore, in the context of our empirical application, the stationary-consideration set assumption is very reasonable, as shown later.

2.2. Modeling brand choice and parameter heterogeneity

We follow the large extant literature on discrete brand choice models and assume that choice of a brand conditioned on the parameters and the consideration set is governed by a multinomial logit model. In order to model heterogeneity in parameters (so that, for example, marketing mix parameters in this model are household specific), the model includes random effects and a distribution of random effects across the population.

Let $y_{it} = j$ denote the event that household i on purchase occasion t chooses brand j and let x_{itj} denote the vector of attributes. Then the conditional probability that brand j is chosen, given C_{is}, is specified as

$$\Pr(y_{it} = j | \beta, b_i, C_{is}) = \begin{cases} \dfrac{\exp(x'_{itj}(\beta + b_i))}{\sum_{k \in C_{is}} \exp(x'_{itk}(\beta + b_i))} & \text{if } j \in C_{is}, \\ 0 & \text{otherwise,} \end{cases}$$

where β is a vector of brand choice parameters to be estimated and b_i is a vector of random effects that represent parameter heterogeneity. Note that if j is not in the consideration set the brand choice probability is zero whereas if C_{is} is a singleton set consisting of j, then the brand choice probability is one. We assume that the random effects (which represent deviations from the fixed effects β) are drawn independently from a normal population with mean zero and (unknown) covariance matrix D, i.e.,

$$b_i \sim \mathrm{N}(0, D), \forall i.$$

The assumption of normality is common in the econometrics and statistical literatures and allows for the easy interpretation of the matrix D and for the possibility that the random effects are correlated. The square root of the diagonal elements of D, the standard deviations of the random effects, are important parameters that measure the extent of parameter heterogeneity across the population.

It is important to recognize that the model with both types of consumer heterogeneity cannot be estimated by standard means, say by maximizing its likelihood function. To see this, let $L_i(\beta,D)$ denote the contribution to the likelihood from the ith household and let $L(\beta,D) = \prod_{i=1}^{n} L_i(\beta, D)$ denote the full likelihood. Then

$$L_i(\beta, D) = \int_{p_i}\int_{b_i}\left[\prod_{t=1}^{T_i}\sum_{s=1}^{S}\Pr(y_{it}|C_{is}, \beta, b_i)p_{is}\right]\phi(b_i|0, D)\pi(p_i|\alpha)\,\mathrm{d}b_i\,\mathrm{d}p_i,$$

where $\phi(b_i|0, D)$ denotes the multivariate normal density function with mean zero and covariance matrix D,$\Pr(y_{it}|C_{is}, \beta)$ denotes the multinomial logit probability of the brand (given the subset identified by C_{is}) and p_{is}, as noted earlier, denotes the probability of C_{is} given p_i which is integrated over p_i, $p_i : \sum_{s=1}^{S} p_{is} = 1$ (using the Dirichlet density, $\pi(p_i|\alpha)$). Note that the term in square brackets represents a mixture distribution, mixed over all the S possible subsets of the J brands. The integrals over the density of b_i and p_i arise from the assumed parameter heterogeneity of the parameters and consideration set probabilities.

It is not difficult to see that this likelihood function, which requires several multiple integrations over b_i and p_i, is quite complex and intractable. An additional difficulty arises from the fact that the model contains a mixture component and such models often generate multiple modes. Thus, the maximum likelihood estimate is a somewhat undesirable summary of the likelihood function, even if it were available. Hence, we now turn our attention to the question of estimation and describe a unified, approximation free Bayesian simulation-based method to estimate the parameter vector β, the random effects covariance matrix D and $\{p_i\}$, the household-specific probability distribution over the subsets.

2.3. Markov chain Monte Carlo estimation

Our approach makes use of recent developments in Markov chain Monte Carlo (MCMC) methods (such as the Gibbs sampler and the Metropolis–Hastings algorithms) that were brought to the attention of Bayesian statisticians by Tanner and Wong, 1987 and Gelfand and Smith, 1990. MCMC methods provide a powerful tool kit for simulating otherwise intractable posterior distributions (see Chib and Greenberg, 1995, 1996 for detailed discussions). The main idea is to simulate the posterior distribution, thus leading to a sample of draws on the parameters. All posterior inferences are then based on this sample.

Since our approach to estimation (described below) is Bayesian, we complete the specification of the model by specifying prior distributions for the remaining unknowns in the model, the parameters, β and D. We assume that

$$\beta \sim \mathrm{N}(\beta_0, B_0),$$

$$D^{-1} \sim \mathrm{Wishart}(v_0, R_0),$$

a normal distribution for β and a Wishart distribution for D^{-1} with degrees of freedom parameter v_0 and scale matrix R_0. The parameters of these prior distributions can be specified to accord with prior information; in our application (which consists of a seven dimensional β vector) we will choose values that represent weak and diffuse prior information, thus allowing the data to dominate the results. Specifically, we let

$$\beta_0 = 0, B_0 = 3I_7, v_0 = 9, R_0 = (\tfrac{1}{9})I_7.$$

These hyperparameter values ensure that the distributions are proper but weak. For example, the implied prior distribution of D is such that the prior mean of the diagonal elements is approximately 30 with standard deviation of around 200 while that of the off-diagonal terms has a prior mean close to zero with a prior standard deviation of around 100. The model is completed with these priors and depicted in Fig. 1.

2.4. Simulation of posterior distribution

Let $y_i = (y_{i1}, \ldots, y_{iT_i})$ denote the choices made by the ith household in the panel and let $y = (y_1, \ldots, y_n)$ denote the sample data on choices made by all households. For reasons that will become apparent shortly, we define the parameter space to include all unknowns and let $\psi = (\beta, D, \{b_i\}, \{C_i\}, \{p_i\})$. Then, from Bayes theorem which states that $\pi(\psi|y) \propto \pi(\psi) \times p(y|\psi)$, we define the posterior density of interest to be

$$\pi(\beta, D, \{b_i\}, \{C_i\}, \{p_i\}|y) \propto \pi(\beta, D, \{b_i\}, \{C_i\}, \{p_i\}) \times \Pr(y|\beta, \{b_i\}, \{C_i\}),$$

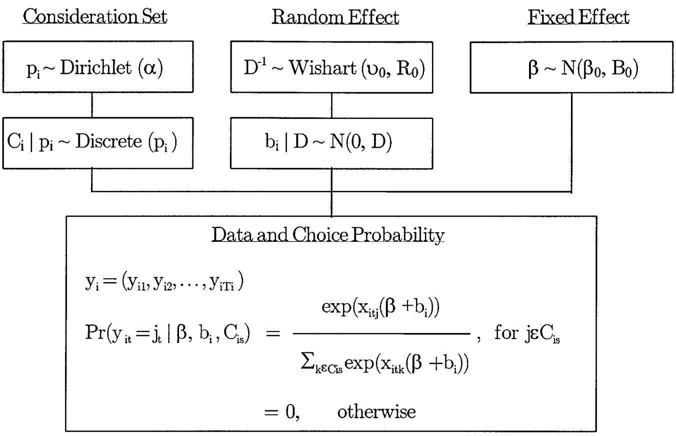

Fig. 1. Data generation process for ith household.

a product of the density on ψ prior to seeing the data and the density of the choices conditioned on the parameters, the random effects and the consideration sets. The normalizing constant of this posterior density (referred to as the marginal likelihood in the literature) is not directly available but, fortunately, in the context of our simulation method, its value is not required. The main reason for defining the parameter space and the posterior density in this manner is that it does not require any integration over $\{b_i\}$, $\{C_i\}$ or $\{p_i\}$. This strategy of enlarging the parameter space to simplify simulations was introduced by Tanner and Wong, 1987 and exploited for the first time in the context of discrete choice models by Albert and Chib, 1993.

With data augmentation the joint posterior density has a simple conditional structure that leads to a tractable recursive simulation algorithm. Consider for instance the conditional posterior distribution of each element of ψ given the data and the remaining elements of ψ:

$$\Pr(\{C_i\}|y, \psi_{-\{C_i\}}), \quad \pi_p(\{p_i\}|y, \psi_{-\{p_i\}}), \quad \pi_\beta(\beta|y, \psi_{-\beta}),$$

$$\pi_b(\{b_i\}|y, \psi_{-\{b_i\}}), \quad \text{and} \quad \pi_D(D|y,\psi_{-D}),$$

where, for example, $\psi_{-\beta} = (D, \{b_i\}, \{C_i\}, \{p_i\})$ denotes all the parameters excluding β. These distributions, which we henceforth refer to as the full conditional distributions, are easily derived and simulated (as we show next). The simulation algorithm proceeds by cycling through the five conditional distributions given above, where the simulations are conditioned on the most recent values of the parameters. Thus, given the current draw at the gth iteration $\beta^{(g)}$, $\{b_i^{(g)}\}$, $D^{(g)}$, $\{p_i^{(g)}\}$, $\{C_i^{(g)}\}$, the next draw in the sequence is obtained by simulating

$$\begin{array}{lll} \{C_i^{(g+1)}\} & \text{from} & \{C_i\}|y, \beta^{(g)}, \{b_i^{(g)}\}, \{p_i^{(g)}\}, \\ \{p_i^{(g+1)}\} & \text{from} & \{p_i\}|y, \{C_i^{(g+1)}\}, \\ \beta^{(g+1)} & \text{from} & \beta|y, \{C_i^{(g+1)}\}, \{b_i^{(g)}\}, \\ \{b_i^{(g+1)}\} & \text{from} & \{b_i\}|y, \beta^{(g+1)}, \{C_i^{(g+1)}\}, \\ D^{(g+1)} & \text{from} & D|\{b_i^{(g+1)}\} \end{array} .$$

Note that the simulation sequence (which block comes first and which comes second, etc.) is arbitrary. In Fig. 2, we present the simulation sequence used in our data analysis. In presenting these full conditional distributions we have utilized the conditional independence structure that is implicit in our model. For example, $\pi_D(D|y, \psi_{-D})$ depends on only $\{b_i\}$ and not the other elements in the conditioning. Likewise, $\pi_\beta(\beta|y, \psi_{-\beta})$ depends on only y and $\{C_i\}$ and $\{b_i\}$, and not on $\{p_i\}$, D. Furthermore, it can be seen that each of the full conditional distributions of $\{C_i\}$, $\{p_i\}$ and $\{b_i\}$ factor into independent distributions – thus, for example,

$$\{C_i\}|y, \beta, \{b_i\}, \{p_i\} = \prod_{i=1}^{n} C_i|y_i, \beta, b_i, p_i,$$

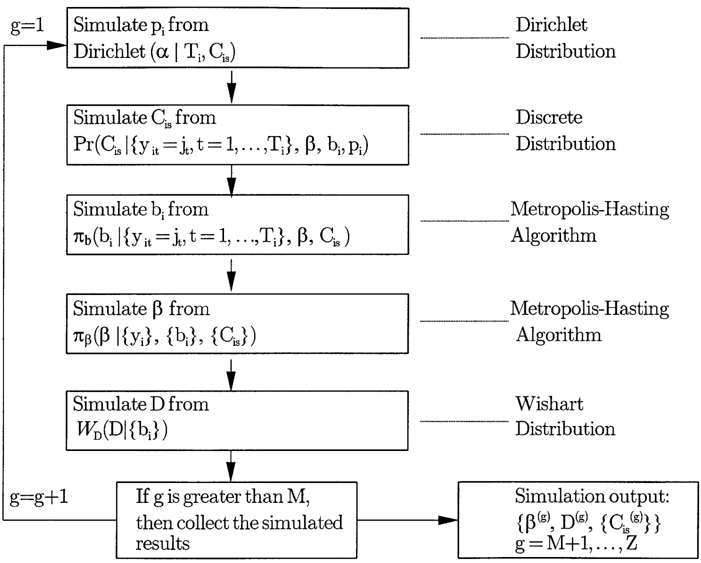

Fig. 2. Simulation blocks.

depending only on information on the ith household. As a result, consideration sets can be simulated one household at a time. These conditional independencies are the source of the simplifications that are exploited by our Gibbs sampling algorithm.

Under regularity conditions, satisfied in this application, it can be shown that if the full conditional distributions are successively sampled, then the simulated trajectory, after ignoring an initial transient phase (discarding the first M simulations), is a sample from the joint posterior distribution $\pi(\psi|y)$.

In the following section, we show how the consideration set C_i is simulated. This section is important in that it reveals how the specification of the model in conjunction with the logic of Bayesian updating provides valuable information about the posterior distribution on the subsets. Because the other full conditional distributions are derived by more routine calculations, the derivations of the remaining full conditional distributions are deferred to Appendix.

2.5. Simulation of C_{is}

The objective is to calculate the mass function $\Pr(C_{is}|y_i, \beta, b_i, p_i)$, $s = 1, 2, \ldots, S$ for each household in the sample. By Bayes theorem and Fig. 1, we have

$$\Pr(C_{is}|y_i, \beta, b_i, p_i) = \frac{\prod_{t=1}^{T_i} \Pr(y_{it} = j_t|\beta, b_i, C_{is}) p_{is}}{\sum_{k=1}^{S} \prod_{t=1}^{T_i} \Pr(y_{it} = j_t|\beta, C_{ik}) p_{ik}},$$

where the denominator is just the normalizing constant. To understand how this probability mass function is computed, let us consider an example. Suppose that $J = 3$, $T_i = 5$ and $y_i = (A, A, B, B, B)$. Thus, brand A is selected on the first two time purchase occasions and brand C on the last three purchase occasions. Now under the assumption that the household has the same consideration set for all purchase occasions, it must be the case that the household's consideration set is either $C_i = 4$ i.e., $\{AB\}$, or $C_i = 7$ i.e., $\{ABC\}$. None of the other subsets include both A and B and thus have a conditional probability of zero. Then, the conditional probability that $C_i = 4$ is proportional to

$$\Pr(y_{i1} = A|\beta, b_i, C_{i4}) \times \Pr(y_{i2} = A|\beta, b_i, C_{i4}) \times \Pr(y_{i3} = B|\beta, b_i, C_{i4}) \times \Pr(y_{i4} = B|\beta, b_i, C_{i4}) \times \Pr(y_{i5} = B|\beta, b_i, C_{i4}) \times p_{i4},$$

whereas the probability that $C_i = 7$ is proportional to

$$\Pr(y_{i1} = A|\beta, b_i, C_{i7}) \times \Pr(y_{i2} = A|\beta, b_i, C_{i7}) \times \Pr(y_{i3} = B|\beta, b_i, C_{i7}) \times \Pr(y_{i4} = B|\beta, b_i, C_{i7}) \times \Pr(y_{i5} = B|\beta, b_i, C_{i7}) \times p_{i7},$$

where each of the terms is simply a multinomial logit probability. The key point to note is that once we condition on the observed sequence of brand choices, a great deal of information is revealed about the allowable consideration sets; Bayes theorem then provides the mechanism to update our prior beliefs to posterior beliefs. Hence, conditional on the values of y_i, β, b_i, and p_i, the mass function $\Pr(C_{is}|y_i, \beta, b_i, p_i)$, $s = 1, 2, \dots, S$ is then used to simulate C_{is}.

For the remaining parameter components, $\{p_i\}$, β, $\{b_i\}$, and D, the simulations are from the respective full conditional distributions (see the Appendix).

2.6. Implementation notes

The Markov chain Monte Carlo sampling process just described is run for G cycles. It furnishes a sample on the parameters $\{\beta^{(g)}, D^{(g)}\}$, the consideration sets $\{C_1^{(g)}, \dots, C_n^{(g)}\}$ and the probabilities $\{p_1^{(g)}, \dots, p_n^{(g)}\}$ along with the random effects, as the iterations are conducted for $g = 1, \dots, G$. In our application, G is chosen to be 12 000 in which the first 3000 iterations are discarded and every third iteration of the remaining 9000 iterations is saved to reduce the serial correlation of the sampled draws.

On the basis of this simulated sample, various posterior estimates are calculated. For example, the posterior mean and variance of β is estimated as sample averages:

$$\bar{\beta} = G^{-1} \sum_{g=1}^{G} \beta^{(g)}, \quad \mathrm{Var}(\beta|y) \approx G^{-1} \sum_{g=1}^{G} (\beta^{(g)} - \bar{\beta})(\beta^{(g)} - \bar{\beta})'$$

Likewise, the posterior mean of D (the random effects variance) is

$$\bar{D} = G^{-1} \sum_{g=1}^{G} D^{(g)}.$$

For G large, these estimates converge to the corresponding posterior moments. As another example, the posterior distribution of the consideration sets,

$$\Pr(C_i = s|y),$$

is estimated by the sample average

$$\Pr(C_i = s|y) = G^{-1} \sum_{g=1}^{G} I(C_i^{(g)} = s),$$

where I is the indicator function that takes the value 1 if $C_i^{(g)} = s$ and the value zero otherwise.

Posterior distributions of functions of the parameters are computed in the same fashion. For instance, the posterior distribution of the choice probability sensitivity with respect to the kth marketing mix, $\eta_k^{(g)}$, is calculated by storing (at each iteration) $\eta_{i,k}^{(g)}$ evaluated at the current values of $\beta^{(g)}$, $C_i^{(g)}$ and $b_i^{(g)}$.

Specifically, let

$$\eta_{i,k}^{(g)} = \frac{\partial \Pr(y_i | \beta^{(g)}, C_i^{(g)}, b_i^{(g)})}{\partial x_k}$$

and

$$\eta_k^{(g)} = n^{-1} \sum_{i=1}^{n} \eta_{i,k}^{(g)}.$$

Then, our estimate of the posterior distribution of the sensitivity with respect to the kth covariate is based on the sample $\{\eta_k^{(g)}, g = 1, \ldots, G\}$.

3. Ketchup scanner data example

In this section we estimate the model using scanner panel data set (from Nielsen). The data set combines purchase history in the ketchup category for a large number of households and store environment records from the Sioux Falls, SD covering the period 1987 to mid 1988. We focus on the results as they relate to consideration set formation in the presence of both parameter and consideration set heterogeneity. After presenting these results, we then make a comparative assessment of this model is relation to several important special cases, namely models with parameter heterogeneity but no consideration set formation and models with consideration set formation but no parameter heterogeneity. We also defer to Section 3.4 discussion of the estimates of the fixed effects parameters β, estimates of various elasticities, and estimates of the random effects variances D.

3.1. Data

We utilize a randomly selected sub-sample of 402 households who consistently purchased during the calibrated period (1987.01–1988.14) as well as the holdout period (1988.15–1988.34), giving rise to a total of 5076 purchases. Part of this sample (4347 observations) is used for estimation while the remainder (for a total of 729 purchase records) is set aside for model validation.

The following features of the data prove useful in the subsequent analysis. First, the ketchup market contains three major brands, Hunt's, Del Monte, and Heinz (denoted A, B, and C) and a private label brands (denoted D) after excluding minor brands. In terms of market share (measured in sales units), Hunt's, Del Monte, Heinz, and Private represent 13.9, 7.1, 72.7, and 6.3% of the market, respectively. Second, the average regular shelf-price for the four brands, across 13 stores and over 86 weeks, is 3.4, 3.8, 4.8, and 3.1 cents/oz, respectively. Third, three forms of promotions are used in the market: price discount, store display, and feature advertising. Based on the frequencies of all three forms of promotion for a given store and across stores, Heinz is found to be the most promoted brand, with the remaining brands receiving substantially less promotional support. Overall Hunt's and Del Monte are promoted at about the same level of frequency while Heinz is promoted about three times more often than either Hunt's or Del Monte. Private label has the least promotional frequency at about 20% of Hunt's or Del Monte's level.

Turning now to the analysis of the data, we begin by pointing out that suppose households consider all brands, then we would expect an average household purchase history to display brand share proportions that are similar to those at the aggregate level. However, if consumers have different consideration sets, then we would expect household purchase histories to be quite different across households. In order to examine this conjecture, we consider the composition of brands purchased by all 402 households to see which pattern dominates. Specifically, we compute the distribution of the number of brands purchased by each household. The results are presented in Fig. 3 for the 15 (2^{4-1}) possible choice sets. Note that only 10.5% of 402 households purchased all four brands while almost 36% of them purchased exclusively one brand (Heinz accounts for 33.8%). The rest can be grouped into those that selected between two brands (31.5%) and those that selected from among three brands (22%). Clearly, the percentage distribution across the possible choice sets is far from even. This evidence provides some partial support for the hypothesis of consideration set heterogeneity in this market.

3.2. Variables

The following covariates are used to model brand choice, conditioned on the consideration set and the household-specific parameter vector.

1. brand-specific indicator variable for each brand j;
2. logarithm of shelf price (cents/oz) for each brand j at purchase occasion t;
3. store display indicator for brand j at purchase occasion t;
4. feature advertising indicator for brand j at purchase occasion t;
5. past Purchase indicator for each brand j – equal to one if brand j was purchased by the ith household at the last occasion and 0 otherwise.

3.3. Estimate of consideration set distribution

Consider now the posterior distribution of the household-specific consideration vector C_i. From the simulation procedure we obtain 3000 vectors of $C^{(g)}$ where each element in $C^{(g)}$ represents the index of a particular household's consideration set. In order to present the results obtained, we focus our discussion on a randomly selected sub-set of 24 households and report (in Table 1)

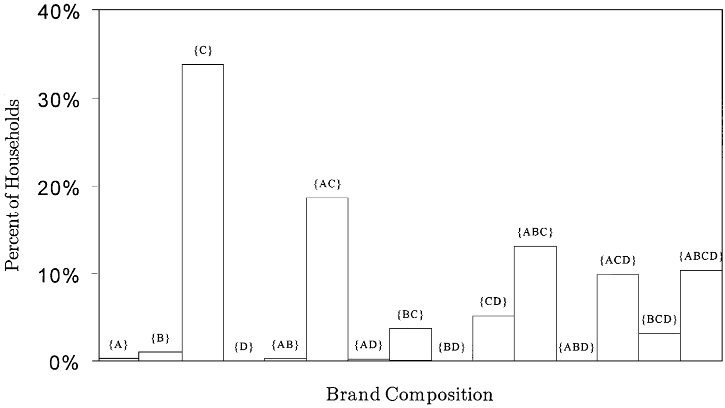

Fig. 3. Brand composition in purchase history of 402 households using Nielsen scanner panel data.

the frequencies of C_{is} and the household's purchase history. For example, the first row of Table 1 shows that household No. 54 (HH-54) made eight purchases and these were exclusively of brand A. To see how our prior information is revised by the data, note that under the prior each consideration set has equal probability. Given the observed history of purchases for this household, however, the posterior distribution of C_i is concentrated on the following subsets: $\{A\}$, $\{AB\}$, $\{AC\}$, $\{AD\}$, $\{ABC\}$, $\{ABD\}$, $\{ACD\}$, or $\{ABCD\}$. Moreover, because $C_{i1} = .753$, it is inferred that the most likely consideration set for this household is $\{A\}$. This is consistent with what may be expected from the purchase history of this household.

Now, consider the households that only purchased brand C (the second block of households in Table 1). Some interesting results are highlighted: (1) the probability mass on C_{i3} (corresponding to the subset $\{C\}$) is an increasing function of the number of purchases; and (2) the probability mass on C_{i3} is not necessarily the same for households with the same number of purchases. The former is the consequence of the Bayesian updating mechanism for p_i in which when the number of observations increases, one of the underlying Dirichlet parameters (α_s) will be assigned more weight. The latter phenomenon arises from the fact that the probability mass on each subset reflects the combined influence of the various components of the hierarchical model. Thus, for a household such as HH-170 in row 3, the model ascribes a low posterior probability on the subset $\{C\}$ because household choice was determined by the influence of the marketing mix variables and parameter heterogeneity and not because its consideration set was $\{C\}$. Further, validity for this connection may be obtained by taking the C-only households and running a simple regression of $\Pr(C_{i3}|y_i)$ against the proportion of purchases made on promotion (recall that brand C is heavily promoted). The regression results reported in Fig. 4 indicate an inverse

Table 1
Posterior distribution of consideration set (sample households)

HH-ID	Proportion $C(is)$ {A}	{B}	{C}	{D}	{AB}	{AC}	{AD}	{BC}	{BD}	{CD}	{ABC}	{ABD}	{ACD}	{BCD}	{$ABCD$}	First 10 purchases
54	0.753	0	0	0	0.130	0.021	0.053	0	0	0	0.007	0.025	0.006	0	0.004	*A A A A A A A*
52	0	0	0.437	0	0	0.084	0	0.191	0	0.111	0.048	0	0.038	0.066	0.025	*C C C C C*
170	0	0	0.171	0	0	0.131	0	0.138	0	0.128	0.115	0	0.112	0.110	0.095	*C C C C C*
332	0	0	0.463	0	0	0.037	0	0.211	0	0.139	0.023	0	0.017	0.093	0.017	*C C C C C C*
107	0	0	0.493	0	0	0.108	0	0.134	0	0.116	0.039	0	0.032	0.047	0.032	*C C C C C C C*
137	0	0	0.706	0	0	0.121	0	0.032	0	0.068	0.016	0	0.032	0.018	0.008	*C C C C C C C C*
233	0	0	0.635	0	0	0.026	0	0.127	0	0.119	0.013	0	0.010	0.065	0.006	*C C C C C C C C C*
272	0	0	0.696	0	0	0.101	0	0.030	0	0.098	0.008	0	0.047	0.017	0.005	*C C C C C C C C C C*
196	0	0	0	0	0	0.450	0	0	0	0	0.288	0	0.157	0	0.104	*C C C A*
339	0	0	0	0	0	0.809	0	0	0	0	0.045	0	0.129	0	0.016	*C A C A A A*
168	0	0	0	0	0	0.539	0	0	0	0	0.217	0	0.150	0	0.093	*C A C C C C C*
209	0	0	0	0	0	0.370	0	0	0	0	0.242	0	0.238	0	0.150	*C C C C A A C C*
247	0	0	0	0	0	0.528	0	0	0	0	0.211	0	0.193	0	0.068	*C C A C C C A C C*
283	0	0	0	0	0	0	0	0.379	0	0	0.230	0	0	0.244	0.147	*C C B C C C C*
275	0	0	0	0	0	0	0	0	0	0.721	0	0	0.176	0.084	0.019	*C C D C C D C*
388	0	0	0	0	0	0	0	0	0	0	0.793	0	0	0	0.207	*B C A C A*
9	0	0	0	0	0	0	0	0	0	0	0.840	0	0	0	0.160	*C C A B A A*
155	0	0	0	0	0	0	0	0	0	0	0.887	0	0	0	0.113	*C C C A B A C C*
396	0	0	0	0	0	0	0	0	0	0	0	0	0.674	0	0.326	*C A D A*
305	0	0	0	0	0	0	0	0	0	0	0	0	0.692	0	0.308	*C C D C A A A*
194	0	0	0	0	0	0	0	0	0	0	0	0	0	0.900	0.100	*C B D C D C*
240	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.00	*C B C B A C D*
376	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.00	*C D A A B C C*
131	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.00	*A B C C D C C A*

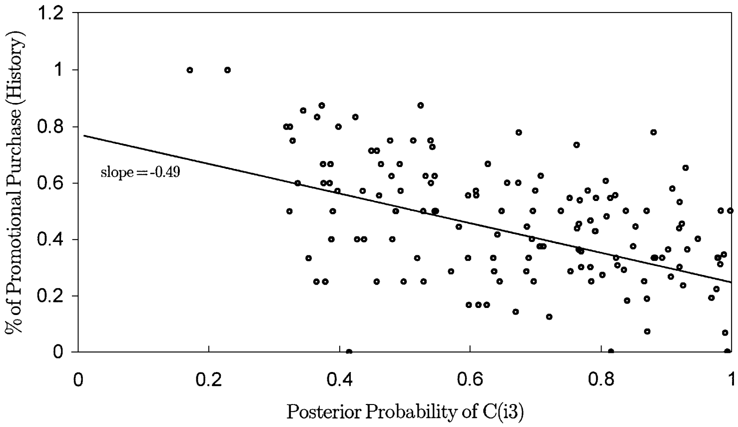

Fig. 4. Posterior probability of consideration set, $\Pr(C_{is}|y_i)$, vs. percentage of promotional purchases.

relationship, as expected. Thus, a single-brand purchase history may reflect the influence of promotional activity, not the influence of the consideration set.

Although the above discussion is phrased in terms of C_{is}, it is also useful to have a simple heuristic to sort households based on $\Pr(C_{is}|y_i)$. To demonstrate this segmentation scheme, we choose a cut-off probability value of 0.5. That is, if $\Pr(C_{is}|y_i)$ is greater than 0.5 for the ith household, then we assume s is its consideration set. Since $\sum_{s=1}^{S}\Pr(C_{is}|y_i) = 1$, one or none of the all possible subsets will be identified. On the basis of this criterion we are able to 'identify' the consideration set of 75% of the households, or 300 in total. In Fig. 5 we compare the two distributions – one based on the actual brand compositions from Fig. 3 and the second based on the posterior probability criterion. The shapes of these frequency distributions are similar. In addition, it should be noted that according to the above criterion, only 59% of those households that purchased only brand C are classified as households whose consideration set is $\{C\}$. In other words, among those who are observed to purchase only brand C (Heinz), only 59% of them we can infer that they are very likely to consider only brand C whereas the remaining 41% may consider multiple brands but ending up buying brand C because of deals.

3.4. Comparative assessment

We now provide a comparative assessment of the results from the full hierarchical model (MNL_CP) with several cases that have been fit in the literature. The special cases are

- Multinomial logit model with no parameter or consideration set heterogeneity (MNL).

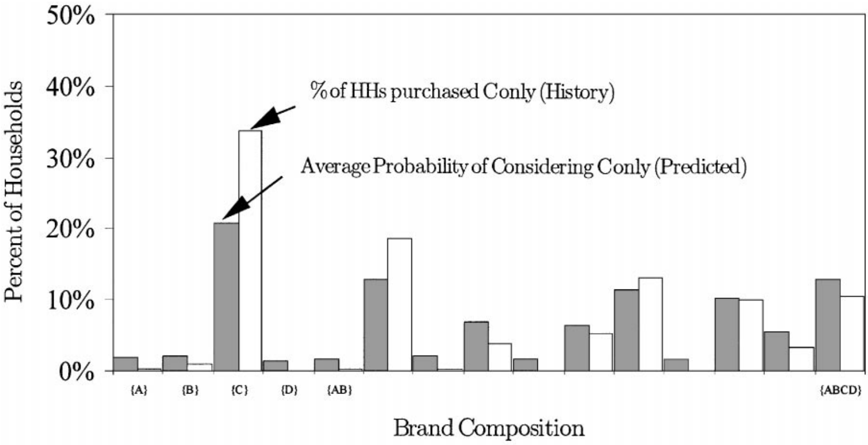

Fig. 5. Segmentation of households by consideration set using $\Pr(C_{is}|y_i) > 0.5$.

- Multinomial logit model with parameter heterogeneity (MNL_P).
- Multinomial logit model with consideration set heterogeneity (MNL_C).

Interestingly, each of the special cases can be estimated according to the simulation method developed in Section 2 by merely suppressing the component that is absent from the full hierarchical model. For example, to estimate the model MNL_C one merely has to delete the simulation step in which the random effect are generated. Software for estimating these models is available from the authors.

3.4.1. Estimates of the brand choice parameters

The posterior distribution of the fixed effects parameters β from the four models is summarized in Table 2. First, note that the signs of the estimates from these models are all consistent with earlier marketing brand choice studies. Second, in the MNL_C model, the coefficient estimates of $\bar{\beta}$ for the marketing mix variables are larger than those in the MNL model while those for the brand-specific and past purchase (i.e., brand re-enforcement) covariates are smaller. This finding is consistent with the previous studies of Stopher, 1980 and Swait and Ben-Akiva, 1986, the latter in the context of the binary logit model. To explain this interesting result in the context of our model, we rely on the fact that $E(\beta|y)$ (which we estimate by $\bar{\beta}$) can be expressed by the law of iterated expectations as

$$\int E(\beta|y, \{C_i\})\pi(C_1, \dots, C_n|y, \{p_i\})\pi(p_1, \dots, p_n|y)\, d\{C_i\}\, d\{p_i\},$$

Table 2
Posterior of brand choice parameter and random effect

Model Covariate	MNL Mean	Std. dev	MNL_C Mean	Std. dev	MNL_P Mean	Std. dev	MNL_CP Mean	Std. dev
Hunt's	0.487	0.081	0.070	0.085	0.973	0.212	0.296	0.164
Del Monte	0.034	0.089	− 0.150	0.111	− 0.159	0.235	− 0.238	0.177
Heinz	1.846	0.115	1.247	0.142	3.227	0.260	1.941	0.204
ln(Price)	− 1.260	0.192	− 1.911	0.250	− 1.445	0.398	− 2.308	0.326
Display	0.554	0.106	0.596	0.108	0.710	0.204	0.716	0.166
Feature	1.332	0.073	1.389	0.084	1.970	0.166	1.917	0.137
Past purchase	1.292	0.045	0.555	0.047	0.836	0.144	0.217	0.103
					Random Effect			
D(Hunt's)					2.616	0.451	1.021	0.257
D(Del Monte)					2.944	0.593	0.828	0.253
D(Heinz)					3.143	0.777	1.150	0.333
D(ln(Price))					4.368	1.685	1.764	0.924
D(Display)					1.400	0.294	0.841	0.270
D(Feature)					1.956	0.442	1.359	0.307
D(Last purchase)					1.794	0.397	0.517	0.105

a mixture of $E(\beta|y, \{C_i\})$ using the posterior distribution of the subsets and the associated probabilities. For those households that only bought brand C, for example, (roughly 34% of all households), and had a high posterior probability of the single consideration set $\{C\}$, marketing mix variables have little or no impact (by definition) and the contribution to $E(\beta|y, \{C_i\})$ is negligible. Thus, the posterior mean of β is automatically based on data from households that have multi-brand purchase histories (and are by definition, more sensitive to marketing mix variables) than from those that have single-brand purchase histories (and are less sensitive to marketing mix variables).

The standard logit model does not account for such non-responsive, one-brand consumers, and instead assumes that such consumers consider all four brands. As a result, the influence of the marketing mix covariates is understated in the standard logit model while those of the brand-specific dummies and last purchase variable are overstated (if the influence of the marketing variables declines, the unusual market share distribution seen in Fig. 3 as well as the strong brand reenforcement phenomenon from brand C buyers can only be captured if the brand constants increase). Similar logic can be advanced to explain why the coefficient estimates for the brand dummies and past purchase variables must be smaller in the MNL_C model. These results are resistant to the inclusion as parameter heterogeneity (see the results from the MNL_CP model).

Another interesting finding contained in Table 2 is that the importance of the past purchase variable (although statistically significant) diminishes with the introduction of consideration set heterogeneity. This is because a single-brand consideration set leads to persistence in same-brand purchases, thus eliminating the need for the past purchase indicator. Therefore, the latter variable is only important for individuals with multi-brand consideration sets.

3.4.2. Estimates of consideration set probabilities

We now consider the estimates of consideration set probabilities from the two models MNL_CP and MNL_C. In this connection, it is interesting to see whether the introduction of parameter heterogeneity to the latter model affects the estimates of the posterior consideration set probabilities $\Pr(C_{is}|y_i)$. To check the degree of robustness, we plot in Fig. 6, the distribution of $\Pr(C_{is}|y_i)$ from each model for each of the 402 households in the sample. It appears that the consideration set probabilities are quite similar. Simple regressions fail to reject the hypothesis that the two sets of estimates are identical.

3.4.3. Estimates of parameter heterogeneity

We now consider the estimates of parameter heterogeneity from the two models MNL_CP and MNL_P. The extent of heterogeneity is measured by the square root of the diagonal elements of D (denoted $D_{ii}^{1/2}$). The posterior sample of $D_{ii}^{1/2}$ is obtained from that of $\{D_{ii}^{(g)}\}$ by taking the point-wise square-root of

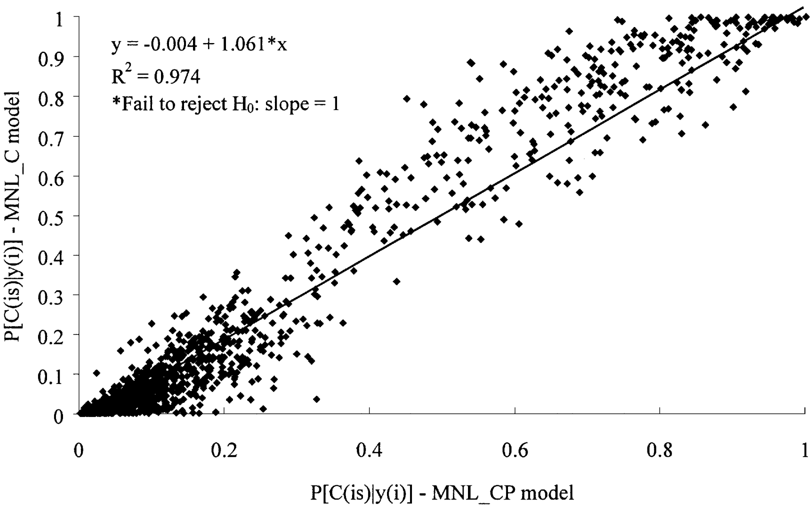

Fig. 6. Scatter plot and regression fit for $\Pr(C_{is}|y_i)$ from model MNL_C vs. $\Pr(C_{is}|y_i)$ from Model MNL_CP.

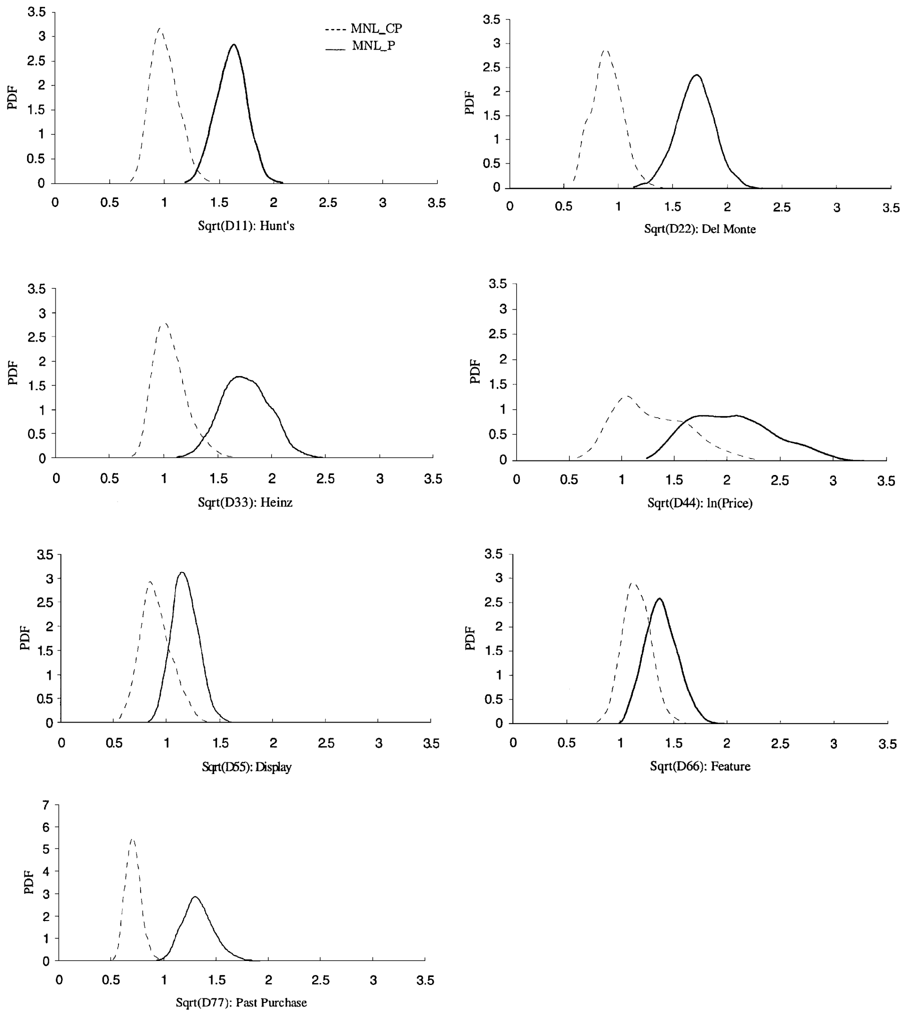

Fig. 7. Posterior distribution of squared root D(ii) using MCMC Algorithm in model MNL_P and model MNL_CP.

each draw. The marginal posterior distribution and the posterior moments of $D_{ii}^{1/2}$ are then obtained from the transformed sample. The results are reported in Table 2 and Fig. 7.

Both models indicate the presence of parameter heterogeneity, in line with previous findings (Allenby and Lenk, 1994). However, we obtain the very interesting new result that when one controls for variation in consideration set, the extent of parameter heterogeneity diminishes considerably, particularly for brand constants and past purchase. The trace of posterior mean matrix of D is 18.22 and 7.48, respectively, for the MNL_P model and the MNL_CP model. The reduction in magnitude ranges from 1.43 times (Feature) to 3.55 times (Del Monte) while the average reduction ratio is 2.55.

3.4.4. Estimates of the impact of marketing mix

The hierarchical model provides important information about the impact of marketing mix variables on consumer response and market share. These impacts are computed by evaluating the response during each iteration of the simulation and averaging the resulting values. The results are summarized in Table 3. The numbers in each row represent the percentage change in aggregate demand when brand j's price increases by 1%.

From the diagonal elements of the price sensitivity matrix in Table 3, we find that relative to the other brands, consumers are more sensitive to a change in the price of Heinz. This is an interesting finding because Heinz has the highest price in the market and also has the highest frequency of promotion. We conjecture that frequent promotions may have increased the awareness of Heinz's price and the expectation of future promotions. Consequently, consumers are more price sensitive for Heinz.

By viewing the pattern of all cross-sensitivities, brands can be divided into two tiers: Hunt's and Heinz in one tier and Del Monte and Private Label in another. Hunt's and Heinz appear to mainly compete with each other.

Table 4 provides the own/cross demand sensitivity with respect to feature. The results in this Table are quite similar to those in Table 3. In summary, we find (1) other things being equal, the biggest gain in market share to promotion occurs for Heinz; it also experiences the biggest reduction in market share when a competing brand is promoted; (2) Hunt's and Heinz compete against each other more than they compete with Del Monte and Private brand.

Though the sensitivity patterns are consistent across all models, the magnitudes are quite different. We find that the estimates from the MNL and MNL_P models are quite similar as are those from the MNL_C and the MNL_CP models. It appears that the effect of marketing variables is most pronounced when consideration set heterogeneity is controlled.

4. Methodological and substantive implications

The hierarchical consideration set-parameter heterogeneity model proposed and estimated in this paper provides a number of useful insights. We find that once consideration set heterogeneity is introduced, the brand specific intercepts and the importance of past purchase effects are lowered, and the effects of marketing mix variables are increased. Such a model, therefore, leads to

Table 3
Aggregate price sensitivity of demand

		Hunt's	Del Monte	Heniz	Private
Price sensitivity					
MNL	Hunt's	− 0.119	0.014	0.091	0.014
	Del Monte	0.014	− 0.071	0.049	0.008
	Heinz	0.091	0.049	− 0.188	0.048
	Private	0.014	0.008	0.048	− 0.069
MNL_C	Hunt's	− 0.154	0.018	0.117	0.019
	Del Monte	0.018	− 0.077	0.049	0.010
	Heinz	0.117	0.049	− 0.224	0.058
	Private	0.019	0.010	0.058	− 0.087
MNL_P	Hunt's	− 0.125	0.018	0.082	0.025
	Del Monte	0.018	− 0.065	0.034	0.013
	Heinz	0.082	0.034	− 0.172	0.055
	Private	0.025	0.013	0.055	− 0.094
MNL_CP	Hunt's	− 0.154	0.020	0.111	0.023
	Del Monte	0.020	− 0.078	0.047	0.012
	Heinz	0.111	0.047	− 0.219	0.062
	Private	0.023	0.012	0.062	− 0.097

Table 4
Aggregate feature sensitivity of demand

		Hunt's	Del Monte	Heniz	Private
Feature sensitivity					
MNL	Hunt's	0.124	− 0.015	− 0.095	− 0.014
	Del Monte	− 0.015	0.074	− 0.051	− 0.008
	Heinz	− 0.095	− 0.051	0.196	− 0.050
	Private	− 0.014	− 0.008	− 0.050	0.072
MNL_C	Hunt's	0.113	− 0.013	− 0.085	− 0.014
	Del Monte	− 0.013	0.056	− 0.036	− 0.007
	Heinz	− 0.085	− 0.036	0.164	− 0.043
	Private	− 0.014	− 0.007	− 0.043	0.064
MNL_P	Hunt's	0.119	− 0.016	− 0.087	− 0.016
	Del Monte	− 0.016	0.062	− 0.037	− 0.009
	Heinz	− 0.087	− 0.037	0.163	− 0.039
	Private	− 0.016	− 0.009	− 0.039	0.065
MNL_CP	Hunt's	0.108	− 0.013	− 0.083	− 0.013
	Del Monte	− 0.013	0.053	− 0.033	− 0.007
	Heinz	− 0.083	− 0.033	0.154	− 0.039
	Private	− 0.013	− 0.007	− 0.039	0.059

a different assessment of the relative importance of core product attributes and customer loyalty vs. marketing mix in explaining market share. Furthermore, we find (see Table 2) that a model that only allows for parameter heterogeneity (MNL_P) provides different conclusions than our new model. First, the fixed effect parameters (except for display and feature) are substantially different. This points to the fact that omitting consideration set heterogeneity understates the impact of marketing variables and overstates the impact of preferences, even when parameter heterogeneity is captured. Second, Fig. 6 shows that the results from the pure consideration set model (MNL_C) are fairly robust to the introduction (or elimination) of parameter heterogeneity.

The second set of insights relate to effectiveness of promotions. In general terms, a model with consideration set heterogeneity implies that promotions have a bigger impact. This is similar to the result documented by Chintagunta et al., (1991) for a model with parameter heterogeneity. In addition, from Table 3 it is clear that the response to promotion is predicated on the size and composition of the consideration set. Specifically, the increase in the choice probability (due to a promotion) for a brand in the consideration set depends positively on the size of the consideration set. This once again highlights the need for segmenting the market in order to assess the impact of promotions. Following the analyses in Siddarth et al., 1995 and Bronnenberg and Vanhonacker, 1996, one can conduct clout and vulnerability analysis by segmenting the households based on their consideration sets. This would identify the segments that are most likely to respond to more targeted promotional strategies.

Finally, by clustering households according to their C_{is} values, a manager can identify broad features of households that do or do not consider his brand in the brand choice decision. Analysis of the these segments might suggest specific differential marketing strategies, such as an extension of a product line, targeted sampling, targeted couponing, etc., directed at these segments.

5. Conclusions

This paper provides an approach for fitting brand choice models with parameter and consideration set heterogeneity. The model for consideration sets is non-parametric (unlike the parametric specifications in previous work) and involves the consideration of all possible subsets of the available brands. The resulting model is estimated by a practical Markov chain Monte Carlo method. Our methods can be applied to large dimensional models and the size of the model that can be fit is limited by hardware constraints, not computational complexity. These constraints will clearly become less severe in the near future. In the meantime, our non-parametric model of consideration set formation can serve as a useful benchmark against which other consideration set models may be evaluated.

Our empirical analysis yielded the following insights of relevance to marketing researchers:

- There is considerable consideration set heterogeneity among the members of the panel.
- A choice model that does not include consideration set heterogeneity yields predictably misleading coefficient estimates.
- When consideration set heterogeneity is incorporated, the impact of brand intercepts and loyalty variables on choice probabilities decreases while the impact of the marketing mix increases. This is true even when parameter heterogeneity is incorporated.
- The extent of parameter heterogeneity is much smaller when consideration set heterogeneity is included in the model.
- The promotional response of households depends on their consideration set even if the underlying choice parameters are identical.
- Segmenting the households by their consideration set may suggest refinements of the marketing mix and differential marketing strategies, directed at the relevant segments.

Appendix A

A.1. Simulation of p_i

The objective is to compute the density $\pi(p_i|y_i, C_{is})$ for each household in the sample. It is straightforward to see that this is just a revised Dirichlet distribution with the prior parameter α_s updated by the number of purchase occasions on which C_{is} is selected. Accordingly, the vector p_i is simulated from

$$D(\alpha_1, \dots, \alpha_s + T_i, \dots, \alpha_S).$$

A vector p_i from this distribution is simulated by letting

$$p_{i1} = \frac{z_1}{\sum_{j=1}^{S} z_j}, \dots, p_{iS} = \frac{z_S}{\sum_{j=1}^{S} z_j},$$

where $z_l \sim \text{Gamma}(\alpha_l, 1)(l \neq s)$ and $z_s \sim \text{Gamma}(\alpha_s + T_i, 1)$. For simulation of the Gamma distribution, see Chib and Greenberg, 1996.

A.2. Simulation of β

The objective is to calculate the density $\pi(\beta|y, \{b_i\}, \{C_i\})$. From Bayes theorem, the desired density is given by

$$\pi(\beta|y, \{C_i\}, \{b_i\}) \propto \phi(\beta|\beta_0, B_0) \times \prod_{i=1}^{n} \prod_{t=1}^{T_i} \Pr(y_{it} = j_t|\beta, C_{is}, b_i),$$

where the second term is the joint density of all the observations given the consideration sets of each household. It is clear that the conditional density of β does not belong to a recognizable family of distributions. Nonetheless, this density can be simulated quite easily by the Metropolis–Hastings algorithm, a Markov chain Monte Carlo algorithm (see Chib and Greenberg, 1995), by drawing candidates from a suitable density and then accepting or rejecting the candidates according to a specified probability. We adopt the random walk version of the M–H algorithm. In this version, given the current draw $\beta^{(g)}$ in the simulation, a candidate draw $\beta^{(c)}$ is obtained as

$$\beta^{(c)} = \beta^{(g)} + \mathrm{chol}(V)'u,$$

where $u \sim \mathrm{N}(0, I)$, V is a pre-specified matrix (in our implementation, we let V be proportional to the negative inverse of the observed information matrix of the multinomial logit likelihood, although any other suitable choice would also suffice) and *chol* denotes the choleski decomposition. Then, $\beta^{(c)}$ is accepted as the next item in the chain with probability given by

$$\min\left\{\frac{\pi(\beta^{(c)}|y, \{C_i\}, \{b_i\})}{\pi(\beta^{(g)}|y, \{C_i\}, \{b_i\})}, 1\right\},$$

where $\pi(\cdot|\cdot)$ is the density given above (knowledge of the normalizing constant is not required). If the candidate is rejected, we set $\beta^{(g+1)} = \beta^{(g)}$ and proceed with the simulation of the next item in the sequence.

A.3. Simulation of $\{b_i\}$ and D

The simulation of the random effects is analogous to that of β, requiring the use of a Metropolis–Hastings step. It is easy to see that the full conditional of b_i (for each i) is proportional to

$$\pi(b_i|y, C_i, \beta, D) \propto \phi(b_i|0, D) \times \prod_{t=1}^{T_i} \Pr(y_{it} = j_t|\beta, C_{is}, b_i)$$

and this cannot be simulated directly. Let the candidate draw be given by

$$b_i^{(c)} = b_i^{(g)} + \mathrm{chol}(D)'u,$$

where u is another standard normal draw. Then, $b_i^{(c)}$ is accepted as the next item with probability

$$\min\left\{\frac{\pi(b_i^{(c)}|y, C_i, \beta, D)}{\pi(b_i^{(g)}|y, C_i, \beta, D)}, 1\right\}.$$

If the candidate value is rejected, we set $b_i^{(g+1)}$ to be current value of b_i. Note that because the chain can generate repetitions, it is sometimes necessary to fine tune the increment variance D so that the acceptance rate is between 15% and 50%. A similar remark applies to the simulation of β and the matrix V used there.

Finally, we simulate D by first simulating D^{-1} from $D^{-1}|\{b_i\}$ and then taking the inverse of the simulated draw. This is because it can be shown through some simple calculations (suppressed here) that

$$D^{-1}|\{b_i\} \sim \text{Wishart}\left(v_0 + n, \left(R_0^{-1} + \sum_{i=1}^{n} b_i b_i'\right)^{-1}\right).$$

References

Alba, J.W., Chattaopadhyay, A., 1985. The effects of context and part-category cues on the recall of competing brands. Journal of Marketing Research 22, 340–349.

Albert, J., Chib, S., 1993. Bayesian analysis and binary and polychotomous response data. Journal of the American Statistical Association 88, 669–679.

Andrew, R.L., Srinivasan, T.C., 1995. Studying consideration effects in empirical choice models using scanner panel data. Journal of Marketing Research 32, 30–41.

Belonax, J.A., Mittelstaedt, R.A., 1978. Evoked set size as a function of number of choice criterion and information variability. Advances in Consumer Research, Association for Consumer Research, Ann Arbor, MI.

Ben-Akiva, M., Boccara, B., 1995. Discrete choice models with latent choice sets. International Journal of Research in Marketing 12, 9–24.

Bronnenberg, B.J., Vanhonacker, W.R., 1995. Limited choice sets, local price response, and limited measures of price competition. Journal of Marketing Research 33, 163–173.

Chib, S., Greenberg, E., 1995. Understanding the Metropolis–Hastings algorithm. The American Statistician 49, 327–335.

Chib, S., Greenberg, E., 1996. Markov chain Monte Carlo simulation methods in econometrics. Econometric Theory 12, 409–431.

Chintagunta, P.K., Jain, D.C., Vilcassim, N.J., 1991. Investigating heterogeneity in brand preferences in logit models for panel data. Journal of Marketing Research 28, 417–429.

Gelfand, A.E., Smith, A.F.M., 1990. Sampling-based approaches to calculating marginal densities. Journal of the American Statistical Association 85, 398–409.

Gensch, D.H., Soofi, E.S., 1995. Information-theoretic estimation of individual consideration set. International Journal of Research in Marketing 12, 25–38.

Gonul, F., Srinivasan, K., 1993. Modeling multiple sources of heterogeneity in multinomial logit models: methodological and managerial issues. Marketing Science 12, 213–229.

Guadagni, P.M., Little, J.D.C., 1983. A logit model of brand choice calibrated on scanner data. Marketing Science 2, 203–238.

Hauser, J.R., Wernerfelt, B., 1990. An evaluation cost model of consideration sets. Journal of Consumer Research 16, 393–408.

International Journal of Research in Marketing, Vol. 1, 12, North-Holland, Amsterdam.

Kamakura, W.A., Russell, G.J., 1989. A probabilistic choice model for market segmentation and elasticity structure. Journal of Marketing Research 26, 379–390.

Parkinson, T.L., Reilly, M., 1979. An information processing approach to evoked set formation. Advances in Consumer Research 6, 227–231.

Roberts, J.H., Lattin, J.M., 1991. Development and testing of a model of consideration set composition. Journal of Marketing Research 28, 429–440.

Shocker, A.D., Ben-Akiva, M., Boccara, B., Nedungadi, P., 1991. Consideration set influences on consumer decision making and choice: Issues, models, and suggestions. Marketing Letters 2, 181–197.

Siddarth, S., Bucklin, R.E., Morrison, D.G., 1995. Making the cut: Modeling and analyzing choice set restriction in scanner panel data. Journal of Marketing Research 32, 255–266.

Stopher, P., 1980. Captivity and choice in travel-behavior models. Transportation Journal of ASCE 106, 427–435.

Swait, J., Ben-Akiva, M., 1986. Analysis of the effects of captivity on travel time and cost elasticities. Behavioral Research for Transport Theory 119–134.

Swait, J., Ben-Akiva, M., 1987. Empirical test of a constrained discrete choice model: Mode choice in Sao-Paulo. Brazil. Transportation Research B 21B, 103–115.

Tanner, M.A., Wong, W.H., 1987. The calculation of posterior distributions by data augmentation. Journal of the American Statistical Association 82, 528–549.