TIMOTHY J. GILBRIDE, GREG M. ALLENBY, and JEFF D. BRAZELL*

Preference heterogeneity is a major research stream in marketing aimed at quantifying and understanding the diversity of demand for product attributes and attribute levels. In experimental settings, in which consumers are presented with simple descriptions of product offerings, continuous distributions of heterogeneity, such as the multivariate normal, provide a useful representation of preference. However, in more complex cases in which respondents have value for only a few of the benefits associated with an offering or cognitive constraints that result in selective attention to a subset of the information available, continuous distributions of heterogeneity do not reflect the possibility that a subset of the variables has nonzero effect sizes for different respondents. Identifying which attributes are used in a brand choice decision is closely related to the statistical procedure of variable selection. This article extends variable selection methods to accommodate heterogeneity across consumers and data contexts, conditions frequently encountered in marketing studies. The authors apply the methods to a discrete-choice conjoint study in which data are collected in both full-profile and partial-profile formats.

Models for Heterogeneous Variable Selection

Marketing models of choice regard the utility of an offering as a function of product attributes and benefits and the importance of these features to the consumer. On the one hand, the potential number of attributes and benefits associated with an offering can be large. On the other hand, consumers may find value in only a small number. Consumer heterogeneity arising from variation in motivations, expertise, and/or perception leads to differences in the preference for attributes and benefits. Cognitive constraints, the use of simplifying heuristics, and elements of the choice task can further lead to consumers focusing on only a small subset of the potential attributes. Although marketing models provide a measure of importance of attributes and benefits, current methods of analysis are not sufficiently sensitive to the possibility that respondents may value or attend to only a few features of an offering.

For example, consider alternative forms of conjoint analysis. Data collection methods range from pairwise trade-off tasks, in which respondents choose between two offerings that differ on just two attributes, to full-profile analysis, in which respondents are presented with completely described offerings. The former method is easy to perform and can be administered over the telephone, whereas the latter method is more complicated because of the rich description of the offering and often requires written description. Pairwise trade-off data more likely involve careful consideration of all the attributes presented to respondents, which may not reflect their decision process in the marketplace. Similarly, it is questionable whether the attributes used in response to full-profile conjoint analysis, or any experimental setting, are all used by consumers in the course of their everyday lives.

The importance of product attributes is driven by factors such as consumers' motivations and beliefs about the product efficacy (Dickson 1982; Fennell 1978; Yang, Allenby, and Fennell 2002). Consumers who lack a particular motivation (e.g., bad breath) or believe that current offerings are deficient at delivering relief will not value particular attributes and benefits (e.g., breath freshening) of a brand. In addition, cognitive constraints may prevent consumers from using all the attributes presented in a brand choice situation (see Bettman, Luce, and Payne 1998). Consumers' limited resources (e.g., memory capacity) imply that attention is selective and that some attributes do not factor into the decision.

Individual-level estimates of choice model parameters can point to the attributes and benefits desired by the respondent if sufficient data are available. However, marketing data are characterized by severe data limitations at the individual level. It is rare to have more than two dozen

---

*Timothy J. Gilbride is an assistant professor, Department of Marketing, Mendoza College of Business, University of Notre Dame (e-mail: tgilbrid@nd.edu). Greg M. Allenby is a professor and Helen C. Kurtz Chair in Marketing, Department of Marketing and Logistics, Fisher College of Business, Ohio State University (e-mail: allenby.1@osu.edu). Jeff D. Brazell is President and Chief Executive Officer of The Modellers, LLC, Salt Lake City (e-mail: jeffb@themodellers.com).

ISSN: 0022-2437 (print), 1547-7193 (electronic)

Journal of Marketing Research
Vol. XLIII (August 2006), 420–430

observations per respondent in conjoint analysis and applied demand modeling. Although this data limitation is often addressed with heterogeneity distributions, in which information is borrowed across respondents, the resulting shrinkage to the average effect size is not sensitive to the reality that respondents may have nonzero utility or focus only on a small number of product features.

This article develops statistical models for determining which respondents are using which attributes in a brand choice decision. Although the statistical model cannot tell whether consumers ignore a product attribute because it has no intrinsic value or as a result of cognitive constraints, it provides the tools for investigating these hypotheses. Our research extends the notion of heterogeneity into the realm of qualitative differences, in which some attributes are used in the product choice decision and others are not used. Identifying which attributes are used in a brand choice decision is closely related to the statistical procedure of variable selection. In many statistical analyses, there are a large number of potential predictor variables, and there is uncertainty about which variables are redundant or irrelevant. Variable selection attempts to identify the best subset of variables to include in the model. However, the statistics literature has focused on variable selection at the aggregate level (i.e., selecting the best subset of variables for the whole sample under study).

This article extends Bayesian variable selection methods in three ways that are relevant to marketing researchers. First, George and McCulloch's (1993) stochastic search variable selection (SSVS) procedure for aggregate linear models is extended to individual-level, random coefficient models. The "heterogeneous variable selection model" uses a distribution of heterogeneity in which each parameter comes from a distribution with mass concentrated at zero and away from zero. This represents the qualitative situation in which different consumers use different subsets of product attributes. Second, we derive and provide a practical methodology for applying the heterogeneous variable selection model to discrete choice data. Third, we propose a model for determining whether the choice context results in a person attending to different variables. The "contextual variable selection model" allows an individual to use different subsets of product attributes in different choice situations. These models improve on existing methods by providing a flexible framework for investigating variable selection at the individual level while avoiding the need to evaluate marginal posterior densities during estimation.

The article proceeds as follows: We provide a basic background of George and McCulloch's (1993) SSVS model, and then we introduce the heterogeneous and contextual models. We discuss estimation procedures, and full details of the estimation algorithms appear in the appendix. We then introduce an empirical application that involves a discrete choice conjoint study. The proposed models provide new insights into the nature of demand, compared with models that assume that all respondents use all product attributes. The article concludes with a discussion and outlines avenues for further research.

HETEROGENEOUS VARIABLE SELECTION

The models introduced in this section address three types of variable selection issues that are pertinent to marketing researchers. First, we develop variable selection models to incorporate heterogeneous selection in which respondents differ in the variables they use. This model makes use of the panel structure that is found in many marketing data sets. Second, we modify the model for discrete choice data. Third, we introduce a contextual selection model that allows respondents to use different variables in different choice contexts.

Heterogeneous Variable Selection Model

The heterogeneous variable selection model is an extension of the Bayesian variable selection procedure that George and McCulloch (1993, 1997) suggest. This model has a distribution of heterogeneity with mass concentrated either at zero or away from zero for each parameter. For marketing managers, knowing who uses which product attributes offers greater opportunity for customizing products and targeting analysis. In this subsection, we give the rudiments of the SSVS model for aggregate-level regression, show how heterogeneity can be introduced into this model, and then derive the heterogeneous variable selection model for discrete choice data.

Consider the normal regression model with T observations and J predictor variables:

$$y|\beta, \sigma \sim \text{Normal}(X\beta, \sigma^2 I) \tag{1}$$

with priors

$$\sigma^2 \sim IG(\kappa, \kappa\psi) \text{ and} \tag{2}$$

$$\beta \sim \text{Normal}(0, D_\tau I D_\tau), \tag{3}$$

where I is the identity matrix, IG(*, *) is the inverse gamma distribution, and Normal(*, *) is the multivariate normal distribution. The variable selection problem is to choose which columns of X to include in the model or, alternatively, which values of β to set to 0. In the SSVS setup, variable selection is accomplished through the J × J matrix D_τ, formed as diag[τ]. Here, τ is a vector of length J, and $\tau_j \in \{c, d\}$, where c is some small constant and d is some large constant set by the researcher. When $\tau_j = c$, the prior for β_j is concentrated in the area around 0, corresponding to the situation in which predictor j is excluded from the model. When $\tau_j = d$, a diffuse prior is used.

If we let the prior probability that variable j is selected, $\Pr(\tau_j = d) = ½$, George and McCulloch (1993) propose the following Gibbs sampler to draw from the posterior distributions:

(i) $[\beta|\tau, \sigma^2, \text{data}] = \text{Normal}\{[\sigma^{-2}X'X + (D_\tau I D_\tau)^{-1}]^{-1}\sigma^{-2}X'y,$

$$[\sigma^{-2}X'X + (D_\tau I D_\tau)^{-1}]^{-1}\};$$

(ii) $[\sigma^2|\beta, \text{data}] = \text{Inverted Gamma}\left(\dfrac{n+\kappa}{2}, \dfrac{|Y - X'\beta|^2 + \kappa\psi}{2}\right)$; and

(iii) $[\tau_j|\beta, \tau_{-j}] = \Pr(\tau_j = d)$, which is Bernoulli with the probability

$$\frac{[\beta|\tau_j = d, \text{data}]}{[\beta|\tau_j = d, \text{data}] + [\beta|\tau_j = c, \text{data}]},$$

where the bracket notation "[ ]" following each step number denotes a distribution and the fraction in Step iii involves the evaluation of multivariate normal densities with different covariance matrices, $D_\tau I D_\tau$, where τ = d, c. Note that this method does not set any β_j exactly to zero but draws it from a distribution with mass concentrated at 0, as determined by the value of τ. Analysis of the posterior distribution of τ indicates the most promising subset of variables to include in the model. George and McCulloch (1993, 1997) and George (2000) provide details on the derivation of this model, alternative approaches and estimation algorithms, and a guide to recent research on Bayesian variable selection methods.

We incorporate heterogeneity into the variable selection model by specifying the random-effects distribution as

$$\beta_h \sim \text{Normal}(C_{\tau h}\bar{\beta}, C_{\tau h}V_\beta C_{\tau h}), \tag{4}$$

where h indexes the respondent and $C_{\tau h}$ is a J × J matrix, formed as diag$[\tau_h]$ and $\tau_{hj} \in \{c, 1\}$, where c is some small constant set small enough so that $C_{\tau h}\bar{\beta}$ is near zero when $\tau = c$. The advantage of this specification is that it leads to standard Bayesian updating formulas conditional on τ because $C_{\tau h}^{-1}\beta_h \sim \text{Normal}(\bar{\beta}, V_\beta)$. Heterogeneity in τ_{hj} is accommodated by assuming $\tau_{hj} = 1$ with probability θ_j, and it equals c with probability $1 - \theta_j$, with prior probability $\pi(\theta_j) \sim \text{Beta}(a, b)$. As is standard in Bayesian analysis of the linear regression model, the covariance matrix V_β is specified with an inverted Wishart prior, and the prior for $\bar{\beta}$ is assumed to be distributed normal.

Standard Bayesian methods can be used to estimate the model. Steps i–iii are modified in the obvious ways; for example,

$$\begin{aligned}&[\beta_h\,|\sigma^2, \tau_h, \bar{\beta}, V_\beta, \text{data}] =\\ &\text{Normal}\{b, [\sigma^{-2}X_h{}'X_h + (C_{\tau h}V_\beta C_{\tau h})^{-1}]^{-1}\},\end{aligned} \tag{i′}$$

where $b = [\sigma^{-2}X_h{}'X_h + (C_{\tau h}V_\beta C_{\tau h})^{-1}]^{-1}\ [\sigma^{-2}X_h{}'y_h + (C_{\tau h}V_\beta C_{\tau h})^{-1}C_{\tau h}\bar{\beta}]$, with additional steps added and $\beta_h^* = C_{\tau h}^{-1}\beta_h$ used to draw $\bar{\beta}$ and V_β.

The matrix $C_{\tau h}$ permits inferences on which variables are being used by each individual, and it simplifies the process of obtaining draws of the hyperparameters $\bar{\beta}$ and V_β. However, note that these parameters represent different quantities between variable selection and standard models. In standard hierarchical models, $\bar{\beta}$ and V_β fully describe the distribution of heterogeneity; in heterogeneous variable selection models, θ_j (the proportion of respondents selecting variable j) is also needed to understand heterogeneity.

Marketing research studies frequently involve discrete choice data from either experimental studies, such as conjoint, or natural settings, such as supermarket scanner panel data. The model is now derived for multinomial discrete choice data. Let $y_{hik} = 1$ if person h selects brand i on choice occasion k. If z_{hik} represents the latent underlying level of utility for y_{hik}, then $y_{hik} = 1$ if $z_{hik} > z_{hnk}$ for all N alternatives in the choice set. As is common in marketing studies, latent utility is represented as

$$z_{hik} = \sum_{j=1}^{J}\beta_{hj}x_{hijk} + \varepsilon_{hik}, \tag{5}$$

where x_{hijk} represents the value of variable (or product attribute) j for brand i facing person h on choice occasion k. If $\varepsilon \sim \text{EV}(0, 1)$, the standard multinomial logit choice probability results.

Estimating the discrete choice model specified in Equation 5 requires that attention is paid to drawing the individual-level τ_h and β_h. In heterogeneous multinomial logit models, the estimates of β_h are obtained by sampling from $[\beta_h|\bar{\beta}, V_\beta, \text{data}]$, typically using a random walk Metropolis–Hastings (M–H) algorithm. However, for the heterogeneous variable selection model, the variable selection parameter τ_h must be added to the conditioning arguments: $[\beta_h|\tau_h, \bar{\beta}, V_\beta, \text{data}]$. As a practical matter, when the chain enters the state $\tau_{hj} = c$, it may become "stuck" and navigate the posterior space very slowly. At the extreme, in a full conditional setup, if $c = 0$, the chain becomes reducible; at $c = 0$, the conditional distribution $\beta_{hj}|\tau_{hj}$ is degenerate at $\beta_{hj} = 0$. Sampling from the posterior distribution $[\tau_h|\beta_h, \theta, \bar{\beta}, V_\beta]$ presents similar difficulties.

To ensure proper mixing of the Markov chain Monte Carlo (MCMC) algorithm, an independence chain M–H algorithm is proposed, and β_h and τ_h are drawn together. Random walk M–H algorithms use the current value of a parameter and randomly perturb it to obtain new candidate values to evaluate. In contrast, independence chain algorithms draw candidates without regard to the current values. As Chib and Greenberg (1995) suggest, we use the prior distribution—in this case, the distribution of heterogeneity of β_h and τ_h—to generate candidates. The prior $\pi(\beta_h, \tau_h|\bar{\beta}, V_\beta, \theta)$ is formed as $\pi(\beta_h|\tau_h, \bar{\beta}, V_\beta) \times \pi(\tau_h|\theta)$. Moves to new values of β_h and τ_h are evaluated jointly in the acceptance probability of the M–H algorithm. Values from the posterior distribution $[\beta_{hj}, \tau_{hj}|\beta_{h-j}, \tau_{h-j}, \theta, \bar{\beta}, V_\beta, \text{data}]$ are drawn for each j and h as follows:

1. Set $\tau_{hj}^{(n)} = 1$ with probability θ_j;
2. Draw candidate $\beta_{hj}^{(n)}$ from $\text{Normal}(C_{\tau h}^{(n)}\bar{\beta}, C_{\tau h}^{(n)}V_\beta C_{\tau h}^{(n)})$, where $C_{\tau h}^{(n)}$ is formed with $\tau_{hj}^{(n)}$; and
3. Accept $\tau_{hj}^{(n)}$ and $\beta_{hj}^{(n)}$ with the following probability:

$$\Pr(\text{accept}) = \min\left[\frac{L_h\left(\beta_{hj}^{(n)}, \tau_{hj}^{(n)}\right)}{L_h\left(\beta_{hj}^{(o)}, \tau_{hj}^{(o)}\right)}, 1\right],$$

where $\tau_{hj}^{(o)}$ and $\beta_{hj}^{(o)}$ are the previous values of τ_{hj} and β_{hj} and $L_h(*)$ is the likelihood of the data for respondent h.

Note that because the prior is used to generate the candidate, it is not used to evaluate the acceptance probability. Full details on the estimation algorithm appear in the Appendix.

This heterogeneous variable selection model differs from George and McCulloch's (1993, 1997) SSVS models in two important respects for marketing researchers. First, variable selection is conducted at the individual level; second, the current model is developed for discrete choice data. Alternative approaches to conducting aggregate-level variable selection have been proposed. Raftery, Madigan, and Hoeting (1997) introduce a model in which $c = 0$, and a special sampling "neighborhood" is defined and searched in each step of the sampler. Geweke (1996) proposes a model in which β_j and τ_j are drawn together, but it requires integrating out β_j from the appropriate distribution (for an application to aggregate-level logistic regression, see Chen, Ibrahim, and Yiannoutsos 1999). Green (1995) and Phillips and Smith (1995) offer theoretical representations and practical algorithms in the most general cases of model selection. In these cases, it is usually necessary to calculate normalizing constants in the posterior distributions in assessing the probability of moving from one model to another in the sampler. The models we propose herein conduct variable selection at the individual level using discrete choice data, and they avoid the evaluation of complex marginal posterior distributions.

Contextual Variable Selection Model

Many studies in marketing collect information across different contexts. Data may include responses from experiments, surveys, and actual brand choice, each with different response formats and the possibility that respondents use different sets of variables. Similar to the heterogeneous variable selection model, the contextual variable selection model provides individual-level estimates of attribute weights (β_h values) and allows the product attributes used to be individual specific. In the heterogeneous variable selection model, a respondent uses the same subset of product attributes for all choice tasks. In the contextual variable selection model, respondents may use different subsets of variables in different choice tasks. The data are pooled across contexts, and the model identifies which variables are used in each setting.

The contextual variable selection model assumes that two sets of data are available to the researcher or that the researcher can partition the data into different choice contexts. Let y_h^a represent the vector of choice data for household h in the choice context A, and let y_h^b represent choices in the other context. Latent utility is represented as

$$\text{(6a)} \qquad z_{hik}^a = \sum_{j=1}^{J} \beta_{hj}\gamma_{hj}x_{hijk} + \varepsilon_{hik}, \text{ and}$$

$$\text{(6b)} \qquad z_{hik}^b = \sum_{j=1}^{J} \beta_{hj}\lambda_{hj}x_{hijk} + \varepsilon_{hik},$$

where the ε are distributed as EV(0, 1). In this model, γ_{hj} and λ_{hj} are scalars equal to 0 or 1 and are distributed Bernoulli(ϕ_j) and Bernoulli(θ_j) across individuals. If $\gamma_{hj} = 0$, individual h does not use attribute j in the first choice context. If $\lambda_{hj} = 0$, individual h does not use attribute j in the second choice context. The proportion of respondents using attribute j in each data set is ϕ_j and θ_j, respectively.

Variable selection in the contextual model is handled through both the distribution of heterogeneity and the indicator variables in the representation of latent utility. The β_h are again assumed to be distributed Normal($C_{\tau h}\bar{\beta}$, $C_{\tau h}V_\beta C_{\tau h}$), where $C_{\tau h}$ is a J × J matrix formed as diag[τ_h]. Here, τ_h is a vector of length J, and $\tau_{hj} \in \{c, 1\}$, where c is some small constant. The value of τ_{hj} is determined by the following deterministic relationship:

$$\text{(7)} \qquad \tau_{hj} = 1 \text{ if } \gamma_{hj} + \lambda_{hj} \geq 1; \text{ otherwise, } \tau_{hj} = c.$$

If variable j is used in either the first or the second choice context by respondent j, the prior is centered at $\bar{\beta}_j$. If respondent h does not use variable j in either choice context, β_{hj} is drawn from a distribution with the mean and variance for item j set very close to 0.

Estimating the model is analogous to estimating the heterogeneous variable selection model. An independence chain M–H algorithm is used, and the candidate generating distribution $\pi(\beta_h, \gamma_h, \lambda_h|\bar{\beta}, V_\beta, \phi, \theta)$ is formed as $\pi(\beta_h|\gamma_h, \lambda_h, \bar{\beta}, V\beta) \times \pi(\gamma_h|\phi) \times \pi(\lambda_h|\theta)$. Moves to new values of β_h, γ_h, and λ_h are evaluated jointly in the acceptance probability of the M–H algorithm. The algorithm is similar to the one outlined previously, and full details are in the Appendix.

Qualitative shifts in the individual-level decision strategy are represented by the parameters γ and λ and their relationship to τ. Modeling these individual-level effects is possible through the distribution of heterogeneity for the parameters (e.g., ϕ and θ) and letting the MCMC method integrate over possible decision strategies for each individual. This is analogous to Gilbride and Allenby's (2004) use of parameters to operationalize individual-level conjunctive, disjunctive, or compensatory screening rules in a discrete choice model.

Different models and different estimation algorithms result from specifying ϕ_j and θ_j:

- $\phi_j = \theta_j = 1$ implies that $\gamma_{hj} = \lambda_{hj} = 1$ for all h and j, and the standard heterogeneity model with no variable selection results.
- $\phi_j = \theta_j$ implies that $\gamma_{hj} = \lambda_{hj}$ for all h and j, and this is equivalent to the heterogeneous variable selection model we discussed previously.
- $\phi_j = 1$, and no restrictions on θ_j implies that respondents use all the variables in one choice context but only a subset in the other choice context.
- ϕ_j and θ_j unrestricted implies that different subsets of variables are used in each choice context.

We discuss estimation algorithms for this model and another model with restrictions implied by our empirical example in the next section.

We conducted studies with simulated data sets to investigate the efficacy and properties of the MCMC algorithms. (A full report on these studies is available on request.) The simulation results show that the models are identified, the true parameters can be recovered to within sampling error, and the chains converge. Although the precision of individual-level estimates in a hierarchical model will always be data dependent, it is instructive to consider the recovery of these parameters in the simulated data. The simulated data set for the heterogeneous variable selection model with a continuous dependent variable included 100 individuals, with 20 observations per person, and five explanatory variables. For the discrete choice heterogeneous variable selection and contextual variable selection models, the data consisted of 500 individuals, with 20 observations each; choice sets consisted of four alternatives and five explanatory variables.

Table 1 summarizes the recovery of individual-level parameters. For τ, λ, and γ, the numbers in the table can be interpreted as a "hit rate." For example, for each observation in the posterior distribution, we record whether the sampled value of $\tilde{\tau}_{hj}$ is equal to the actual value used in the simulation τ_{hj}. We average these results across a sample of draws from the posterior for all variables and all individuals. Recovery of these latent variables is good, ranging from 72% to 97%, indicating that the model is capable of identifying which respondents use which variables. These recovery statistics correspond to all variables and all respondents. We conducted a similar analysis at the individual level to determine the proportion of the true β_{hj} in the sample that are contained in the 95% posterior highest density of the posterior distribution. The analysis indicates that the reported intervals are accurate. Moreover, we find that though true values of β_{hj} close to zero, but not equal to zero, are more likely to be misclassified, the 95% highest posterior density still includes the true value with a high probability.

Table 1

RECOVERY OF INDIVIDUAL-LEVEL PARAMETERS FROM SIMULATED DATA

	Probability That β_{hj} *Is in 95% hpd of* $\tilde{\beta}_{hj}$[a]	$p(\tilde{\tau}_{hj} = \tau_{hj})$	$p(\tilde{\gamma}_{hj} = \gamma_{hj})$	$p(\tilde{\lambda}_{hj} = \lambda_{hj})$
HVS/continuous	.946	.970		
HVS/discrete	.949	.797		
CVS	.923	.871	.740	.723

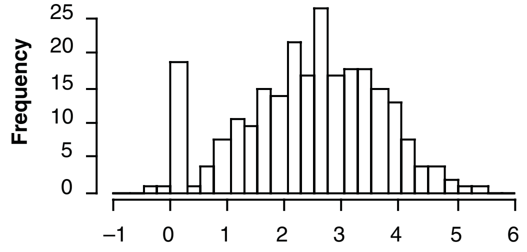
[a]This analysis includes $\beta_{hj} \approx 0$ and β_{hj} away from 0; hpd = the highest posterior density.

Notes: Parameters with ~ indicate draws from the posterior distribution, parameters without ~ indicate the actual value used in the simulation. HVS = heterogeneous variable selection, and CVS = contextual variable selection.

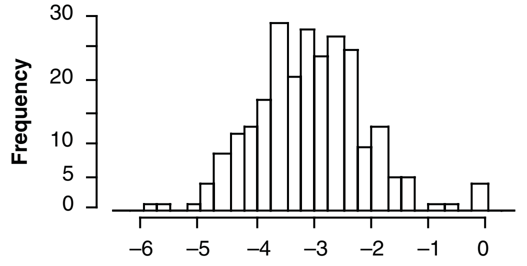
Figure 1

POSTERIOR DISTRIBUTIONS OF SELECTED β_{hj} SIMULATED DATA: DISCRETE HETEROGENEOUS VARIABLE SELECTION MODEL

True Value = 2.37



True Value = –3.65



True Value = .00

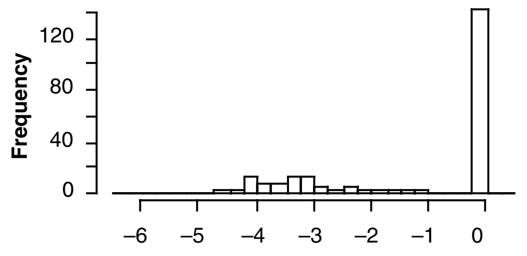


Figure 1 shows the posterior distribution of selected β_{hj} for one individual from the discrete choice heterogeneous variable selection simulation. The posterior distributions are not normal or symmetric, which complicates the choice of a summary statistic. A mass at zero appears in each of the distributions, corresponding to the discrete/continuous nature of the random-effects distribution. The distribution displayed at the top of Figure 1 shows overlap of the continuous distribution and zero, and the bottom two posterior distributions have components that overlap to a lesser extent. The true values of the coefficients used to generate the data in the simulation appear above each distribution.

If the analyst's objective is to classify an individual as either using or not using an attribute (a 0/1 loss function), the mode of the posterior distribution should be used. In these simulated examples, focusing on the mode (the highest spike) would lead to the correct conclusion. In other situations, more complicated loss functions related to market shares and profitability projections are appropriate. In these instances, the analyst should integrate over the entire posterior distribution of all parameters to form inferences.

Heterogeneous variable selection models incorporate a flexible distribution of heterogeneity, which explicitly allows for the possibility that a respondent may be using only a subset of product attributes. The individual-level posterior distributions can be used to determine who is using which attributes. In data sets with more variables and fewer observations per respondent, the posterior modes are likely to be less pronounced than those we illustrate here with simulated data. However, this uncertainty is incorporated into marketing analyses by integrating over the full posterior distribution of all parameters when calculating quantities such as predicted choice probabilities. In the next section, we show that models with these flexible distributions of heterogeneity have better predictive accuracy.

EMPIRICAL APPLICATION

Data

We illustrate the heterogeneous variable selection and contextual variable selection models with a discrete choice conjoint study. The study is sponsored by a product manufacturer that identified a competitive set of three brands and 15 additional product attributes of interest. Because of the proprietary nature of the data, we do not reveal the product, brands, and the majority of product attributes. We screened the participants recruited for this study to ensure that they were suitable prospects for the product. A total of 111 respondents participated in the study. We administered the questionnaire with personal computers.

Respondents evaluated 19 buying scenarios. Each scenario included three products, one for each brand name. Of the 19 buying scenarios, 10 are "full-profile" tasks, in which each alternative is described on all 16 product attributes, including the brand name, and 9 are "partial-profile" tasks, in which each alternative is described on 9 attributes, including the brand name. The product attributes selected for each partial-profile task followed an experimental design that ensured that the partial-profile tasks alone could be used to measure partworths. In the partial-profile task, we instructed respondents to assume that the alternatives were alike on all the omitted product attributes. Half of the respondents saw the partial-profile followed by the full-profile scenarios; we reversed the order for the other half of the respondents. In each buying scenario, we asked respondents to indicate which product they "would be most likely to choose." Thus, the data available for analysis represent multinomial outcomes across two different choice contexts.

Table 2 presents the product attributes, levels, indicator variables, and the orthogonal coding scheme used in the study. The coding scheme results in 28 β_{hj} parameters per individual to estimate from the data. We obtain partworths by multiplying and adding the appropriate β_{hj} values across the listed coefficients; the orthogonality constraint makes the partworths sum to zero across the levels of the attribute. The unique coding scheme facilitates both partial-profile analysis and variable selection. When an attribute is not included in a partial-profile choice set, the values of the indicator variables are set to zero. Similarly, in the variable selection models, the appropriate β_{hj} values are set to 0. In contrast, in standard "dummy variable" coding, for an attribute with three levels, both indicator variables are set to 0 to indicate the third level. Thus, standard dummy variable coding cannot be used in this type of analysis.

Table 2

ATTRIBUTES, ATTRIBUTE LEVELS, AND ORTHOGONAL CODING

Attribute	*Number of Levels*	*Indicator*
Attribute 1 (brand)	3	x_1, x_2
Attribute 2 (price)	3	x_3, x_4
Attribute 3	3	x_5, x_6
Attribute 4	3	x_7, x_8
Attriubte 5 (durability)	3	x_9, x_{10}
Attribute 6	3	x_{11}, x_{12}
Attribute 7	3	x_{13}, x_{14}
Attribute 8	2	x_{15}
Attribute 9	3	x_{16}, x_{17}
Attribute 10	2	x_{18}
Attribute 11	2	x_{19}
Attribute 12	3	x_{20}, x_{21}
Attribute 13	2	x_{22}
Attribute 14	3	x_{23}, x_{24}
Attribute 15	3	x_{25}, x_{26}
Attriubte 16	3	x_{27}, x_{28}

Example of Orthogonal Coding: Three Levels

Attribute Level	x_j	x_{j+1}
1	–.333	–.333
2	.666	–.333
3	–.333	.666

Example of Orthogonal Coding: Two Levels

Attribute Level	x_k
1	.5
2	–.5

Estimation

We calibrated the models with nine partial-profile choice tasks and eight full-profile choice tasks per respondent; we held out two full-profile choice tasks per respondent for predictive testing. This results in 1887 observations for calibration and 222 for holdout validation.

We fit five models to the data. In-sample and predictive fit statistics appear in Table 3. Model 1 uses a standard heterogeneity distribution with no variable selection ($\phi = \theta = 1$). Model 2 is the heterogeneous variable selection model ($\phi = \theta$) that assumes that the same variables are used in both choice contexts. Model 3 is the unconstrained contextual variable selection model (ϕ and θ unconstrained) that allows the variables used in each choice context to be independently determined. Model 4 assumes that all variables are used in the partial-profile context and that variable selection occurs with the full-profile data ($\phi = 1$, θ unrestricted). Model 5 is a hybrid model, reflecting a logical relationship between the variables selected in the full-profile and partial-profile contexts. If a variable is not used in the partial-profile context, it is also not used in the full-profile context, but if the variable is used in the former context, it could be used in the latter.

Details of the MCMC estimation algorithms appear in the Appendix. We find that the chains converge quickly for all models; we assess convergence by starting the chains from multiple starting points and inspecting the time series plots of parameters. We ran each chain for 50,000 iterations and used a sample of every tenth iteration from the last

Table 3

COMPARISON OF MODELS AND FIT

	Variable Selection				*Out of Sample*	
Model	*Full Profile*	*Partial Profile*	*Restrictions*[a]	*LMD*	*Hit Rate*	*Hit Probability*
1. Base model	No	No	$\phi = \theta = 1$	–890.8	.50	.45
2. Heterogeneous variable selection	Yes	Yes	$\phi = \theta$	–1,118.0	.56	.49
3. Contextual variable selection	Yes	Yes	None	–903.2	.53	.47
4. Full-profile variable selection	Yes	No	$\phi = 1$	–854.8	.54	.48
5. Hybrid model	Yes	Yes	If $\gamma_{hj} = 0$, then $\lambda_{hj} = 0$. If $\gamma_{hj} = 1$, then $\lambda_{hj} \in \{0, 1\}$.	–1,007.2	.58	.49

[a] We present restrictions in terms of Equations 6a, 6b, and 7; γ determines variable selection in the partial-profile choice sets and $\gamma \sim \text{Bernoulli}(\phi)$, and λ determines variable selection in the full-profile choice sets and $\lambda \sim \text{Bernoulli}(\theta)$.

Notes: LMD = log-marginal density of the data calculated with Newton and Raftery's (1994, p. 121) importance sampling method.

25,000 to estimate the moments of the posterior distribution.

Model Fit Statistics

Table 3 displays in-sample and out-of-sample fit statistics for the five models. We find evidence that consumers use different decision strategies across the two choice contexts. The model with the best in-sample fit is the full-profile variable selection model, and the model with the best predictive fit is the hybrid model. We measured in-sample fit using the importance-weighted estimate of the log-marginal density that Newton and Raftery (1994) suggest. We measured out-of-sample fit by the hit rate and the hit probability. Hit rate is the posterior mean of correct predictions for the two holdout choices for each individual, averaged across respondents. Hit probability is the posterior mean of the predicted probability for the selected alternative.

The variable selection models have better out-of-sample predictive fits, despite several of them having lower log-marginal densities. The predictions from a variable selection model result in a form of model averaging in which the weights used to form the average are determined by the ratio of the posterior densities in the M–H algorithm. Several studies have investigated the superior predictive fit resulting from model averaging, using both linear and logarithmic scoring rules (Hoeting et al. 1999; Madigan and Raftery 1994; Raftery, Madigan, and Hoeting 1997; for an example in the marketing literature, see Yang and Allenby 2000).

The full-profile variable selection model fits the calibration data better than the baseline model, and each of the variable selection models provides better out-of-sample predictive results. The variable selection models result in a 6%–16% improvement in predictive accuracy over the base model. We present results and implications of the full-profile variable selection model in the next subsection.

Parameter Estimates and Implications

Table 4 compares the posterior means of $\bar{\beta}$ estimated from the baseline model and the full-profile variable selection model; the corresponding estimates of the partworths and θ, the variable selection hyperparameters, appear in Table 5. The attributes listed in Table 3 result in 16 values of λ_{hj} for each individual and 16 hyper-parameters θ_j. The partworth estimates from the variable selection model have larger absolute values than the estimates from the baseline model. This is due to variable selection, and comparisons of these estimates must include reference to θ. These results suggest that when respondents are faced with the full-profile of 16 attributes, they use only 45.5%, or 7.3, attributes on average on which to base their decision. Without variable selection, the baseline model differentially shrinks all the estimates toward 0.

Similarly, the amount of unexplained heterogeneity in the full-profile variable selection and the baseline models differs. (The posterior estimates of V_β are available on request.) Of the 28 diagonal elements of V_β in the baseline model, 25 are smaller in magnitude than the corresponding elements in the full-profile variable selection model. The full-profile variable selection model has a greater number of off-diagonal elements significantly different from 0. The baseline model, which does not consider who uses particular attributes, shrinks the partworth estimates toward zero and understates the amount of heterogeneity in the sample.

Table 4

COMPARISON OF POSTERIOR ESTIMATES OF $\bar{\beta}$

	Baseline Model: All Attributes		*Full-Profile Variable: Selection Model*	
Beta	*Posterior M*[a]	*Posterior SD*	*Posterior M*	*Posterior SD*
1	.158	(.19)	.117	(.21)
2	–.528	(.19)	–.733	(.23)
3	.248	(.22)	.528	(.29)
4	.686	(.21)	1.246	(.25)
5	.923	(.22)	1.225	(.34)
6	1.329	(.22)	1.773	(.33)
7	.612	(.23)	1.046	(.38)
8	.664	(.22)	.956	(.33)
9	2.288	(.27)	2.647	(.33)
10	2.733	(.29)	3.286	(.39)
11	.337	(.21)	.456	(.31)
12	.561	(.21)	.659	(.25)
13	–.161	(.22)	–.243	(.24)
14	–1.278	(.23)	–1.674	(.37)
15	.458	(.21)	.633	(.23)
16	.741	(.21)	.968	(.27)
17	1.740	(.21)	2.229	(.28)
18	.061	(.19)	–.006	(.24)
19	.634	(.18)	1.080	(.27)
20	.918	(.21)	1.365	(.29)
21	.860	(.22)	1.334	(.26)
22	.131	(.19)	.113	(.23)
23	.559	(.21)	.426	(.26)
24	.920	(.21)	1.195	(.25)
25	.329	(.22)	.697	(.29)
26	.350	(.21)	.728	(.29)
27	.186	(.23)	.254	(.31)
28	.712	(.24)	1.233	(.34)

[a]Posterior mean and (posterior standard deviation) are presented.

The results from a market simulation highlight the differences in the models. For each respondent in the study, we drew a sample of 500 simulated realizations from the posterior distribution of β_h for the baseline model and β_h and λ_h for the full-profile variable selection model. Recall that for the full-profile variable selection model, λ_h is a vector indicating which attributes are used by individual h in full-profile choice tasks. The simulation is designed to measure the sensitivity in choice probability for Attribute 1 (brand) and Attribute 16. In the variable selection model, approximately 61% of respondents use the brand attribute, whereas only 29% use Attribute 16. Figure 2 displays the posterior average difference in choice probability between selected levels of the attributes across the models. For brand, we measure importance as the difference between the choice probability of Brand 2 and Brand 3, holding all other attributes constant. We selected Brand 2 and Brand 3 because they represented the highest and lowest predicted market share. For Attribute 16, importance is the difference between the "high" and the "low" level of the attribute, averaged across brands and holding the other attributes fixed.

Figure 2 shows that the baseline model does not yield the same measures of attribute importance "on average" as the full-profile variable selection model. It might be argued that individual-level variable selection is unimportant if ignoring it yields the same market-level predictions. However, Figure

Table 5
COMPARISON OF PARTWORTHS AND ATTRIBUTES SELECTED

		Baseline Model	*Full-Profile Variable Selection*	
Attribute	*Level*	*Partworth*[a]	*Partworth*	θ[b]
1	1	.123	.205	
	2	.282	.322	.611 (.11)
	3	–.405	–.527	
2	1	–.311	–.591	
	2	–.064	–.064	.430 (.11)
	3	.375	.655	
3	1	–.751	–.999	
	2	.172	.226	.550 (.11)
	3	.578	.774	
4	1	–.426	–.667	
	2	.187	.379	.407 (.12)
	3	.239	.288	
5	1	–1.674	–1.977	
	2	.615	.669	.647 (.07)
	3	1.059	1.308	
6	1	–.299	–.372	
	2	.038	.084	.410 (.12)
	3	.262	.287	
7	1	.480	.639	
	2	.318	.396	.625 (.11)
	3	–.798	–1.035	
8	1	.229	.317	.516 (.15)
	2	–.229	–.317	
9	1	–.827	–1.066	
	2	–.086	–.098	.575 (.09)
	3	.913	1.163	
10	1	.030	–.003	.379 (.15)
	2	–.030	.003	
11	1	.317	.540	.337 (.13)
	2	–.317	–.540	
12	1	–.593	–.899	
	2	.325	.465	.463 (.11)
	3	.267	.434	
13	1	.066	.056	.355 (.15)
	2	–.066	–.056	
14	1	–.493	–.540	
	2	.066	–.114	.433 (.12)
	3	.427	.655	
15	1	–.226	–.475	
	2	.102	.222	.243 (.11)
	3	.124	.253	
16	1	–.299	–.495	
	2	–.113	–.242	.292 (.09)
	3	.413	.737	

[a] We calculated partworths using posterior means of $\bar{\beta}$ from Table 1.
[b] Poseterior mean and (posterior standard deviation) for θ presented.

2 shows that this is not the case. The baseline model indicates that the posterior average difference in choice probability between Brand 2 and Brand 3 is 15.3%. Averaging over respondents who use and do not use brand in their product choice, the variable selection model indicates that the average difference in choice probability is 12.8%. For Attribute 16, the baseline model overstates the average change in choice probability by almost two times, compared with the full-profile variable selection model. These results indicate that optimizing product configuration using the baseline model will result in a different configuration than the variable selection model. The superior in-sample fit and predictive accuracy of the full-profile variable selection model argues for its use over the baseline model.

Figure 2 also highlights opportunities for target marketing. Focusing on the results of the variable selection model, the average importance of Attribute 16 is relatively low. However, for the 29% of the market that uses Attribute 16 (e.g., $\lambda = 1$), moving from the "low" to the "high" level results in an average change in choice probability of 12.7%. Variable selection models provide a mechanism for identifying the "extremes" of the distribution of heterogeneity and focusing managerial attention on those segments, as Allenby and Ginter (1995) suggest.

DISCUSSION

Marketing researchers have procedural and statistical methods of obtaining measures of attribute importance for each individual on each attribute. In laboratory or experimental choice settings, studies can be designed to help focus respondents' attention and processing of the attributes. Statistical methods of modeling heterogeneity shrink poorly measured individual-level parameters to the overall or group-level mean. However, it appears erroneous to assume that consumers use all the product attributes in all brand choice situations. This article demonstrates that improved inference and predictive accuracy can be obtained by modeling which attributes consumers are actually using in different choice situations.

Variable selection in brand choice studies may be the result of consumers having no intrinsic value for a product attribute or cognitive constraints that result in selective processing of attribute information. This article provides three methodological improvements over standard statistical models of variable selection. First, we extend variable selection to the individual as opposed to the aggregate level of analysis. Second, we develop a practical method for applying the heterogeneous variable selection model to discrete choice data. Third, we introduce a flexible family of models that allows the variables that an individual uses to vary according to the choice context.

We introduce tractable algorithms for estimating the proposed variable selection models. We use MCMC methods to explore the posterior space of included variables and parameter estimates. Estimating variable selection models at the individual level presents challenges that are not encountered in the aggregate-level models discussed in the statistics literature. Specifically, the proposed algorithms obtain meaningful distributions of heterogeneity and avoid the evaluation of complex marginal posterior distributions.

The empirical application demonstrates that the variable selection models result in better prediction and inference. Applying variable selection to the full-profile choice sets fits the in-sample data better than the standard model, which supports the hypothesis that consumers adopt different decision strategies in different choice contexts. All the variable selection models had better out-of-sample fit, with a 6%–16% improvement in predictive accuracy. The study also shows that ignoring variable selection leads to biased parameter estimates and different conclusions about the importance of individual product attributes. These differences would result in different optimal product designs.

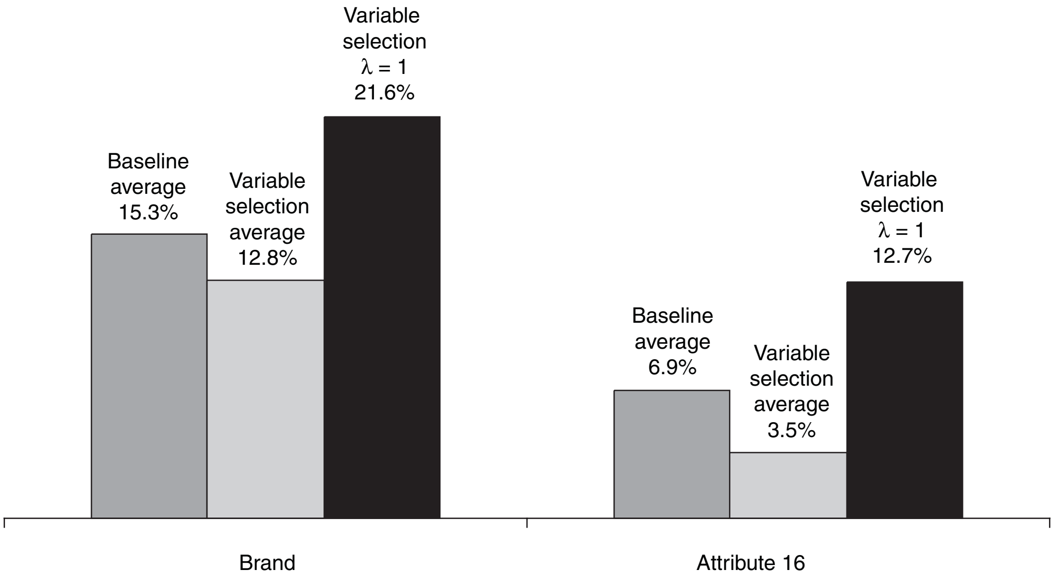
There are several opportunities for extending the empirical application. First, this study included a change in context, moving from partial- to full-profile choice situations. The heterogeneous variable selection model may perform better in data sets that consist of only full-profile choice tasks. Second, the variable selection models can be

Figure 2

COMPARISON OF ATTRIBUTE IMPORTANCE

BASELINE MODEL AND FULL-PROFILE VARIABLE SELECTION MODEL

CHANGE IN AVERAGE POSTERIOR CHOICE PROBABILITY



extended by adding another layer to the hierarchical model and estimating a statistical relationship between the variables selected and explanatory variables, such as demographics, situation specific variables, and so forth.

This article extends existing Bayesian variable selection models to identify the most promising subset of variables at the individual and context level. The variable selection models yielded useful managerial insights and, through model averaging, improved predictive results. These methods should be useful to managers marketing products with a large number of attributes who are concerned that different people use different subsets of attributes in their brand choice decisions.

APPENDIX: ESTIMATION ALGORITHMS

In the empirical application we discuss in this article, we study 16 product attributes. The orthogonal coding results in 28 parameters. We perform "variable" selection at the attribute level. Let m = 1, …, 16 index the 16 attributes. Let j = 1, …, 28 index the β_h parameters. Let j′ signify the set of parameters that map onto attribute m, as identified in Figure 1. For example, for m = 1, j′ = {1, 2} and $\tau_{h1} = \tau_{h2}$ by definition. A similar convention applies for λ_{hj} as well. In addition, the notation –k indicates all members of the set K, except k.

Heterogeneous Variable Selection with Discrete Choice Data

(A1) $y_{hik} = 1$ if $z_{hik} > z_{hnk}$ for all N alternatives in the choice set,

and

$$\text{(A2)}\quad z_{hik} = \sum_{j=1}^{J} \beta_{hj} x_{hijk} + \varepsilon_{hik} \quad \varepsilon \sim \text{EV}(0, 1),$$

where h indexes the individual, i the brand, and k the choice occasion. Equation A2 results in the standard multinomial logit choice probability. Let l_h represent the likelihood of the observed choices for person h.

The priors are represented as

$\beta_h \sim \text{Normal}(C_{\tau h}\bar{\beta}, C_{\tau h}V_\beta C_{\tau h})$;
$\tau_{hj'} = 1$, with probability θ_m, and $\tau_{hj'} = c$, with probability $1 - \theta_m$;
$\bar{\beta} \sim \text{Normal}(0, 100I)$;
$V_\beta \sim \text{IW}(\nu, \Delta)$; and
$\theta_m \sim \text{Beta}(a, b)$,

where IW is the inverted Wishart distribution, with $\nu = J + 8$ and $\Delta = \nu I$, and a = b = 3. Consistent with previous work in this area, the probability that any variable is selected is independent of the other variables in the model.

The following steps describe an MCMC with the posterior distribution of all model parameters as the stationary distribution.

1. Generate $\beta_{hj'}, \tau_{hj'} | \bar{\beta}, V_\beta, \theta, y_h, X_h$, for j′ = 1, …, M and h = 1, …, H.

We use an independence chain M–H step. Let $\tau_{hj'}^{(n)}$ represent a new candidate vector (scalar), and let $\tau_{hj'}^{(o)}$ represent the old vector (scalar) from the previous iteration of the chain. We use similar notation for $\beta_{hj'}^{(n)}$ and $\beta_{hj'}^{(o)}$.

- Set $\tau_{hj'}^{(n)} = 1$, with probability θ_m; otherwise, $\tau_{hj'}^{(n)} = c$.
- Draw $\beta_{hj'}^{(n)}$ from $\text{Normal}(C_{\tau h}^{(n)}\bar{\beta}, C_{\tau h}^{(n)}V_\beta C_{\tau h}^{(n)})$, where this distribution is conditional on all other values of $\beta_{h-j'}$.

Accept the new values $\tau_{h'}^{(n)}$ and $\beta_{hj'}^{(n)}$ with the probability

$$\text{Pr(accept)} = \min\left[\frac{L_h\left(\beta_{hj'}^{(n)}, \tau_{hj'}^{(n)}\right)}{L_h\left(\beta_{hj'}^{(o)}, \tau_{hj'}^{(o)}\right)}, 1\right].$$

Note that because the prior (in this case, the distribution heterogeneity) is used to generate the candidate, it is not used to evaluate the acceptance probability.

2. Generate $\bar{\beta}|\{\tau_h\}, \{\beta_h\}, V_\beta$.

Form β_h^* by $C_{\tau h}^{-1}\beta_h$.

$$\bar{\beta} \sim \text{Normal}\{\bar{b}, [(V_\beta/H)^{-1} + (100I)^{-1}]^{-1}\}.$$

$$\bar{b} = \{[(V_\beta/H)^{-1} + (100I)^{-1}]^{-1}\}\left[V_\beta^{-1}\sum_{h=1}^{H}\beta_h^* + (100I)^{-1}(0)\right].$$

3. Generate $V_\beta|\{\tau_h\}, \{\beta_h\}, \bar{\beta}$.

Form β_h^* by $C_{\tau h}^{-1}\beta_h$.

$$V_\beta \sim IW\left[\nu + H, \Delta + \sum_{h=1}^{H}(\beta_h^* - \bar{\beta})'(\beta_h^* - \bar{\beta})\right].$$

4. Generate $\theta_m|\{\tau_{hj'}\}$, for m = 1, …, M.

Let $s_{hm} = 1$ if $\tau_{hj'} = 1$, and 0 if otherwise.

$$\theta_m \sim \text{Beta}\left(a + \sum_{h=1}^{H}s_{hm}, H - \sum_{h=1}^{H}s_{hm} + b\right).$$

Contextual Variable Selection Model

$$\text{(A3)} \qquad z_{hik}^p = \sum_{j=1}^{J}\beta_{hj}\gamma_{hj}x_{hijk} + \varepsilon_{hik} \quad \varepsilon \sim EV(0, 1),$$

$$\text{(A4)} \qquad z_{hik}^f = \sum_{j=1}^{J}\beta_{hj}\lambda_{hj}x_{hijk} + \varepsilon_{hik} \quad \varepsilon \sim EV(0, 1), \text{ and}$$

$$\text{(A5)} \qquad \tau_{hj'} = 1 \text{ if } \gamma_{hj'} + \lambda_{hj'} \geq 1; \text{ otherwise, } \tau_{hj'} = c.$$

Equations A3 and A4 result in multinomial logit choice probabilities for the partial- and full-profile choice sets represented by L_h^p and L_h^f. The full individual likelihood is $L_h = L_h^p \times L_h^f$. We represent changes to the priors from the heterogeneous model as follows:

$\gamma_{hj'} \sim \text{Bernoulli}(\phi_m)$,
$\lambda_{hj'} \sim \text{Bernoulli}(\theta_m)$,
$\phi_m \sim \text{Beta}(a, b)$, and
$\theta_m \sim \text{Beta}(a, b)$.

1. Generate $\beta_{hj'}, \gamma_{hj'}, \lambda_{hj'}|\bar{\beta}, V_\beta, \phi, \theta, y_h, X_h$, for j′ = 1, …, M and h = 1, …, H.

- Set $\lambda_{hj'}^{(n)} = 1$, with probability θ_m; otherwise, $\lambda_{hj'}^{(n)} = 0$.
- Set $\gamma_{hj'}^{(n)} = 1$, with probability ϕ_m; otherwise, $\gamma_{hj'}^{(n)} = 0$.
- $\tau_{hj'}^{(n)} = 1$ if $\gamma_{hj'}^{(n)} + \lambda_{hj'}^{(n)} \geq 1$; otherwise, $\tau_{hj'}^{(n)} = c$.
- Draw $\beta_{hj'}^{(n)}$ from $\text{Normal}(C_{\tau h}^{(n)}\bar{\beta}, C_{\tau h}^{(n)}V_\beta C_{\tau h}^{(n)})$, where this distribution is conditional on all other values of $\beta_{h-j'}$.

Accept the new values $\lambda_{hj'}^{(n)}$, $\gamma_{hj'}^{(n)}$, and $\beta_{hj'}^{(n)}$ with the probability

$$\Pr(\text{accept}) = \min\left[\frac{L_h\left(\beta_{hj'}^{(n)}, \tau_{hj'}^{(n)}\right)}{L_h\left(\beta_{hj'}^{(o)}, \tau_{hj'}^{(o)}\right)}, 1\right].$$

The remaining steps are similar to the heterogeneous variable selection model.

Full-Profile Variable Selection Model

Model setup is identical to Equations A3–A5. However, both γ and φ are set equal to 1 for all h and j′.

1a. Generate $\beta_h|\lambda_h, \bar{\beta}, V_\beta, y_h, X_h$, for h = 1, …, H.

We used a random walk M–H algorithm to draw from the posterior distribution. A candidate vector $\beta_h^{(n)}$ is formed as $\beta_h^{(n)} = \beta_h^{(o)} + \zeta$, where $\zeta \sim \text{Normal}(0, wI)$ and w is chosen to ensure a 50% acceptance rate. The new parameter is accepted with the probability

$$\Pr(\text{accept}) = \min\left\{\frac{L_h\left(\beta_h^{(n)}\right) \times \exp\left[-\frac{1}{2}\left(\beta_h^{(n)} - \bar{\beta}\right)'V_\beta^{-1}\left(\beta_h^{(n)} - \bar{\beta}\right)\right]}{L_h\left(\beta_h^{(o)}\right) \times \exp\left[-\frac{1}{2}\left(\beta_h^{(o)} - \bar{\beta}\right)'V_\beta^{-1}\left(\beta_h^{(o)} - \bar{\beta}\right)\right]}, 1\right\}.$$

1b. Generate $\lambda_{hj'}|\lambda_{h-j'}, \theta_{jm}, \beta_h, y_h^f, X_h$, for j = 1, …, M and h = 1, …, H.

We use a griddy Gibbs (Tanner 1993) algorithm. The griddy Gibbs is useful because the support for $\lambda_{hj'}$ takes on only two values, (0, 1). Let $\lambda_{hj'}^{(1)}$ represent $\lambda_{hj'} = 1$, and let $\lambda_{hj'}^{(0)}$ represent $\lambda_{hj'} = 0$.

$$\Pr(\lambda_{hj'} = 1) = \left\{\frac{L_h^f\left(\lambda_{hj'}^{(1)}\right) \times \theta_m}{\left[L_h^f\left(\lambda_{hj'}^{(1)}\right) \times \theta_m\right] + \left[L_h^f\left(\lambda_{hj'}^{(0)}\right) \times (1 - \theta_m)\right]}\right\}.$$

Note that only the likelihood for the full-profile tasks is used in drawing λ_h. The remaining steps are similar to the heterogeneous variable selection model.

Hybrid Model

The hybrid model is set up identical to Equations A3–A5. We modify priors as follows:

$\lambda_{hj'} = 1|\gamma_{hj'} = 1$, with probability θ_m.
$\lambda_{hj'} = 0|\gamma_{hj'} = 0$, with probability 1.

1a. Generate $\beta_{hj'}, \gamma_{hj'}|\lambda_{hj'}, \bar{\beta}, V_\beta, \theta, \phi, y_h^p, y_h^f, X_h$, for j′ = 1, …, M and h = 1, …, H.

We use an independence chain M–H step.

- Set $\gamma_{hj'}^{(n)} = 1$, with probability 1, if $\lambda_{hj'} = 1$.
- Set $\gamma_{hj'}^{(n)} = 1$, with probability $[(1 - \theta_m) \times \varphi_m]/[1 - (\theta_m \times \varphi_m)]$, if $\lambda_{hj'} = 0$; otherwise, $\gamma_{hj'}^{(n)} = 0$. (Derivation of the probability is available on request.)
- $\tau_{hj'}^{(n)} = 1$, if $\gamma_{hj'}^{(n)} + \lambda_{hj'} \geq 1$; otherwise, $\tau_{hj'}^{(n)} = c$.
- Draw $\beta_{hj'}^{(n)}$ from $\text{Normal}(C_{\tau h}^{(n)}\bar{\beta}, C_{\tau h}^{(n)}V_\beta C_{\tau h}^{(n)})$, where this distribution is conditional on $\beta_{h-j'}$.

Accept the new values $\gamma_{hj'}^{(n)}$ and $\beta_{hj'}^{(n)}$ with the probability

$$\Pr(\text{accept}) = \min\left[\frac{L_h\left(\beta_{hj'}^{(n)}, \tau_{hj'}^{(n)}\right)}{L_h\left(\beta_{hj'}^{(o)}, \tau_{hj'}^{(o)}\right)}, 1\right].$$

1b. Generate $\lambda_{hj'}|\lambda_{h-j'}, \phi_m, \gamma_{hj'}, \beta_h, y_h^f, X_h$, for j′ = 1, …, M and h = 1, …, H.

We use a modified griddy Gibbs sampler. Let $\lambda_{hj'}^{(1)}$ represent $\lambda_{hj'} = 1$ and $\lambda_{hj'}^{(0)}$ represent $\lambda_{hj'} = 0$.

If $\gamma_{hj'} = 0$, then $\lambda_{hj'} = 0$.

If $\gamma_{hj'} = 1$, then

$$\Pr(\lambda_{hj} = 1) = \frac{L^f_h(\lambda^{(1)}_{hj'}) \times \theta_m}{\left[L^f_h(\lambda^{(1)}_{hj'}) \times \theta_m\right] + \left[L^f_h(\lambda^{(0)}_{hj'}) \times (1 - \theta_m)\right]}.$$

The remaining steps are similar to the heterogeneous variable selection model.

REFERENCES

Allenby, Greg M. and James L. Ginter (1995), “Using Extremes to Design Products and Segment Markets,” *Journal of Marketing Research*, 32 (November), 392–403.

Bettman, James R., Mary Frances Luce, and John W. Payne (1998), “Constructive Consumer Choice Processes,” *Journal of Consumer Research*, 25 (December), 187–217.

Chen, Ming-Hui, Joseph G. Ibrahim, and Constantin Yiannoutsos (1999), “Prior Elicitation, Variable Selection and Bayesian Computation for Logistic Regression Models,” *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 61 (1), 223–42.

Chib, Siddhartha and Edward Greenberg (1995), “Understanding the Metropolis-Hastings Algorithm,” *The American Statistician*, 49 (November), 327–35.

Dickson, Peter R. (1982), “Person-Situation: Segmentation’s Missing Link,” *Journal of Marketing*, 46 (3), 56–64.

Fennell, Geraldine (1978), “Consumers’ Perceptions of the Product Use Situation,” *Journal of Marketing*, 42 (April), 38–47.

George, Edward I. (2000), “The Variable Selection Problem,” *Journal of the American Statistical Association*, 95 (December), 1304–1308.

——— and Robert E. McCulloch (1993), “Variable Selection via Gibbs Sampling,” *Journal of the American Statistical Association*, 88 (September), 881–89.

——— and ——— (1997), “Approaches for Bayesian Variable Selection,” *Statistica Sinica*, 7 (2), 339–73.

Geweke, J. (1996), “Variable Selection and Model Comparison in Regression,” in *Bayesian Statistics*, Vol. 5, J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith, eds. Oxford: Oxford University Press, 609–620.

Gilbride, Timothy J. and Greg M. Allenby (2004), “A Choice Model with Conjunctive, Disjunctive, and Compensatory Screening Rules,” *Marketing Science*, 23 (Summer), 391–406.

Green, Peter J. (1995), “Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination,” *Biometrika*, 82 (December), 711–32.

Hoeting, Jennifer A., David Madigan, Adrian E. Raftery, and Chris T. Volinsky (1999), “Bayesian Model Averaging: A Tutorial,” (with discussion), *Statistical Science*, 14 (November), 382–417.

Madigan, D. and Adrian. E. Raftery (1994), “Model Selection and Accounting for Model Uncertainty in Graphical Methods Using Occam’s Window,” *Journal of the American Statistical Association*, 89 (December), 1535–46.

Newton, Michael A. and Adrian E. Raftery (1994), “Approximating Bayesian Inference with the Weighted Likelihood Bootstrap,” *Journal of the Royal Statistical Society. Series B*, 56 (1), 3–48.

Phillips, David B. and Adrian F.M. Smith (1995), “Bayesian Model Comparison via Jump Diffusions,” in *Markov Chain Monte Carlo in Practice*, W.R. Gilks, S. Richardson, and D.J. Spiegelhalter, eds. London: Chapman & Hall, 215–39.

Raftery, Adrian E., David Madigan, and Jennifer A. Hoeting (1997), “Bayesian Model Averaging for Linear Regression Model,” *Journal of the American Statistical Association*, 92 (March), 179–91.

Tanner, Martin A. (1993), *Tools for Statistical Inference*. New York: Springer-Verlag.

Yang, Sha and Greg M. Allenby (2000), “A Model for Observational, Structural, and Household Heterogeneity in Panel Data,” *Marketing Letters*, 11 (2), 137–49.

———, ———, and Geraldine Fennell (2002), “Modeling Variation in Brand Preference: The Roles of Objective Environment and Motivating Conditions,” *Marketing Science*, 21 (Winter), 14–31.