Quantitative Marketing and Economics, 3, 311–335, 2005.

Modeling Simultaneity in Survey Data

TIMOTHY J. GILBRIDE
Mendoza College of Business, University of Notre Dame, Notre Dame, IN 46556

SHA YANG
Stern School of Business, New York University, 44 West Fourth Street, New York, NY 10012

GREG M. ALLENBY
Fisher College of Business, Ohio State University, 2100 Neil Ave. Columbus, OH 43210

Abstract. Responses to questions in a survey can reflect a behavior process that influences multiple response items. Respondent ratings of brand attributes, for example, can be affected by past purchases by making a brand more salient, or by respondents attributing higher performance to justify their purchases. When multiple response items are influenced by a common underlying process, there is simultaneity in the data. This paper proposes an approach to model the simultaneity in different survey responses by using common parameters and structural relationships motivated by behavioral theories on how consumers respond to surveys. Specifically, the proposed models show how brand usage and attribute perception responses are jointly determined by justification, order, and brand halo effects in two brand positioning studies. We detect a significant tendency for respondents to inflate their reported beliefs for particular brands as well as the selected brand across five countries in an international survey as well as in a domestic study.

Key words. Bayesian methods, endogeneity, survey research, data augmentation

JEL Classification: C35, C53, D12, M31

1. Introduction

Consumer response to survey questions reflects a complex process that researchers in marketing are just beginning to understand. Consumers are known to be heterogeneous in their use of response scales (Baumgartner and Steenkamp, 2001, Rossi et al., 2001), to judge stimuli using relative versus absolute scales (Janiszewski et al., 2003), and to report greater willingness to purchase offerings than their purchase history displays (Wertenbroch and Skiera, 2002). By modeling the observed responses as being the result of a behavioral process, researchers are able to draw inferences about respondent beliefs and preferences that corrects for effects that may not be present at the time of purchase, leading to improved prediction of marketplace behavior.

Simultaneity occurs when the data analyst cannot treat observations on different phenomena as independent. In economics, observations on price and quantity demanded are not treated as independent observations but as arising from a model with both common parameters and a structural relationship. In other words, price and quantity are endogenous and are jointly determined by the system under study. This paper proposes an approach to model the simultaneity in survey responses using common parameters and structural relationships motivated by behavioral theories on how consumers respond to surveys.

Our analysis involves use of a multi-attribute utility model for brand choice and preference using survey data on attribute perceptions and importance. We show that attribute perceptions (i.e., brand beliefs) are endogenously related to other survey questions, and that modeling these simultaneous effects leads to improved predictions and different inferences than that obtained with standard models. In particular, we find that when simultaneous effects are not modeled, stated attribute importances are not predictive of brand use within the multi-attribute utility model. This result is consistent with the belief that self-explicated weights have low validity, and that attribute-level importances are better obtained using techniques such as conjoint analysis (Cattin and Wittin, 1982). The predictive validity of self-explicated importance weights improves greatly once simultaneous effects are taken into account.

Researchers have typically focused on modeling the observed responses to a subset of questions without incorporating information from other parts of the survey. It is reasonable to expect that beliefs or attitudes may be reflected across different survey questions, and that the reporting process may result in distorted measures of attitudes and beliefs. Beliefs about brand attributes, for example, are reflected in both reported brand choice and reported brand beliefs. One may inflate the reported attribute ratings for the chosen brand in order to justify one's choice or simply because they are more salient and accessible in memory. In these situations, the survey responses for brand use and its drivers are simultaneous because they are the result of a latent process with common components.

The model is illustrated with two sets of data. The first set of data is an international survey from five countries, where we document simultaneity between brand ratings data and brand choice data. Our analysis indicates that brand positioning inferences across a large number of product attributes for single-use cameras are substantially biased in favor of the market share leader and/or the first brand evaluated by respondents. We find that the simultaneous model coupled with self-explicated importance weights leads to predictive improvements in holdout samples in each country compared to traditional approaches. This result suggests that respondents are able to say how important the attributes are, but have difficulty providing perceptions of the brand attributes without being influenced by justification, order, or brand halo effects.

The second data set is from a U.S. based survey on brand beliefs and consideration for digital cameras. The respondents rated 7 brands on 14 attributes and indicated for each brand whether or not it would be considered on their next purchase occasion. Self-explicated attribute importance weights are not collected in this data set, and instead are treated as unknown parameters. Changes to the basic model set-up and estimation algorithms are discussed when attribute importance weights are estimated from the data. Again we find strong evidence of an endogenous relationship between reported brand beliefs and consideration and show that inferences on brand perceptions are different when endogeneity is accounted for the model.

The paper is organized as follows. In Section 2 we detail the proposed model and place it in the context of behavioral theories and other empirical models. Our Bayesian approach to estimation is outlined in Section 3. Section 4 describes the data, models, and results for the single-use camera study. Section 5 reports on the digital camera study. Section 6 offers a general discussion and directions for further research.

2. Modeling simultaneity

The proposed model represents observed responses to different survey questions as arising from a common set of parameters. Let survey responses y and x each be a function of a common parameter α and other question-specific parameters θ_y and θ_x. If the likelihood is $\pi(x, y \mid \theta_x, \theta_y, \alpha) = \pi(x \mid \theta_x, \alpha) \times \pi(y \mid \theta_y, \alpha)$ then x and y are simultaneously determined, or jointly determined by the system of study due to the common parameter α. If one were to model and estimate either $\pi(x \mid \theta_x, \alpha)$ or $\pi(y \mid \theta_y, \alpha)$ separately, then the resulting estimate for α is inefficient. Other forms of simultaneity result in a structural relationship between x and y that can lead to inconsistent estimates unless the system is fully specified. If the true likelihood is $\pi(x, y \mid \theta_x, \theta_y, \alpha) = \pi(y \mid \theta_y, \alpha) \times \pi(x \mid y, \theta_x, \alpha)$, then using x to obtain better estimates of α requires specification of the structural relationship between x, y and α. Ignoring the structural relationship results in inconsistent parameter estimates because the model is misspecified.

The existence of a common parameter also defines an endogenous relationship between x and y. As discussed by Engle et al. (1983), a necessary condition for x to be considered exogenous is that it must not share common parameters with y, nor can there be cross parameter restrictions between θ_y and θ_x. In general, ignoring an endogenous relationship between x and y will lead to inefficient parameter estimates. Inconsistent parameters will result if the endogeneity involves structural relationships (see also Yang et al., 2003).

To illustrate these concepts and to set the stage for our empirical examples, consider a survey to measure brand perceptions and relate them to brand preference. Let x_{hij} be the reported rating by household h of brand i on attribute j. This may be either a binary yes/no or Likert-type fixed scale format. Let $y_{hi} = 1$ if household h reports using brand i on the last usage occasion and 0 otherwise. Finally, let b_{hj} be the reported importance of attribute j by household h. Thus the observed data consists of attribute importance, b_{hj}, brand beliefs, x_{hij}, and brand choice, y_{hi}. Modifications to the models when self-explicated attribute importance weights, b_{hj}, are not available and/or the researcher wishes to estimate β_j are discussed after the basic model set-up.

We model the simultaneity in reported brand beliefs and reported brand choice using the following set-up. Following standard latent variable techniques we assume the observed responses are a function of latent variables x^*_{hij} and y^*_{hi}. Specifically, for binary brand ratings data, let $x_{hij} = 1$ if $x^*_{hij} > 0$ and for brand usage let $y_{hi} = 1$ if $y^*_{hi} > y^*_{hk}$ for all brands in the set K. Modifying this set-up for fix-scaled x_{hij} is straightforward. Now, let

$$x^*_{hij} = \alpha_{ij} + \varepsilon_{hij} \quad \text{for all } i, j \tag{1}$$

$$y^*_{hi} = \sum_{j=1}^{J} b_{hj}\alpha_{ij} + \eta_{hi} \quad \text{for all } i \tag{2}$$

where ε, η are distributed independently across respondents as N(0, 1) for identification. In this model, brand ratings and brand choice data are driven by a common perception of the level of attribute j for brand i, α_{ij}, plus idiosyncratic and individual error terms. In this instance, estimating only equation (1) or (2) will result in inefficient estimates of α, but not an inconsistent estimate. The model addresses the issue of endogeneity by considering the consumer choice process and it anticipates a single observation per respondent, which is typical in survey research

The basic model represented in (1) and (2) can be modified to allow for more complicated response models. Respondents may "inflate" their responses to the brand belief questions for the selected brand ($y_{hi} = 1$). This can be represented as:

$$x^*_{hij} = \alpha_{ij} + \delta y_{hi} + \varepsilon_{hij} \quad \text{for all } i, j \tag{3}$$

$$y^*_{hi} = \sum_{j=1}^{J} b_{hj}\alpha_{ij} + \eta_{hi} \quad \text{for all } i \tag{4}$$

In this model, brand choices and reported brand choices are a function of b_{hj}, the self-explicated importance weights and α_{ij}, the derived measures of brand perceptions. However, reported brand beliefs are a function of both α_{ij} and y_{hi}. Ignoring the role of y in equation (3) and basing inferences on just x results in inconsistent estimates for α. If only equation (4) is used as the basis for inferences, then the estimates for α are inefficient (but not inconsistent). Recall that (4) is based on a single multinomial outcome per respondent whereas (3) adds $(i \times j)$ additional observations per respondent.

As discussed by Maddala (1983) and Maddala and Lee (1976) equations (3) and (4) form a recursive system and in general, certain conditions are required for the system to be logically consistent and for the parameters to be identified. Note that the system would be logically inconsistent if (4) were specified as

$$y^*_{hi} = \sum_{j=1}^{J} b_{hj}x^*_{hij} + \eta_{hi} \quad \text{for all } i. \tag{4$'$}$$

This can be seen by substituting in $\alpha_{ij} + \delta y_{hi} + \varepsilon_{hij}$ for x^*_{hij}; in simple terms, the latent y^*_{hi} is now a function of its realized value y_{hi}, in addition to the realized value being a function of the latent variable. All parameters are identified in equations (3) and (4) because the α_{ij}'s are common and the equations for x^*_{hij} do not contain all the variables included in the equation for y^*_{hi}. For a complete discussion of these conditions, see Maddala (1983).

Several theoretical arguments support the response model specified for brand beliefs in (3). First, respondents may feel a need to justify their choices by inflating the ratings of the chosen brand. The ability to "justify" one's choices has been cited as a rationale in experimental choice situations involving dominated and non-dominated alternatives by Pettibone and Wedell (2000) and Simonson (1989). Brown and Feinberg (2004) review other theoretical arguments for "bolstering." Similar to the justification argument, in order to reach a favorable conclusion about what they consume, consumers may interpret information in a biased fashion. Alternatively, having chosen a particular brand may make information more cognitively accessible for that brand as compared to others. Brown and Feinberg present experimental evidence documenting that choice does change reported evaluations and present an alternative representation of simultaneity by allowing correlated errors between pre- and post-choice measures. By comparison, the models proposed in this paper represent simultaneous effects as common parameters and structural shifts in the mean of the joint distribution of the data, as opposed to common parameters in the covariance matrix.

The proposed model does not provide insight as to the mechanism that causes an inflated response to the brand belief questions. Instead, the model tests for the existence of such an effect and controls for the effect in measuring the parameters of interest. For convenience and without claiming that it is the only possible explanation, we will refer to δ in equation (3) as measuring the "justification" bias.

Consumers' responses to questionnaire items may also be influenced by the order in which questions or responses are presented. This well documented effect (e.g. Nisbett and Wilson, 1977) suggests that respondents are more likely to respond or evaluate the first item or items in a list than the last items. Consider the following model:

$$x^*_{hij} = \alpha_{ij} + \delta y_{hi} + \gamma I(\cdot)_{hij} + \varepsilon_{hij} \quad \text{for all } i, j \tag{5}$$

$$y^*_{hi} = \sum_{j=1}^{J} b_{hj}\alpha_{ij} + \eta_{hi} \quad \text{for all } i \tag{6}$$

where $I(\cdot)_{hij}$ represents an indicator function equal to 1 if the argument $(\cdot)$ is satisfied and 0 otherwise. $I(\cdot)_{hij}$ may equal one for attributes that are at the beginning of the list presented to a particular respondent. Alternatively, $I(\cdot)_{hij}$ may equal one for brands that are evaluated first by a particular respondent. With multiple indicator functions, one may specify and test various order effects as long they are not perfectly correlated. The parameter γ will be referred to as the "order effect."

Consumers may not be able to distinguish their overall impression of a brand from their evaluation of specific characteristics of a brand. This is referred to as a "halo" effect. Dillon et al. (2001) propose a model utilizing only brand ratings data that decomposes a brand rating x_{hij} into a "brand specific association" and a "global brand impression." Incorporating this insight into the proposed model yields:

$$x^*_{hij} = \alpha_{ij} + \delta y_{hi} + \sum_{k=1}^{K-1} \lambda_k I(k = i)_{hij} + \varepsilon_{hij} \quad \text{for all } i, j \tag{7}$$

$$y^*_{hi} = \sum_{j=1}^{J} b_{hj}\alpha_{ij} + \eta_{hi} \quad \text{for all } i \tag{8}$$

In this model, the α_{ij} correspond to the brand specific associations, λ_k are global brand associations, and K is the set of all brands. For identification purposes, the λ_k can be measured for all but one brand. The value of λ_k must therefore be interpreted with respect to the omitted brand.

The preceding models assume that self-explicated attribute importance weights, b_{hj}, are available and incorporated into the model. In instances where b_{hj} are not available and/or the researcher prefers to estimate common attribute weights, it is straightforward, at least conceptually, to alter the models as follows:

$$x^*_{hij} = \alpha_{ij} + \delta y_{hi} + \varepsilon_{hij} \quad \text{for all } i, j \tag{3a}$$

$$y^*_{hi} = \sum_{j=1}^{J} \beta_j(\alpha_{ij} + \varepsilon_{hij}) + \eta_{hi} \quad \text{for all } i \tag{4a}$$

The change is illustrated using the "justification bias" model but easily extends to the other models. Inclusion of the error term ε_{hij} in the brand choice equation is necessary to ensure enough individual level variation to identify the vector of β's. Note that including $\alpha_{ij} + \varepsilon_{hij}$ as opposed to x^*_{hij} ensures the logical consistency of the model as discussed earlier. However, as compared to standard data augmentation models, the model must be specified and estimated in terms of ε and η as opposed to x^* and y^*. This is discussed in greater detail below.

We have illustrated a variety of models incorporating simultaneity in survey response data. The simultaneity involves common parameters in a model of how consumers fill-out surveys. In all instances, greater efficiency is obtained by jointly modeling the brand choice and brand ratings responses. Our approach also provides consistent parameter estimates if behavioral response patterns such as justification bias, order bias, or brand halo effects are present. The next section turns attention to obtaining parameter estimates from the joint distribution of the data.

3. Model estimation

Model estimation is facilitated by the Bayesian data augmentation technique proposed by Tanner and Wong (1987). The model parameters are augmented with the latent variables y^* and x^*, and regarded as objects of inference similar to all other unobservables in the model (see Rossi and Allenby, 2003). Without data augmentation, the likelihood function must be evaluated through integration. While this is straightforward in a simple model, it can quickly become intractable in multivariate and multinomial applications. Data augmentation requires only the ability to draw from truncated distributions and in this set-up, to do regression analysis. Detailed treatments of data augmentation can be found in Albert and Chibb (1993) and McCulloch and Rossi (1994). Algorithms and details for drawing the latent values in a probit context are available in Rossi et al. (1996). As noted above, estimating the endogeneity models with derived importance weights, β_j's, requires writing the models in terms of ε and η and altering the standard algorithms. These changes are introduced after the discussion on estimating the model with self-explicated importance weights, b_{hj}'s.

We discuss estimation for the "justification" model as given in equations (3) and (4). Extending the method to the other models is straightforward and hopefully apparent from the presentation. The simultaneity in observed responses is due to the common parameter α, and the model can be written in hierarchical form:

$$y \mid y^* \tag{9a}$$

$$x \mid x^* \tag{9b}$$

$$y^* \mid b, \alpha, \sigma_\eta^2 = 1 \tag{9c}$$

$$x^* \mid y, \alpha, \delta, \sigma_\varepsilon^2 = 1 \tag{9d}$$

Estimation is carried out using Markov chain Monte Carlo methods by sequentially generating draws from the full conditional distributions of all model parameters:

$$y^* \mid y, b, \alpha, \sigma_\eta^2 = 1 \tag{10a}$$

$$x^* \mid x, y, \alpha, \delta, \sigma_\varepsilon^2 = 1 \tag{10b}$$

$$\alpha, \delta \mid x^*, y^*, \sigma_\varepsilon^2 = 1, \quad \sigma_\eta^2 = 1 \tag{10c}$$

Two practical issues present themselves in estimating the model. The first is drawing the latent variables y^* and x^* in equations (10a) and (10b). Conditioning on the observed data and the parameters α, δ, and $\sigma_\varepsilon^2 = 1$, $\sigma_\eta^2 = 1$, for each individual, draws of x^* and y^* are obtained from a truncated normal distribution. The truncation points correspond to the censoring mechanisms in (9a) and (9b). Specifically, $x_{hij}^* > 0$ if $x_{hij} = 1$ and $x_{hij}^* < 0$ if $x_{hij} = 0$. For brand usage $y_{hi}^* > y_{hk}^*$ for all brands in the set K if $y_{hi} = 1$, otherwise $y_{hi}^* < \max\{y_{hk}^*\}$ where the $\max\{y_{hk}^*\}$ corresponds to the chosen brand. Drawing from a truncated normal distribution is facilitated by the reverse CDF method as explained in McCulloch and Rossi (1994) and Allenby et al. (1995).

Second, given the latent variables x^* and y^*, estimates for α and δ in (10c) can be obtained via standard conjugate Bayesian regression methods (see Zellner, 1971). Consideration is needed when determining the design matrix in order to obtain estimates of the common parameters. This is illustrated in Figure 1 for a simplified example involving 3 brands and 4 attributes for a single respondent. The number of columns in the design matrix is equal to [(# attributes × # of brands) + # of behavioral parameters]. In our simplified example, which

Attribute Importance Data: b_{hj}

Attr. 1	3
Attr. 2	4
Attr. 3	1
Attr. 4	5

Brand Choice Data: y_{hi}

Brand 1	0
Brand 2	1
Brand 3	0

Brand Ratings Data: x_{hij}

	Brand 1	Brand 2	Brand 3
Attr. 1	1	1	0
Attr. 2	0	1	0
Attr. 3	0	0	1
Attr. 4	0	1	0

(Brand ratings data is not used to specify the design matrix and is included only for completeness)

Design Matrix for a single respondent

	α_{11}	α_{12}	α_{13}	α_{14}	α_{21}	α_{22}	α_{23}	α_{24}	α_{31}	α_{32}	α_{33}	α_{34}	δ
x_{h11}	1	0	0	0	0	0	0	0	0	0	0	0	0
x_{h12}	0	1	0	0	0	0	0	0	0	0	0	0	0
x_{h13}	0	0	1	0	0	0	0	0	0	0	0	0	0
x_{h14}	0	0	0	1	0	0	0	0	0	0	0	0	0
x_{h21}	0	0	0	0	1	0	0	0	0	0	0	0	1
x_{h22}	0	0	0	0	0	1	0	0	0	0	0	0	1
x_{h23}	0	0	0	0	0	0	1	0	0	0	0	0	1
x_{h24}	0	0	0	0	0	0	0	1	0	0	0	0	1
x_{h31}	0	0	0	0	0	0	0	0	1	0	0	0	0
x_{h32}	0	0	0	0	0	0	0	0	0	1	0	0	0
x_{h33}	0	0	0	0	0	0	0	0	0	0	1	0	0
x_{h34}	0	0	0	0	0	0	0	0	0	0	0	1	0
y_{h1}	3	4	1	5	0	0	0	0	0	0	0	0	0
y_{h2}	0	0	0	0	3	4	1	5	0	0	0	0	0
y_{h3}	0	0	0	0	0	0	0	0	3	4	1	5	0

Figure 1. Determining the design matrix, simplified example for a single respondent.

includes a "justification" parameter, the number of columns equals $[(4 \times 3) + 1] = 13$. The number of rows for each individual is equal to [(# attributes × # of brands) + # of brands]. Note that the column for the justification parameter δ has one's only in the rows corresponding to the attribute ratings for the chosen brand. Also, the attribute importance data b_{hj} is replicated in the appropriate columns for each row representing brand choice responses. The vector of dependent variables is obtained by vectorizing and stacking the appropriate x^* and y^* for each individual. Data for the entire sample is obtained simply by stacking the matrices and vectors across individuals. In practical applications involving many attributes, brands, and respondents this will yield a relatively large regression problem.

The error structure required to estimate common importance weights, β_j's, does not lend itself to the standard data augmentation techniques. Recall the model:

$$x^*_{hij} = \alpha_{ij} + \delta y_{hi} + \varepsilon_{hij} \quad \text{for all } i, j \tag{3a}$$

$$y^*_{hi} = \sum_{j=1}^{J} \beta_j(\alpha_{ij} + \varepsilon_{hij}) + \eta_{hi} \quad \text{for all } i \tag{4a}$$

Equations (3a) and (4a) involve cross-equation restrictions for the error terms ε. As a result, x^*_{hij} cannot be drawn from a truncated distribution implied solely by x_{hij}, y_{hi}, α_{ij}, δ, and $\sigma^2_\varepsilon = 1$. In fact, the implied error space for ε_{hij} is dependent on the set of all α_{ij}, y_{hi}, ε_{hij}, and η_{hi} for person h. This difficulty is overcome by noting that it is equally valid to augment the parameter space with ε and η as opposed to x^* and y^*. As a result the model is represented as:

$$y \mid y^* \tag{11a}$$

$$x \mid x^* \tag{11b}$$

$$\eta \mid \beta, \alpha, \varepsilon, \sigma^2_\eta = 1 \tag{11c}$$

$$\varepsilon \mid y, \beta, \alpha, \delta, \eta, \sigma^2_\varepsilon = 1 \tag{11d}$$

where y^* and x^* result from the (now) deterministic functions specified in (3a) and (4a).

Estimation is carried out by generating draws from the following distributions:

$$\eta \mid y, \beta, \alpha, \varepsilon, \sigma^2_\eta = 1 \tag{12a}$$

$$\varepsilon \mid x, y, \beta, \alpha, \delta, \eta, \sigma^2_\varepsilon = 1 \tag{12b}$$

$$\alpha \mid x, y, \beta, \delta, \varepsilon, \eta \tag{12c}$$

$$\delta \mid x, y, \alpha, \varepsilon \tag{12d}$$

$$\beta \mid y, \alpha, \varepsilon, \eta \tag{12e}$$

The values of η are drawn using standard truncated normal distributions. Each ε_{hij} is drawn using a modified rejection sampler using a truncated $N(0, \sigma^2_\varepsilon = 1)$ as the generating distribution with the truncation point determined by x_{hij} and $\alpha_{ij} + \delta y_{hi.}$ The rejection step tests the drawn ε_{hij} against y_{hi} using $\{\varepsilon_{hi-j}\}$, η_{hi}, and the current values of β and α. The notation "-j" implies all other elements of the set other than "j". The process is repeated until a value of ε_{hij} is accepted. Values of β, α, and δ are drawn using a random walk

Metropolis-Hastings algorithm, but conditional on $\{\eta\}$ and $\{\varepsilon\}$, the likelihood is simply an indicator function. Full details of the estimation algorithm are contained in the appendix. Results from simulation studies testing the algorithms detailed in this section are available from the authors upon request.

Estimating an endogeneous response model is often not feasible using standard maximum likelihood methods. Theoretically, with ε and $\eta \sim N(0,1)$ the likelihood function can be built up independently for each observed response for each respondent when self-explicated importance weights are used. However, deriving the first and second derivatives for use in the method of scoring or Newton-Raphson involves several probabilities per common parameter. These complicated functions must then be numerically estimated either directly, or for the gradient, approximated by the outer-product method, which can be numerically challenging. When importance weights β are to be estimated from the data, a complicated partition of the error space must be specified, and probabilities simulated for use in a simulated maximum likelihood estimation procedure, further increasing the computational burden. Brand perception studies may involve many brands and/or many attributes; one of our empirical studies involves 4 brands and 126 attributes resulting in 504 common parameters. At minimum, this requires a very large dataset for the asymptotic properties of maximum likelihood estimators to apply; under realistic sample sizes, the likelihood may simply be too diffuse or too "flat" to obtain meaningful MLE or simulated MLE estimates.

Two empirical studies are presented to illustrate the models and estimation algorithms. The first data set involves an international study of single-use cameras involving a large number of product attributes and self-explicated attribute importance weights. The second study looks at consumers' perceptions of digital cameras and derives the attribute importance weights from the data.

4. Single-use cameras

Data from an international brand positioning study of single-use (disposable) cameras is used to illustrate the presence of simultaneity in survey data, and the biasing effect it has on inferences. Table 1 lists the countries in the study, the number of respondents in each

Table 1. Single-use camera study.

	Italy	U.K.	France	U.S.	Japan
No. of respondents	363	317	306	992	752
Actual market share					
Brand 1	53.7%	45.4%	38.2%	51.0%	41.9%
Brand 2	42.1%	40.1%	46.4%	42.2%	50.0%
Brand 3	2.5%	3.8%	7.5%	6.8%	8.1%
Brand 4	1.7%	10.7%	7.8%		
Hit rate: Max $\sum x_i$	57.9%	52.1%	52.3%	69.4%	50.1%
Hit rate: Max $\sum bx_i$	57.9%	51.7%	52.6%	69.4%	50.1%
Average no. of x_i indicated					
Chosen brand	30.8	29.8	30.2	27.6	30.8
Other brands	10.3	9.9	10.1	13.8	15.4

country, and the market share for each brand as reported in the survey. The brand names are disguised for confidentiality. In each country, eight different versions of the questionnaire were used. Depending on the version of the questionnaire, each respondent saw between 62 and 64 of a total of 126 product attributes studied. In the presentation that follows, J is understood to index the set of product attributes evaluated by the respondent. Respondents were part of a multi-country panel maintained by a professional market research company; the research company managed the rotation of attributes across the different versions of the questionnaire as well as translation into native languages.

A primary goal of the study is to understand how the brands are perceived relative to each other on the 126 product attributes. Each respondent provided the following information: a rating on a 1 to 5 scale of how important each attribute was in the last usage situation (b_{hj}), which brand was used (y_{hi}), and whether the brand "definitely had" each attribute on a binary yes/no scale (x_{hij}). Additional information on usage occasion, brand familiarity, purchase location, and general brand impressions was also collected.

The single-use camera study presents a challenging data collection environment. While asking each respondent to rate 62 to 64 attributes is extreme, the current study is certainly not unheard of in applied survey research. Particularly in exploratory research, it is not uncommon for managers to generate a long list of product attributes. It seems reasonable to expect that in situations where the list of attributes is long and salience low for any individual on any one attribute, that behavioral response patterns such as justification bias, order bias, and/or brand halo effects may distort the reported brand beliefs.

Table 1 illustrates a possible simultaneous relationship between brand ratings and brand choice survey responses. Reported in Table 1 are two predictions of a respondent's preferred brand (y_{hi}): (i) a weighted prediction that combines the stated importance of each attribute with the reported attributes ($\sum_j x_{hij} b_{hj}$); and (ii) an unweighted prediction that simply computes the number of attributes ($\sum_j x_{hij}$). The brand with the highest total is the predicted brand, and the hit rates reported in Table 1 indicate the frequency that the predicted brand was the brand actually used by the respondent. Surprisingly, the hit rates for the unweighted predictions are roughly the same as that for the weighted predictions in each country. Historically, researchers have interpreted this sort of result to imply that respondents cannot provide valid measures of attribute importance (e.g. Day, 1972). In fact, this was an early motivation for the development of conjoint analysis (Green and Rao, 1971).

However, upon closer inspection of the data, we find that the number of times a brand was reported to have "definitely had" an attribute is strongly related to whether the brand was actually chosen by the respondent. On average, respondents indicated that the brand he/she chose "definitely had" two to three times as many attributes as the brands that weren't chosen. The question is: did respondents choose the reported brand because it had so many attributes, or did respondents report the brand had so many attributes because they chose the brand? If it is the latter, then the x_{hij} cannot be relied upon to represent brand perceptions as they exist when the consumer is choosing a camera.

To investigate the presence of simultaneity, four models were fit to the data from each country in the study. Approximately 10% of respondents were excluded from the calibration sample and used as a hold-out sample for validation. *Model 1* corresponds to equations (1) and (2) and represents the base case endogeneity model where brand ratings and brand choice responses arise from common perceptions of brand attributes. *Model 2* corresponds to equations (3) and (4) and controls for a possible "justification" bias or the tendency to inflate the brand ratings for the chosen brand. Although the attributes and the order in which they were presented was rotated across the different versions of the questionnaire, the order of the brands (i.e., the column headings in the survey) was fixed. This led to the hypothesis that respondents may be going through the long list of attributes for the first brand listed, and paying less attention to the subsequent brands. Accordingly, *Model 3* includes a "1st position" bias and equations (5) and (6) are modified as:

$$x^*_{hij} = \alpha_{ij} + \delta y_{hi} + \gamma I(i = 1)_{hij} + \varepsilon_{hij} \quad \text{for all } i, j \tag{5$'$}$$

$$y^*_{hi} = \sum_{j=1}^{J} b_{hj}\alpha_{ij} + \eta_{hi} \quad \text{for all } i \tag{6$'$}$$

Note that brand $i = 1$ was always listed in the first position on all versions of the survey. Finally, *Model 4* reflects the "brand halo" effect in equations (7) and (8). Models 1 through 4 all offer improved efficiency over alternatives specifications that rely on only the brand ratings or the brand choice data. Models 2 through 4 also correct for any inconsistencies that may result from ignoring a structural relationship between the observed x's and y's and/or other specific behavioral response patterns in the data.

The model specification is finalized by assuming priors for all parameters. Diffuse but proper priors are specified for α, δ, γ, and λ. Let ψ represent the vector of parameters, then the prior is specified as $\psi \sim N(0, 100\mathrm{I})$ where I is an identity matrix of the proper dimension.

Under these assumptions, a relatively simple algorithm is used to obtain parameter estimates. The following illustration is for Model 2 with the justification parameter. On each iteration, for each individual, each value of x^*_{hij} and y^*_{hi} is drawn as:

$$\text{If} \quad x_{hij} = 1, x^*_{hij} \sim TN(\alpha_{ij} + \delta y_{hi}, \sigma^2_\varepsilon = 1, x^*_{hij} > 0). \tag{13a}$$

$$\text{If} \quad x_{hij} = 0, x^*_{hij} \sim TN(\alpha_{ij} + \delta y_{hi}, \sigma^2_\varepsilon = 1, x^*_{hij} < 0). \tag{13b}$$

$$\text{If} \quad y_{hi} = 1, y^*_{hi} \sim TN\left(\sum_{j=1}^{J} b_{hj}\alpha_{ij}, \sigma^2_\eta = 1, y^*_{hi} > y^*_{hk} \quad \text{for all brands in the set } K\right) \tag{13c}$$

$$\text{If} \quad y_{hi} = 0, y^*_{hi} \sim TN\left(\sum_{j=1}^{J} b_{hj}\alpha_{ij}, \sigma^2_\eta = 1, y^*_{hi} < y^*_{hk} \quad \text{where } y_{hk} = 1\right). \tag{13d}$$

Let q equal the stacked vector of x^* and y^* across all respondents and Z represent the stacked design matrix across all respondents as described in the previous section. Then draws from the posterior distribution of ψ are obtained from:

$$\psi \sim N(\psi^*, (Z'Z + V^{-1})^{-1}) \tag{14a}$$

$$\psi* = (Z'Z + V^{-1})^{-1}(Z'q + V^{-1}\bar{\Psi}) \tag{14b}$$

Table 2. Single-use camera study—Posterior mean of selected parameter estimates (posterior standard deviation).

			Italy	UK	France	US	Japan
Model 2	*Justification*	δ	0.885 (0.01)	0.634 (0.01)	0.567 (0.01)	0.593 (0.01)	0.295 (0.01)
Model 3	*Justification*	δ	0.442 (0.01)	0.379 (0.01)	0.407 (0.01)	0.424 (0.01)	0.365 (0.01)
	1st position	γ	1.531 (0.01)	1.043 (0.01)	0.770 (0.01)	0.756 (0.01)	−0.543 (0.01)
Model 4	*Justification*	δ	0.295 (0.01)	0.305 (0.01)	0.303 (0.01)	0.354 (0.01)	0.061 (0.01)
	Brand 1 halo	λ_1	1.806 (0.01)	1.203 (0.01)	0.788 (0.01)	1.030 (0.01)	0.010 (0.01)
	Brand 2 halo	λ_2	0.230 (0.01)	0.390 (0.01)	0.078 (0.01)	0.207 (0.01)	0.621 (0.01)
	Brand 3 halo	λ_3	0.018 (0.01)	0.510 (0.01)	−0.069 (0.01)		

where $\bar{\Psi} = 0$ and $V = 100\text{I}$ as noted above. Although ψ can include up to 508 elements, when Z is constructed for each country, there are between 70,000 and 200,000 "observations" available for estimation.

This algorithm is a simple Gibbs sampling scheme as described by Gelfand et. al. (1990). The MCMC chain was run for 20,000 iterations for Models 1–3 and 40,000 iterations for Model 4. Convergence was assessed by examining time series plots of the parameter draws across iterations. The last 3,000 iterations from each chain were used for inference.

Selected parameter estimates are reported in Table 2. For sake of brevity, the estimates of α_{ij} are not presented. For all model specifications and across all five countries a positive justification bias (δ) is estimated, suggesting that choosing a brand results in inflated brand ratings data. This effect appears to be pervasive across countries and cultures. Focusing on Model 4, the magnitude of the justification bias is similar across countries, with the exception of Japan. Although the justification bias is positive with posterior mass well away from 0, the posterior mean of 0.06 is much less than the posterior means of the U.S. and European countries at ≈0.30.

Similarly, the parameter estimates for the 1st position bias, γ, and brand halo effects, λ, exhibit a different pattern for Japan than the other countries. Japan is the only country with a negative value for the posterior mean for the 1st position bias in Model 3 and a relatively low value for the brand halo effect for brand 1 in Model 4. This may be explained by the fact that brand 1 in all versions of the survey is an American brand while brand 2 is a Japanese brand. We note, however, that since the same brand occupied the 1st position in all versions of the survey, the brand halo effect is confounded with a position effect as described by Nisbett and Wilson (1977). Although the survey design does not allow one to separate the 1st position bias from the brand halo effect, the parameter estimates support the need to consider specific behavioral response models when analyzing survey data.

Predictive results for alternative models are reported in Table 3. The alternative models include the four endogeneity models discussed above, a simple model that uses the reported attribute importances (b_{hj}) and brand beliefs (x_{hij}), and a standard multinomial probit model that uses reported brand beliefs and estimates the importance weights β. In all models, choice probabilities are computed for each respondent, and then averaged to obtain the predicted market share. Under the column headed "Prob $\Sigma \text{b}_{hj} \text{x}_{hij}$", an individual's choice probability for brand i, p_{hi} is calculated as:

$$p_{hi} = \frac{\sum_j b_{hj} x_{hij}}{\sum_i \sum_j b_{hj} x_{hij}} \tag{15}$$

Table 3. Single-use camera study—Predicted market share using hold-out sample.

		Reported	Prob $\sum b_{hj} x_{hij}$	Probit Model	Model 1	Model 2	Model 3	Model 4
Italy	Brand 1	58.3%	68.2%	59.6%	67.5%	65.5%	57.4%	56.5%
$n = 36$	Brand 2	36.1%	21.5%	23.6%	30.9%	30.4%	35.0%	34.5%
	Brand 3	5.5%	7.8%	10.5%	1.1%	1.5%	2.7%	3.3%
	Brand 4	0.0%	2.5%	6.3%	0.5%	2.6%	4.9%	5.7%
	MAD		7.3	6.3	4.8	4.8	2.4	2.8
U.K.	Brand 1	46.9%	55.3%	48.7%	54.8%	53.4%	46.3%	48.5%
$n = 32$	Brand 2	37.5%	16.6%	20.9%	37.8%	38.0%	43.8%	39.7%
	Brand 3	3.1%	9.5%	8.8%	1.8%	2.2%	2.9%	3.2%
	Brand 4	12.5%	18.5%	21.6%	5.6%	6.3%	7.0%	8.6%
	MAD		10.4	8.3	4.1	3.5	3.1	1.9
France	Brand 1	45.2%	52.5%	43.2%	43.1%	43.2%	35.5%	37.6%
$n = 31$	Brand 2	41.9%	27.8%	36.4%	44.5%	41.9%	46.0%	41.6%
	Brand 3	6.5%	11.7%	13.1%	6.5%	8.9%	11.2%	11.8%
	Brand 4	6.5%	8.3%	7.3%	5.8%	6.0%	7.4%	9.0%
	MAD		7.1	3.8	1.4	1.3	4.9	3.9
U.S.A.	Brand 1	53.5%	49.9%	51.2%	53.0%	51.5%	47.6%	48.5%
$n = 99$	Brand 2	38.4%	33.3%	35.4%	39.6%	39.9%	42.8%	41.2%
	Brand 3	8.1%	15.6%	13.4%	7.4%	8.7%	9.6%	10.4%
	MAD		5.4	3.5	0.81	1.4	3.9	3.4
Japan	Brand 1	40.0%	19.0%	29.2%	37.8%	38.3%	41.6%	43.4%
$n = 75$	Brand 2	53.3%	61.7%	45.8%	54.8%	53.6%	50.7%	47.2%
	Brand 3	6.6%	19.2%	25.0%	7.4%	8.1%	7.7%	9.4%
	MAD		14.0	12.3	1.5	1.2	1.8	4.1

This measure uses the actual responses for each individual and does not use any parameter estimates from the data.

A multinomial probit model is also fit to see if conditioning on the x_{hij}'s and estimating the β_j's yields a better representation of the data than conditioning on the b_{hj}'s and modeling the x_{hij}'s. The model is represented as follows:

$$y_{hi} = 1 \quad \text{if} \quad y^*_{hi} > y^*_{hk} \quad \text{for all brands in the set } K \tag{16a}$$

$$y^*_{hi} = \sum_{j=1}^{J} \beta_j w_{hij} + \eta_{hi} \quad \text{for all } i \tag{16b}$$

where $\eta \sim N(0, 1)$ and w_{hij} are effects-coded such that if $x_{hij} = 1, w_{hij} = 0.5$, else $w_{hij} = -0.5$. Diffuse but proper priors were specified for β and standard methods used to estimate the parameters. A sample of 300 draws from the posterior distribution of β was used to calculate choice probabilities for the hold-out sample. Choice probabilities were calculated using the GHK simulator (Keane, 1994; Hajivassilious et al., 1996).

Predictions for models 1–4 are also presented in Table 3. Similar to the probit model, a sample from the posterior distribution of α_{ij} was used to compute the distribution of the latent variable y^*:

$$y^*_{hi} = \sum_{j=1}^{J} b_{hj}\alpha_{ij} + \eta_{hi} \quad \text{for all } i \tag{17}$$

with $\eta \sim N(0,1)$. The predictive market shares were estimated using the GHK simulator to obtain the choice probabilities, which were averaged across draws and respondents.

The results show strong evidence that explicitly modeling the simultaneous relationship between survey responses leads to improved predictions of market place behavior. In all five countries, modeling the simultaneity yielded a smaller mean absolute deviation (MAD) between the reported and predicted market shares than either models that condition on the brand belief data. Only in France and the U.S. did the out-of-sample predictions from the multinomial probit model compare to those for the endogeneity models. In-sample fit using the MAD metric (not shown, available from the authors) favored Model 4 over the multinomial probit model in all five countries. As discussed earlier, models that estimate attribute weights (e.g. the multinomial probit) are motivated by the viewpoint that respondents cannot provide meaningful responses to those questions. These results challenge that viewpoint and suggest that in survey research it's not necessarily the attribute weights, the b_{hj}'s that are suspect, it's the brand beliefs, the x_{hij}'s.

Figure 2 displays the effect of modeling behavioral response patterns on the brand beliefs (α_{ij}) using Model 4 for the U.S. data. Plotted is the actual frequency that the brand "definitely

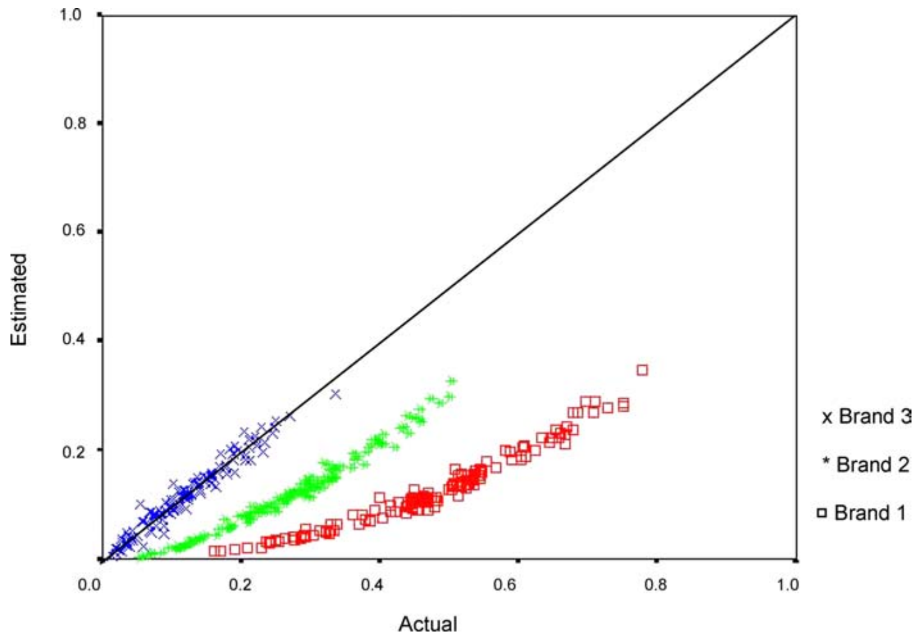

Figure 2. Single use camera study—Actual and estimated brand belief frequency, by brand for the U.S.

had" the attribute (x_{hij}) on the horizontal axis, versus the estimated frequency of the brand belief after controlling for the justification (δ) and halo (λ) effects. This estimated frequency is calculated as Pr ($x^*_{ij} = \alpha_{ij} + \varepsilon_{hij} > 0$). Also displayed in the plot is a 45 degree line that indicates agreement between the actual and estimated frequencies. Each brand is represented by a different symbol in the plot.

The estimated frequency of brand beliefs is lower for brands 1 and 2, and the same for brand 3. This result is consistent with the parameter estimates reported in Table 3, where the justification and brand halo effects have positive algebraic signs. The actual versus estimated beliefs for brand 1 are the most different, and the effects have a non-linear influence on the reported beliefs. Moreover, the effects are large, with frequency decreasing from 0.6 to 0.2 for many attributes. The impact of the justification and halo effects are also present, but not as dramatic for brand 2.

Figure 3 demonstrates the impact of justification and halo effects on brand positioning analysis. The figure displays "snake plots" used by practitioners to conduct gap analysis and identify brand strengths. The upper portion of Figure 3 displays the actual frequency that a brand "definitely had" an attribute, and the lower portion displays the estimated frequency, for the five most important attributes as identified by the average brand importance rating, b_{hj}. Relying on just reported x_{hij}, Brand 1 appears to dominate on all the most important product attributes. However, when justification and halo effects are accounted for, no single brand dominates the market on any one of the most important attributes.

The results imply that use of the raw brand belief data, x_{hij}, leads to misleading results in brand positioning analysis. Justification and brand halo effects were found across all five countries, and ignoring these effects can result in incorrect inferences. The data augmentation method outlined in Section 3 was shown to be viable even in a study involving over 500 parameters. The next study involves a much smaller set of attributes and does not include self-explicated attribute weights.

5. Digital cameras

The second empirical example uses data from a U.S. based study on consumer perceptions of a variety of product categories related to photography. The data is collected as part of an ongoing tracking study and is administered by a professional market research company via the Internet. Data from a consecutive 18-month period were provided for analysis. A random sample of 319 responses was selected from survey participants who provided information on digital cameras. For confidentiality, the brands and positioning attributes are not disclosed. Each respondent indicated for each of 7 brands whether they "would most likely" consider that brand on their next purchase occasion. Respondents also rated each brand on 14 brand positioning attributes and indicated whether or not the attribute "is strongly associated" with the brand. For each respondent for each brand, the data consists of binary indicator of whether or not the brand would be considered, and a vector of 14 other binary variables indicating whether or not the brand possesses certain attributes. The actual percentage of respondents indicating they would consider each brand is displayed in Table 5; Table 6 summarizes the attribute perception data for each brand. This data is used to investigate the simultaneity in brand perception and brand consideration questions.

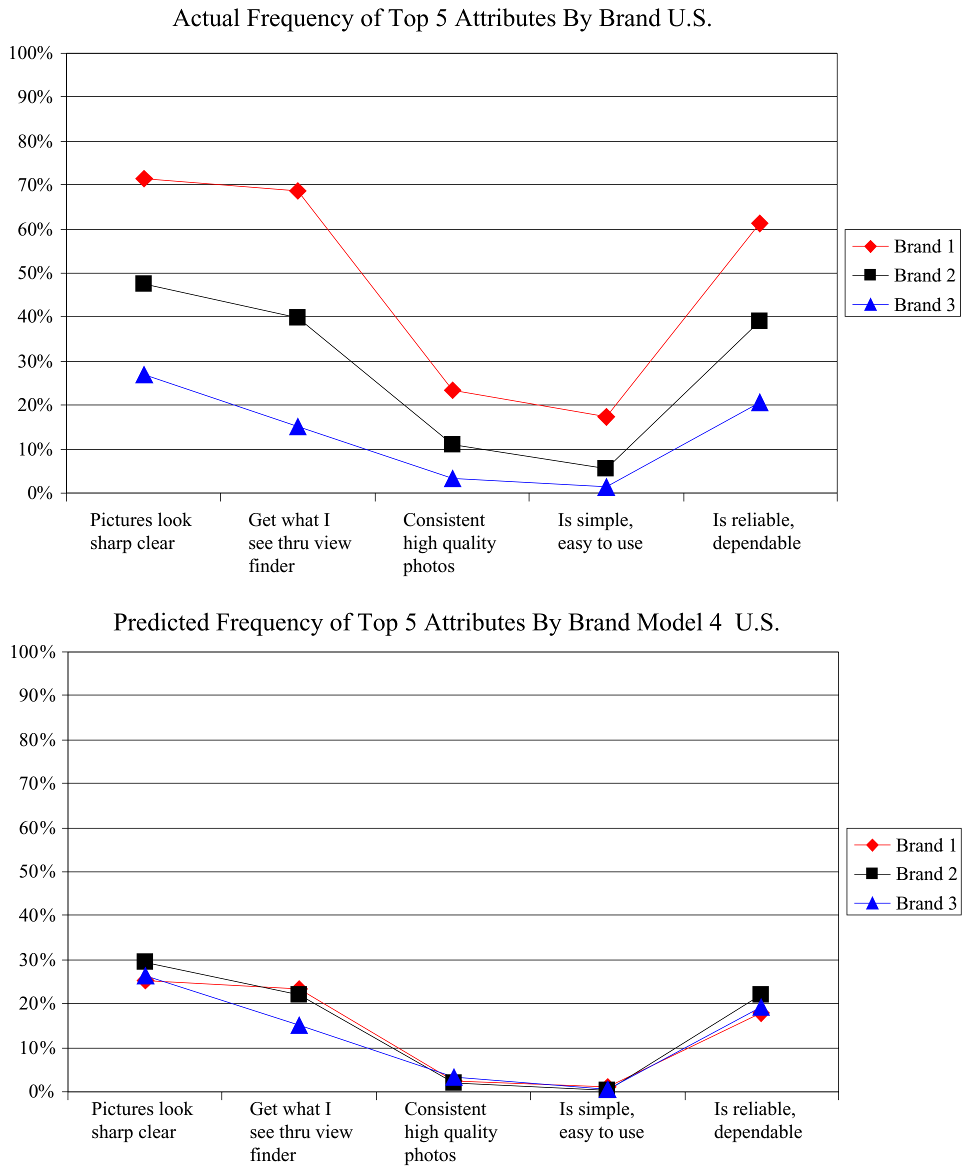

Figure 3. Single use camera study—Brand positioning analysis.

Three different simultaneity models are investigated. Unlike the single-use camera study, self-explicated attribute importance weights, b_{hj} are not available and common weights, β_j, are estimated. Using notation similar to the earlier presentation, let $x_{hij} = 1$ if person h indicates that attribute j is strongly associated with brand i, otherwise $x_{hij} = 0$. Let $y_{hi} = 1$ if person h indicates that they would most likely consider brand i on their next purchase occasion, otherwise $y_{hi} = 0$. Note that this is a binary outcome for consideration as compared to the single-use camera study that had a multinomial brand choice outcome. Using standard latent variable techniques let $x_{hij} = 1$ if $x^*_{hij} > 0$ and $y_{hi} = 1$ if $y^*_{hi} > 0$. Model 1a represents the relationship between brand perception and brand consideration as:

$$x^*_{hij} = \alpha_{ij} + \varepsilon_{hij} \quad \text{for all } i, j \tag{1a}$$

$$y^*_{hi} = \sum_{j=1}^{J} \beta_j(\alpha_{ij} + \varepsilon_{hij}) + \eta_{hi} \quad \text{for all } i \tag{2a}$$

where the common parameters α capture endogeneity. In contrast to equation (2), equation (2a) includes the error term ε_{hij} in the model for brand consideration (see equations (3a) and (4a) above).

Two additional models incorporate a structural relationship between responses to consideration and brand perception questions. *Model 2a* incorporates a justification bias and is represented by equations (3a) and (4a) above. The digital camera study was administered via computer and the presentation of all brands and attributes were randomly rotated across respondents. Therefore, no analog to Model 3 with a 1st position bias from the single-use camera study is estimated for this data. However, *Model 4a* incorporates brand halo effects and is represented as:

$$x^*_{hij} = \alpha_{ij} + \delta y_{hi} + \sum_{k=1}^{K-1} \lambda_k I(k = i)_{hij} + \varepsilon_{hij} \quad \text{for all } i, j \tag{5a}$$

$$y^*_{hi} = \sum_{j=1}^{J} \beta_j(\alpha_{ij} + \varepsilon_{hij}) + \eta_{hi} \quad \text{for all } i \tag{6a}$$

Model estimation is carried out using data augmentation and MCMC methods described in Section 3 and detailed in the appendix. Two additional considerations are required to estimate Models 1a, 2a, and 4a. First, as in all binary probit models, the parameters β are identified only up to a multiplicative constant e.g. $\beta_j\sqrt{\sigma^2_{y*}}$ where σ^2_{y*} = variance (y^*_{hi}). From equations (2a), (4a), and (6a) the variance $(y^*_{hi}) = \sum \beta_j^2 + 1$. We follow the approach suggested by Edwards and Allenby (2004) and estimate the unidentified model, and then normalize the posterior draws of β by dividing each by $\sqrt{\sigma^2_{y*}}$.

Performing data augmentation at the level of ε and η, as opposed to x^* and y^*, results in an MCMC chain that is highly autocorrelated. In general, Metropolis-Hastings (M-H) algorithms yield autocorrelated samples and caution must be used in calculating posterior moments from samples generated using this method (Geyer, 1992). In the current model, conditional on the augmented error terms, the likelihood function used to draw posterior values of α, δ, λ, or β in the M-H step reduces to an indicator function. This rather uninformative likelihood exacerbates the autocorrelation between draws. A practical approach to dealing with autocorrelated draws from M-H algorithms is to retain every nth value from the MCMC chain. We adopt this approach and allow the sampler to complete many iterations in order to achieve coverage of the posterior distribution of the parameters.

The MCMC chain was run for 1,000,000 iterations for Models 1a and 2a, and 2,000,000 iterations for Model 4a. Every 50th iteration was retained for Models 1a and 2a; every 100th iteration was retained for Model 4a. The last 5,000 draws from each chain were used for inference on the posterior moments. Convergence was assessed by examining time series plots of the parameter draws across iterations.

In addition to the models incorporating simultaneity, a standard binary probit model was fit to the data. The model is represented as:

$$y_{hi} = 1 \quad \text{if } y^*_{hi} > 0 \tag{18a}$$

$$y^*_{hi} = \sum_{j=1}^{J} \beta_j w_{hij} + \eta_{hi} \tag{18b}$$

where $\eta \sim N(0, 1)$ and w_{hij} are effects-coded such that if $x_{hij} = 1$, $w_{hij} = 0.5$, else $w_{hij} = -0.5$. Standard data augmentation and MCMC methods are used to estimate the binary probit model. Note that since the variance$(y^*_{hi}) = 1$, this model does not have the identification problem for β as described for Models 1a, 2a, and 4a.

Table 4 shows posterior means for selected parameters for the four models. The values of β in Table 4 for Models 1a, 2a, and 4a have been normalized for the error variance as discussed above. Consistent with the single-use camera study, a relatively large justification bias is detected in Model 2a and also in Model 4a where brand halo effects are also measured.

Table 5 compares the in-sample fit of predicted consideration against the actual frequency for each brand across the four models. Similar to the single-use camera study, a sample from the posterior distribution of the parameters was used to calculate Pr $(y^*_{hi} > 0)$. Note that for Models 1a, 2a, and 4a, $y^*_{hi} \sim N(\sum_{j=1}^{J} \beta_j \alpha_{ij}\ , \sum \beta_j^2 + 1)$. Based on the mean absolute deviation (MAD) between the aggregate actual and predicted consideration, the results favor Model 4a. Since no individual level data is used to make predictions, out-of-sample fit would match in-sample fit assuming both samples are equivalent representations of the population. If the data used to calibrate the models is not representative, then the out-of-sample predictions will be less accurate.

Table 6 and Figure 4 illustrate the biasing effects of justification and brand halo effects on inferences of brand perceptions. Table 6 shows the actual frequency of brand attribute ratings and the estimated brand attribute ratings using Model 4a. The most striking disparity between the actual and estimated brand perceptions is for brands 3 and 5. Table 5 shows that brands 3 and 5 have the lowest percentage of respondents who would consider them, 15.4% and 13.5%, roughly one-half to one-third of the other brands. Yet the actual frequencies for the brand attribute ratings in the top half of Table 6 do not reflect this order of magnitude of discounting. The simultaneity models rescale brand perceptions to be consistent with a common model of brand consideration and a model of behavioral responses that result in reported brand perceptions. Accordingly, Table 4 shows that brands 3 and 5 have relatively high brand halo effects of 1.8 and 1.2. After accounting for these brand halo effects and the justification bias, Table 6 and Figure 5 show that the estimated frequency for the brand attribute data is shrunk down to less than 4% for each attribute for these two brands. Thus for brand 3 and 5, virtually all of the reported brand perceptions are driven by justification and brand halo effects.

Table 4. Digital camera study—Posterior means of parameters (posterior standard deviation).

Binary Probit		Model 1a		Model 2a		Model 4a	
Attributes	β's	Attributes	β's	Attributes	β's	Attributes	β's
1	0.005 (0.08)	1	0.020 (0.05)	1	−0.130 (0.07)	1	−0.077 (0.05)
2	**0.242** (0.08)	2	**0.175** (0.05)	2	0.058 (0.04)	2	**0.092**(0.04)
3	0.148 (0.09)	3	0.042 (0.06)	3	0.031 (0.03)	3	0.079 (0.04)
4	−0.052 (0.09)	4	−0.045 (0.05)	4	− **0.145** (0.05)	4	−**0.114**(0.04)
5	**0.576** (0.08)	5	**0.259** (0.04)	5	**0.260** (0.02)	5	**0.174** (0.04)
6	**0.299** (0.09)	6	**0.125** (0.04)	6	**0.134** (0.03)	6	**0.089** (0.04)
7	0.006 (0.09)	7	−0.017 (0.05)	7	−0.049 (0.04)	7	−0.041 (0.05)
8	**0.177** (0.09)	8	**0.127** (0.05)	8	0.000 (0.05)	8	−0.004 (0.04)
9	**0.415** (0.09)	9	**0.261** (0.04)	9	**0.217** (0.03)	9	**0.237** (0.04)
10	0.035 (0.09)	10	0.025 (0.06)	10	0.026 (0.04)	10	−0.017 (0.04)
11	−0.012 (0.09)	11	−0.052 (0.04)	11	−**0.116** (0.04)	11	−**0.122** (0.04)
12	0.150 (0.09)	12	0.069 (0.04)	12	0.000 (0.03)	12	−0.019 (0.04)
13	**0.384** (0.08)	13	**0.237** (0.05)	13	**0.113** (0.03)	13	**0.132** (0.04)
14	−0.018 (0.09)	14	−0.046 (0.04)	14	−**0.070** (0.02)	14	−**0.084** (0.04)
				Justification		Justification	
				δ	**1.071** (0.02)	δ	**1.057** (0.02)
						Brand halo's	
						λ_1	0.000
						λ_2	**0.442** (0.11)
						λ_3	**1.811** (0.06)
						λ_4	−**0.501** (0.04)
						λ_5	**1.176** (0.07)
						λ_6	**0.201** (0.05)
						λ_7	0.163 (0.16)

Estimates in bold have more than 95% of the posterior mass away from 0. λ_1 set $= 0$ in Model 4a for identification. Estimates for β for Model 1a, 2a, and 4a are normalized for the error variance, see text for a full explanation.

Table 5. Digital camera study—In–sample fit comparison.

Brand	Actual consider (%)	**Predicted** Binary Probit (%)	Model 1a (%)	Model 2a	Model 4a (%)
1	45.8	48.9	50.5	43.3	46.4
2	31.3	30.8	23.7	32.4	30.8
3	15.4	32.6	23.7	31.9	18.4
4	44.5	39.2	37.4	38.6	46.2
5	13.5	25.2	13.8	26.2	18.8
6	37.6	37.1	33.2	38.8	37.7
7	35.4	41.8	38.0	37.2	37.3
	MAD	6.4	5.0	6.0	1.9

Table 6. Digital camera study—Actual and estimated attribute ratings, Model 4a.

	Attributes													
Brands	1 (%)	2 (%)	3 (%)	4 (%)	5 (%)	6 (%)	7 (%)	8 (%)	9 (%)	10 (%)	11 (%)	12 (%)	13 (%)	14 (%)
	Actual frequency													
1	49.2	59.2	57.4	51.7	35.7	46.7	56.7	58.0	50.5	51.7	30.4	34.5	57.1	46.1
2	33.5	32.3	21.9	22.6	17.6	21.0	24.8	31.7	28.8	27.6	19.7	23.8	30.1	25.4
3	36.7	37.3	33.2	34.5	15.7	24.8	32.9	31.7	31.3	25.4	24.1	23.5	35.7	31.3
4	42.0	49.2	31.3	32.9	27.6	29.2	32.6	44.2	40.4	37.3	25.7	29.8	40.4	32.6
5	29.2	24.1	17.6	20.1	11.0	13.2	19.1	19.7	21.6	18.2	19.4	23.2	21.6	22.9
6	33.2	45.1	27.6	29.8	25.4	31.7	32.0	45.1	34.5	35.4	21.9	27.0	35.4	23.5
7	50.8	55.2	40.1	34.2	28.5	35.1	32.9	41.4	45.8	36.1	43.9	47.0	37.9	52.0
	Estimated													
1	31.2	42.2	39.6	33.8	18.5	28.4	39.2	40.4	33.3	33.8	14.6	17.6	40.2	27.9
2	19.8	18.1	10.0	11.6	6.5	9.5	12.6	17.9	14.8	14.5	9.5	11.6	16.0	13.0
3	3.5	3.5	2.8	3.2	0.5	1.4	2.8	2.5	2.4	1.5	1.5	1.3	3.2	2.5
4	43.4	51.3	30.6	33.1	25.3	27.9	32.4	45.0	40.5	37.8	25.9	29.0	40.7	32.4
5	2.9	1.9	1.0	1.4	0.4	0.5	1.2	1.3	1.5	1.2	1.4	1.8	1.5	1.8
6	24.1	36.3	18.4	21.4	15.7	22.1	22.8	36.4	24.4	25.9	14.6	18.0	25.5	15.3
7	37.1	41.8	25.8	21.0	15.1	21.2	19.7	27.7	31.6	22.6	29.6	33.1	23.4	38.6

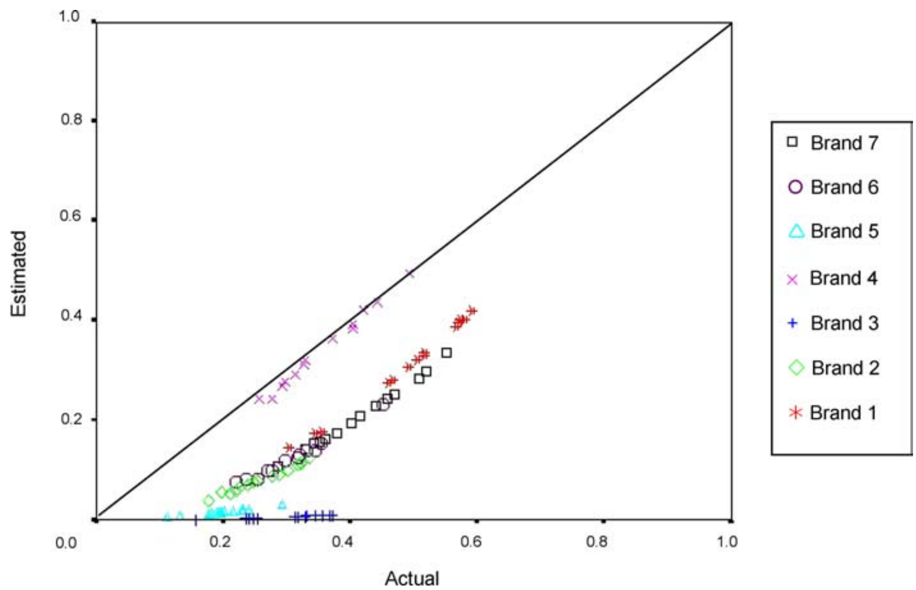

Figure 4. Digital camera study—Actual and estimated brand belief frequency.

6. Discussion

This paper provides a general methodology for simultaneously modeling survey response data by relating different measures to common parameters. We illustrate the method by modeling brand choice/consideration and reported brand attribute ratings as functions of common parameters and control for justification, order, and brand halo biases in the attribute data. If only the brand choice/consideration responses were used to measure brand attributes, there would have been far fewer "observations" per parameter. Greater statistical efficiency is achieved by using both brand attribute ratings and brand choice/consideration data. Consistent parameter estimates are obtained by explicitly modeling the structural relationship between reported brand choice/consideration and reported brand attribute ratings. The proposed approach incorporates common parameters and structural parameters that shift the mean of the distribution of the data as opposed to instrumental variable approaches which model endogeneity through correlated error terms.

The empirical results demonstrate the importance of incorporating simultaneity and specific behavioral response models in the analysis of survey data. In the single-use camera study, across five different countries, respondents demonstrated a justification bias wherein brand ratings for the chosen brand were inflated. A strong brand halo/1st position bias was also measured in all five countries. Results from the single-use camera study call into question the view that self-explicated attribute importance weights are suspect but hold brand perception data sacrosanct. Strong justification bias and brand halo effects were also measured in the digital camera study where both attribute importance weights and brand perceptions were estimated with the model. In both studies, very different conclusions were drawn about attribute based brand positioning after controlling for survey response biases. The simultaneous models also provided better in-sample and out-of-sample fit for predicted brand choice and brand consideration than standard models. Inference was facilitated by MCMC methods and data augmentation. These Bayesian methods reduce a complicated multivariate binomial and multinomial probit model to taking draws from truncated normal distributions and doing regression or evaluating indicator functions.

The analysis of survey data frequently proceeds with simple summaries of the responses, or latent variable models that deal with the discreteness of using fixed-point rating scales (see Rossi et al., 2001). Our analysis illustrates the benefit of employing quantitative models that better reflect how consumers respond to survey questions, which makes use of multiple sets of responses. Survey response models with endogeneity offer a fruitful area of research for marketers in representing the true data generating mechanism.

Appendix: Estimation algorithm

This appendix details the estimation algorithm used for *Model 2a* in the digital camera study, a model that controls for a possible "justification" bias in brand ratings. In this model, attribute importance weights, β_j's, are estimated from the data. Each respondent indicated if they would be "most likely to consider" each of 7 brands and provided dichotomous ratings of each of the 7 brands on 14 attributes. The model is represented as:

$$x_{hij} = 1 \quad \text{if} x^*_{hij} > 0, \quad \text{otherwise}\, x_{hij} = 0 \tag{A1}$$

$$y_{hi} = 1 \quad \text{if} y^*_{hi} > 0, \quad \text{otherwise}\, y_{hi} = 0 \tag{A2}$$

$$x^*_{hij} = \alpha_{ij} + \delta y_{hi} + \varepsilon_{hij} \quad \varepsilon \sim N\big(0, \sigma^2_\varepsilon = 1\big) \tag{A3}$$

$$y_{hi}^{*} = \sum_{j=1}^{J} \beta_j(\alpha_{ij} + \varepsilon_{hij}) + \eta_{hi} \quad \eta \sim N\big(0, \sigma_\eta^2 = 1\big) \tag{A4}$$

where h indexes the respondent, i the brand, and j the product attribute. $x_{hij} = 1$ means that respondent h indicated that brand i has attribute j; $y_{hi} = 1$ means that respondent h indicated that he/she would most likely consider brand i if he/she was looking to purchase a digital camera. Diffuse priors are specified for the parameters:

$$\beta \sim \text{MVN}(0, 100I) \quad \alpha \sim MVN(0, 100I) \quad \delta \sim N(0, 100)$$

where β is a vector of length J (14 in this data set) and α is a vector of length $I \times J$ (7 $\times 14 = 98$). As noted in the text, data augmentation is used to estimate the model. However, the augmented variables are ε and η as compared to other applications where x^* and y^* are the augmented variables. The following steps describe an MCMC chain with the posterior distribution of all model parameters as the stationary distribution.

1. *Generate* $\varepsilon_{hij} \mid x_{hij}, y_{hi}, \alpha, \beta, \delta, \{\varepsilon_{hi-j}\}, \eta_{hi}, \sigma_\varepsilon^2 = 1$ for $h = 1, \ldots, H, i = 1, \ldots, I,$ and $j = 1, \ldots J$

   A rejection sampler is used to obtain a value of ε_{hij} consistent with the observed data and all other parameters and augmented variables.

   (a) if $x_{hij} = 1$ then $\varepsilon_{hij} \sim TN(0, \sigma_\varepsilon^2 = 1, \varepsilon_{hij} > -\alpha_{ij} - \delta y_{hi})$
   if $x_{hij} = 0$ then $\varepsilon_{hij} \sim TN(0, \sigma_\varepsilon^2 = 1, \varepsilon_{hij} < -\alpha_{ij} - \delta y_{hi})$
   (b) Test ε_{hij}: form $y_{hi}^* = \sum_{j=1}^{J} \beta_j(\alpha_{ij} + \varepsilon_{hij}) + \eta_{hi}$
   if $y_{hi} = 1$ and $y_{hi}^* < 0$ then reject, and return to a)
   if $y_{hi} = 0$ and $y_{hi}^* > 0$ then reject, and return to a)
   (c) else, accept ε_{hij}

2. *Generate* $\eta_{hi} | y_{hi}, \{\varepsilon_{hij}\}, \alpha, \beta$ for $i = 1, \ldots I$ and $h = 1, \ldots, H$

   if $y_{hi} = 1$ then $\eta_{hi} \sim TN(0, \sigma_\eta^2 = 1, \eta_{hi} > \sum_{j=1}^{J} -\beta_j(\alpha_{ij} + \varepsilon_{hij}))$
   if $y_{hi} = 0$ then $\eta_{hi} \sim TN(0, \sigma_\eta^2 = 1, \eta_{hi} < \sum_{j=1}^{J} -\beta_j(\alpha_{ij} + \varepsilon_{hij}))$

3. *Generate* $\alpha_{ij} | \{x_{hij}\}, \{y_{hi}\}, \{\varepsilon_{hij}\}, \{\eta_{hi}\}, \{\alpha_{i-j}\}, \beta, \delta$ for $i = 1, \ldots I$ and $j = 1, \ldots, J$

   A random walk Metropolis-Hastings algorithm is used to draw a sample from the posterior distribution of α. Let $\alpha_{ij}^{(n)} = \alpha_{ij}^{(\circ)} + e$ where $\alpha_{ij}^{(n)}$ is new candidate value for α_{ij} and $\alpha_{ij}^{(\circ)}$ is the value from the previous iteration of the chain. $e \sim N(0, \sigma_e)$ where σ_e is chosen so that the acceptance rate is between 25 and 50%.

   The posterior distribution of $\alpha_{ij} \propto [l(\alpha_{ij}) \times \pi(\alpha_{ij})]$ where $l(\alpha_{ij})$ is the likelihood and $\pi(\alpha_{ij})$ is the prior. The likelihood of $\alpha_{ij}^{(n)}$ conditional on the data, the augmented error terms, and the other parameters reduces to an indicator function; either the candidate value of α_{ij} is consistent with the data, or it is not. Let $l(\alpha_{ij}^{(n)}) = I(\alpha_{ij}^{(n)}) = 1$ if $\alpha_{ij}^{(n)}$ is consistent with the observed data, otherwise $l(\alpha_{ij}^{(n)}) = I(\alpha_{ij}^{(n)}) = 0$. The Metropolis-Hastings algorithm proceeds in two steps:

(a) For $h = 1,\ldots H$

(i) form $y_{hi}^* = \sum_{\substack{k=1 \\ k \neq j}}^{J-1} \beta_k(\alpha_{ik} + \varepsilon_{hik}) + \beta_j(\alpha_{ij}^{(n)} + \varepsilon_{hij}) + \eta_{hi}$
if $y_{hi} = 1$ and $y_{hi}^* < 0$ then $\mathrm{I}(\alpha_{ij}^{(n)}) = 0$
if $y_{hi} = 0$ and $y_{hi}^* > 0$ then $\mathrm{I}(\alpha_{ij}^{(n)}) = 0$
(ii) form $x_{hij}^* = \alpha_{ij}^{(n)} + \delta y_{hi} + \varepsilon_{hij}$
if $\mathrm{x}_{hij} = 1$ and $x_{hij}^* < 0$ then $\mathrm{I}(\alpha_{ij}^{(n)}) = 0$
if $\mathrm{x}_{hij} = 0$ and $x_{hij}^* > 0$ then $\mathrm{I}(\alpha_{ij}^{(n)}) = 0$
(iii) else, $\mathrm{I}(\alpha_{ij}^{(n)}) = 1$

(b) Accept $\alpha_{ij}^{(n)}$ with probability:

$$\min: \left(\frac{I\left(\alpha_{ij}^{(n)}\right) \times \exp\left(\left[-(2 \times 100)^{-1}\left(0 - \alpha_{ij}^{(n)}\right)\right]\right)^2}{I\left(\alpha_{ij}^{(o)}\right) \times \exp([-(2 \times 100)^{-1}\left(0 - \alpha_{ij}^{(o)}\right)^2}, 1 \right)$$

4. *Generate* $\delta \mid \{x_{hij}\}, \{y_{hi}\}, \{\varepsilon_{hij}\}, \alpha$

A random walk Metropolis-Hastings algorithm is used. Let $\delta^{(n)} = \delta^{(o)} + u$ where $u \sim N(0, \sigma_u)$ and σ_u is chosen to so that the acceptance rate is approximately 50%. Similar to drawing α, the likelihood conditional on the other parameters and augmented variables reduces to an indicator function $l(\delta^{(n)}) = I(\delta^{(n)})$. The Metropolis-Hastings algorithm proceeds in two steps:

(a) For $h = 1,\ldots H$, $i = 1,\ldots,I$, and $j = 1,\ldots J$

(i) form $x_{hij}^* = \alpha_{ij} + \delta^{(n)} y_{hi} + \varepsilon_{hij}$
if $\mathrm{x}_{hij} = 1$ and $x_{hij}^* < 0$ then $\mathrm{I}(\delta^{(n)}) = 0$
if $\mathrm{x}_{hij} = 0$ and $x_{hij}^* > 0$ then $\mathrm{I}(\delta^{(n)}) = 0$
(ii) else, $\mathrm{I}(\delta^{(n)}) = 1$

(b) Accept $\delta^{(n)}$ with probability:

$$\min: \left(\frac{I\left(\delta^{(n)}\right) \times \exp([-(2 \times 100)^{-1}\left(0 - \delta^{(n)}\right)^2}{I\left(\delta^{(o)}\right) \times \exp([-(2 \times 100)^{-1}\left(0 - \delta^{(o)}\right)^2}, 1 \right)$$

5. *Generate* $\beta_j | \{y_{hi}\}, \{\varepsilon_{hij}\}, \{\eta_{hi}\}, \alpha, \beta_{-j}$ for $j = 1, \ldots, J$

Again, a random walk Metropolis-Hastings algorithm is used. Let $\beta_j^{(n)} = \beta_j^{(o)}$+ w where w $\sim$ N(0, σ_w) and σ_w is chosen to so that the acceptance rate is between 25% and 50%. The likelihood conditional on the other parameters and augmented variables reduces to an indicator function $l(\beta_j^{(n)}) = I(\beta_j^{(n)})$. The Metropolis-Hastings algorithm proceeds in two steps:

(a) For $h = 1, \ldots h$ and $i = 1, \ldots, I$

(i) form $y_{hi}^* = \sum_{k=1, k \neq j}^{J-1} \beta_k(\alpha_{ik} + \varepsilon_{hik}) + \beta_j^{(n)}(\alpha_{ij} + \varepsilon_{hij}) + \eta_{hi}$
if $y_{hi} = 1$ and $y_{hi}^* < 0$ then $I(\beta_j^{(n)}) = 0$
if $y_{hi} = 0$ and $y_{hi}^* > 0$ then $I(\beta_j^{(n)}) = 0$
(ii) else, $I(\beta_j^{(n)}) = 1$

(b) Accept $\beta_j^{(n)}$ with probability:

$$\min : \left(\frac{I\left(\beta^{(n)}\right) \times \exp([-(2 \times 100)^{-1}\left(0 - \beta^{(n)}\right)^2}{I\left(\beta^{(o)}\right) \times \exp([-(2 \times 100)^{-1}\left(0 - \beta^{(o)}\right)^2}, 1 \right)$$

Acknowledgment

The authors thank Michael Lotti and Eastman Kodak Company for providing access to the data.

References

Albert, James H. and Siddhartha Chib. (1993). "Bayesian Analysis of Binary and Polychotomous Response Data." Journal of the American Statistical Association 88, 669–679.

Allenby, Greg M., Nereraj Arora, and L. James Ginter. (1995). "Incorporating Prior Knowledge in the Analysis of Conjoint Studies." Journal of Marketing Research 32 (May), 152–162.

Baumgartner, Hans and E.M. Jan-Benedict Steenkamp. (2001). "Response Styles in Marketing Research: A Cross-National Investigation." Journal of Marketing Research 38 (May), 143–156.

Brown, Christina L. and Fred Feinberg. (2004). "Bolstering: How Does Choice Distort Product Evaluations?" Working paper, The University of Michigan.

Cattin, Philippe and Dick R. Wittink. (1982). "Commercial Use of Conjoint Analysis: A Survey." Journal of Marketing 46 (Summer), 44–53.

Day, George S. (1972). "Evaluating Models of Attitude Structure." Journal of Marketing Research 9 (August), 279–286.

Dillon, William R., Thomas J. Madden, Amna Kirmani, and Soumen Mukherjee. (2001). "Understanding What's in a Brand Rating: A Model for Assessing Brand and Attribute Effects and Their Relationship of Brand Equity." Journal of Marketing Research 38 (November), 415–429.

Edwards, Yancy D. and Greg M. Allenby. (2003). "Multivariate Analysis of Multiple Response Data." Journal of Marketing Research 40 (August), 321–334.

Engle, Robert F., David F. Hendry, and Jean-Francois Richard. (1983). "Exogeneity." Econometrica 51, 277–304.

Gelfand, Alan E., Susan E. Hills, Amy Racine-Poon, and F. M. Adrian Smith. (1990). "Illustration of Bayesian Inference in Normal Data Models Using Gibbs Sampling." Journal of the American Statistical Association 85, 972–985.

Geyer, Charles J. (1992). "Practical Markov Chain Monte Carlo." Statistical Science 7 (November), 473–483.

Green, Paul E. and Vithala R. Rao. (1971). "Conjoint Measurement for Quantifying Judgmental Data." Journal of Marketing Research 8 (August), 355–363.

Hajivassiliou, V., D. McFadden, and P. Rudd. (1996). "Simulation of Multivariate Normal Rectangle Probabilities and their Derivatives." Journal of Econometrics 72, 85–134.

Janiszewski, Chris, Tim Silk, and Alan D. J. Cooke. (2003). "Different Scales for Different Frames: The Role of Subjective Scales and Experience in Explaining Attribute Framing Effects." Journal of Consumer Research 30, 311–325.

Keane, M. (1994). "A Computationally Practical Simulation Estimator for Panel Data." Econometrica 62, 95–116.

Maddala, G.S. (1983). Limited Dependent and Qualitative Variables in Econometrics. New York: Cambridge University Press.

Maddala, G.S. and Lung-Fei Lee, (1976). "Recursive Models with Qualitative Endogenous Variables." Annals of Economic and Social Measurement 5/4, 525–545.

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

Nisbett, R. and T. D. Wilson. (1977). "Telling More Than We Know: Verbal Reports on Mental Processes." Psychological Review 84, 231–259.

Pettibone, Jonathan C. and Douglas H. Wedell. (2000). "Examining Models of Nondominated Decoy Effects across Judgment and Choice." Organizational Behavior and Human Decision Processes 81 (March), 300–328.

Rossi, Peter E. and Greg M. Allenby. (2003). "Bayesian Statistics and Marketing." *Marketing Science* 22, 304–328.

Rossi, Peter E., Zvi Gilula, and Greg M. Allenby. (2001). "Overcoming Scale Use Heterogeneity: A Bayesian Hierarchical Approach." Journal of the American Statistical Association 96, 20–31.

Rossi, Peter E., Robert E. McCulloch, and Greg M. Allenby. (1996). "The Value of Purchase History Data in Target Marketing." Marketing Science 15, 321–340.

Simonson, Itamar. (1989). "Choice Based on Reasons: The Case of Attraction and Compromise Effects." Journal of Consumer Research 16 (September), 158–174.

Tanner, M. A. and W. Wong (1987) "The Calculation of Posterior Distributions by Data Augmentation." Journal of the American Statistical Association 82, 528–549.

Wertenbroch, Klaus and Bernd Skiera. (2002). "Measuring Consumers Willingness to Pay at the Point of Purchase." Journal of Marketing Research 39, 228–241.

Yang, Sha, Yuxin Chen, and Greg Allenby. (2003). "Bayesian Analysis of Simultaneous Demand and Supply." Quantitative Marketing and Economics 1 (September), 251–275.

Zellner, Arnold. (1971). An Introduction to Bayesian Inference in Econometrics. New York: John Wiley & Sons, Inc.