ANAND V. BODAPATI and AIMEE DROLET*

In the usual multinomial choice model, consumers choose to use "actual value" information; that is, utilities are continuous functions of product attributes (e.g., choices depend on actual magnitudes of price differences). The authors propose an alternative model in which consumers use only "ordered value" information; that is, utilities are functions only of the relative orderings of the attributes' values across alternatives (e.g., choices depend only on the price ordering and not on actual prices). The ordered value model is attractive because it fits well with psychological evidence that consumers often favor decision mechanisms that are cognitively less demanding. Using a supermarket shopper panel data set, the authors evaluate four models in which (1) all consumers use actual values; (2) all consumers use ordered values; (3) some consumers use actual values all the time, and some consumers use ordered values all the time; and (4) all consumers use both actual values and ordered values but with different propensities. In the analysis, the ordered value model finds stronger support than the actual value model: Model 1 outperforms Model 2; in the two hybrid choice models (Models 3 and 4), ordered value processing is more prevalent than actual value processing. These results suggest that consumers in some product categories engage more heavily in ordered value processing than in actual value processing.

A Hybrid Choice Model That Uses Actual and Ordered Attribute Value Information

The traditional assumption derived from both marketing and economics is that consumers' predefined preferences or utility functions determine both their attitudinal and behavioral responses to changes in products' actual attribute values (for a recent review, see Gustafsson, Herrmann, and Huber 2000). Techniques that are designed to predict consumer choices, notably conjoint analysis, are founded on this assumption. However, contrary to this assumption, growing research suggests that consumers frequently do not have predefined preferences but instead construct their preferences as needed to make choices (for a review, see Bettman, Luce, and Payne 1998).

To construct their preferences, consumers use a variety of choice rules. The use of different rules is contingent on various factors of the decision problem (e.g., framing, context) and of the decision maker (e.g., expertise) (Payne, Bettman, and Johnson 1992, 1993; Simonson and Tversky 1992). These factors can systematically and often dramatically influence choice outcomes, even though, normatively speaking, they should not. For example, the framing of choices (e.g., in terms of "lives saved" versus "lives lost"; Tversky and Kahneman 1991) can cause large preference reversals, even though the consequences of choices are formally identical.

These rules all transform information inputs into final goal states of knowledge, but they can vary in the type of attribute value information they use (Payne, Bettman, and Johnson 1993). Some rules use actual attribute values (e.g., \$.10, 10X magnification power) as inputs. For example, the weighted-additive rule considers the actual values of all alternatives on all the relevant attributes. Other rules make use of ordered values, or values that are defined by the relative positions of considered options along attributes (e.g., higher magnification, most expensive). For example, the lexicographic rule chooses the alternative with the relatively best value on the most important attribute.

Rules using actual values as inputs are in accord with the idea that "fundamental" values underlie consumer choices. Such rules require that consumers assess the true or psycho-

*Anand V. Bodapati is Assistant Professor of Marketing (e-mail: bodapati@ucla.edu), and Aimee Drolet is Associate Professor of Marketing (e-mail: adrolet@anderson.ucla.edu), Anderson School of Management, University of California, Los Angeles. The authors acknowledge the many thoughtful suggestions of the *JMR* reviewers. In addition, the authors thank Don Morrison, Itamar Simonson, and Pat Suppes for comments on previous drafts. The authors are especially grateful to Randy Bucklin for his invaluable feedback and guidance at various stages of this research.

logical significance of a product's actual values or actual value differences between a product and other considered products. However, for consumers, these tasks are difficult, even for simple products and even when full information is available (Ariely, Loewenstein, and Prelec 2003; Frederick and Fischhoff 1998). In contrast, consumers can make relative valuations easily. Accordingly, much research shows that consumers rely heavily on rules that use ordered values as information inputs. Such rules convert actual values to more psychologically meaningful, better-defined ordered values (e.g., relative attribute importance) (Drolet, Simonson, and Tversky 2000).

If choices are based on ordered rather than actual values, adjustments in considered products' actual attribute value profiles will presumably not have an appreciable effect on choice shares as long as products' relative positions along attributes are preserved. If preserved, consumers should be less responsive to actual value changes than traditional models, such as the multinomial logit (MNL) model, predict. Because these models assume that consumers assess the psychological value of actual value differences, even minor adjustments in products' actual values should affect preferences. In contrast, if choices are based on ordered values and the relative positions of products on attributes are not preserved, even slightly perceptible actual value changes can have a dramatic effect on choice shares. This (disproportionately large) response is at odds with what traditional models predict.

To illustrate, consider three situations in which a consumer chooses between two products, A and B. In Situation 1, the price of Product A is \$7.55, and the price of Product B is \$7.54; the price difference between A and B is \$.01. In Situation 2, Product A is \$7.54, and Product B is \$7.55; the difference between A and B is –\$.01. In Situation 3, Product A is \$7.52, and Product B is \$7.55; the price difference between A and B is –\$.03. The MNL model predicts that the increase in Product A's choice share when moving from Situation 1 to Situation 2 will be approximately equal to the increase in Product A's choice share when moving from Situation 2 to Situation 3 because the corresponding shifts in actual value difference in price are equal. However, intuition and experimental research (e.g., Drolet and Bodapati 2003) suggest that Product A's choice share will increase more when moving from Situation 1 to Situation 2 than from Situation 2 to Situation 3. Although the two moves are quantitatively equivalent, they differ qualitatively. The move from Situation 1 to Situation 2 alters which product (A or B) is priced lower. The move from Situation 2 to Situation 3 does not. Consumers' heavy use of decision rules based on ordered values suggests that consumers would pay greater attention to the ordering of the prices than to their actual values and that there would be a greater increase in Product A's choice share if the ordering of the two products' prices is altered.

Nevertheless, consumers do not base all of their choices solely on ordered values. If they did, we would predict that the change in the choice share of Product A when moving from Situation 2 to Situation 3 would be zero. The move from Situation 2 to Situation 3 preserves the ordering of prices and alters only the magnitude of the price difference. Similar to the MNL model's prediction that consumers respond only to actual value differences, this prediction appears equally incorrect in light of empirical evidence that even order-preserving changes in price have an effect on choice shares. Consumers typically do not consider only the ordering of prices, because even order-preserving changes alter choice if the changes are large enough.

In short, it appears that models that predict choices on the basis of only one kind of attribute value processing (actual or ordered) predict choices less effectively than do models that represent decision making as involving the use of both actual value processing and ordered value processing. In this article, we develop an econometric model that allows consumers to choose either actual or ordered value information. We estimate the hybrid model using real consumer purchase data from an Information Resources Inc. (IRI) scanner panel data set. The results indicate that the hybrid model predicts choices better than do models that allow consumers to use only one type of attribute value information, actual or ordered. This article concludes with a discussion of the implications of this research.

MODEL

Our proposed model assumes that on any one choice occasion, a consumer may operate in one of two decision modes, actual value mode (AVM) or ordered value mode (OVM). We assume that a consumer who operates in OVM evaluates each product only according to its position along attributes relative to those of other considered products. In contrast, a consumer who operates in AVM evaluates the numerical differences between the actual attribute values of a product and the actual attribute values of other products. The probability of choosing a product i in AVM or OVM is given by $p(\text{choice} = \text{i}|\text{AVM})$ and $p(\text{choice} = \text{i}|\text{OVM})$, respectively; we derive algebraic expressions subsequently. Because we do not ordinarily know in which of the two modes the consumer operates on any given occasion, we posit a probability p_{AVM} for the consumer operating in AVM. The overall probability of the consumer's choosing product i is a weighted average of $p(\text{choice} = \text{i}|\text{AVM})$ and $p(\text{choice} = \text{i}|\text{OVM})$; the weight is governed by p_{AVM}.

Choice Probabilities for AVM

For a consumer operating in AVM, the utility for product i from a set of I products and described on a space of A attributes as $(x_{i1}, x_{i2}, \ldots, x_{iA})$ is as follows:

$$\text{(1)} \qquad U_{i_{\text{AVM}}} = \sum_{a=1}^{A} b_a x_{ia} + e_i,$$

where e_i is an error term. The consumer chooses the option I, which maximizes utility $U_{i\text{AVM}}$, and the probability that the consumer chooses i is as follows:

$$\text{(2)} \quad p(u_i > u_j, \forall j \neq i) = p\left(\sum_{a=1}^{A} b_a x_{ia} + e_i > \sum_{a=1}^{A} b_a x_{ja} + e_j, \forall j \neq i\right).$$

If we assume that the stochastic components $e_1, e_2, \ldots, e_i$ are drawn from the extreme value distribution, the choice probability is given by the following:

(3) $$p(\text{choice} = i|\text{AVM}) = p(u_i > u_j, \forall j \neq i) = \frac{\exp\left(\sum_{a=1}^{A} b_a x_{ia}\right)}{\sum_{j=1}^{I} \exp\left(\sum_{a=1}^{A} b_a x_{ja}\right)}.$$

This is the classical MNL model. In this model, any change in an attribute value x_{ia} alters the choice probability, even if the change preserves the ordered values of product i.

Choice Probabilities for OVM

To describe a consumer operating in OVM, we modify the preceding probability (Equation 3) to make it immune to attribute value changes that do not alter the ordering of considered products' attribute values. We use the "signum" function. By definition, this function takes the value of 1 if its argument is strictly positive, –1 if its argument is strictly negative, and 0 if its argument is exactly zero. We define the OVM utility for product i as follows:

(4) $$U_{i_{OVM}} = \sum_{a=1}^{A} b'_a \left[\sum_{j=1}^{I} \text{signum}(x_{ia} - x_{ja})\right] + e_i.$$

Correspondingly, the choice probabilities are given by the following:

(5) $$p(\text{choice} = i|\text{OVM}) = \frac{\exp\left\{\sum_{a=1}^{A} b'_a \left[\sum_{j=1}^{I} \text{signum}(x_{ia} - x_{ja})\right]\right\}}{\sum_{k=1}^{I} \exp\left\{\sum_{a=1}^{A} b'_a \left[\sum_{j=1}^{I} \text{signum}(x_{ka} - x_{ja})\right]\right\}}.$$

In this model, any change that preserves the ordering of products' values along attribute a leaves the values in $\{\text{signum}\ (x_{ia} - x_{ja}),\ i \neq j\}$ unchanged. Thus, the value of $U_{i_{OVM}}$ is unchanged. Conversely, any change that alters the ordering results in at least one value in $\{\text{signum}\ (x_{ia} - x_{ja}),\ i \neq j\}$ being changed and, thus, a change in at least one of the $U_{i_{OVM}}$.

An important point of difference between the AVM model and the OVM model is brought out in a comparison of the two expressions for the utilities. In Equation 1 for AVM, the contribution of a certain attribute to the overall utility is absolute and does not depend on the attribute's value in other products. This is consistent with the idea in economic utility theory that fundamental values underlie consumer choices and that these fundamental values are not idiosyncratic to context. In contrast, Equation 4 for OVM assumes that the contribution of a certain attribute to the overall utility is assessed only in the context of the other products' values for that attribute. This reflects the position in behavioral decision theory that preferences are often not preexisting but rather constructed according to the context (see Tversky and Simonson 1993). Because the utilities in the OVM are context specific in this way, the OVM model's choice probabilities do not have the property of independence from irrelevant alternatives (for a discussion, see Amemiya 1985).

To determine the decision mode of a consumer to predict choice in a given situation, the probability of operating in AVM (p_{AVM}) is given by a simple logistic transformation of c_h, which is a score that represents the tendency of consumer h of H consumers (h = 1, 2, …, H) to operate in AVM:

$$p\text{AVM} \equiv F(c_h),$$

where we use F to denote the usual logistic function:

(6) $$F(s) = \frac{e^s}{(1 + e^s)}.$$

Consumers differ in their tendency to operate in AVM, so different consumers have different values of c_h.[1] Because we consider only two possible decision modes (AVM and OVM), the probability that the consumer is operating in OVM is the complement of the previous probabilities: $p_{OVM} = 1 - p_{AVM}$, or $1 - F(c_h)$, or $F(-c_h)$. Averaging across the two possibilities, we obtain the overall probability that the consumer chooses product i:

(7) $$p(\text{choice} = i;\ b, b', h) = p(\text{choice} = i|\text{AVM};\ b)p_{AVM} + p(\text{choice} = i|\text{OVM};\ b')p_{OVM}.$$

Then,

(8) $$p(\text{choice} = i;\ b, b', h) = p(\text{choice} = i|\text{AVM};\ b)F(c_h) + p(\text{choice} = i|\text{OVM};\ b')F(-c_h).$$

The total number of parameters in the model is 1 + 2A: A parameters for the b vector in the model for $p(\text{choice} = i|\text{AVM};\ b)$, another A parameters for the b′ vector in the model for $p(\text{choice} = i|\text{OVM};\ b')$, and one parameter corresponding to c_h.

To understand the behavior of this hybrid model, it is helpful to consider a simple example. Assume that there are only two products (1 and 2) with only one attribute and that values of $x_{11} - x_{21}$ vary uniformly between –5 and 5. The attribute value difference between Product 1 and Product 2 would be $x_{11} - x_{21}$. Let b_1, the weight of Attribute 1, equal 2. Figure 1 depicts the AVM probability of choosing Product 1 as a function of $x_{11} - x_{21}$. The model gives p_{AVM} as $p(\text{choice} = 1|\text{AVM}) = F[2(x_{11} - x_{21})]$. Compare this response curve with the curve for a consumer operating in OVM (see Figure 2). Here, Product 1 is chosen with a probability of 1 if $x_{11} > x_{21}$, and Product 2 is chosen with a probability of 1 if $x_{11} < x_{21}$. Consider the attribute weight b_1 to be 1.5. The choice probability if $x_{11} > x_{21}$ would be F(1.5), and the choice probability if $x_{11} < x_{21}$ would be F(–1.5). If $x_{11} = x_{21}$, the probability of choosing x_1 over x_2 would be .5.

Our model posits that the choice probability is an intermediate function between these two curves. The intermediacy is governed by $p_{AVM} = F(Z)$, or the likelihood that the consumer is operating in AVM in a given situation. For example, if Z = 1, then Z > 0, F(Z) > .5, and the consumer

[1]The score c_h can be a sum of fixed effects and random effects. The fixed effects can be a function of observed consumer characteristics (e.g., demographic characteristics). In the empirical work in this article, we model only a random effects influence on c_h.

Figure 1
PROBABILITY OF CHOOSING PRODUCT 1 UNDER AVM

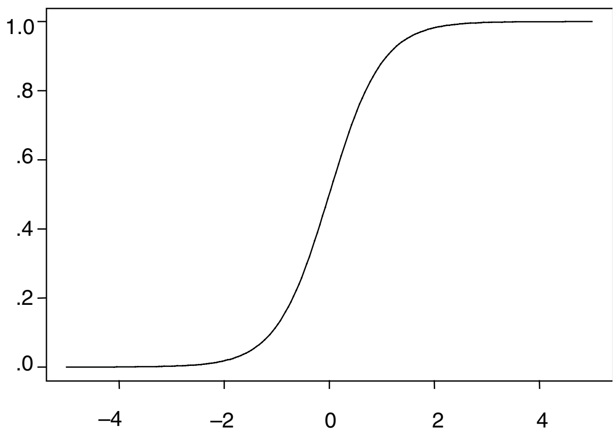


Figure 2
PROBABILITY OF CHOOSING PRODUCT 1 UNDER OVM

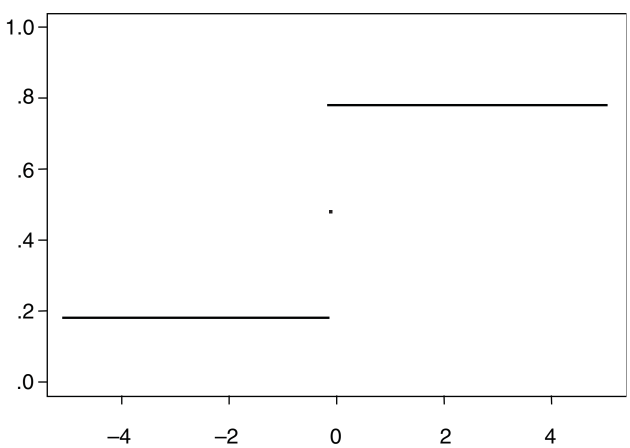


Figure 3
PROBABILITY OF CHOOSING PRODUCT 1 UNDER THE AVM–HEAVY MIXTURE MODEL

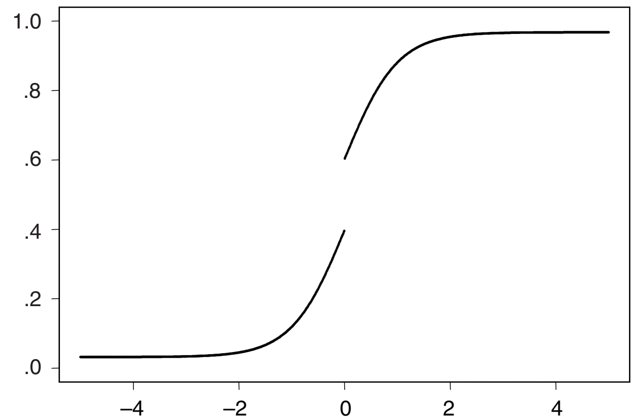
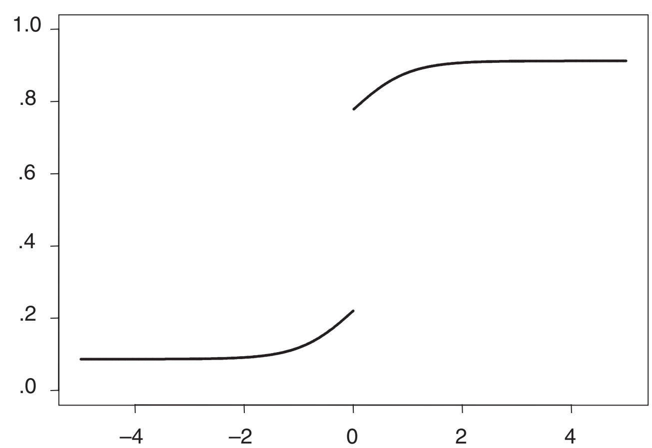


Figure 4
PROBABILITY OF CHOOSING PRODUCT 1 UNDER THE OVM–HEAVY MIXTURE MODEL



is more likely to be in AVM than in OVM. For this example, we average the response curves in Figures 1 and 2 by weights F(1) = .731 and F(–1) = .269, and we obtain the choice probabilities depicted in Figure 3.

Consider another value for Z. If Z = –1, then Z < 0, F(*Z*) < .5, and the consumer is less likely to be in AVM than in OVM. Here, we average the response curves by the reverse weights F(–1) = .269 and F(1) = .731 (see Figure 4). Both of the response curves (Figures 3 and 4) show a discontinuity at $x_{11} - x_{21} = 0$. This is a consequence of the discontinuity in the model for $p(\text{choice} = 1|\text{OVM})$. Consumers who operate in OVM switch choices in the vicinity of $x_{11} - x_{21} = 0$ no matter how small the difference is from zero. This behavior is reflected in the jump in choice probability. The jump is greater in Figure 4 because the likelihood of the consumer responding in OVM is greater in that situation. Our model always has such a discontinuity except in the special case in which Z = ∞. In this case, the model reduces to the usual logit model, which does not allow ordered value processing.

Likelihood Function with Panel Data

Assume that on each of t choice occasions (t = 1, 2, …, T), we observe choice y_t, where y_t can be one of the I choice options; the collection of observed choices on all T_h occasions of the hth consumer is denoted by Y_h. Let the 1 + 2A parameters in our hybrid model be collectively denoted by $\theta = [b, b', c_h]$. Because at least some of the A attributes of each of the I options vary from occasion to occasion, the $p(\text{choice} = i|\text{AVM}; b)$, the $p(\text{choice} = i|\text{OVM}; b')$, and the consequent $p(\text{choice} = i)$ also vary from occasion to occasion. The occasion-specific values of these are denoted as $p_t(\text{choice} = i|\text{AVM}; b)$, $p_t(\text{choice} = i|\text{OVM}; b')$, and the consequent $p_t(\text{choice} = i|b, b', c_h)$, respectively.[2]

[2]For the sake of notational brevity, we have omitted the product attribute values.

We develop two forms of the likelihood function for the T_h observed choices. We denote the decision mode of the consumer on the tth occasion as m_t. We consider m_t a Bernoulli random variable with $m_t = 1$ or 0 according to whether the decision mode on the tth occasion is AVM or OVM. The expectation of m_t is $E(m_t) = F(c_h)$. If we observe the values of m_t for each of the T_h occasions, the likelihood for the collection of choices observed for the consumer can be written as

$$(9)\quad L(\theta; Y_h) = \prod_{t=1}^{T_h} [m_t p_t(y_t|AVM; b) + (1 - m_t)p_t(y_t|OVM; b')].$$

We do not actually observe the values of $m_1, m_2, ..., m_{T_h}$, so the likelihood we consider is the expectation of the $L(\theta; Y_h)$ over the realizations of the $\{m_t\}$. We distinguish between two cases: (1) time invariant and (2) time varying. Both forms of the likelihood posit that the probability of operating in AVM is given by $F(c_h)$. The difference between the two is related to their views of the stability of the decision mode. In contrast to the time-varying version, the time-invariant version views the decision mode as stable.

Time-invariant version. In the time-invariant version, if a consumer operates in AVM on one occasion, the consumer operates in AVM on all occasions. In the time-invariant version, the $\{m_t\}$ realizations are related in that they are all constrained to be equal. Either $m_t = 1\ \forall t$ or $m_t = 0\ \forall t$ while maintaining each expected value at $E(m_t) = F(c_h)$. Accordingly, we call this version time invariant because the value of m_t within a consumer does not vary over time. We denote the expectation of $L(\theta; Y_h)$ in this situation as $L_{\text{time-invariant}}(\theta; Y_h)$. If the common value taken by the $\{m_t\}$ is denoted by just m, Equation 9 simplifies to the following:

$$(10)\quad L(\theta; Y_h) = m\prod_{t=1}^{T_h} [p_t(y_t|AVM; b)] + (1 - m)\prod_{t=1}^{T_h} [p_t(y_t|OVM; b')].$$

Because $E(m_t) = F(c_h)$, the expectation of Equation 10 is as follows:

$$(11)\quad L_{\text{time–invariant}}(\theta; Y_h) = F(c_h)\prod_{t=1}^{T} p_t(y_t|AVM; b) + F(-c_h)\prod_{t=1}^{T} p_t(y_t|OVM; b').$$

Time-varying version. In the time-varying version, the $\{m_t\}$ realizations are unrelated in that they are independent draws of Bernoulli variables with common expectation $F(c_h)$. We call this version time varying because the value of m_t within a consumer can vary over time. We denote the expectation of $L(\theta; Y_h)$ in this situation as $L_{\text{time-varying}}(\theta; Y_h)$. Because the $\{m_t\}$ are i.i.d. with common expectation $F(c_h)$, the expectation of $L(\theta; Y_h)$ can be written as

$$(12)\quad L_{\text{time-varying}}(\theta; Y_h) = \prod_{t=1}^{T_h} [F(c_h)p_t(y_t|AVM; b) + F(-c_h)p_t(y_t|OVM; b'].$$

In both versions of the likelihood, the term $F(c_h)$ can be interpreted as the occasion-specific prior probability of the consumer operating in AVM. As we formulated, the prior probability is a constant and does not depend on any known characteristics of the consumer.[3]

Modeling Heterogeneity

The collection of parameters in θ vary from consumer to consumer because different consumers have different tastes for the A attributes and different propensities to operate in AVM. To a large extent, these differences are due to unobserved heterogeneity. We model this unobserved heterogeneity with a hierarchical Bayes model and estimate the model by Markov chain Monte Carlo (MCMC) methods.

We denote the parameter vector for household h(= 1, …, H) as θ_h. We assume that the H household-level parameter vectors $\theta_1, \theta_2, …, \theta_H$ are independent draws from a common Gaussian generating distribution. We denote the mean and variance of this generating distribution as μ and Σ, respectively, with the two being collectively designated as Θ. The number of elements contained in θ is designated as $d = 2A + 1$. Thus:

$$(13)\quad \Theta = [\mu, \Sigma].$$

$$(14)\quad N(\theta; \Theta) = \frac{\exp-\frac{1}{2}(\theta - \mu)^T\Sigma^{-1}(\theta - \mu)}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}}$$

$$\theta_h \sim N(\theta; \Theta).$$

The likelihood based on Y_h would be as follows:

$$(15)\quad L(\Theta; Y_h) = \int L(\theta; Y_h)N(\theta; \Theta)d\theta.$$

The likelihood in the integrand can be based on either the time-invariant or the time-varying likelihood form. Using the previous expression, the likelihood for the full collection of choices observed over all consumers is as follows:

$$(16)\quad L(\Theta) = \prod_{h=1}^{H} L(\Theta; Y_h).$$

Inferences can be made about Θ either by drawing random variates from the Bayesian posterior density of Θ or by maximizing the previous expression to obtain the maximum likelihood estimate of Θ. In this article, we chose the former approach and used MCMC to construct draws from posterior density of Θ. We used the WinBUGS program to construct the draws.

Summary

We propose that all consumers exhibit average choice behavior that is intermediate to AVM and OVM, and the position of intermediacy is governed by the consumer-specific term c_h. Furthermore, we propose that a consumer's

[3]The model could be extended to control for consumer characteristics by replacing c_h with a linear function of the vector of characteristics. In addition, because $F(c_h)$ is not time varying, the occasion-specific prior probability is the same for all occasions. However, if c_h is replaced by a time-varying quantity, we would obtain a model in which the occasion-specific prior probability varies with time as a function of the marketing environment.

decision mode can be viewed as either time invariant, in the sense that it does not change from occasion to occasion, or time varying, in the sense that it can vary from occasion to occasion with stochastic independence. If a consumer operates in AVM, the choice outcome is governed by b_s. If a consumer operates in OVM, the choice outcome is governed by b'_s. Finally, whereas all consumers exhibit this hybrid behavior in choice making, the parameters of the hybrid model $\theta_n = [b_h, b'_h, c_h]$ can vary from consumer to consumer.

EMPIRICAL APPLICATION

To evaluate the usefulness of our hybrid model, we must address the question of how prevalent OVM is compared with the commonly assumed AVM. We denote the mean and standard deviation of c_h in the Gaussian generating distribution as μ_c and σ_c (these are contained in overall model parameter Θ). From these two terms, we can gauge the overall prevalence (OP) of OVM decision making. A key quantity of interest is the expectation of $F(c_h)$ over the generating distribution:

$$\text{(17)} \qquad OP = \int F(-c_h)\frac{\exp-\frac{1}{2}\frac{(c_h-\mu_c)^2}{\Sigma_{cc}}}{\sqrt{2\pi\Sigma_{cc}}}dc_h.$$

As we previously discussed, the interpretation of $F(c_h)$ differs in the time-invariant versus time-varying likelihood form. Accordingly, the interpretation of OP also differs between the two forms. In the time-invariant case, $F(c_h)$ can be viewed as the posterior data-based probability of $m = 1$, the event that the consumer operates in AVM on all occasions viewed against the event that the consumer operates in OVM on all occasions. Recall that in the Kamakura-Russell latent-class model, the estimate for the fraction of consumers who belong to a certain segment is given by the average consumer-level posterior probability for that segment. Similarly, the average of the $F(c_h)$ gives the fraction of consumers for whom $m = 1$ (AVM) rather than $m = 0$ (OVM). For this reason, the OP term for the time-varying model corresponds to the fraction of consumers for whom the OVM dominates.

In the time-varying version, $F(c_h)$ represents the expected fraction of occasions on which consumer h operates in AVM. Applying the law of iterated expectation from probability theory, the overall fraction of occasions (across all consumers) in which AVM is exhibited is given by the average value of $F(c_h)$. For this reason, the OP term for the time-varying model corresponds to the fraction of occasions on which the OVM dominates.

Model Estimation

We estimated our model using real consumer purchase data in the bacon product category. We drew our observations from an IRI scanner panel data set that was collected in a city in the Midwest region of the United States over a two-year period. We used only those consumers who restricted their bacon purchases to the three leading brands: Oscar Mayer, Wilson Corn King, and Lazy Maple, which we henceforth refer to as Brand 1, Brand 2, and Brand 3, respectively. We also restricted our observations to purchases of the 16-ounce size.[4] We were left with 252 panelists who collectively made 2950 purchases. Our model studied the choice made among the three brands on each of the 2950 purchase occasions.

The utility function used five attributes: a dummy term corresponding to Brand 2, a dummy term corresponding to Brand 3, a price term giving the price in dollars, a “feature” term that was 1 when the product was feature advertised and 0 when the product was not, and a “display” term that was also a 0–1 variable to represent existence of special displays for the product.

We estimated four different forms of the choice model. In the first, we allowed only AVM. This form would be equivalent to a hierarchical Bayes model with Gaussian random effects, the standard model in the modern econometric choice-modeling literature. In the second, we allowed only OVM. The third and fourth forms are dual-mode models in which we allowed both AVM and OVM. In the third model, we used the time-invariant version of the likelihood, and in the fourth model, we used the time-varying version.

The models with only AVM or only OVM have 5 parameters from μ and 15 from Σ. The dual-mode models have 11 parameters from μ and 66 from Σ. Note that for the time-invariant version of the dual-mode model, the likelihood function does not give any information about covariance terms for response variables that are drawn from the two different modes. As a result, the inference for such covariances is driven mainly from the inverse-Wishart hyperprior that is specified for Σ.

We estimated the four models by expressing each model as a “directed graphical model,” which is a version of a Bayesian network that is restricted so that it is acyclic. This model is then expressed as a WinBUGS program and is estimated by the software.[5] (The WinBUGS model statements used to specify the four models are available on request.) We followed Gelman and Rubin’s (1992) prescription for running the MCMC and assessing convergence. We used ten widely dispersed starting values to initiate ten parallel Markov chains. We discarded the first 10,000 draws from each chain, and we used the next 10,000 draws to assess convergence and to make inferences. We assessed convergence by establishing that variance of the draws pooled across chains was equal to the within-chain variance, as called for by the Gelman-Rubin diagnostic.

We used the mean of the posterior draws to compute the point estimates for μ and Σ. We computed deviance information criterion (DIC) scores (Spiegelhalter et al. 2002) for each of the four models based on the posterior draws. The DIC score is similar to other model assessment scores, such as the Akaike information criterion (AIC) score (Akaike 1973), the Bayesian information criterion (BIC) score (Schwarz 1978) and the Mallows Cp score (Mallows 1973), in that it credits a model for goodness-of-fit that is achieved on the estimation data but punishes the model for the number of parameters (for a further discussion of these related model assessment scores, see Burnham and Anderson 2002). The overall score trades off between goodness-of-fit and number of parameters; the various model assessment

[4]These are not strong restrictions. The three brands collectively account for 68.85% of the purchases in the bacon category; the largest brand outside of these three brands has a share of only 3.35%, and each of the other omitted brands has considerably smaller market shares. Furthermore, more than 85% of the purchases in this data set are in the 16-ounce size.

[5]For information on the WinBUGS program and for explanations about why its conditional-density sampling algorithms require the graphical model to be acyclic, see Spiegelhalter and colleagues (2003) and Gilks, Richardson, and Spiegelhalter (1996).

scores differ primarily in how this trade-off is done. As Spiegelhalter and colleagues (2002) discuss, AIC, BIC, and Mallows Cp scores are not considered suitable for hierarchical Bayes models, because there is ambiguity about the effective number of parameters. However, the DIC score fixes that problem. Thus, it is the measure of choice in our article.

In terms of the DIC scores, the best model is the dual-mode model of the time-variant form (DIC = 1736.1), followed by the dual-mode model of the time-invariant form (DIC = 1800.4). The OVM-only model is third best (DIC = 1848.8), and the AVM-only model performs the worst (DIC = 1900.1). Use of the BIC score produces the same ranking.

Model assessment tools, such as the DIC and its previously mentioned cousins, assign a score to a model based on theoretically calculated values for the expected goodness-of-fit on new, unobserved data ("holdout" data) under the assumption that the new data come from the same distribution as the data used in the estimation of the model (i.e., the estimation data). As an alternative to using the theoretically calculated values, it is possible to compute empirical values for goodness-of-fit on new data if holdout data have been set aside for this purpose. However, as Efron (2004) argues, empirical goodness-of-fit measures such as cross-validation have high variance, and the theoretically calculated measures are more accurate measures of model correctness. This is particularly true when, as in the present situation, the responses are discrete valued rather than continuous and the sample size available for any holdout analysis is small.

The model estimates for the four models appear in Table 1. We estimated the full covariance matrix Σ for all four models. Because of space limitations, we report only square roots of the diagonal terms of Σ. Table 1 lists the estimates of the across-household mean and standard deviation for each of the model parameters and the standard errors for each of these estimates.

As we previously discussed, a main focus of our attention is on the prevalence of choices based on ordered values as measured by the OP index. The mean and variance of c_h for the two dual-mode models imply an OP propensity index of 55.8% for the time-invariant case and 62.2% for the time-varying case. This implies an estimate of 62.2% for the fraction of purchase occasions on which the consumer operates in OVM and an estimate of 55.8% for the fraction of households in which the OVM dominates. These results suggest that OVM choice making is widely prevalent in real-life purchasing.

To understand the different managerial implications for each of the four models, we used the model estimates to make the predictions on market share as considered from the viewpoint of Brand 3. We held the prices of Brand 1 and Brand 2 fixed at their median values, which are $2.49 and $1.49, respectively. We then considered how the market share of Brand 3 varies as its price is varied over 141 price points, from $1.29 to $2.69 in increments of $.01. For the sake of simplicity, we assumed that each brand is neither featured nor displayed in the shopping environment. We obtained the predictions of the market share in full Bayesian fashion. We used draws produced by WinBUGS using MCMC for the posterior distribution of the household to compute the posterior distribution of that household's purchase probability for Brand 3 for each price level. We then took the average from this posterior density as a point estimate for the household's purchase probability for Brand 3. The average of all such point estimates across all households was then taken to be the point estimate for the market share of Brand 3. We did this calculation for each of the 141 price points for each of the four models. The curves depict-

Table 1

MODEL ESTIMATES

		Actual Coefficients					*Ordered Coefficients*				
Model	c_h	*Brand 2*	*Brand 3*	*Price*	*Feature*	*Display*	*Brand 2*	*Brand 3*	*Price*	*Feature*	*Display*
AVM Only											
Mean (μ)		–.98	–2.55	–.94	2.12	.60					
		(.36)	(.44)	(.16)	(.17)	(.14)					
Standard deviation (σ)		2.92	2.84	.65	.35	.23					
		(.44)	(.44)	(.20)	(.14)	(.13)					
OVM Only											
Mean (μ)							–.43	–2.52	–.25	.82	.24
							(.40)	(.54)	(.06)	(.08)	(.06)
Standard deviation (σ)							3.12	3.22	.24	.34	.18
							(.41)	(.50)	(.08)	(.10)	(.06)
Dual Mode: Time Invariant											
Mean (μ)	–.24	–1.30	–2.75	–1.29	2.14	.65	–.87	–3.17	–.32	.76	.22
	(.11)	(1.30)	(.83)	(.41)	(.13)	(.22)	(.90)	(1.39)	(.17)	(.12)	(.08)
Standard deviation (σ)	1.27	2.35	2.67	.66	.27	.22	3.11	3.23	.34	.33	.21
	(2.61)	(.87)	(.7465)	(.32)	(.15)	(.14)	(.70)	(1.28)	(.16)	(.11)	(.11)
Dual Mode: Time Varying											
Mean (μ)	–.95	–.72	–2.34	–1.08	2.27	.65	–.18	–1.87	–.36	1.16	.26
	(.48)	(.50)	(.96)	(.88)	(.22)	(.18)	(.63)	(.71)	(.35)	(.22)	(.14)
Standard deviation (σ)	.26	3.73	2.89	.66	.38	.31	3.36	2.96	.46	.45	.28
	(3.44)	(.79)	(.69)	(.28)	(.17)	(.15)	(.41)	(.74)	(.24)	(.15)	(.18)

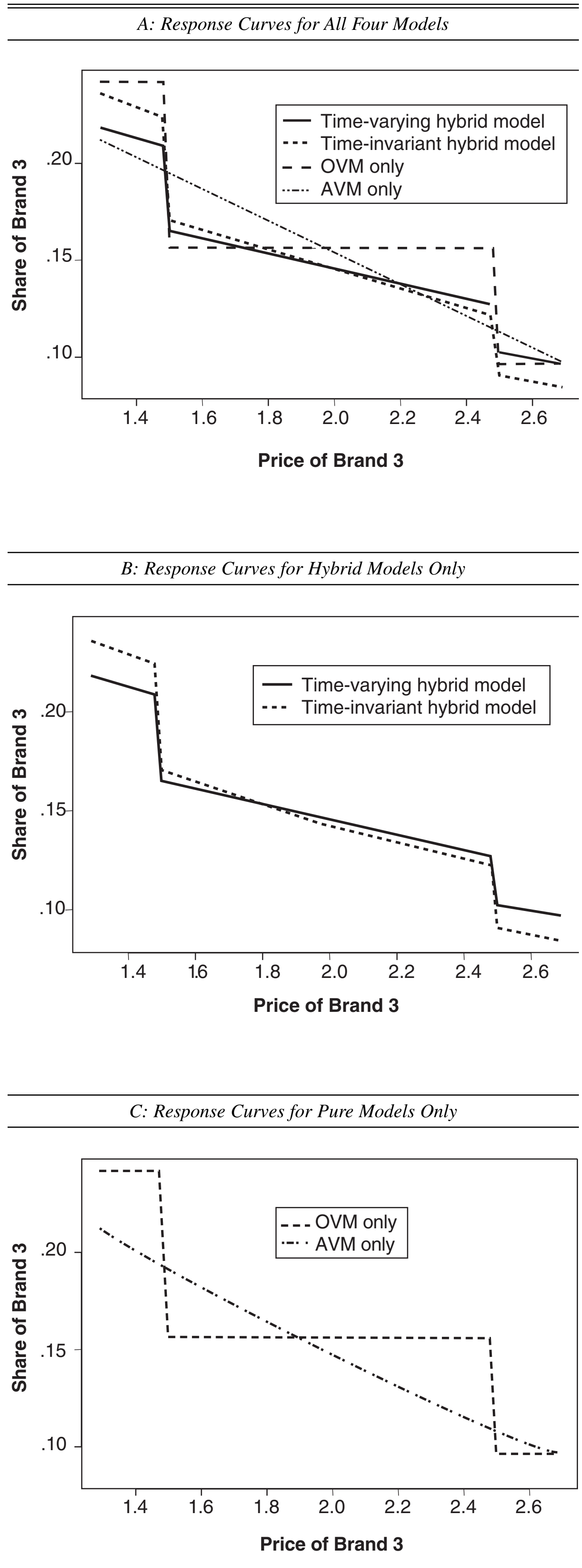
Notes: Standard errors are in parentheses.

ing how Brand 3's market share changes with price for the four models appear in Figure 5.

Figure 5, Panel A, shows the share curves from all four models. Because the curves can be difficult to distinguish, we also show two separate panels; Panel B shows the two curves for the two hybrid models, and Panel C shows the curves for the two pure models. These pictures make clear the differences in predictions from the four models. The pure AVM model, which on visual inspection performs the worst, is linear in this region. Contrast this with the pure OVM model, which is flat everywhere except for discontinuities near $1.49 (the price of Brand 2) and near $2.49 (the price of Brand 1); here, changes in the price of Brand 3 cause changes in the price ordering. Note that the AVM curve exhibits no such discontinuities. The two hybrid curves possess characteristics of both the pure models: They show discontinuities near these two prices, but they are not flat. As can be observed, the share predictions of the time-invariant hybrid model are more extreme than are those of the time-varying hybrid model. The hybrid models' better performance is consistent with the idea that (1) even small price changes can have large effects on shares if the changes alter the ordering and (2) even order-preserving changes can have large effects if the changes are large enough.

Figure 5

EFFECT OF PRICE SHARE FOR BRAND 3



DISCUSSION

In this article, we proposed a probabilistic choice model that takes into account that consumers may rely on both ordered value information and actual value information when making choices. We estimated this model using real purchase data. The results show that the hybrid model predicts choices better than do models that allow choices to be based on only one type of attribute value information (actual or ordered).

The current research contributes to the econometric choice-modeling literature, which has not focused much on consumers' use of ordered values. The findings of this research support the view expressed in prior research, which has similarly suggested that comparative information can be usefully incorporated into formal choice models (see, e.g., the generic cross-effects choice model in Elrod, Louviere, and Davey [1992] and the context-dependent preference model in Tversky and Simonson [1993]). Our hybrid model improves on traditional choice models in terms of both its descriptive ability and its predictive ability. Traditional models are "as-if" models and do not reflect actual underlying psychological processes. The goal of these models is not to describe choice processes but rather to predict choice outcomes well. However, if consumers rely heavily on ordered values in their decision making, these traditional models may not predict choice outcomes well.

Indeed, our estimation results are consistent with experimental results (e.g., Ariely, Loewenstein, and Prelec 2003) in that they suggest that choice making based on ordered values is prevalent in real-life purchasing; we base this conclusion on the heavy weight of the ordered component of the hybrid model and the assumption that our model is correctly specified. Some adjustments to traditional models can successfully approximate choice behavior that is inconsistent with normative assumption. For example, by changing the weights given to attributes, the standard logit choice model can approximate lexicographic choice rule use. However, the standard logit choice model cannot be (easily) adjusted to accommodate the disproportionately large effect of a small change in actual values on choice shares if consumers rely on ordered values and these ordered values are changed by a change in actual values. Our hybrid model can accommodate this situation.

Note also that, at the same time, our estimation results suggest that the hybrid model represents an improvement over a model that includes only choices based on ordered values. Some experimental research implies that decision making is mostly based on ordered values. For example, Drolet, Simonson, and Tversky (2000) examine whether consumers are able to predict their choices solely on the basis of products' ordered values without knowing their actual values. One group of respondents was presented with choice sets in different categories with options described only in relative terms (e.g., binoculars Pair A has greater magnification power and higher price than the other two pairs; binoculars Pair B has intermediate values). The choice sets included both categories in which consumers tend to select the compromise option (e.g., portable grills that vary in weight and cooking area) and categories in which consumers tend to avoid the compromise option (e.g., dental insurance plans that vary in coverage and annual premium). The respondents' task was to try to predict whether they would have chosen the compromise option in these categories if they were presented with sets with actual values. A second group of respondents made choices from sets in the same categories and with the same attributes, but they were presented with the options' actual values (e.g., binoculars Pair A had 15-times magnification and cost $99). A comparison across choice sets of the compromise choice predictions in the set without actual values and the actual share of the compromise option in these sets among respondents in the second group revealed a high correlation of .89. In other words, even without knowing the options' actual values, respondents were accurate in predicting the likelihood of choosing the compromise option from each choice set. These results imply that consumers base their choices primarily on ordered values. Accordingly, models that include only ordered value–based choices might predict as well as a hybrid model that includes actual value–based and ordered value–based choices. Nevertheless, our results are inconsistent with this suggestion.

The current research also has limitations that may be addressed in future studies. First, our hybrid model is restrictive in the sense that it assumes that all attributes are evaluated either in AVM or in OVM. The signum function is applied only when an attribute is processed in OVM. The model could be generalized by having a multinomial choice model in which the utility contribution of each attribute is given by a weighted sum of an absolute component and a signum component, such as the one we have in the OVM model. Alternatively, the model could be generalized by allowing for applications of the signum function to arbitrary subsets of the attributes; the estimation algorithm would then search for the subset of attributes in which the application of the signum function produces the best fit to the data. There are costs and benefits to each approach.

Second, our estimates show that the b and b′ are not proportional, even after adjusting for the differently scaled corresponding variables in AVM and OVM models. This implies that the relative importance of one attribute versus others is not comparable in AVM versus OVM. A possible explanation for this nonproportionality is that the ease with which certain attributes are evaluated differs between decision modes, and thus the relative weight given to certain attributes differs between decision modes. For example, it may be that continuous attributes (e.g., price) are (psychologically) easier to evaluate in relative terms than are categorical attributes (e.g., feature), causing continuous attributes to be weighted relatively more in OVM.

Third, this article considers only the brand choice decision and ignores other aspects of buying behavior, such as the purchase-incidence decision or the quantity decision. It is unclear how ordered value processing affects these other decisions. Further research is necessary to investigate this.

Fourth, our model is restricted in that it allows for only main effects of price and the other predictor terms. This raises an important consideration about testing for AVM versus OVM. It can be demonstrated that an AVM model with higher-order effects may be able to mimic the OVM quite well if there is little price variation. Indeed, in the extreme case (albeit an unrealistic one), if the other brands' attributes are fixed, a brand's OVM utility function is composed of nothing but step functions, which can be approximated exactly with an AVM model with appropriately nonlinear form. However, if there is a large enough price variation, it becomes much more difficult for a nonlinear AVM to mimic OVM because the locations of the step functions vary too much. Our product category has a heavy level of promotions, and thus we expect that the OVM model behaves differently from an AVM model even if the AVM model has a lot of nonlinearity. To reject the hypothesis that what we are observing as OVM in our data analysis is not merely AVM behavior with nonlinearity, we estimated an AVM model with a linear and a quadratic term for price. This model yielded a DIC score of 1889.2. The DIC score for this model, similar to the DIC score for the AVM model with only a linear term, is substantially worse than the DIC score for the OVM model. Moreover, there is little improvement in the DIC score over the AVM model with only a linear term. This is consistent with efforts in the literature in which modifications to the linear form (e.g., by using the logarithm of the price or by adding a quadratic term) have not often helped model performance.

Finally, more research is necessary to specify the factors that cause consumers to base choices on actual versus ordered value information. Marketing-mix factors and consumer-specific factors would be expected to influence consumers' use of attribute value information. For example, marketing communications that fixate on specific attribute values (e.g., $.99-only stores) rather than attribute importance (e.g., Volvo and the attribute safety) might lead consumers to focus on products' actual rather than ordered values. Significant price variation may also lead consumers to focus on actual values. In addition, consumers' appreciation of actual value information (e.g., due to their level of product knowledge) might also affect whether they choose to use actual or ordered values. All of these factors could be incorporated into an enlarged model. Note, however, that consumer choice making is probably not discrete (i.e., only in OVM or AVM) but rather a mixture. More research is necessary to specify the circumstances under which consumers use only actual or ordered values and, thus, the circumstances under which different model specifications work well as approximations of choice behavior and predict choice outcomes well.

REFERENCES

Akaike, H. (1973), "Information Theory and an Extension of the Maximum Likelihood Principle," in *Proceedings of the Second International Symposium on Information Theory*, B.N. Petrov and F. Csaki, eds. Budapest: Akademiai Kiado, 267–81.

Amemiya, Takeshi (1985), *Advanced Econometrics*. Cambridge, MA: Harvard University Press.

Ariely, Dan, George Loewenstein, and Drazen Prelec (2003), "Coherent Arbitrariness: Stable Demand Curves Without Stable Preferences," *The Quarterly Journal of Economics*, 118 (1), 73–105.

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

Burnham, Kenneth and David Anderson (2002), *Model Selection and Multi-Model Inference*. New York: Springer-Verlag.

Drolet, Aimee and Anand V. Bodapati (2003), "Use of Attribute Value Information in Choice and Sensitivity to Product Change," working paper, Anderson School of Management, University of California, Los Angeles.

———, Itamar Simonson, and Amos Tversky (2000), "Indifference Curves that Travel with the Choice Set," *Marketing Letters*, 11 (August), 199–209.

Efron, B. (2004), "The Estimation of Prediction Error: Covariance Penalties and Cross-Validation," *Journal of the American Statistical Association*, 3 (467), 619–32.

Elrod, Terry, Jordan J. Louviere, and Krishnakumar S. Davey (1992), "An Empirical Comparison of Ratings-Based and Choice-Based Conjoint Models," *Journal of Marketing Research*, 29 (August), 368–77.

Frederick, Shane and Baruch Fischhoff (1998), "Scope (In)sensitivity in Elicited Valuations," *Risk Decision and Policy*, 3 (2), 109–123.

Gelman, Andrew and Donald Rubin (1992), "Inference from Iterative Simulation Using Multiple Sequences," *Statistical Science*, 7 (4), 457–511.

Gilks, W.R., S. Richardson, and D.J. Spiegelhalter, eds. (1996), *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall.

Gustafsson, Anders, Andreas Herrmann, and Frank Huber (2000), *Conjoint Measurement: Methods and Applications*. New York: Springer Verlag.

Mallows, C.L. (1973), "Some Comments on Cp," *Technometrics*, 15 (4), 661–75.

Payne, John, James Bettman, and Eric Johnson (1992), "Behavioral Decision Research: A Constructive Processing Perspective," *Annual Review of Psychology*, 43, 87–131.

———, ———, and ——— (1993), *The Adaptive Decision Maker*. Cambridge, UK: Cambridge University Press.

Schwarz, G. (1978), "Estimating the Dimension of a Model," *Annals of Statistics*, 6 (2), 461–66.

Simonson, Itamar and Amos Tversky (1992), "Choice in Context: Tradeoff Contrast and Extremeness Aversion," *Journal of Marketing Research*, 29 (August), 281–95.

Spiegelhalter, D., N. Best, B. Carlin, and A. van der Linde (2002), "Bayesian Measures of Complexity and Fit," *Journal of the Royal Statistical Society, Series B*, 64 (4), 583–639.

———, A. Thomas, N. Best, and D. Lunn (2003), "WinBUGS User Manual Version 1.4," technical report, Medical Research Council Biostatistics Unit, Institute of Public Health, Cambridge University.

Tversky, Amos and Daniel Kahneman (1991), "The Framing of Decisions and the Psychology of Choice," *Science*, 211 (4481), 453–58.

——— and Itamar Simonson (1993), "Context-Dependent Preferences," *Management Science*, 39 (10), 1179–89.