



Marketing Science

Publication details, including instructions for authors and subscription information:
<http://pubsonline.informs.org>

A Nonparametric Approach to Identifying Latent Relationships in Hierarchical Models

Thomas S. Shively, Greg M. Allenby, Robert Kohn,

To cite this article:

Thomas S. Shively, Greg M. Allenby, Robert Kohn, (2000) A Nonparametric Approach to Identifying Latent Relationships in Hierarchical Models. Marketing Science 19(2):149-162. <https://doi.org/10.1287/mksc.19.2.149.11807>

Full terms and conditions of use: <https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

© 2000 INFORMS

Please scroll down for article—it is on subsequent pages



With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

A Nonparametric Approach to Identifying Latent Relationships in Hierarchical Models

Thomas S. Shively • Greg M. Allenby • Robert Kohn

Department of Management Science and Information Systems, The University of Texas at Austin, Austin, Texas 78712
shively@mail.utexas.edu

Max M. Fisher College of Business, Ohio State University, Columbus, Ohio 43210

Australian Graduate School of Management, University of New South Wales, Sydney, New South Wales 2052, Australia

Abstract

This paper provides a method for nonparametrically modeling the relationship between consumer preference for product features, such as reliability or durability, and covariates that describe consumers and how they use the product. This relationship is of interest to firms designing and delivering products to a market because the extent to which consumers are sensitive to particular features determines the potential profitability of product offerings, and affects decisions relating to appropriate distribution outlets and advertising strategies. The successful identification of these relationships also aids in efficiently targeting marketing activities to specific segments of the consumer population.

The relationship between consumer preference for product features and observable covariates is important but is typically unknown. In addition, these relationships are often deeply embedded in a model hierarchy and are not observed directly. For example, in models of household choice, the observed outcomes are multinomial with probabilities driven by latent utilities or values that consumers place on the choice alternatives. These utilities are in turn a function of characteristics, such as price and product features, which are differentially valued. Of primary interest is the relationship between consumer sensitivity to product characteristics and readily observed covariates such as household demographics or aspects of product usage. Because the relationships of interest are not directly observed, it is difficult to draw inferences about them without formal statistical models.

This paper presents a three-level hierarchical Bayes model for modeling binary consumer preferences as a function of observable covariates. The hierarchical model nonparametrically estimates the relationships between consumer preferences for product features and the covariates without assuming a specific functional form. A nonparametric model is particularly useful in the exploratory analysis of consumer data in which the primary purpose of the analysis is to generate further questions rather than provide specific answers to well-posed questions. This type of analysis is frequently encountered in marketing where a series of studies are commissioned to better understand the nature of demand. The first level of the hierarchy in the Bayesian model relates the binary consumer choice to the sensitivities of the consumer to product attributes such as brand name, price, reliability,

and durability. The second level of the hierarchy models the heterogeneity across consumers using functions that relate attribute sensitivities to observable covariates. This level of the hierarchy also allows each respondent to have unique demand coefficients by introducing random effect components. The third level of the hierarchy specifies a smoothness prior for each of the unknown functions used in the second level. The approach is flexible and works well both when the unknown function can be closely approximated by a linear function and when it cannot be. A Bayesian model selection technique is used to determine which functions can be modeled using a linear function and which ones should be modeled nonparametrically to provide the necessary flexibility to estimate the function accurately.

The proposed methodology is illustrated using data from a survey of consumer preferences for features of marine outboard engines that was collected as part of a consulting project. Our analysis focuses on measuring consumer preferences for engine features and their relationships to two variables related to boat length and engine size. Consumer preferences for engine features were obtained through a national survey conducted over the telephone. Preferences were elicited by means of a pairwise evaluation in which respondents chose between two engines that were identical in every respect except for two engine features. The methodology can be modified to allow for more complex comparisons such as conjoint data collected in full profiles.

The application of a Bayesian model selection procedure indicates that 4 of the 28 covariate relationships in the model are nonlinear, while the other 24 are linear. The preferences associated with these four functions are involved in 56% of the pairwise comparisons in the study. Therefore, in practice, if the nonlinear functions are not properly estimated there is the potential to draw misleading inferences regarding 56% of the pairwise choices. Firms can use the estimates of the functions relating preferences to covariates in a number of ways. First, they can use the covariates to determine the total number of consumers who have high demand for a particular product feature, and then they can target communication efforts to those individuals. Alternatively, the empirical results can be used as a basis of subsequent analysis to obtain a more complete characterization of a market segment.

(Consumer Preferences; Cubic Smoothing Spline; Hierarchical Bayes Model; Markov Chain Monte Carlo; Random Effects)

1. Introduction

What is the relationship between consumer preference for product features, such as reliability or durability, and covariates that describe consumers and how they use the product? There is no universally correct answer to this question because the relationship depends on the product and the benefits it offers. Yet this relationship is of interest to firms designing and delivering products to a market. The extent to which consumers are sensitive to particular features determines the potential profitability of product offerings, and affects decisions relating to appropriate distribution outlets and advertising strategies. In addition, the successful identification of these relationships aids in efficiently targeting marketing activities to specific segments of the consumer population. The relationship between consumer preference for product features and observable covariates is important but is typically unknown.

Further complicating the identification of relationships between determinants of demand and observable covariates is the fact that consumer preferences and sensitivities are often deeply embedded in a model hierarchy and are not observed directly. For example, in models of household choice, the observed outcomes are multinomial with probabilities driven by latent utilities or values that consumers place on the choice alternatives. These utilities are in turn a function of characteristics, such as price and product features, which are differentially valued. Of primary interest is the relationship between consumer sensitivity to product characteristics and readily observed covariates such as household demographics or aspects of product usage. Because the relationships of interest are not directly observed, it is difficult to draw inferences about them without formal statistical models.

In addition, these inferences are often based on very limited amounts of information. For example, household purchases recorded in scanner panel data sets typically contain less than a dozen or so entries in most product categories (see Allenby and Lenk 1994, 1995). Data obtained through consumer surveys are similarly limited because the quality of responses starts to deteriorate after about 20 minutes of interviewing. This data limitation precludes the precise estimation of many model parameters of interest to firms. As a result, it is often not possible to analyze latent relationships by first obtaining point estimates of parameters

and then relating them to other variables because large standard errors mask the true latent relationship.

This paper presents a hierarchical Bayesian approach for modeling binary consumer preferences as a function of observable covariates. The model estimates the effects of the covariates nonparametrically, that is, without assuming a particular functional form for the regression functions. The estimates of the regression functions are cubic smoothing splines. A number of authors aim to achieve such flexibility by using low-order polynomials, for example quadratics or cubics, but such low-order polynomials cannot capture the shape of all plausible functional forms. Our approach is more flexible and works well both when the unknown function can be closely approximated by a low-order polynomial and when it cannot be. A model selection technique is used to determine which functions can be modeled using a low-order polynomial and which ones require a smoothing spline to provide the necessary flexibility to model the relationship. The model also allows each respondent to have unique demand coefficients by introducing random effects components.

The proposed methodology is particularly useful in the exploratory analysis of consumer data in which the primary purpose of the analysis is to generate further questions rather than to provide specific answers to well-posed questions. This type of analysis is frequently encountered in marketing where a series of studies are commissioned to better understand the nature of demand. For example, variation in consumer demand for automobile features occurs mainly because consumers use automobiles to solve very different problems. Families with small children are motivated by a very different set of needs than those of a teenager. It is often not possible to specify exactly how these needs translate into demand for product features, and whether there exist observable covariates that could be used to identify groups of consumers (i.e., segments) with particularly high demand. Instead, marketers often work in reverse by first measuring demand for features and then attempting to understand how variation in demand is related to observed covariates, and, hopefully, to a set of underlying needs. Hence, results from earlier studies are used to refine the collection of data in later studies, with the goal of

eventually identifying and understanding why specific groups of consumers demand the products they do.

The paper is organized as follows. Section 2 introduces the empirical application that is a field study of consumer preferences for a product. Section 3 presents the statistical model and the priors. This section includes setting priors for the unknown regression functions. Section 4 discusses model estimation as well as the variable selection technique used to determine which functions should be estimated nonparametrically. Section 5 presents and interprets the estimation results when the model is applied to the data on consumer preferences. Section 6 contains some concluding remarks. The appendix provides the Markov chain Monte Carlo sampling schemes used to implement the estimation and variable selection techniques.

2. Empirical Example

The proposed methodology is illustrated using data from a survey of consumer preferences for features of marine outboard engines that was collected as part of a consulting project. Outboard engines are used in a variety of activities, including fishing, water skiing, and leisure boating. Our analysis focuses on measuring consumer preferences for engine features (for example, brand name, price, reliability, and durability) and their relationships to two variables related to boat length and engine size. We use boat length and engine size as the covariates in the model because they represent aspects of product usage that are expected to be related to demand for product attributes. Marketing studies typically find that covariates representing previous product usage contain more explanatory power than, for instance, demographic variables. There are 10 engine features used in our analysis, and they are presented in Table 1.

Consumer preferences for engine features were obtained through a national survey conducted over the telephone. Respondents were randomly chosen from boat registration lists. Preferences were elicited by means of a pairwise evaluation in which respondents chose between two engines that were identical in every respect except:

The first engine was *high* on feature A and *low* on feature B; and

The second engine was *low* on feature A and *high* on feature B

where “low” and “high” represent levels of performance on the feature. The specific wording used in the evaluation is provided in Table 1. Note that “high” always represents better, so “high” on price means a lower price. Using this convention makes the empirical results in §5 easier to interpret. For proprietary reasons, values of the performance levels cannot be revealed and are referred to as low and high. Specific performance levels and brand names were used in the actual survey. Pairwise evaluations were used in the study because the survey was administered over the telephone, which limited the complexity of the questions that could be asked. The methodology can be modified to allow for more complex comparisons (e.g., conjoint data collected in full profiles) and for covariates other than boat length and engine size (e.g., demographic variables such as age or income level).

Our analysis involves 682 respondents for whom boat length and engine size are available and who used outboard engines on freshwater lakes and streams. Each respondent reported their preferences for 13 to 15 pairwise evaluations involving a subset of the product attributes in Table 1, resulting in a total of 9,421 observations. A randomized design was used in which respondents were randomly assigned to one of five groups, and each group performed a different subset of evaluations. For example, respondents providing preferences about price and fuel economy did not provide information about emissions. As discussed below, inferences for these respondents about sensitivity to emissions were obtained through the random effects distribution for these respondents and estimated covariate relationships. Each respondent answered questions involving either five or six attributes. The validity of using fractionated designs to draw inferences about individual effects is discussed in detail by Lenk et al. (1996).

The goal of our analysis is to determine how sensitive a consumer is to various engine features and to determine whether boat length and engine size are related to these sensitivities. This information can be used to identify current owners who have high demand for a particular product feature and can also be used to identify usage situations that require higher or

Table 1 Description of the Data

Sample Size:	
682 Respondents	
9,421 Observations	
Attributes:	
1. Brand Name	6 brands of outboard engines were studied, denoted by: A, B, C, D, E, F
2. Price	x% less (more) expensive
3. Fuel Economy	x% more (less) fuel efficient
4. Reliability	Starts quickly every (x% of the) time
5. Durability	x% decreased (increased) risk of mechanical failure
6. Vibration and Noise	x% less (more) vibration and noise
7. Acceleration	Gets on "plane" x% faster (slower)
8. Speed	x% faster (slower) for the same horsepower
9. Emissions	x% less (more) smoke
10. Technology	Cutting-edge (standard) technology
Covariates:	
Boat Length (feet)	
Engine Size (horsepower)	

lower levels of engine performance. The relationship of interest is embedded in the middle of a model hierarchy, and standard diagnostic tests such as residual analysis are not available. Therefore, it is often useful to employ exploratory techniques to uncover patterns in the data, and to then examine these patterns more rigorously in a subsequent study by narrowing the focus of the study to a specific activity for which the brand is relevant.

3. Statistical Model

A three-level hierarchical Bayes model was used to estimate the sensitivity of the respondents to the engine features in Table 1 and to estimate the relationships of these sensitivities to boat length and engine size. The model also includes a random component that captures the extent of unobservable heterogeneity in the

population. The relationships between attribute sensitivities and the covariates are estimated using a non-parametric estimation technique, so no restrictive assumptions are placed on the functional forms of these relationships.

The observed binary choices w are defined as follows:

$$w_{ij} = \begin{cases} 1 & \text{if the first engine is preferred,} \\ 0 & \text{if the second engine is preferred,} \end{cases}$$

where i indexes the choices for the j th respondent. The probability that $w_{ij} = 1$ is modeled as

$$p(w_{ij} = 1 | x_{ij1}, x_{ij2}, \beta_j) = \Phi((x_{ij1} - x_{ij2})' \beta_j), \quad (1)$$

where Φ is the standard normal cumulative distribution function so that a probit link function is used. For each index pair (i, j) , the vectors x_{ij1} and x_{ij2} correspond to a comparison of the first and second engines. Each element of these two vectors is either zero or one. For example, suppose that a respondent's preference for brand A is compared to that for brand B, with the price of brand A lower (better) than that for brand B. Then x_{ij1} has a one for the first (brand A) and sixth (price) elements and zeros elsewhere, while x_{ij2} has a one for the second (brand B) element and zeros elsewhere. If a respondent's preference for speed and durability was being determined, then x_{ij1} has a one for the twelfth (speed) element, and zeros elsewhere; while x_{ij2} has a one for the ninth (durability) element, and zeros elsewhere. Only five of the six brands are included in the x_{ij1} and x_{ij2} to avoid multicollinearity. The vector $\beta_j = (\beta_{1j}, \dots, \beta_{14j})'$ is a 14×1 vector, and its l th element is the sensitivity of the j th respondent to the l th engine feature. We note that for a given combination of i and j the vector $x_{ij1} - x_{ij2}$ has at most three nonzero values corresponding to the features of the two engines.

The second level of the hierarchy models the heterogeneity across respondents using the following: (1) functions that relate sensitivities (β_{lj}) to the covariates; these functions are assumed common to all respondents; (2) respondent-specific random effects. Let $z_j = (z_{1j}, z_{2j})'$ be the vector of covariates for the j th respondent, where z_{1j} represents boat length in feet and z_{2j} represents engine size measured in horsepower. The l th element of β_j is modeled as

$$\beta_{lj} = f_{1l}(z_{1j}) + f_{2l}(z_{2j}) + \zeta_{1lj}, \quad (2)$$

$l = 1, \dots, 14$, and $j = 1, \dots, 682$. The model allows for a different function for each product feature and each covariate. Each function is modeled by the smoothness prior described below. The vector $\zeta_j = (\zeta_{1j}, \dots, \zeta_{14j})'$ is a respondent specific random effect that captures the effects of variables other than boat length and engine size. We model the ζ_j as independent multivariate $N(0, D)$ random variables. The diagonal elements of D indicate the amount of unexplained heterogeneity in the respondent preferences. The off-diagonal elements with positive entries indicate pairs of features which tend to be jointly preferred, while negative entries indicate pairs for which either one or the other tend to be preferred, but not both. The prior distribution for the matrix D is inverse Wishart $IW(I, 15)$.

The third level of the hierarchy specifies the prior on each of the functions f_{lk} . It will be convenient to express each function as

$$f_{lk}(z_{kj}) = \mu_{lk} + \gamma_{lk} z_{kj} + g_{lk}(z_{kj}), \quad (3)$$

with $g_{lk}(0) = g_{lk}^{(1)}(0) = 0$, where $g_{lk}^{(1)}(0)$ is the first derivative of g_{lk} evaluated at zero. Thus, g_{lk} is the nonlinear part of f_{lk} and has zero intercept. Only μ_{11} is estimated, with μ_{12} set to zero because the individual μ_{lk} are not identified unless they have informative priors. That is, only the first function f_{11} has a nonzero intercept with the remainder having zero intercept. For conciseness, we write μ_{11} as μ_1 . Diffuse priors are placed on μ_l , $l = 1, \dots, 14$, and γ_{lk} , $l = 1, \dots, 14$, and $k = 1, 2$.

We assume that the g_{lk} are independent a priori and that they all have the same prior. It is thus sufficient to discuss the prior for a typical function $g = g_{lk}$. Without loss of generality we assume that the argument z of g is nonnegative. The prior distribution on the function g is defined by the integral equation

$$g(z) = \tau \int_0^z (z - s)W(ds), \quad (4)$$

where $W(s)$ is a Wiener process with $W(0) = 0$ and $\text{Var}\{W(s)\} = s$. This prior is discussed by Wahba (1978) for cubic spline smoothing and more generally by Hastie and Tibshirani (1990) and Green and Silverman (1994) for penalized likelihood estimation. Kalyanam

and Shively (1998) use this prior to fit flexible functional forms to data in a nonhierarchical model. Unlike most of Bayesian analysis, where a prior is placed on a discrete number of parameters, Equation (4) is a prior on a curve and has the following properties: (1) The function g is smooth and has a continuous first derivative because $g^{(1)}(z) = dg(z)/dz = W(z)$; (2) for $s > 0$, the increment $g^{(1)}(z + s) - g^{(1)}(z)$ in the first derivative is not predictable from $g(u)$, $u \leq z$, because $g^{(1)}(z + s) - g^{(1)}(z) = W(z + s) - W(z)$ is independent of $g(u)$; (3) the second derivative of g is diffuse because $d^2g(z)/dz^2 = dW(z)/dz$ and $\text{Var}\{dW(z)/dz\} = \infty$; (4) if $\tau^2 = 0$, then g is identically zero and the corresponding function f is linear in z ; (5) the posterior mean of g is a cubic smoothing spline.

To complete the specification of the priors on the functions g_{lk} we assume that the scale factors τ_{lk}^2 corresponding to the functions g_{lk} are a priori independent and have flat priors. These priors are used to allow the data to determine the "smoothness" of the function estimates. Simulation results in Wood and Kohn (1998) indicate that the use of these priors provides good function estimates in binary regression for a wide range of functional forms.

4. Estimation and Variable Selection

This section provides a brief discussion of the method used to estimate the parameters and function values in the hierarchical model. It also outlines the variable selection procedure used to determine whether each function should be estimated as a linear function or estimated nonparametrically. To successfully implement the estimation and variable selection procedures it is important to take advantage of the structure of the model to reduce the computational time required to run the estimation and variable selection sampling schemes. Details of the sampling schemes used to implement these procedures are given in the appendix.

The smoothing parameters τ_{lk}^2 , the function values $f_{lk}(z_{kj})$, the random effects vectors ζ_j , and the elements of the covariance matrix D are estimated by their posterior means. A Markov chain Monte Carlo sampling scheme is used to estimate the parameters and function values because it is computationally intractable to obtain analytically the posterior moments of interest. The

practical problem that arises in the estimation is the large amount of computation required to estimate the 28 functions using 9,421 observations. There are two insights that reduce the computational burden. First, we exploit the fact that there are only 16 distinct values for boat length and 43 distinct values for engine size. Figure 1 provides histograms that summarize the data regarding the number of respondents having boats and engines of various sizes. As we show in the appendix, by averaging across observations with the same covariate values we can reduce the problem of generating 682 function values (i.e., one for each respondent) for each of the 28 functions to one of generating function values for only the distinct covariate values. These distinct values can then be used to construct the entire f_{lk} vector. Second, we exploit the structure of the design vector $(x_{ij1} - x_{ij2})'$ because, as noted in §3, at least 11 of the 14 elements of $(x_{ij1} - x_{ij2})'$ take on a value of zero. This type of structure is common in pairwise evaluations and is frequently employed in marketing studies. It allows us to construct a sampling scheme that substantially reduces the number of observations that must be cycled through each iteration by using only those observations that contain information about a specific function value. Details are given in the appendix.

Variable Selection

A variable selection method is used to determine which functions in Equation (2) should be estimated as linear functions and which functions should be estimated nonparametrically. Each function f_{lk} can be written as

$$f_{lk}(z_{kj}) = \mu_l + \gamma_{lk}z_{kj} + \tau_{lk} \int_0^{z_{kj}} (z_{kj} - s)W_{lk}(ds). \quad (5)$$

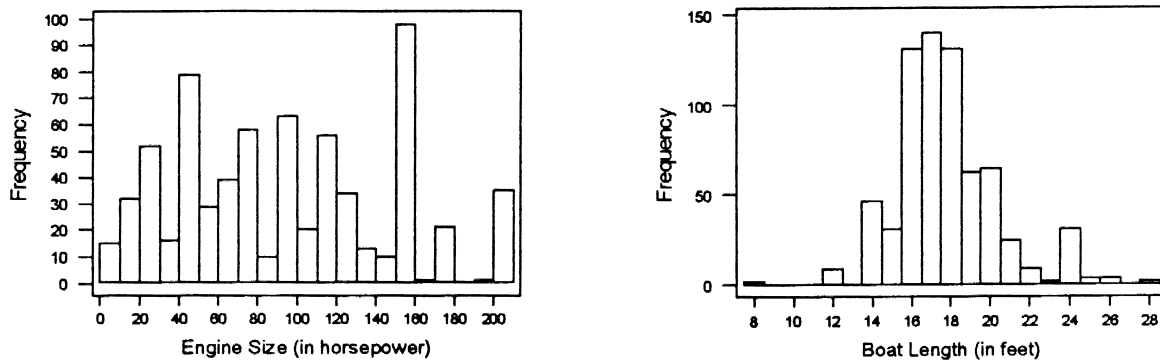
(If $k = 2$, then μ_l is set to 0.) Note that if $\tau_{lk}^2 = 0$, then f_{lk} is a linear function, while if $\tau_{lk}^2 > 0$, then f_{lk} should be estimated nonparametrically. Therefore, we compute the posterior probability $\text{pr}(\tau_{lk}^2 > 0 \mid w)$ where w is a 9421×1 vector containing the w_{ij} s. The higher the probability, the stronger the evidence that a linear function is inadequate to model f_{lk} .

To compute $\text{pr}(\tau_{lk}^2 = 0 \mid w)$ and $\text{pr}(\tau_{lk}^2 > 0 \mid w)$, let M_{lk} be an indicator variable such that if $\tau_{lk}^2 = 0$, then $M_{lk} = 0$, and if $\tau_{lk}^2 > 0$, then $M_{lk} = 1$. We use the following priors for M_{lk} and τ_{lk}^2 in our variable selection technique. First, the prior for each M_{lk} is $\text{pr}(M_{lk} = 0) = \text{pr}(M_{lk} = 1) = 1/2$. The prior distributions for the τ_{lk}^2 are the following: If $M_{lk} = 0$, then $\tau_{lk}^2 = 0$ with probability one (i.e., a point mass at 0). If $M_{lk} = 1$, then a data generated prior is used for τ_{lk}^2 . To obtain this prior, the full model is estimated (i.e., all functions for boat length and engine size are estimated nonparametrically) and estimates of the posterior mean and variance of $\log(\tau_{lk}^2) \mid w$, denoted $\hat{\mu}_{\tau_{lk}}$ and $\hat{\sigma}_{\tau_{lk}}^2$, are obtained. We then use the lognormal prior for τ_{lk}^2 :

$$p(\tau_{lk}^2) = (2\pi)^{-1/2} \frac{1}{\tau_{lk}^2} (c\hat{\sigma}_{\tau_{lk}}^2) \cdot \exp\left(\frac{-1}{2(c\hat{\sigma}_{\tau_{lk}}^2)} (\log(\tau_{lk}^2) - \hat{\mu}_{\tau_{lk}})^2\right),$$

where c is an appropriate multiple of $\hat{\sigma}_{\tau_{lk}}^2$. The rationale for using these priors as well as how to choose c is given by Shively et al. (1999).

Figure 1 Histograms for Engine Size and Boat Length for All Respondents



The variable selection algorithm therefore requires two sampling schemes. The first sampling scheme is used to compute $\hat{\mu}_{\tau_{lk}}$ and $\hat{\sigma}_{\tau_{lk}}^2$ for each function f_{lk} , and therefore provides the data generated priors for τ_{lk}^2 . The second sampling scheme then uses these priors to compute the posterior probabilities $\text{pr}(M_{lk} = 0 \mid w)$ and $\text{pr}(M_{lk} = 1 \mid w)$. The two sampling schemes are given in the appendix.

5. Results for Empirical Example

The application of the variable selection procedure described in § 4 to the model in Equations (1)–(3) indicates that four of the 28 covariate relationships are nonlinear, with posterior probabilities greater than 0.50 that a nonlinear function is required to model the relationship (i.e., posterior odds greater than one). The four nonlinear functions relate engine size to preferences for brand E, reliability, durability, and emissions (the corresponding posterior probabilities that the functions are nonlinear are 1.0, 0.814, 0.919, and 0.745, respectively). The functions relating engine size to the preferences for the other 10 product attributes are linear, and the functions relating all the preferences to boat length are also linear.

The model was re-estimated with the four nonlinear functions estimated nonparametrically and the other 24 functions estimated as linear functions. The estimates of the nonlinear functions are plotted in Figures 2 through 5. The intercepts and slope coefficients for the linear functions for the other 24 functions are reported in Table 2. A model in which all 28 functions are restricted to be linear was also estimated for comparison purposes. The estimates of the four linear functions relating preferences to brand E, reliability, durability, and emissions are plotted as the dashed lines in Figures 2 through 5.

A Bayesian procedure that imposes a penalty for extra parameters can be used to compare the fit of the nonparametric model with four nonlinear functions to the fully linear model. More specifically, the posterior probabilities for the two models are used to determine which model provides a better fit to the data. The calculation of the posterior probability for a specific model imposes a penalty for each parameter in the model because the parameters are integrated out. The

Figure 2 Nonlinear and Linear Relationships Between Respondent Preference for Brand E and Engine Size (in Horsepower)

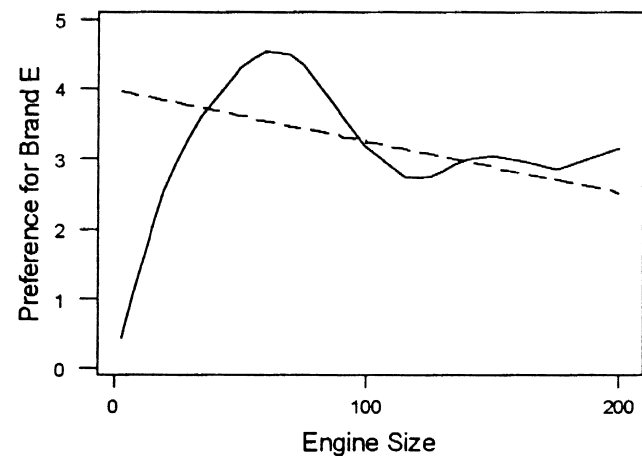


Figure 3 Nonlinear and Linear Relationships Between Respondent Preference for Reliability and Engine Size (in Horsepower)

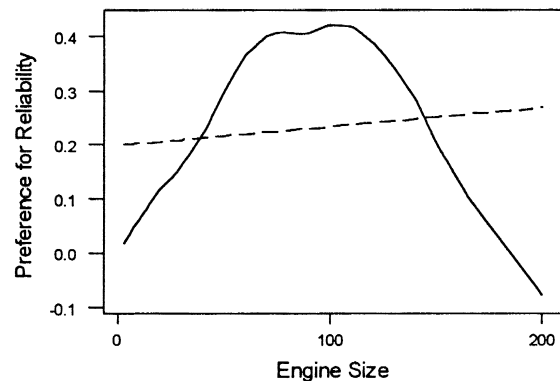


Figure 4 Nonlinear and Linear Relationships Between Respondent Preference for Durability and Engine Size (in Horsepower)

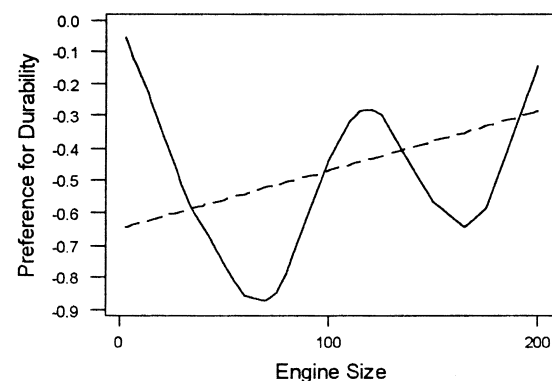


Figure 5 Nonlinear and Linear Relationship Between Respondent Preference for Emissions and Engine Size (in Horsepower)

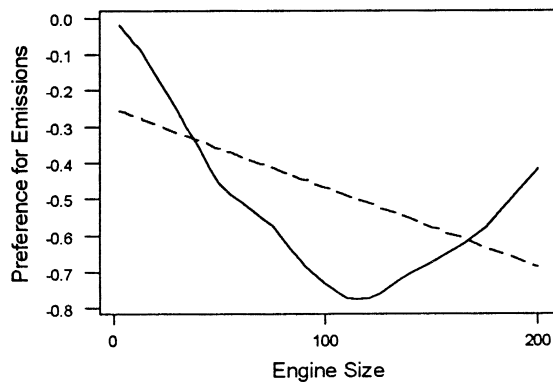


Table 2 Posterior Means of the Coefficients for the 14 Product Attributes for the Linear (Nonspline) Functions

Note: Posterior standard deviations are reported in parentheses

Product Feature	Intercept	Boat Length (z_1)	Engine Size (z_2)
Brand A	1.674 (0.868)	-0.031 (0.054)	-0.0063 (0.0035)
Brand B	2.958 (1.104)	-0.120 (0.067)	-0.0037 (0.0041)
Brand C	1.259 (1.361)	-0.147 (0.084)	0.0040 (0.0055)
Brand D	0.057 (1.818)	-0.037 (0.113)	-0.0161 (0.0075)
Brand E	1.251 (2.261)	-0.287 (0.117)	Spline Function
Price	-0.162 (0.882)	0.023 (0.056)	0.0008 (0.0030)
Fuel Economy	-0.288 (0.902)	0.093 (0.056)	0.0005 (0.0032)
Reliability	0.211 (0.799)	0.042 (0.049)	Spline Function
Durability	1.098 (0.887)	0.077 (0.055)	Spline Function
Vibration	0.185 (0.819)	0.092 (0.051)	-0.0014 (0.0029)
Acceleration	-0.008 (0.862)	0.065 (0.054)	-0.0001 (0.0029)
Speed	-0.102 (0.795)	0.052 (0.049)	0.0033 (0.0028)
Emissions	0.250 (0.900)	0.127 (0.054)	Spline Function
Technology	-2.898 (0.931)	0.197 (0.058)	0.0005 (0.0031)

posterior probabilities are 1.0 for the nonparametric model and 0.0 for the fully linear model, so there is very strong evidence that the nonparametric model provides a substantially better fit to the data even after penalizing for its extra complexity.

Figures 2 through 5 indicate how preferences for various engine features are related to the size of the outboard engine currently owned by the respondent. The solid curves representing the nonlinear function estimates indicate substantial variation in preferences for each of the four features, with respondents differing in their estimated preferences for product features by a half unit or more. Because a probit link function was used, a difference of a half unit in the function estimates can translate to differences up to 0.19 in the choice probability, while a difference of one unit in the function estimates can translate to differences up to 0.38. This is discussed in more detail below.

The linear function estimates in Figures 2 through 5 do not show as much variation in preferences as the nonlinear function estimates, particularly in Figures 3 and 4, and therefore provide a different interpretation of the preferences for each attribute among respondents. For example, in Figure 3 the nonlinear function estimate varies by a half unit, while the linear function estimate varies by less than 0.1 unit.

The nonlinear function estimate in Figure 2 indicates that brand E is most preferred by respondents with engines between 50 and 100 horsepower and is not valued as highly by those boat owners with engines larger than 100 horsepower. Figure 3 indicates that engine reliability (engine starts quickly) is a feature that is preferred most by respondents with moderately sized engines. Figure 4 displays a highly nonlinear relationship between durability and engine size, in which a group of respondents with engines between 100 and 140 horsepower have higher demand for this feature while those with both slightly smaller and slightly larger engines have less extreme preference. Finally, Figure 5 indicates that engine emissions (smoke) are of much greater concern to respondents with small engines than to those with medium or large engines.

By allowing for a flexible functional form between preferences and the covariates, the spline allows for more abrupt changes in the relationship than can be represented by a low-order polynomial. For example,

a cubic polynomial is insufficient to capture the relationship between durability ($l = 9$) and engine size ($k = 2$) in Figure 4. This is confirmed when we compute posterior probabilities that the spline term $g_{lk} = g_{92}$ in Equation (3) is still required when quadratic and cubic terms are included in the model for β_{9j} . The posterior probability that τ_{92}^2 is greater than zero is 0.78, which indicates that a cubic polynomial is not flexible enough to model the relationship between durability and engine size. The advantage of the spline methodology is that it allows a user to determine if a low-order polynomial is sufficient to model each relationship. If it is not, the nonparametric spline model allows the data to specify the appropriate functional form. This model provides a more flexible method of uncovering groups of respondents with similar covariate values who have heightened demand for particular product features. These groups are often referred to as segments in the marketing literature (see Kotler 1997).

Figure 6 shows how the choice share for an engine involved in a specific pairwise comparison varies as engine size varies. This figure summarizes the results for a pairwise trade-off involving the attributes of emissions and durability. The first engine in the pairwise comparison has 10% lower emissions (attribute

13) than the second engine, while the second engine has 10% better durability (attribute 9) than the first. We set boat length to 17 feet (because this is the most common boat length and in the middle of the boat length range) and the random effects vector ζ to $(0, 0, \dots, 0)'$ to represent a typical respondent.

The solid line in Figure 6 is the estimate of the probability of choosing the first engine obtained using the linear model (i.e., all 28 functions are estimated linearly). The long dashed line is the estimated probability obtained when 24 functions are estimated linearly and the remaining four are estimated nonparametrically. The short dashed lines are the 95% confidence bands obtained from the model with all 28 functions estimated linearly. Figure 6 shows that the estimated choice share function from the nonparametric model falls outside the 95% confidence bands at each of the peaks and valleys in the function. Many of the probability functions for pairwise trade-offs involving durability show similar behavior with the nonparametric estimates falling outside the 95% confidence bands from the linear model.

Figure 6 also shows a practical difference between the probabilities given by the linear model and those given by the nonparametric model. For example, for an engine size of 110 horsepower, the difference in the probabilities between the linear and nonlinear estimates of the probabilities is approximately 0.15; while for an engine size of 170 horsepower, the difference is approximately 0.20.

To provide evidence that the variability in the function estimates are not a result of idiosyncratic noise, we reran the estimation procedure four times, with a different quarter of the data omitted each time, and compared the estimated choice share functions. For example, there are 682 respondents in the sample, so for the first run the model is estimated with the first 25% of the data omitted. For the second run, the second 25% of the data are omitted, and so on. Figure 7 provides the four estimates of the choice share function for the pairwise comparison involving emissions and durability. The results for this pairwise comparison are provided because the function relating the sensitivity of durability to engine size is the most nonlinear and is therefore most susceptible to having been affected by idiosyncrasies in the data. The four function estimates

Figure 6 Estimated Choice Share as a Function of Engine Size (in Horsepower) from Linear and Nonparametric Models for Engine 1 in a Pairwise Comparison.

Note: Engine 1 has 10% lower emissions than Engine 2, and Engine 2 has 10% better durability than Engine 1. (The solid line is the estimate from the linear model, the long-dashed line is the estimate from the nonparametric model, and the two short-dashed lines are 95% confidence bands from the linear model.)

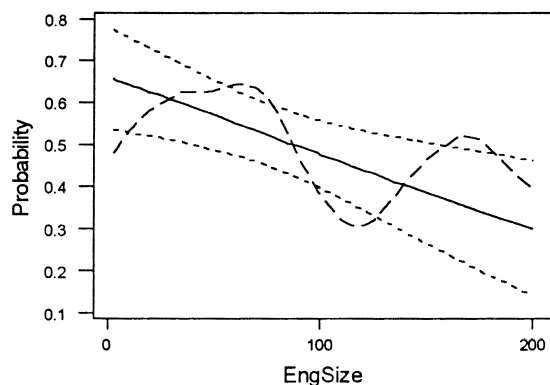
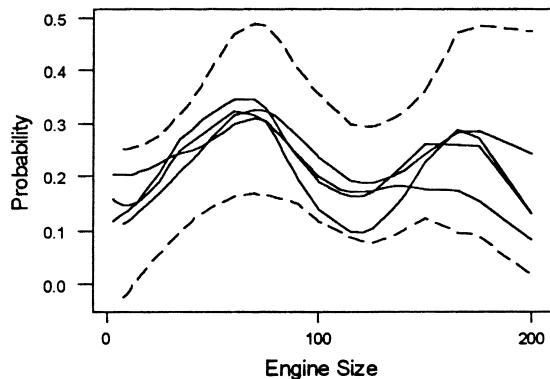


Figure 7 Estimated Choice Share from the Nonparametric Model as a Function of Engine Size (in Horsepower) for Four Data Sets Where a Different 25% of the Full Data Set Is Removed Each Time

Note: Engine 1 in the Pairwise Comparison has 10% lower emissions than Engine 2, and Engine 2 has 10% better durability than Engine 1. (Dashed lines are 95% confidence bands.)



all show the same general shape with bumps for engine size values of approximately 70 and 160 horsepower. The dashed lines are the 95% confidence bands obtained when the choice share function is estimated with the first 25% of the data omitted. Plots of the estimated choice share functions for other pairwise comparisons involving brand E, reliability, durability, and emissions (i.e., the four nonlinear functions in the model) also show similar shapes for each of the four data sets.

The high posterior probabilities that the functions in Figures 2 through 5 are nonlinear, along with the results discussed above relating to Figure 7, indicate it is unlikely that the nonlinearity in the functions in Figures 2 through 5 is a result of random fluctuations in the data. This possibility can never be ruled out entirely, but the evidence in the data suggests it is unlikely. Some of the nonmonotonicity is difficult to explain (e.g., the function in Figure 4 that relates respondent preference for durability to engine size). It is possible that omitted covariates are driving the nonlinearity. However, this indicates one of the strengths of the nonparametric modeling procedure as an exploratory data analysis technique, because it suggests avenues for further exploration of the data (as discussed below).

An out-of-sample validation procedure was used to

validate the model, with the first 75% of the observations used for estimation and the remaining 25% used for out-of-sample validation. There was little difference between the validation results for the parametric and nonparametric models, even though the variable selection and graphical results show strong evidence of nonlinearity in four of the functions. We believe the reason is the following: Detecting differences between the linear and nonlinear models using model validation is difficult because the differences, while real, are large for only a limited range of *Engine Size* values. For example, in Figure 6 where the pairwise comparison is between emissions and durability, only a small number of respondents in the validation sample fall in the regions where the differences between the linear and nonparametric choice share functions differ by more than 0.10. This means it will be difficult to detect differences between the two models because the inherent random variability in the dependent variable will likely overwhelm the differences because of the small number of pairwise comparisons that will have substantial differences in the choice probabilities. However, the population of boat owners is very large, and in particular, the number of boat owners that have substantial differences between the choice probabilities implied by the two models is large. This implies the differences in the probabilities will translate into large and practically significant differences in the number of boat owners in the population actually interested in an attribute such as durability.

The covariance matrix (D) of the random effects distribution is available from the authors on request. The diagonal elements of the covariance matrix are large (ranging from 1.4 to 3.4), indicating a substantial amount of unexplained heterogeneity in respondent preferences. Many of the off-diagonal elements are nonzero, particularly for the engine features not associated with the name of the manufacturer. For example, the covariance between reliability and durability is 1.63, which corresponds to a correlation of 0.83. Positive covariances indicate pairs of features which tend to be evaluated similarly after accounting for the variability explained by the covariates, with either both features important to a particular respondent or neither being important. Many of the posterior means of

the covariance terms are more than two posterior standard deviations from zero, indicating the importance of including a random effects distribution in the model specification.

6. Discussion

Firms can use the estimates of the functions relating preferences to covariates in a number of ways. First, they can use the covariates to determine the total number of consumers who have high demand for a particular product feature, and then target communication efforts to those individuals. For example, consider the respondents in Figure 4 whose engine size is between 100 and 140 horsepower and who prefer high durability. A manufacturer considering improvements in the durability of their outboard engine would be critically interested in the number of boat owners with engines in the 100–140 horsepower range, and whether it is possible to identify them through sources such as state licensing agencies.

Alternatively, the empirical results can be used as a basis of subsequent analysis to obtain a more complete characterization of a market segment. Further analysis of our questionnaire reveals that 29 percent of the respondents with 100–140 horsepower engines report using their boats primarily for nonfishing activities (e.g., recreational boating, water skiing, and camping). In contrast, respondents with engines in the 35–100 horsepower range and the 140–170 horsepower range report this type of primary usage only 19 and 13 percent of the time, respectively. A cross-tabulation of these three segments (35–100 horsepower, 100–140 horsepower, and 140–170 horsepower) against primary boat usage indicates a statistically significant relationship with a p -value of 0.02. The 100–140 horsepower segment also tends to have a higher proportion of first-time boat owners (23%) than the other two groups (18% and 12%, p -value = 0.06) and tend to own multi-purpose (62%) rather than single-purpose boats (32% and 42%, p -value = 0.00).

Subsequent analysis would attempt to understand why these differences exist. Why, for example, do owners of engines in the 100–140 horsepower range have a heightened sensitivity to durability, and in what way is this related to nonfishing activities? Do

these boaters come in contact with submerged objects, or use their engines for long periods of time? Do they use their boats in remote locations or in extreme climatic environments? Durability does not have a precise meaning and may translate into different engineering specifications depending on the context and environment of engine use. Furthermore, it is possible that consumers within a specific context may express a variety of concerns and interests about the engines. Some water skiers, for example, may want to build their endurance and need to maintain high speeds for extended periods of time without engine overheating. Other skiers may be more sensitive to acceleration and require engines that stand up to the strain of frequent starts and stops. Insights into the questions raised by our exploratory analysis can be obtained from additional analyses that focus on a more limited scope of boating activities.

The value of using the smoothness prior in Equation (4) is that it results in a functional specification that flexibly summarizes relationships that are embedded in the middle of a model hierarchy. These relationships are of critical interest to firms attempting to relate covariates to model-based estimates of preferences and sensitivities. These relationships are difficult to specify a priori because of the lack of any guiding theory that would help specify an appropriate functional form. The smoothness prior, therefore, offers a useful approach to identifying these latent relationships without imposing any functional form assumptions that obscure the true latent relationship.¹

Appendix: Sampling Schemes for Estimation and Variable Selection

This appendix provides the sampling schemes used to estimate the unknown parameters and functions of the model and to do variable selection. It also provides a detailed description of how the sampling scheme for estimation is implemented to make the estimation technique computationally feasible in a large scale problem such as the one in this paper.

A Gibbs sampler is used to estimate all parameters of interest

¹The authors thank the area editor and two reviewers for their helpful comments and suggestions. They improved the paper considerably. Tom Shively's work was partially supported by a Faculty Research Committee grant from the Graduate School of Business at the University of Texas at Austin. Robert Kohn's work was partially supported by an Australian Research Council grant.

because it is computationally intractable to obtain analytically the posterior moments of interest. For very readable introductory accounts of the Gibbs sampler, see Gelfand and Smith (1990) and Casella and George (1992). To implement the Gibbs sampler we follow Albert and Chib (1993) and McCulloch and Rossi (1994) and introduce the additional latent variables y_{ij} such that

$$y_{ij} = (x_{ij1} - x_{ij2})' \beta_j + \epsilon_{ij},$$

where the ϵ_{ij} are $N(0,1)$ random variables and independent for all i and j , and are also independent of β_j . Let $w_{ij} = 1$ if $y_{ij} > 0$, and let $w_{ij} = 0$ otherwise. It is readily checked that w_{ij} satisfies Equation (1).

To describe the sampling scheme it is necessary to introduce some extra notation. Let I_j be the number of questions asked of respondent j (I_j is either 13, 14, or 15 for each j), $J = 682$ be the number of respondents, $K = 2$ be the number of covariates, and $L = 14$ be the number of product features. Let $w = \{w_{ij}: i = 1, \dots, I_j; j = 1, \dots, J\}$, $y = \{y_{ij}: i = 1, \dots, I_j; j = 1, \dots, J\}$, $\gamma = \{\gamma_{lk}: l = 1, \dots, L; k = 1, \dots, K\}$, $z_k = \{z_{kj}: j = 1, \dots, J\}$, $g_{lk} = \{g_{lk}(z_{kj}): j = 1, \dots, J\}$, $g = \{g_{lk}: l = 1, \dots, L; k = 1, \dots, K\}$, $\mu = \{\mu_l, \dots, \mu_L\}$, $\zeta = \{\zeta_l, \dots, \zeta_L\}$, and $\tau^2 = \{\tau_{lk}^2: l = 1, \dots, L; k = 1, \dots, K\}$. For conciseness we use the notation g_{lk} in two senses. First, g_{lk} is a function with argument z_{kj} . Second, for the purposes of the sampling scheme, we use g_{lk} to denote the set of function values $\{g_{lk}(z_{kj}): j = 1, \dots, J\}$.

Sampling Scheme 1

(A) Choose initial values $\gamma^{[0]}$, $g^{[0]}$, $\mu^{[0]}$, $\zeta^{[0]}$, $(\tau^2)^{[0]}$ and $D^{[0]}$. These values are either fixed or generated from some distribution.

Let $A = \{w, y, \gamma, g, \mu, \zeta, \tau^2, D\}$.

(B) Generate from the following conditional distributions:

- (i) $p(y_{ij} | A - \{y_{ij}\})$, for $i = 1, \dots, I_j$ and $j = 1, \dots, J$;
- (ii) $p(\mu_l, \gamma_{lk}, g_{lk} | A - \{\mu_l, \gamma_{lk}, g_{lk}\})$ for $l = 1, \dots, L$ and $k = 1; p(\gamma_{lk}, g_{lk} | A - \{\gamma_{lk}, g_{lk}\})$ for $l = 1, \dots, L$ and $k = 2, \dots, K$;
- (iii) $p(\tau_{lk}^2 | A - \{\tau_{lk}^2\})$ for $l = 1, \dots, L$ and $k = 1, \dots, K$;
- (iv) $p(\zeta_j | A - \{\zeta_j\})$ for $j = 1, \dots, J$;
- (v) $p(D | A - \{D\})$.

If f_{lk} is restricted to be a linear function, then set $\tau_{lk}^2 = 0$ in step B(ii).

Steps B(i)–B(iii) are similar to those given in the algorithm by Wood and Kohn (1998), with two exceptions. First, in Step B(ii), we exploit the specific structure of the design vector $(x_{ij1} - x_{ij2})'$ and the numerous repeated values in the covariates to improve the speed of the algorithm. Second, in Step B(ii) we use an eigenvalue decomposition that speeds up the algorithm as a result of the structure of the model. The implementation of Step B(ii) in the sampling scheme is outlined below. Steps B(iv) and (v) are taken from an algorithm given in Allenby and Ginter (1995).

Implementation of Step B(ii)

We exploit the structure of the design vector $x_{ij1} - x_{ij2}$, and the multiple repeated values in z_1 and z_2 to generate the random variates in this step in a computationally efficient manner. The repeated values in z_1 (a similar argument holds for z_2) allow observations with the same z_1 values to be averaged. This reduces the problem of generating $J = 682$ function values for g_{1l} to one of generating function values for only the distinct z_1 values and then using these values to

construct the entire g_{1l} vector. We will outline how to generate $(\mu_l, \gamma_{1l}, g_{1l}) | A - \{\mu_l, \gamma_{1l}, g_{1l}\}$.

To construct an algorithm for generating these random variates, we express β_{1j} as

$$\begin{aligned} \beta_{1j} &= f_{11}(z_{1j}) + f_{12}(z_{2j}) \\ &= [\mu_l + \gamma_{11}z_{1j} + g_{11}(z_{1j})] + [\gamma_{12}z_{2j} + g_{12}(z_{2j})]. \end{aligned}$$

Let $\tilde{x}_{ij}(l)$ be the l th element of $x_{ij1} - x_{ij2}$, and

$$\begin{aligned} \tilde{y}_{ij} &= y_{ij} - [\tilde{x}_{ij}(1)\beta_{1j} + \dots + \tilde{x}_{ij}(l-1)\beta_{1,j} \\ &\quad + \tilde{x}_{ij}(l+1)\beta_{1,j} + \dots + \tilde{x}_{ij}(14)\beta_{1,j}] \\ &\quad - \tilde{x}_{ij}(l)[\gamma_{12}z_{2j} + g_{12}(z_{2j})] + \epsilon_{ij}. \end{aligned}$$

Then

$$\tilde{y}_{ij} = \tilde{x}_{ij}(l) [\mu_l + \gamma_{11}z_{1j} + g_{11}(z_{1j})] + \epsilon_{ij}. \quad (\text{A.1})$$

Therefore, generating $(\mu_l, \gamma_{1l}, g_{1l}) | A - \{\mu_l, \gamma_{1l}, g_{1l}\}$ is equivalent to generating $(\mu_l, \gamma_{1l}, g_{1l}) | \tilde{y}$, where $\tilde{y} = \{\tilde{y}_{ij}: i = 1, \dots, I_j, j = 1, \dots, J\}$. If $\tilde{x}_{ij}(l)$ is zero, then the ij th observation contains no information regarding γ_{1l} and g_{1l} , and can be omitted when these quantities are generated. Further, there are only 16 distinct values for z_{1j} . (There are 43 distinct values for z_{2j} .) The remaining $\tilde{x}_{ij}(l)$ are all 1 or -1 so $[\tilde{x}_{ij}(l)]^2 = 1$. Multiplying the left and right sides of (A.1) by $\tilde{x}_{ij}(l)$ gives

$$\tilde{y}_{ij}^* = \tilde{x}_{ij}(l)\tilde{y}_{ij} = \mu_l + \gamma_{11}z_{1j} + g_{11}(z_{1j}) + \epsilon_{ij}^*, \quad (\text{A.2})$$

where $\epsilon_{ij}^* = \tilde{x}_{ij}(l)\epsilon_{ij}$ are independent $N(0, 1)$ random variables. Observations with $\tilde{x}_{ij}(l) = 0$ are not included in (A.2).

For the j th respondent, z_{1j} (i.e., the length of the boat currently owned by the respondent) is the same for each question $i = 1, \dots, I_j$ asked of the respondent so the observations within a respondent can be averaged without losing any information regarding μ_l, γ_{1l} , and g_{1l} . In addition, many respondents will have the same z_1 value so the observations can also be averaged across respondents with the same z_1 . This gives

$$\bar{y}_s = \mu_l + \gamma_{11}\bar{z}_{1s} + g_{11}(\bar{z}_{1s}) + \bar{\epsilon}_s, \quad (\text{A.3})$$

$s = 1, \dots, S_{1l}$ where S_{1l} is the number of distinct values that z_{1j} takes on in (A.2), \bar{z}_{1s} , $s = 1, \dots, S_{1l}$, are the distinct values z_{1j} takes on, \bar{y}_s and $\bar{\epsilon}_s$ are the averages of the \tilde{y}_{ij}^* and ϵ_{ij}^* values for which z_{1j} takes on the same value, and $\bar{\epsilon}_s \sim N(0, 1/n_s)$ where n_s is the number of observations in (A.2) for which z_{1j} takes on the s th distinct value. Note that S_{1l} will be different for each l because different $\tilde{x}_{ij}(l)$ elements will be zero for each l . The observations in (A.3) can be arbitrarily ordered, so we will assume that the observations are ordered in ascending order with respect to \bar{z}_{1s} .

(A.3) can be rewritten in matrix notation as

$$\bar{y} = \bar{Z} \begin{pmatrix} \mu_l \\ \gamma_{1l} \end{pmatrix} + \bar{g}_{1l} + \bar{\epsilon}, \quad (\text{A.4})$$

where $\bar{y} = (\bar{y}_1, \dots, \bar{y}_{S_{1l}})'$, \bar{Z} is an $S_{1l} \times 2$ matrix whose s th row is $(1, \bar{z}_{1s})$, and $\bar{\epsilon} = (\bar{\epsilon}_1, \dots, \bar{\epsilon}_{S_{1l}})' \sim N(0, G)$ with $G = \text{diag}(1/n_1, \dots, 1/n_{S_{1l}})$. $\bar{g}_{1l} = \{g_{1l}(\bar{z}_{1s}), \dots, g_{1l}(\bar{z}_{1S_{1l}})\}'$, with $\bar{g}_{1l} \sim N(0, \tau_{1l}^2 V)$ where V is a well-defined $S_{1l} \times S_{1l}$ symmetric matrix with s th value given by

$$V(s, t) = \bar{z}_{1t}^2 \left(\bar{z}_{1t} - \frac{\bar{z}_{1t} + \bar{z}_{1s}}{2} + \frac{\bar{z}_{1s}}{3} \right)$$

for $\bar{z}_{1s} < \bar{z}_{1t}$. Note that the original 9,421 observations have been reduced to S_{1l} observations, where S_{1l} is the number of distinct values the regressor z_1 takes on in (A.2). We note here that for boat length, S_{1l} takes on values ranging from 13 to 16 for the 14 attributes. For engine size, S_{1l} takes on values ranging from 33 to 43. This makes the sampling scheme we employ feasible to implement in practice.

To generate $(\mu_l, \gamma_{1l}, \bar{g}_{1l}) \mid \bar{y}$, an eigenvalue method is used rather than the state space approach used by Wood and Kohn (1998) because for a small number of distinct knots it is computationally faster to use eigenvalues. First, let $G^{-1/2} = \text{diag}(\sqrt{n_1}, \dots, \sqrt{n_{S_{1l}}})$ and multiply both sides of (A.4) by $G^{-1/2}$ to give

$$G^{-1/2}\bar{y} = G^{-1/2}\bar{Z} \begin{pmatrix} \mu_l \\ \gamma_{1l} \end{pmatrix} + G^{-1/2}\bar{g}_{1l} + G^{-1/2}\bar{\varepsilon}, \quad (\text{A.5})$$

where $G^{-1/2}\bar{\varepsilon} \sim N(0, I)$ and $G^{-1/2}\bar{g}_{1l} \sim N(0, \tau_{1l}^2 G^{-1/2} V G^{-1/2})$. Next, let $PRP' = G^{-1/2} V G^{-1/2}$ where P is an $S_{1l} \times S_{1l}$ matrix whose columns contain the eigenvectors of $G^{-1/2} V G^{-1/2}$, and R is a diagonal matrix whose diagonal elements are the eigenvalues of $G^{-1/2} V G^{-1/2}$. Note that the eigenvalues and eigenvectors need to be computed only once, before the sampling scheme starts; they do not need to be computed each iteration. Multiplying both sides of (A.5) by P' gives

$$\hat{y} = \hat{Z} \begin{pmatrix} \mu_l \\ \gamma_{1l} \end{pmatrix} + \hat{g}_{1l} + \hat{\varepsilon},$$

where $\hat{y} = P' G^{-1/2} \bar{y}$, $\hat{Z} = P' G^{-1/2} \bar{Z}$, $\hat{g}_{1l} = P' G^{-1/2} \bar{g}_{1l}$, and $\hat{\varepsilon} = P' G^{-1/2} \bar{\varepsilon}$. Note that $\hat{g}_{1l} \sim N(0, \tau_{1l}^2 R)$ and $\hat{\varepsilon} \sim N(0, I)$.

We now generate $(\mu_l, \gamma_{1l}, \hat{g}_{1l}) \mid \hat{y}$ in two steps. First, generate $(\mu_l, \gamma_{1l}) \mid \hat{y}$, and then generate $\hat{g}_{1l} \mid \hat{y}, (\mu_l, \gamma_{1l})$. To generate $(\mu_l, \gamma_{1l}) \mid \hat{y}$, note that

$$\hat{y} \mid (\mu_l, \gamma_{1l}) \sim N \left(\hat{Z} \begin{pmatrix} \mu_l \\ \gamma_{1l} \end{pmatrix}, I + \tau_{1l}^2 R \right).$$

Standard Bayesian calculations can be used to show that $(\mu_l, \gamma_{1l}) \mid \hat{y}$ is normally distributed with mean and variance given by

$$E \left(\begin{pmatrix} \mu_l \\ \gamma_{1l} \end{pmatrix} \mid \hat{y} \right) = (\hat{Z}'(I + \tau_{1l}^2 R)^{-1} \hat{Z})^{-1} \hat{Z}' \hat{y} \sim (I + \tau_{1l}^2 R)^{-1} \hat{y},$$

$$\text{Var} \left(\begin{pmatrix} \mu_l \\ \gamma_{1l} \end{pmatrix} \mid \hat{y} \right) = (\hat{Z}'(I + \tau_{1l}^2 R)^{-1} \hat{Z})^{-1}. \quad (\text{A.6})$$

To generate $\hat{g}_{1l} \mid \hat{y}, (\mu_l, \gamma_{1l})$, standard multivariate normal calculations can be used to show that $\hat{g}_{1l} \mid \hat{y}, (\mu_l, \gamma_{1l})$ is normally distributed with mean and variance given by

$$E(\hat{g}_{1l} \mid \hat{y}, (\mu_l, \gamma_{1l})) = (\tau_{1l}^2 R)(I + \tau_{1l}^2 R)^{-1} \left(\hat{y} - \hat{Z} \begin{pmatrix} \mu_l \\ \gamma_{1l} \end{pmatrix} \right),$$

$$\text{Var}(\hat{g}_{1l} \mid \hat{y}, (\mu_l, \gamma_{1l})) = (\tau_{1l}^2 R)(I + \tau_{1l}^2 R)^{-1}. \quad (\text{A.7})$$

Finally, to generate $g_{1l} \mid \bar{y}, (\mu_l, \gamma_{1l})$, transform \hat{g}_{1l} to \bar{g}_{1l} using $\bar{g}_{1l} = P \hat{g}_{1l}$, and then expand \bar{g}_{1l} appropriately to give g_{1l} .

We point out here that the computational efficiency described

above where we exploit the specific structure of the design vector $(x_{ij1} - x_{ij2})'$ is available only for conjoint models and does not apply to more standard nonparametric regression models.

The output of the sampling scheme can now be used to estimate the posterior moments of the parameters of interest in the usual way. The results reported in § 5 are based on 20,000 iterations of the sampling scheme chain in which the first 10,000 iterations are discarded. Parameter point estimates and other summary statistics are based on the last 10,000 iterations. Time series plots indicate that the draws converge in distribution well within the initial 10,000 iterations.

Sampling Scheme for Variable Selection

The variable selection algorithm requires two sampling schemes. The first is Sampling Scheme 1, given above. It is used to compute $\hat{\mu}_{\tau_{lk}}$ and $\hat{\sigma}_{\tau_{lk}}^2$ for each function f_{lk} , and therefore provides the data generated priors for τ_{lk}^2 discussed in §4. The sampling scheme given below then uses these priors to compute the posterior probabilities $\text{pr}(M_{lk} = 0 \mid w)$ and $\text{pr}(M_{lk} = 1 \mid w)$.

To describe the second sampling scheme, define $M = \{M_{lk} : l = 1, \dots, L; k = 1, \dots, K\}$.

Sampling Scheme 2

(A) Choose initial values $\gamma^{[0]}, g^{[0]}, \mu^{[0]}, \zeta^{[0]}, M^{[0]}, (\tau^2)^{[0]}$ and $D^{[0]}$. These values are either fixed or generated from some distribution.

Let $A^* = \{w, y, \gamma, g, \mu, \zeta, M, \tau^2, D\}$.

(B) Generate from the following conditional distributions:

- (i) $p(y_{ij} \mid A^* - \{y_{ij}\})$, for $i = 1, \dots, I_j$ and $j = 1, \dots, J$;
- (ii) $p(M_{lk}, \tau_{lk}^2, \mu_l, \gamma_{lk}, g_{lk} \mid A^* - \{M_{lk}, \tau_{lk}^2, \mu_l, \gamma_{lk}, g_{lk}\})$ for $l = 1, \dots, L$ and $k = 1$; $p(M_{lk}, \tau_{lk}^2, \gamma_{lk}, g_{lk} \mid A^* - \{M_{lk}, \tau_{lk}^2, \gamma_{lk}, g_{lk}\})$ for $l = 1, \dots, L$ and $k = 2, \dots, K$;
- (iii) $p(\zeta_j \mid A^* - \{\zeta_j\})$ for $j = 1, \dots, J$;
- (iv) $p(D \mid A^* - \{D\})$.

Steps B(i), B(iii), and B(iv) are the same as in Sampling Scheme 1. To generate $M_{lk}, \tau_{lk}^2, \mu_l, \gamma_{lk}, g_{lk} \mid A^* - \{M_{lk}, \tau_{lk}^2, \mu_l, \gamma_{lk}, g_{lk}\}$ in Step B(ii), we generate from the four conditional distributions:

- (a) $p(M_{lk} \mid A^* - \{M_{lk}, \tau_{lk}^2, \mu_l, \gamma_{lk}, g_{lk}\})$, i.e., generate M_{lk} without conditioning on $\tau_{lk}^2, \mu_l, \gamma_{lk}$, or g_{lk} .
- (b) $p(\tau_{lk}^2 \mid A^* - \{\tau_{lk}^2, \mu_l, \gamma_{lk}, g_{lk}\})$.
- (c) $p(\mu_l, \gamma_{lk} \mid A^* - \{\mu_l, \gamma_{lk}, g_{lk}\})$.
- (d) $p(g_{lk} \mid A^* - \{g_{lk}\})$.

Step (a) requires a numerical integration while, Step (b) uses a Metropolis-Hastings algorithm. The computational details of generating these terms are given in Shively et al. (1999). Steps (c) and (d) are generated the same way as Step B(ii) of Sampling Scheme #1.

If the number of distinct values for the covariates are very large, then the software used to implement the estimation and variable selection techniques may take a long time to run. Because the CPU time is highly dependent on the type of computer used, it is difficult to specify in general when the number of distinct values becomes too large for practical purposes.

References

Albert, J., S. Chib. 1993. Bayesian analysis of binary and polychotomous response data. *J. Amer. Statist. Assoc.* **88** 669–679.

- Allenby, G. M., J. L. Ginter. 1995. Using extremes to design products in segment markets. *J. Marketing Res.* **32** 392–403.
- , P. J. Lenk. 1994. Modeling household purchase behavior with logistic normal regression. *J. Amer. Statist. Assoc.* **89** 1218–1231.
- , ———. 1995. Reassessing brand loyalty, price sensitivity, and merchandising effects on consumer brand choice. *J. Bus. Econom. Statist.* **13** 281–290.
- Blattberg, R., S. Neslin. 1990. *Sales Promotion: Concepts, Methods and Strategies*. Prentice Hall, Englewood Cliffs, NJ.
- Casella, G., E. I. George. 1992. Explaining the Gibbs sampler. *Amer. Statistician* **46** 167–174.
- Gelfand, A. E., A. F. M. Smith. 1990. Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.* **85** 398–409.
- Green, P. J., B. W. Silverman. 1994. *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*. Chapman and Hall, New York.
- Hastie, T. J., R. J. Tibshirani. 1990. *Generalized Additive Models*. Chapman and Hall, New York.
- Kalyanam, K., T. S. Shively. 1998. Estimating irregular pricing effects: a stochastic spline approach. *J. Marketing Res.* **35** 16–29.
- Kotler, Philip. 1997. *Marketing Management*. 9th edition, Prentice Hall, Englewood Cliffs, NJ.
- Lenk, P. J., W. S. DeSarbo, P. E. Green, M. R. Young. 1996. Hierarchical Bayes conjoint analysis: recovery of part-worth heterogeneity from incomplete designs in conjoint analysis. *Marketing Sci.* **15** 173–191.
- McCulloch, R., P. Rossi. 1994. An exact likelihood approach to analysis of the MNP model. *J. Econometrics* **64** 207–240.
- Shively, T. S., R. Kohn, S. Wood. 1999. Variable selection and function estimation in additive nonparametric regression using a data-based prior (with discussion). *J. Amer. Statist. Assoc.* **94** 777–806.
- Wahba, G. 1978. Improper priors, spline smoothing and the problem of guarding against model errors in regression. *J. Royal Statist. Soc. B40* 364–372.
- Wecker, W., C. F. Ansley. 1983. The signal extraction approach to nonparametric regression and spline smoothing. *J. Amer. Statist. Assoc.* **78** 81–89.
- Wood, S., R. Kohn. 1998. A Bayesian approach to robust nonparametric binary regression. *J. Amer. Statist. Assoc.* **93** 203–213.

This paper was received November 30, 1998, and was with the authors 6 months for 2 revisions; processed by Dick Wittink.