

Capturing Flexible Heterogeneous Utility Curves: A Bayesian Spline Approach

Jin Gyo Kim

College of Business Administration, Seoul National University, Gwanak-Gu, Shillim 9 Dong,
Seoul 151-742, Korea, jingkim@snu.ac.kr

Ulrich Menzefricke

Joseph L. Rotman School of Management, University of Toronto, 105 St. George Street,
Toronto, Ontario, Canada, M5S 3E6, menzefricke@rotman.utoronto.ca

Fred M. Feinberg

Stephen M. Ross School of Business, University of Michigan, 701 Tappan Street,
Ann Arbor, Michigan 48109, fein@umich.edu

Empirical evidence suggests that decision makers often weight successive additional units of a valued attribute or monetary endowment unequally, so that their utility functions are intrinsically nonlinear or irregularly shaped. Although the analyst may impose various functional specifications exogenously, this approach is ad hoc, tedious, and reliant on various metrics to decide which specification is “best.” In this paper, we develop a method that yields individual-level, flexibly shaped utility functions for use in choice models. This flexibility at the individual level is accomplished through splines of the truncated power basis type in a general additive regression framework for latent utility. Because the number and location of spline knots are unknown, we use the birth-death process of Denison et al. (1998) and Green’s (1995) reversible jump method. We further show how exogenous constraints suggested by theory, such as monotonicity of price response, can be accommodated. Our formulation is particularly suited to estimating reaction to pricing, where individual-level monotonicity is justified theoretically and empirically, but linearity is typically not. The method is illustrated in a conjoint application in which all covariates are splined simultaneously and in three panel data sets, each of which has a single price spline. Empirical results indicate that piecewise linear splines with a modest number of knots fit these data well, substantially better than heterogeneous linear and log-linear a priori specifications. In terms of price response specifically, we find that although aggregate market-level curves can be nearly linear or log-linear, individuals often deviate widely from either. Using splines, hold-out prediction improvement over the standard heterogeneous probit model ranges from 6% to 14% in the scanner applications and exceeds 20% in the conjoint study. Moreover, “optimal” profiles in conjoint and aggregate price response curves in the scanner applications can differ markedly under the standard and the spline-based models.

Key words: choice models; utility theory; heterogeneity; splines; Bayesian methods; Markov chain Monte Carlo

History: Accepted by Jagmohan S. Raju, marketing; received August 29, 2002. This paper was with the authors 32 months for 3 revisions.

1. Heterogeneous Utility and Flexible Functional Forms

Previous studies of choice models have typically assumed that utility can be linked to covariates linearly, usually via a fixed functional form exogenously imposed by the analyst. These assumptions are motivated largely by methodological convenience in terms of model structure and estimation, as well as post hoc interpretability of the underlying parameterization. There is, however, substantial theoretical and empirical justification from several disciplines in favor of utility functions of various, often irregular, shapes. In the economics literature, many studies on “flexible” functional forms (Wales 1977, Caves and Christensen

1980) in demand systems have supported utility nonlinearity at the aggregate level. Such flexibility is crucial in gauging response to price, in particular, a finding echoed in marketing (Abe 1998, Bell and Lattin 2000) and psychology (Wu and Gonzalez 1996, Gonzalez and Wu 1999).

The estimation of flexible individual-specific functions in empirical studies using choice models has largely been neglected, hampered by methodological difficulties. One difficulty—of needing to exogenously impose a particular functional specification for latent utility—raises a number of problems of its own. The first such problem is ordinarily theory-dependent: Which particular specification should be imposed? Because utility itself is unobserved, there is

seldom any overriding theoretical justification for one functional specification over another, other than pure parsimony. This may be, in part, why linear specifications have come to assume something of a default role. A second problem follows from the first: When theory fails to suggest a particular functional specification, which should be explored? A trial-and-error approach is time consuming and raises questions of overfitting; it also requires specification of functional families from which one might reasonably draw candidates. A final problem concerns how the analyst might choose among various candidate specifications; that is, which test procedures should be used? Because different empirical applications favor different parameterizations, theoretical generalizations across studies may be difficult to come by.

In this paper, we propose a nonparametric, spline-based model to investigate flexible utility functions for choice models, with a particular empirical emphasis on capturing price response. The presented method allows one to (1) model utility functions at the individual rather than at the aggregate level, (2) examine the degree of cross-sectional variation in utility function shape, and (3) better understand the true relationship among specific covariates (particularly, price), observed choice probabilities, and market-level response. Throughout, we stress that it is not strict linearity, *per se*, that our approach is meant to relax. Rather, it is the need to prespecify a particular functional form for individual-level utility.

We test the proposed formulation against those commonly applied in prior literature in a variety of settings: a conjoint application for the design of a small durable and three scanner panel data sets of varying characteristics. The conjoint application, although stemming from a design with fixed within-attribute levels, allows each of six input variables to be splined simultaneously. The scanner applications allow for the assessment of individual-level price response curves, where the spline “knots” are household-specific. In all applications, we find that piecewise-linear splines give rise to curves that appear to capture covariate effects well, particularly for price. Gains from the application of splines are substantial, not only in terms of Bayesian measures of model fit but in the managerially critical metric of predictive accuracy in hold-out samples. Moreover, response estimates systematically differ when strict assumptions of price response linearity are relaxed to monotonicity, as we do here.

The remainder of this paper is organized as follows. Section 2 reviews key literature that bears on the assessment of latent utility, with particular regard to individual-level specifications and price response

linearity. Section 3 incorporates truncated power basis type splines into a heterogeneous probit choice framework, and describes specification issues, identification, and Bayesian estimation. Section 4 presents four applications and the comparative results of a variety of utility specifications on the assessment of price response. Finally, §5 discusses both theoretical and managerial implications of our suggested approach, as well as avenues for future research.

2. Modeling Flexible Utility and Pricing Effects

Choice models typically assume that multiattribute systems link observed choices to covariates (e.g., prices or environmental variables). However, the extant literature has been largely silent on estimating consumer-specific, flexible functions for latent utility. Gonzalez and Wu (1999) found substantial evidence for complex utility shapes at both the aggregate and individual levels; moreover, they detailed the substantial methodological problems in measuring them at the individual level, even in a controlled setting and using parametric representations. To our knowledge, obtaining such measurements using field data not subject to experimental controls is an open question.

Marketing studies provide consistent evidence that response to environmental variables, particularly price, can be nonlinear, with substantial individual-level variation. For example, Kalyanaram and Little (1989) demonstrated the existence of a range of prices in which consumers are very nearly insensitive to price changes, and Gupta and Cooper (1992) reported a similar effect for price discounts. Accounting for complex pricing effects has been among areas to which splines have been successfully applied. Kalyanam and Shively (1998) applied a stochastic spline methodology to weekly unit sales for multiple brands and categories, finding wide shape variation, although their model was not designed to account for household-level effects.

A number of studies have addressed nonlinear utility formulations nonparametrically. Abe (1998) proposed a nonparametric additive model to estimate a binary logit model utility function, albeit only at the aggregate level, finding it to be nonlinear. Briesch et al. (2002) used a nonparametric representation to study whether consumers react differently to price reductions versus larger deal discounts, discovering deviations from linearity in deal effects across four categories. Shively et al. (2000) proposed a nonparametric model for the relationship between consumer preference and a set of explanatory covariates and found that several such relationships (four in 24) in their study were decidedly nonlinear. In conjoint

modeling, the use of piecewise linear, heterogeneous utility functions of multiple attributes is quite common (Lenk et al. 1996, Andrews et al. 2002).

In this paper, we conflate the objectives of these various studies, allowing for complex functional shapes at the individual level, where utilities themselves are latent. This goal is analogous to that of Bell and Lattin (2000), who found that unless response heterogeneity is appropriately accounted for, measurements of important aspects of consumer behavior—in their case, reliance on reference prices—could go dramatically awry. Our empirical applications will examine whether presumptions about the shape of individual-level price response may be altering inferences about market-level strategy. In this way, we seek to free managerial decisions, specifically regarding optimal pricing strategy, from artifacts of specific functional assumptions about how individuals translate product attributes and prices into relative preference. In the next section, we present a spline-based model formulated with this specific goal in mind.

3. Model Specification

3.1. Multinomial Probit Model

Let $y_{ht} = j$ denote the event that individual h ($h = 1, \dots, H$) chooses alternative j ($j = 1, \dots, J$) on choice occasion t ($t = 1, \dots, T_h$). Let \mathbf{x}_{hjt} denote individual h 's k -dimensional vector of discrete explanatory variables (e.g., feature, display) for alternative j on choice occasion t . Throughout, we use “choice occasion” as a generic label for purchase occasion (in scanner panel data) or choice task (in conjoint data). Suppose that there are M continuous explanatory variables $\mathbf{v}_{hjt} = (v_{hjt,1}, \dots, v_{hjt,M})'$. Then, individual h 's utility for alternative j on choice occasion t is assumed to follow an additive regression specification (Hastie and Tibshirani 1990),

$$u_{hjt} = \mathbf{x}'_{hjt}\boldsymbol{\beta}_h + \sum_{m=1}^M f_m^h(\mathbf{v}_{hjtm}) + \varepsilon_{hjt}, \quad (1)$$

where $f_m^h(\bullet)$ is an unknown, possibly nonlinear, function of v_{hjtm} and ε_{hjt} is an error term. A more detailed discussion of our specification for f_m^h is given in §3.2.

To complete the model specification, let $\mathbf{u}_{ht} = (u_{h1t}, \dots, u_{hJt})'$ denote a J -dimensional vector of latent utilities, let $\mathbf{x}_{ht} = (x_{h1t}, \dots, x_{hJt})'$ denote a $J \times k$ covariate matrix, define $\boldsymbol{\pi}_{hjt} = \sum_{m=1}^M f_m^h(v_{hjtm})$, $\boldsymbol{\pi}_{ht} = (\pi_{h1t}, \dots, \pi_{hJt})'$, and let $\boldsymbol{\varepsilon}_{ht} = (\varepsilon_{h1t}, \dots, \varepsilon_{hJt})'$ be a J -dimensional normal random vector with mean vector 0 and covariance matrix $\boldsymbol{\Sigma}_u$. Summarizing, our model is

$$y_{ht} = j \quad \text{if } u_{hjt} = \max(u_{h1t}, \dots, u_{hJt}), \quad \text{such that} \quad (2)$$

$$\mathbf{u}_{ht} = \mathbf{x}_{ht}\boldsymbol{\beta}_h + \boldsymbol{\pi}_{ht} + \boldsymbol{\varepsilon}_{ht}, \quad \boldsymbol{\varepsilon}_{ht} \sim N_J(\mathbf{0}, \boldsymbol{\Sigma}_u),$$

where $N_J(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes a J -dimensional normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. The resulting choice probabilities for the multinomial probit model are given by the J -fold integral,

$$p_{hjt} = p(y_{ht} = j \mid \boldsymbol{\beta}_h, \boldsymbol{\pi}_{ht}, \boldsymbol{\Sigma}_u)$$

$$= \int_{A_{h1t}} \cdots \int_{A_{hJt}} n_J(\mathbf{u}_{ht} \mid \mathbf{x}_{ht}\boldsymbol{\beta}_h + \boldsymbol{\pi}_{ht}, \boldsymbol{\Sigma}_u) d\mathbf{u}_{ht},$$

$$j = 1, \dots, J, \quad (3)$$

where $n_J(\mathbf{u} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes a J -variate normal density for the vector \mathbf{u} with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Furthermore, A_{hjt} is the interval $(-\infty, u_{hjt})$ if $i \neq j$ and $(-\infty, \infty)$ if $i = j$. This specification suffers from two well-known identification problems: location invariance and scale invariance. Our approach to resolving these identification problems is slightly different for the conjoint and scanner panel data applications that we describe in §4.

In the conjoint data application, we let the last option, J , be a “no-choice” option. We set the utility of the no-choice option to $u_{hJt} \equiv 0$; therefore, we must introduce an intercept for each individual, β_{h0} , that measures the individual's intrinsic utility preference of choice versus no choice. Thus, we have

$$u_{hjt} = \beta_{h0} + \mathbf{x}'_{hjt}\boldsymbol{\beta}_{h1} + \pi_{hjt} + \varepsilon_{hjt},$$

$$j = 1, \dots, J - 1, \quad \text{and} \quad (4)$$

$$u_{hJt} \equiv 0,$$

where \mathbf{x}'_{hjt} and π_{hjt} are described after (1) and before (2), and where the covariance matrix for the $(J - 1)$ -dimensional error term, $\boldsymbol{\varepsilon}_{ht} = (\varepsilon_{h1t}, \dots, \varepsilon_{h,J-1,t})'$, is a $(J - 1)$ -dimensional identity matrix, denoted \mathbf{I}_{J-1} .

In the scanner data applications, we suppose that there are alternative-specific intercepts for the J alternatives and also a number of additional binary variables; we will include two such in our applications: feature advertising and display. The portion of the resulting $J \times (J + 2)$ matrix \mathbf{x}_{ht} that corresponds to the alternative-specific intercepts is a $(J \times J)$ identity matrix. To deal with the location identification problem, we arbitrarily pick one of the alternatives (say, J) and drop the J th column of this identity matrix; the corresponding alternative-specific intercept is therefore unnecessary, and thus the other $J - 1$ alternative-specific intercepts are measured relative to alternative J . We note that this will affect interpretations of alternative-specific intercepts across models, as both the intercept in question and the “base” used for identification must jointly be taken into account. There are several methods available to handle the scale invariance problem in the scanner data applications. Chib and Greenberg (1998) and McCulloch et al. (2000) provide thorough discussions.

3.2. Incorporating Splines into the Model

In this section, we first describe our use of splines to approximate the function f_m^h in (1) and then some estimation issues related to splines. To incorporate flexibly shaped utility functions, we use a nonparametric approach without strong functional assumptions to model f_m^h . It is well known that a continuous function may be arbitrarily well approximated by a piecewise polynomial function with a sufficiently large number of knots (cf. Wegman and Wright 1983). Specifically, we let each $f_m^h(\bullet)$ be a spline function with a truncated power basis (Denison et al. 1998, Lindstrom 2002), also known as a “one-sided basis” (Schumaker 1981, p. 112),

$$f_m^h(v_{hjt}) = \sum_{n=1}^{l_m} \gamma_{hmn}(v_{hjt} - s_{hm0})_+^n + \sum_{i=1}^{q_{hm}} \gamma_{hm,1+i}(v_{hjt} - s_{hmi})_+^{l_m} \quad (5)$$

for $v_{hjt} \in [s_{hm,0}, s_{hm,q_{hm}+1}]$, where $w_+ = \max(0, w)$, $w_+^0 = I(w \geq 0)$, q_{hm} is the number of interior knots for individual h for the m th spline function, l_m is the order of the spline, $\{\gamma_{hmi}\}$ are individual-specific spline coefficients, and $\{s_{hmi}\}$, arranged in ascending order, are individual-specific interior knot points with boundary knots, $s_{hm,0}$ and $s_{hm,q_{hm}+1}$, for the m th spline. Here, $I(a)$ is the usual indicator function for the event a .

In our applications, we found linear splines to perform better than splines of higher order, and we restrict our discussion here to linear splines.¹ Thus, we use

$$f_m^h(v_{hjt}) = \gamma_{hm1}(v_{hjt} - s_{hm0})_+ + \sum_{i=1}^{q_{hm}} \gamma_{hm,1+i}(v_{hjt} - s_{hmi})_+ \quad (6)$$

Note that, for identification purposes, we do not include an intercept term in (5) or (6), accommodating one as needed through β_h in (1).

Let us now turn to some estimation issues related to the spline functions. Because the u_{hjt} are unknown latent variables, it is very difficult to choose a reasonable knot configuration for each individual in advance. A key feature of our model is that it endogenously settles on an appropriate knot configuration, leading to the difficult problem of spline estimation with varying knots (see Wegman and Wright 1983

for a review). The linear spline function model in (6) requires the estimation of two sets of parameters: (1) the knot configuration, that is, the number and location of knots, q_{hm} and $s_{hm,1}, \dots, s_{hm,q_{hm}}$; and (2) the spline coefficients, $\gamma_{hm,1}, \dots, \gamma_{hm,1+q_{hm}}$.

We use Markov chain Monte Carlo (MCMC) simulation to estimate the spline functions. Our notation for the parameters that describe the household-specific knot configurations is as follows:

- $q_{hm} \in \{0, 1, \dots, Q_m\}$ denotes the possible number of interior knots for $f_m^h(\bullet)$ given Q_m candidate knots,
- $\mathcal{D}_{hm} = \{D_{hm1}, \dots, D_{hmQ_m}\}$ denotes a set of Q_m candidate interior knots, and
- $\theta_{hm} = \{s_{hm,1}, \dots, s_{hm,q_{hm}}\}$ denotes a set of q_{hm} interior knots chosen from \mathcal{D}_{hm} .

The appendix describes the hierarchical structure for the regression coefficients β_h in (1) and $\beta_h = (\beta_{h0}, \beta'_{h1})'$ in (4), and the prior distributions for all parameters.

3.2.1. Monotonicity and Knot Configuration. Researchers often seek to place constraints on individual-specific utility shapes. These can come about for a variety of reasons, including those imposed by theory, rationality, or previous findings. For example, economic theory suggests that the spline reflecting the effect of price be monotonically decreasing. In our applications, we will thus require

$$f_m^h(v_{hjt}) \geq f_m^h(v_{hjt}^*) \quad \text{if } v_{hjt} \leq v_{hjt}^* \quad (7)$$

Note that (7) leads to several separate inequality constraints that must be jointly satisfied. The multivariate normal random-effects probit, by contrast, allows for the possibility that some individual-level “draws” are nonmonotonic or even nondecreasing.

The knot configuration for each individual at each MCMC iteration is sampled from a collection of pre-specified candidate knot points; we take up how to arrive at this candidate set in our applications. We use a special discrete version of Green’s (1995) reversible-jump Metropolis-Hastings algorithm in our MCMC simulation. Green’s method allows the number and location of knots to vary across iterations by allowing “births” and “deaths” among them, a critical feature to which we will return in our applications.

4. Empirical Applications

The proposed model is illustrated for two widely popular applications of choice models in marketing: choice-based conjoint and scanner panel data. The conjoint application is chosen to demonstrate how to estimate multiple spline functions simultaneously. The three scanner data applications are pursued to show how to solve the varying knot problem for individual-level utility splines where there is a relatively large number of candidate knots.

¹ Details of the MCMC sampler for linear and higher order splines, various prior settings, and all full conditionals are included in the online appendix to this paper (provided in the e-companion), and in Kim et al. (2007). The electronic companion to this paper is available as part of the online version that can be found at <http://mansci.journal.informs.org/>.

The conjoint data set contains six attributes, each with five levels, thus limiting the number of possible knots and predetermining their potential locations. We estimate the spline-based model on three separate scanner data sets, comparing fit and hold-out performance to that of the standard multinomial probit (MNP) model for both linear and log-linear price specifications. We note that the log-price form is directly suggested by economic theory, given optimal consumer budget allocation (Allenby and Rossi 1991). For both types of application, we focus mainly on parameter estimates relevant to, and implications of, splines. Because our empirical results supported linear splines in all four applications, we present only those results; all others are available in the online appendix.

4.1. Application I: Choice-Based Conjoint

We model conjoint choice data for bathroom scales collected by the Optimal Design Engineering Laboratory at the University of Michigan (see Michalek et al. 2005 for further details regarding study design and materials). The design consisted of six attributes—weight capacity, platform aspect ratio, platform area, interval gap between one-pound markings, size of printed numbers, and price—each with five levels. Specific levels for each of these attributes appear in Table 1. Each of 184 subjects completed an online series of 50 fixed tasks, each including three product profiles and a no-choice option; thus, $H = 184$ and $J = 4$. For the purpose of out-of-sample validation, we divided the full data set into training, \mathbf{y} , and prediction data sets, $\tilde{\mathbf{y}}$. The prediction data consisted of the last 10 choice observations for each subject, so that the number of observations in the training data set is $T_h = 40 \forall h$.

Because none of the six attributes is purely categorical, we introduced a separate spline for each, so $M = 6$ in (1). To aid in graphical comparability, we divided by the lowest value for each attribute, leaving the transformed largest values distinct; see the last column in Table 1. Splines were then introduced over these rescaled levels.

4.1.1. Model Comparison. We estimated several different models. First, we estimated a baseline model, M_0 , the random effects probit commonly applied in

choice-based conjoint:

$$u_{hjt} = \beta_{h0} + \mathbf{x}'_{hjt} \boldsymbol{\beta}_{h1} + \varepsilon_{hjt}, \quad j = 1, \dots, J - 1, \quad \text{and} \quad (8)$$

$$u_{hjt} \equiv 0,$$

where $\boldsymbol{\varepsilon}_{ht} = (\varepsilon_{h1t}, \dots, \varepsilon_{h,J-1,t})' \sim N_{J-1}(\mathbf{0}, \mathbf{I})$. Note that M_0 is actually less restrictive than a spline model, as it allows any sort of relationship between the part-worths for a particular attribute. We will examine empirically whether this flexibility helps or hampers hold-out performance.

Because $u_{hjt} \equiv 0$ and β_{h0} is included in the specification for u_{hjt} , we define \mathbf{x}_{hjt} so that the “effect” of the first level of each attribute is zero. Therefore, \mathbf{x}_{hjt} was a 24-dimensional vector, consisting of four dummy variables for the second to the fifth levels for each attribute. Thus, β_{h0} measures the utility of the choice option at the lowest values for all attributes over the no-choice option. The parameters of the baseline model are β_{h0} and $\boldsymbol{\beta}_{h1}$. The relevant hierarchical structure and prior distributions are as in item (1) in the appendix; the chosen values of the priors for M_0 were $g_0 = 0$, $c_0 = 20$, $a_0 = b_0 = 0.5$, $\mathbf{g}_1 = \mathbf{0}_{24}$, $\mathbf{C}_1 = \mathbf{W}_1 = 20\mathbf{I}_{24}$, and $r_1 = 2$. We also estimated the linear spline model S described after (6):

$$u_{hjt} = \beta_{h0} + \sum_{m=1}^6 f_m^h(v_{hjt m}) + \varepsilon_{hjt}, \quad j = 1, \dots, J - 1, \quad \text{and} \quad (9)$$

$$u_{hjt} \equiv 0,$$

where $\boldsymbol{\varepsilon}_{ht} = (\varepsilon_{h1t}, \dots, \varepsilon_{h,J-1,t})' \sim N_{J-1}(\mathbf{0}, \mathbf{I})$, as described previously.

The set \mathcal{D}_{hm} of candidate knots for the m th spline ($m = 1, \dots, 6$) consisted of the three interior levels of the corresponding attribute. The exterior levels (here, the first and the fifth) serve as boundary knots. Because the profile stimuli sets were identical across subjects, so are these candidate knots. Thus, the number of interior knots $q_{hm} \in \{0, 1, 2, 3\} \forall h, m$. In addition, the monotonicity constraint was imposed on the price spline, $f_6^h(\bullet)$, although the other splines were unconstrained.

Table 1 Attributes and Their Levels in the Conjoint Study

m	Attributes	Unit	Levels	Rescaled levels
1	Weight capacity	lbs.	200; 250; 300; 350; 400	1; 1.25; 1.50; 1.75; 2.00
2	Platform aspect ratio	length/width	6/8; 7/8; 1; 8/7; 8/6	1; 1.17; 1.33; 1.52; 1.78
3	Platform area	in. ²	100; 110; 120; 130; 140	1; 1.10; 1.20; 1.30; 1.40
4	Interval mark gap	in.	0.063; 0.094; 0.125; 0.156; 0.188	1; 1.49; 1.98; 2.45; 2.98
5	Size of number	in.	0.75; 1.00; 1.25; 1.50; 1.75	1; 1.33; 1.66; 2.00; 2.33
6	Price	\$	10; 15; 20; 25; 30	1; 1.50; 2.00; 2.50; 3.00

Table 2 Model Performance in the Conjoint Study

Model	Training data		Prediction data	
	Log of integrated likelihood	log of BF_{M_0}	Log of posterior predictive distribution	Correctly predicted: hit rate (%)
M_0 (Heterogeneous traditional probit)	-5,828.70	0	-2,361.80	48.08
S (Heterogeneous linear spline probit)	-4,451.67	1,377.03	-1,895.57	59.42

The parameters for our spline model in the conjoint application are β_{h0} and the parameters characterizing the splines. The appendix describes the hierarchical structure for these parameters and the prior distributions. We chose the prior parameter values $g_0 = 0$, $c_0 = 20$, and $a_0 = b_0 = 0.5$; these are identical to those for M_0 . Furthermore, the prior distributions for the spline coefficients $\gamma_{hm,i}^{(q_{hm}, \theta_{hm})}$ had parameters $a_m = 0$ and $b_m = 10$ ($m = 1, \dots, 6$). Finally, the distribution for the number of internal knots q_{hm} was Poisson (with parameter $\lambda_m = 3$) truncated at 3, the maximum number of internal knots. Empirically, we observed little substantive dependence on λ_m for nearby values or relative to a uniform prior.

After 10,000 iterations, all models seemed to converge; the proportion of sampled quantities that passed the Geweke convergence statistics ranged from 77.6% to 100%. Parameter inferences were based on an additional 30,000 iterations. For all spline models, the subject-specific knot configurations were well mixed, as indicated by examining individual-level trace plots.² To choose between two models, B_1 and B_2 , we calculate the Bayes factor, BF_{B_1, B_2} , the ratio of the two respective integrated likelihoods. For such computations, we applied Genz' (1992, 1993) adaptive Monte Carlo and Lattice Rule methods. Fortran codes for these methods are available from Genz.

To assess the relative performance of the spline models for the prediction data set, we computed two quantities:

1. The predictive density at the observed values \tilde{y} of the prediction data set, $p(\tilde{y} | \mathbf{y}) = \int p(\tilde{y} | \Phi)p(\Phi | \mathbf{y})d\Phi$, where Φ denotes all unknown parameters of the given model, $p(\tilde{y} | \Phi)$ denotes the density of \tilde{y} given the parameters Φ , and $p(\Phi | \mathbf{y})$ denotes the posterior distribution for Φ given the training data set \mathbf{y} .
2. The proportion of correctly predicted purchases for the given model, this so-called "hit rate" was obtained by first sampling latent utilities for each purchase observation in \tilde{y} , given Φ , at each MCMC iteration, computing the proportion of all purchases for which the chosen alternative was the one with

maximal utility and then averaging these proportions across iterations.

Table 2 presents the integrated likelihoods and predictive inferences. By either measure, the linear spline model S was decisively preferred to the traditional heterogeneous probit model. This result may seem surprising, given the great improvement—23.6%—in hold-out hit rate for the spline-based models. However, we believe that it may be explained as follows. Note that the prediction data set relied on part-worth values that were deliberately included in the training set. However, this was relatively less "informative" for some respondents than others because they tended to select the no-choice option or simply avoided certain attribute levels among their choices. In such cases, the spline model could leverage information on other *attribute levels* for that particular respondent, whereas the traditional heterogeneous probit (M_0) could do so only for other *respondents* (through the across-subjects attribute correlation). In other words, the spline model provides an explicit mechanism for a "smoothing" of part-worths so that the values of poorly estimated attribute levels can be improved via interpolation across MCMC runs; M_0 , although less restricted, lacks such a mechanism.

An extreme case—avoided in our study by design—is one in which the prediction stimuli include levels never tested in the training data. In such a case, M_0 might have to resort to ad hoc (linear) interpolation using adjacent values only, whereas spline-based models can interpolate based on all levels for that attribute. We speculate that a spline-based model might perform well as an update mechanism for adaptive conjoint studies, although we have not explored that possibility here.

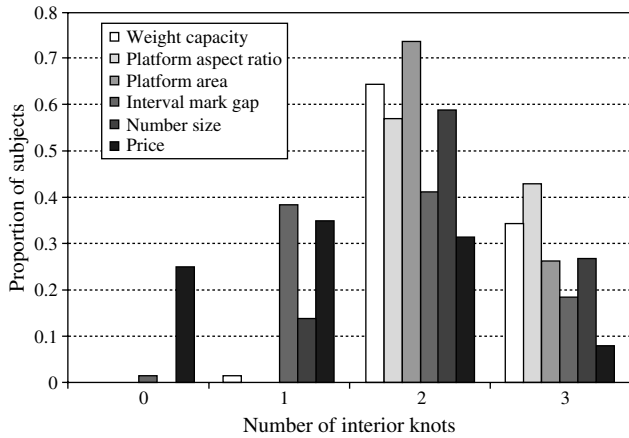
Because the linear spline model performed best, we examine it in greater detail in the following section.

4.1.2. Estimation Results for the Linear Spline Model, S . It is instructive to examine estimates resulting from S , relative to those from M_0 , and their relative implications. For brevity, we consider variation in knot configuration, function shape, and optimal design using various summary measures and tests.

Interior Knot Configurations. The knot configuration was quite varied across subjects, and we explore this issue in greater depth in our scanner applications. Let us temporarily set aside the posterior uncertainty

² As an informal check on convergence, we recomputed all reported quantities using the first and last third of the iterations used for inference; this was done for all four included applications. Differences were substantively minor and never significant.

Figure 1 Conjoint Data—Histogram of Modal Number of Interior Knots Across Subjects



about the number of knots and focus instead on the modal number of interior knots for each subject and attribute. Figure 1 presents a histogram of these subject-specific modal numbers for each of the six attributes.

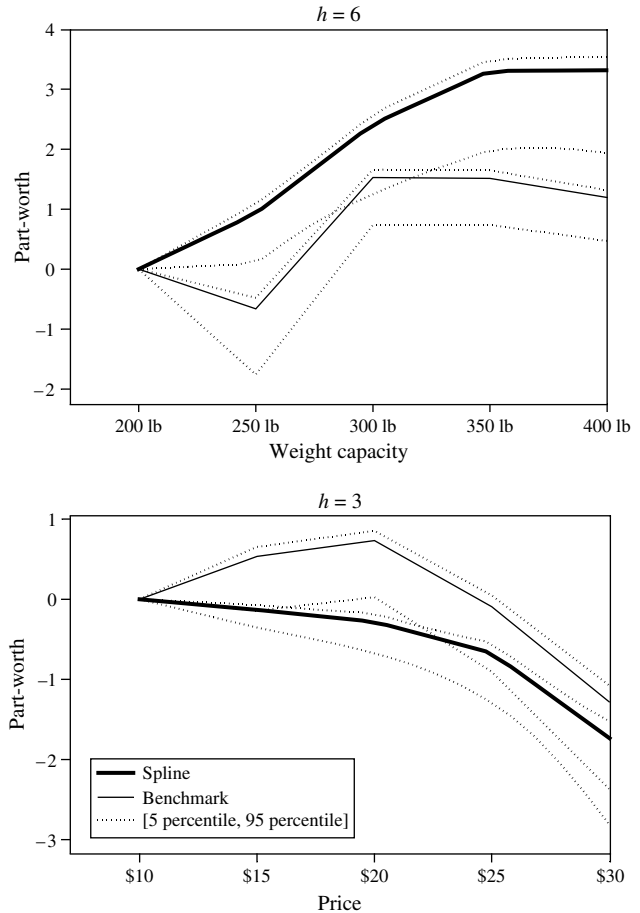
First, consider attributes $m = 4$ and $m = 6$, interval mark gap, and price. The modal number was 0 for 1.6% and 25% of subjects, respectively; for the linear spline model S , this lack of knots implies linear (increasing or decreasing) part-worth curves for relatively few subjects. Conversely, the modal number was greater than zero for a sizable proportion of subjects, 98.4% and 75.0%; these subjects thus appear to have nonlinear part-worth curves for these two attributes. For the other four attributes, there was not even a single subject whose modal number of interior knots was zero, suggesting that all subjects exhibited some degree of nonlinearity for these four attributes. Intriguingly, of all six attributes, price had by far the largest proportion of modal knot zeros, as per Figure 1; for these individuals, a linear specification would apparently be most appropriate for price among the six attributes.

Summarizing, Model S suggests considerable variation in the number of interior knots across subjects, and it provides very strong evidence in favor of nonlinearity in part-worth curves for all attributes (with the possible exception of price for about one quarter of respondents). Of course, M_0 also does not presume linearity, so the superior performance of S cannot be attributed directly to allowing for nonlinear response. But M_0 imposes a sort of foundational nonlinearity: Each attribute level is estimated separately, so near-linear relations across them occur only by happenstance, not as a built-in feature of the model of which the estimation can avail. In practice, the analyst using M_0 is likely to interpolate between adjacent levels when necessary, but not between nonadjacent levels because a distinct utility value is indicated at each

level. Under the linear spline model S , however, as long as the number of internal knots is not at its maximum value (here, three), the utility values at some of these fixed levels would, in fact, be the linear interpolant of its adjacent values. The estimated part-worth curves may thus be smoother under S than under M_0 . Let us now examine this possibility and the shapes of individual part-worth curves under various models.

Disaggregate Spline Part-Worth Curves. It is straightforward to use the MCMC output to estimate individual-specific splines. Because knot number and location can change from one iteration to another, the resulting curve of posterior means is not necessarily a linear spline curve; it is likely smoother. Using the linear spline model, Figure 2 presents posterior mean part-worth curves for two randomly selected subjects for weight capacity and price. The solid and the thin lines in Figure 2 represent the posterior mean part-worth curves under S and M_0 , respectively; dotted lines enveloping the part-worth curves indicate posterior 5th and 95th percentile part-worth curves. At the lowest (base) level of each attribute, the posterior variance of the part-worth curve is zero because,

Figure 2 Conjoint Data—Individual-Level Spline (S) and Benchmark (M_0) Part-Worth Curves



for identification purposes, the curve measures differences relative to this base.

As conjectured, each of the posterior mean part-worth curves under S appears smoother than the analogous curves under M_0 , which often tend to zigzag across attribute levels, as evidenced by $h = 6$ for weight capacity. Except for the price spline under S , no constraints were imposed, so this difference in smoothness (and, on occasion, monotonicity) is a genuine substantive difference between the two models' individual-level part-worth predictions. We report informally that a similar pattern is evident throughout the subject pool.

Whereas it is not meaningful to compare the vertical distances between the two posterior mean part-worth curves in Figure 2 at different levels of an attribute, we can compare the shapes of the two posterior mean part-worth curves: It is smoother under the spline model than model M_0 . One must bear in mind that the posterior variation around such curves can be large, even with many observations per subject. As such, the analyst should make individual-level inferences with caution, perhaps eschewing them entirely in favor of aggregate inferences, which we take up next.

Market-Level Optimal Product Design. For the same two attributes just considered, weight capacity ($m = 1$) and price ($m = 6$), Figure 3 depicts the aggregate posterior mean part-worth curves under S and M_0 . As before, curves are noticeably smoother under S than M_0 . Although we must stop short of deeming this a desirable or correct feature, such smoothness, particularly in aggregate, does square with the intuitive notion of latent utility.

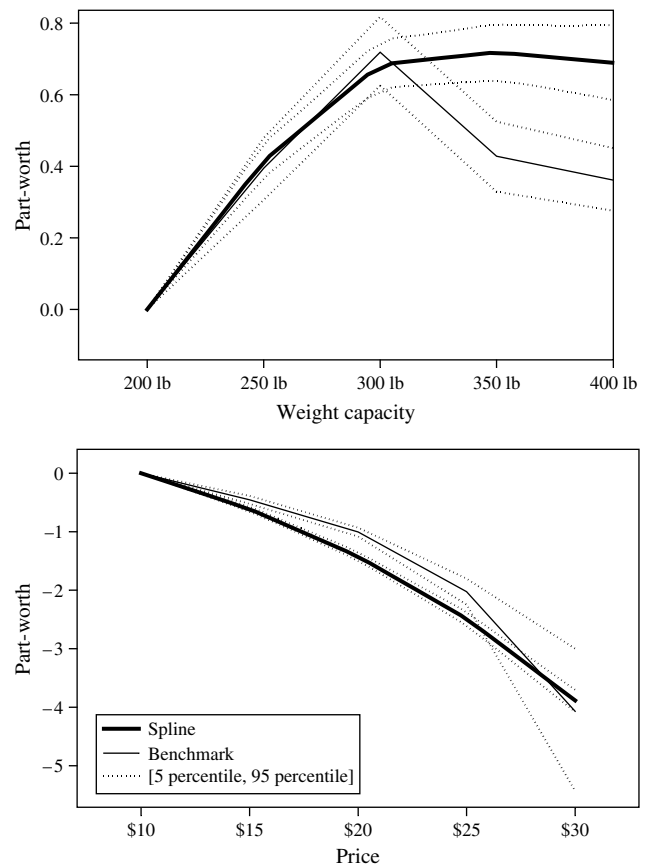
For each of the six attributes, the levels with highest aggregate posterior mean part-worths were {350 lb., 0.875, 130 sq. in., 0.156 in., 1.75 in., \$10} for S and {300 lb., 1, 100 sq. in., 0.156 in., 1.5 in., \$10} for M —that is, the most promising single-product profile differed under the two models. Given the superior hold-out accuracy of S , something so basic as the optimal product design may be mispredicted by the standard model, M_0 .

4.2. Application II: Scanner Panel Data

Estimating price response is often among the chief tasks in scanner studies. Previous investigators studying price and promotional response (e.g., Hardie et al. 1993, Bell and Lattin 2000) took their cue from the difficulty of disentangling reference effects, preference heterogeneity, and the specification of the utility function. We continue in this vein by attempting to free measures of price response from artifacts of overspecification of the functional form for utility.

For three Information Resources, Inc. (IRI) scanner product categories (kitchen paper towels, toilet tissue,

Figure 3 Conjoint Data—Aggregate Spline (S) and Benchmark (M_0) Part-Worth Curves



and ice cream), we compare our spline-based model with a base model exogenously specifying the functional form of price response. For each product category, we selected the top J SKUs in terms of total dollar sales, so that at least 95% of the market was covered. Each data set was divided into two subsets: training data, y , and prediction data, \tilde{y} . For each household, we used the first 80% of the household's n_h purchase observations for model estimation, that is, $T_h = \lceil n_h * 0.8 \rceil$; the remaining $(n_h - T_h)$ purchase observations (i.e., 20%) were reserved for the prediction task. To aid in household-level inference, households were selected if T_h was at least $(J - 1) + 3$, the dimension of β_h when there are $J - 1$ product dummies, a feature ad dummy, a display dummy, and price; we elaborate on these variables next.

Table 3 offers a brief summary of our three data sets. For the paper towel and ice cream data sets, sizes were common across products. Only the toilet tissue data set included two different sizes—two and nine products for one- and four-roll packages, respectively. Despite these different package sizes, we used posted price rather than unit price per roll so that the price spline was expressed in constant dollar units, not dollars per unit. We thus assess the incremental disu-

Table 3 Summary of IRI Scanner Panel Data Sets

Category	Product size	J	H	Observed price range	Total number of purchase observations	
					Training data	Prediction data
Paper towels	1 roll	6	133	\$0.33, \$1.32	2,131	479
Toilet tissue	1 and 4 rolls	11	198	\$0.38, \$2.45	5,123	1,197
Ice cream	64 oz	7	81	\$0.49, \$4.49	1,159	252

tility of each additional monetary unit, size differences being at least partially mitigated via product dummies.³

4.2.1. Model Comparison. We compared several models for the three scanner data sets, with price being the focal and only continuous variable; thus, $M = 1$. Furthermore, we imposed monotonicity on all spline-based price response curves. The predictor variable vector \mathbf{x}_{hjt} needed for the spline model in (1) was a vector of $J - 1$ product dummies, a feature ad dummy, and a display dummy. To ensure identifiability, we did not introduce a dummy variable for the last product, J . The dimension of the design matrix \mathbf{x}_{ht} , required in (2), was $(J \times k)$, with $k = J + 1$. For simplicity, as in the conjoint application, each data set had a common spline order for all households, that is, $l_{h1} = l$. We again estimated the proposed spline models, S_l , for increasing l until there was no further improvement in fit.

As benchmarks, we estimated two models popular in the literature: the linear, M_0 , and log-linear, M_{\log} , price response models. As in our conjoint application, M_0 was the standard probit model, where \mathbf{x}_{hjt} consisted of $J - 1$ product dummies, one feature ad dummy, one display dummy, and price. For both M_0 and M_{\log} , the dimension of the regression coefficient vector was $k = J + 2$.

As noted before, using the approach of McCulloch et al. (2000) and an unrestricted covariance matrix Σ_u , the MCMC simulation did not always converge for all our models (i.e., M_0 , M_{\log} , and S), particularly for the product dummy elements in Σ_β . Because most households in the data sets purchased a sharply restricted subset of the available products, Σ_β and Σ_u for infrequently purchased products may have been weakly identified. As there were no convergence problems for Σ_β when Σ_u was specified to be a correlation

matrix, we use the correlation form for Σ_u in all model comparisons.⁴

The relevant hierarchical structure and the prior distributions are given in the appendix. The prior distributions for μ_β , Σ_β , and Σ_u had the following values:

$$\mathbf{g}_\beta = \mathbf{0}_k, \quad \mathbf{C}_\beta = 20\mathbf{I}_k, \quad r_\beta = 2, \\ \mathbf{W}_\beta = 20\mathbf{I}_k, \quad m = 5, \quad \text{and} \quad \mathbf{C} = 4\mathbf{I}_J.$$

Note that the values for m and \mathbf{C} led to a prior distribution for Σ_u centered at \mathbf{I}_J , with reasonably diffuse priors for the off-diagonal entries of Σ_u ; specifically, the [mean $\pm 2 \times$ std. dev.] interval for the off-diagonal elements under the prior was $[-1.26, 1.26]$, covering the range $(-1, 1)$ of possible correlation values.

To determine the prior distribution for a household's knot locations, we first obtained a household-specific set of candidate knots for the price splines, \mathcal{D}_{h1} , based on the household's observed prices in the training data: After listing the distinct prices for all alternatives and all purchase occasions, we selected Q_{h1} candidate knots by subdividing the distinct prices into $Q_{h1} + 1$ equal parts. By doing so, all intervals bounded by two adjacent interior knots had at least one price observation; this was important, as it is impossible to make inferences regarding intervals devoid of observations. Discretizing the set of candidate knot locations resulted in a discrete proposal distribution for knot locations, simplifying the application of the birth-death steps in our reversible-jump algorithm.

For simplicity, we set the number of candidate knots Q_{h1} to be the same for all households in a given data set: 9, 11, and 10 for the paper towel, toilet tissue, and ice cream data sets, respectively. As a practical complication, price ranges differed in the training and prediction data sets, even for a given household. To ensure that predictions could be made for the prediction data set under the monotonicity constraint across the relevant price range, we let $s_{hm,0}$ and $s_{hm,q_{h+1}}$ be the minimal and maximal observed prices in the whole data set, as opposed to the training

³ There were empirical justifications for this choice as well. First, prices across the two different sizes did overlap: Price ranges were \$0.38–\$0.63 and \$0.48–\$2.45 for one- and four-roll products, respectively; standard tests using the individual prices as inputs did not show means to differ significantly ($p > 0.1$) across the two sizes. Second, all models subsequently reported fit better when actual posted prices were used instead of unit prices: $\text{BF}_{\text{actual price, unit price}}$ ranged from 304.9 to 365.1 across all models.

⁴ We do not believe the correlation matrix to be an especially severe restriction as Σ_u allows for correlated utilities, and does not impose IIA; see, for example, Chib and Greenberg (1998).

data alone. These minimal and maximal values were used only for the prediction tasks; inferences on price splines should be made only in the range of prices observed in the training data set.

The prior distribution for a household’s interior knot number was Poisson with parameter $\lambda_1 = 3$, implying a reasonably diffuse prior. Simulations suggested that inferences were not very sensitive to choice of λ_1 , nor to using a flat prior. Given a value for the interior number of knots, q_{h1} , the prior for a household’s knot locations was a random sample from the $\binom{Q_{h1}}{q_{h1}}$ possible knot configurations from \mathcal{D}_{h1} . Finally, the prior distribution for each spline coefficient $\gamma_{h1,i}^{(q_{h1}, \theta_{h1})}$ was normal with mean $a_1 = 0$ and variance $b_1 = 10$.

For all models, the first 20,000 iterations of the MCMC simulation were a burn-in period; parametric inferences were based on an additional 20,000 iterations. The proportion of parameters among μ_β and Σ_β that passed the Geweke convergence statistics ranged from 82.9% to 91.1% for spline models; the household-specific knot configurations were well mixed.

Table 4 presents model comparison results for the training and prediction data. Two of our models could capture nonlinear price response curves: the log-linear price model M_{log} and the spline model, S . Model M_{log} captures the exogenously specified log-linear response curve suggested by economic theory (Allenby and Rossi 1991), whereas the spline model is considerably more flexible. Across the three data sets, both the linear and quadratic spline models were preferred to the linear price model, M_0 (estimation results for quadratic and higher order splines are available from the authors). The relative performance of the log-linear price model, M_{log} , was not as clear cut. Although M_{log} did indeed perform better than the linear price model in the two paper products data sets (although not ice cream), it was inferior to linear spline models for all three data sets, as assessed by both the predictive density for the prediction data

set and the hit rate. Among exogenously specified price response curves, this offers limited evidence in favor of a log-price formulation over a linear one. Yet both were handily outperformed by the linear spline model. Given that M_0 and M_{log} are ordinarily considered fairly general model formulations, that the linear spline model improves hold-out hit rate as much as 14.3% is persuasive evidence in its favor (e.g., for paper towels and M_{log} versus S , $(31.2 - 27.3)/27.3 = 14.3\%$).

In summary, the linear spline, S , was the best-performing model across all data sets in terms of Bayes factors and both prediction measures. We examine this model and its implications in greater detail in the next section.

4.2.2. Estimation Results for the Linear Spline Model, S . We first consider linear spline model results for the discrete covariates and for the error correlation matrix. The hierarchical model coefficient for feature ad (that is, the relevant element of μ_β) showed the expected sign in all three data sets, as did that for display (with the exception of the ice cream data set). Display activities were rarely observed for ice cream products, which may explain this finding. The estimates of Σ_u showed strong error correlations: Posterior means for off-diagonal elements in Σ_u ranged from -0.783 to 0.912 , -0.632 to 0.572 , and -0.912 to 0.914 for the paper towel, toilet tissue, and ice cream data sets, respectively. Clearly, these data would have been poorly served by a standard logit or an uncorrelated probit specification.

To gauge the prevalence of nonlinearities, we obtained the modal numbers of interior knots for all households in each data set. The most common household-specific modal numbers were 2, 0, and 0 for the paper towel, toilet tissue, and ice cream data sets, respectively. The proportions of households for which these modal numbers were greater than zero were 90.2%, 44.4%, and 42.0%, suggesting that a large proportion of households appears to exhibit some

Table 4 Scanner Data—Model Comparisons

Category	Model	Training data		Prediction data	
		Log of integrated likelihood	Log of BF_{M_0}	Log of posterior predictive distribution	Correctly predicted: hit rate (%)
Paper towels	M_0	-3,260.2	0	-891.7	27.3
	M_{log}	-3,173.6	86.6	-791.4	27.8
	S	-2,922.1	338.1	-718.0	31.2
Toilet tissue	M_0	-8,650.7	0	-2,100.2	25.8
	M_{log}	-8,499.1	151.6	-2,051.2	27.1
	S	-8,315.6	335.2	-2,032.1	27.6
Ice cream	M_0	-1,713.0	0	-366.2	30.5
	M_{log}	-1,751.3	-38.3	-402.3	28.0
	S	-1,567.4	145.7	-360.3	32.3

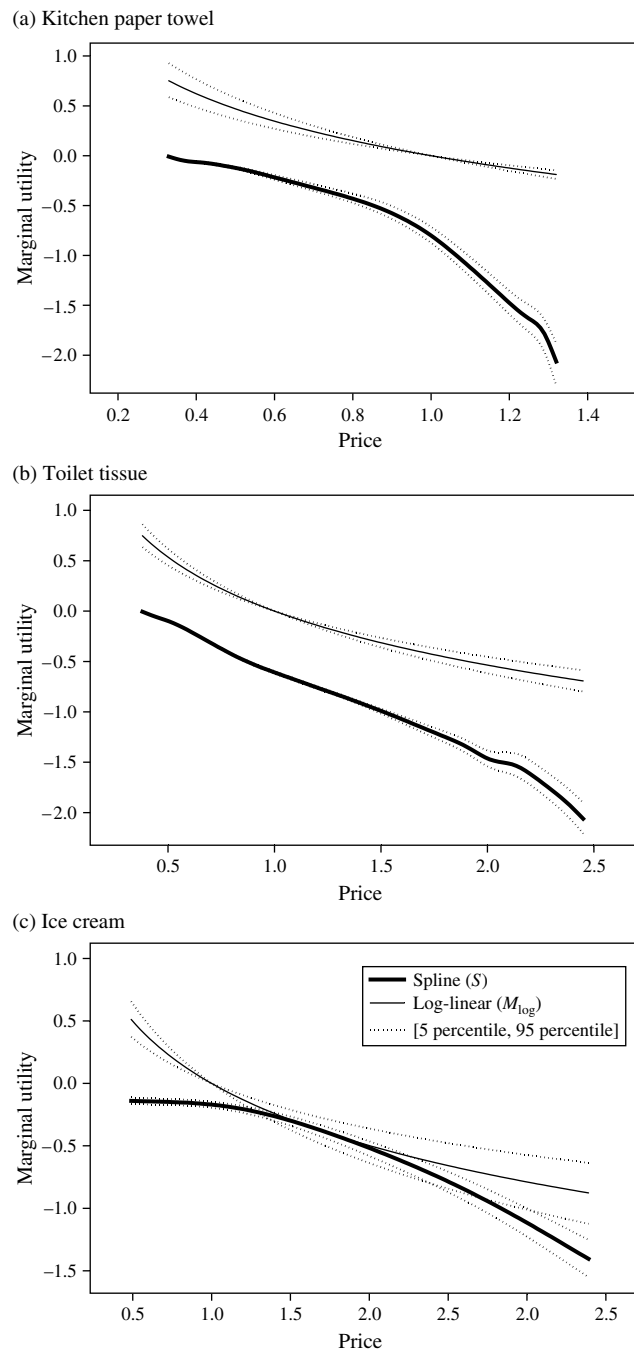
degree of price response nonlinearity. Although we did not systematically study drivers of nonlinearity or “kinkedness,” price response curves for the paper towel data required, on average, significantly ($p < 0.01$) more interior knots than either the toilet tissue or ice cream data.

This raises the issue of how the spline shapes vary across households. The MCMC output readily allows the calculation of household-level response curves, although we must note that posterior variation around these is large for many of them. We report informally that, in spite of the monotonicity constraint, the posterior price splines displayed a remarkable variety of functional shapes, both in terms of concavity versus convexity and degree of kinkedness. Although most individual-level curves appeared to be consistent with a logarithmic or linear specification, many were not, a subject to which we next turn our attention.

Estimated Household-Specific and Aggregate Price Spline Curves. As discussed, we found the shape of the price spline to be quite heterogeneous across households (albeit “noisy”). This finding says little about the market-level price response, as described by an aggregate price spline curve. It is entirely possible that a large proportion of the consumer pool has a household-level price spline curve that is not consistent with either linearity or log-linearity, but that the aggregate price response is well described by one of these functional forms. To investigate this, we calculated the posterior distribution for this aggregate price spline by averaging the H individual price splines at each MCMC iteration. For each of 100 evenly spaced price grid points, Figure 4 presents the posterior means of the average price splines and their associated 90% posterior interval for both the spline model S and under the better fitting of the standard heterogeneous probit models, M_{\log} .

Regardless of the data set, the posterior mean curve for the average price spline does not appear to be well described by linearity or log-linearity. For the paper towel data set, the aggregate price spline curve is clearly concave. For the toilet tissue data, on the other hand, it is consistent with linearity throughout most of its range, with what appear to be inflections for low and high prices. For the ice cream data, the curve based on S is of a very different shape than that based on M_{\log} . Note that these curves are all of the same form under M_{\log} ; they are literally restricted to be as such, even if they accord with certain a priori theories of price response. By contrast, the far more flexible linear spline model, S , allows for rather different aggregate shapes for each of the three data sets. Based on Figure 4 and the superior hold-out performance of S , one might argue that the imposed shape under M_{\log} could well be a misspecification for at least two of these data sets.

Figure 4 Scanner Data—Aggregate Price Splines



We consider these aggregate results telling indications of the power of the nonparametric approach to representing utility at the individual level. However, such aggregate functions provide practitioners with no guidance with regard to effective pricing segmentation schemes. Thus, one might question whether, and how, individual-level price spline results can inform pricing decisions.

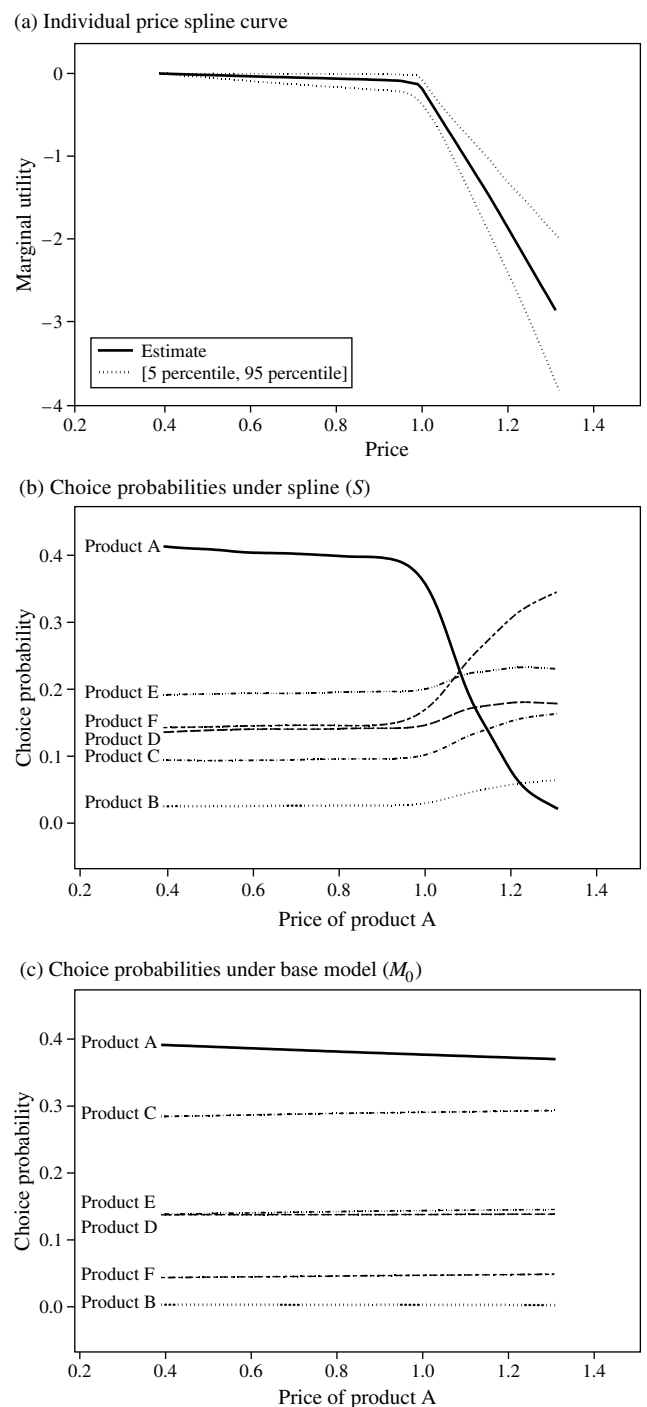
Disaggregate and Aggregate Price Response. Heterogeneity in nonlinear price response raises questions of appropriate pricing practice. Managers would like

to optimize pricing on the “finest” basis possible, that is, on the household level or, failing that, for each of a set of well-delineated segments. Although truly individual-level pricing/promotion is not currently practicable in a traditional supermarket environment, it is becoming increasingly common in electronic transactions. In such settings, managers would wish to know whether there are systematic differences between aggregate- and household-level “optimal” pricing strategies.

Given the price splines, it is straightforward to examine the impact of various sorts of pricing policies.⁵ Doing so requires that household-level price response curves be calculated and averaged across any group or segment of interest. For example, let us consider for illustration household $h = 9$ in the paper towel data, which had a total of $T_9 = 33$ purchase occasions in the training data and a highly kinked posterior price spline, as in Figure 5(a). Given the regression coefficient vector for this household (β_9), the correlation matrix (Σ_{ii}), and the household’s price splines, we can readily obtain choice probabilities for all products as a function of product A’s price. For both the linear spline model and model M_0 , Figures 5(b) and 5(c) plots these choice probabilities for six products. Under M_0 , the posterior mean of the price coefficient for $h = 9$ was -0.0996 , so the household appears to be rather insensitive to price changes. That is, price response (i.e., choice probability curves) are nearly flat as a function of the price for product A. Thus, model M_0 might well suggest a high price for $h = 9$. Under Model S, the household’s choice probabilities also appeared rather insensitive over the price range $[\$0.40, \$0.90]$, but they became considerably more sensitive for higher prices, calling into question the conclusion of M_0 . Such an analysis is, of course, informal, as it sidesteps implementation issues and unintended consequences of targeted pricing policies (Feinberg et al. 2002). Still, group-based pricing could be implemented whenever households can be identified and segmented based on, for example, geodemographics. It would then be a simple matter to estimate segment- or market-level response curves by combining individual-level curves.

Among a store manager’s major decisions is finding the best single price for a product at a given time, that is, at the aggregate level. As in the previous section, one can simply view the entire market as a single segment, aggregating across the respondent pool using draws for β_{ht} , Σ_{ut} , and π_{ht} . As before, we used our

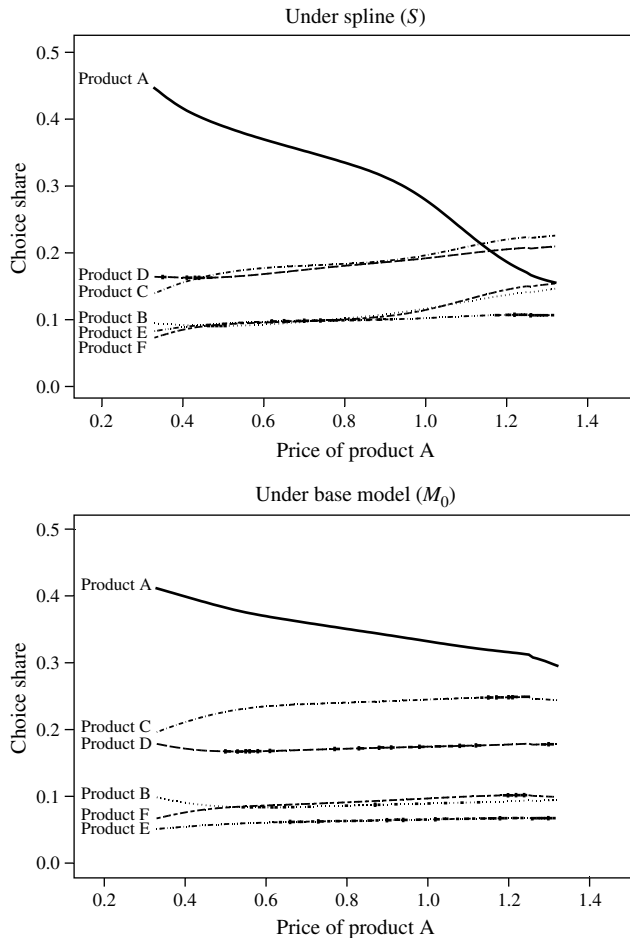
Figure 5 Paper Towels Data—Marginal Utility and Choice Probabilities as a Function of the Price for Product A



paper towel training data set to compute the choice shares of products given different prices for a focal brand. Figure 6 presents a plot of these choice shares for product A. Note that choice probabilities based on the spline model are more sensitive (i.e., steeper) than under the traditional MNP formulation. Moreover, the choice probability for product A displays a notable nonlinearity under S not apparent under M_0 .

⁵ Exploring this issue rigorously requires detailed data on wholesale costs, as well as a model of promotional response, purchase timing, and stockpiling behavior (e.g., Montgomery 1997, Montgomery and Bradlow 1999). In principle, splines can supplement any such model in the manner pursued here, so that household- or segment-level profit calculations follow directly from price response curves.

Figure 6 Paper Towels Data—Aggregate Choice Shares as a Function of the Price for Product A



Given the substantially better hold-out performance of the spline models, one might question pricing suggestions arising from M_0 . In short, the spline model offers systematically and substantively distinct estimates of market-level price response. In turn, the two models would suggest rather different optimal pricing policies, regardless of what else the analyst builds into an optimization model (e.g., wholesale costs and other retailer data). We thus believe that splines may offer a useful method through which to assess price optimization frameworks proposed in marketing, as price splines invoke monotonicity only, as opposed to strict linearity, at the household level.

5. Discussion

In this paper, we present a new approach to estimating utility functions of various shapes at the consumer level. Applying splines, under an additive modeling structure, to a variety of data sets yielded several conclusions:

1. The proposed model performed substantially better than the traditional linear or log-linear specifications. This improvement was apparent in terms of

both Bayes factors and, more important, proportion of correctly predicted hold-out choices.

2. Despite a linear or log-linear appearance to some aggregate utility curves (for price), many individuals had utility curves consistent with neither specification. Moreover, there existed a great deal of heterogeneity in the functional shapes of the splines, with varying degrees of concavity.

3. Market-level response to price differed nontrivially under the traditional log-linear and the linear spline specifications.

The first finding suggests that, all else equal, there is reason to believe that splines can be profitably applied in settings for which the functional nature of the response variable to continuous covariates is unknown. We believe this to be of both theoretical and practical importance. In the vast majority of prior studies, nonlinearities in the utility valuation function have been imposed exogenously and, therefore, must be of a known, prespecified form. Splines can eliminate the sort of guesswork such an approach requires, as well as its attendant biases. Surprisingly, splines offered the greatest incremental value, in terms of hold-out predictive accuracy, in our conjoint application. Although more work is needed to address this issue fully, we believe that the spline model offers a key benefit in being able to “smooth” part-worths across all of a respondent’s levels (for a particular attribute), in addition to the shrinkage afforded by traditional (Bayesian) random-effects conjoint models.

The second finding suggests that even when price response may appear linear or log-linear at the aggregate level, this should not be taken as evidence that individual-level curves are of the same form. Not only did most households have decidedly nonlinear price response, but the shape—both the slope and concavity—of such response varied widely throughout the respondent pool for each of our data sets. One caveat here is that posterior intervals about these curves are often wide enough to call into question any “shapes” based on posterior means alone.

The final finding is especially important: Market-level response was tightly determined in each of our scanner applications, and a rather different functional shape was in evidence across them, in stark contrast to the traditional model, which imposes its shape exogenously. Because optimization frameworks rely on rendering choice share as a function of marketing mix inputs, biases stemming from traditional utility functional forms can lead to costly errors in setting marketing policies.

We do not wish to suggest that “simple” functional specifications have little to recommend them. Indeed, such functions are easier to interpret, their estimators are more efficient, and they are often more accurate when extrapolating outside the range of observed

prices. As general as our modeling framework was, it nevertheless invoked a number of limitations and assumptions. For example, our splines were of the truncated power basis type; more general, although less parsimonious, forms do exist (Schumaker 1981). Although splines do allow flexibility in utility shape, they do not help researchers with another sort of flexibility, that of assessing possible interactions between covariates. In addition, our splines presume continuity; in some applications (far removed from product choice), nonlinear utility functions may exhibit notable jumps, which might potentially be captured through general polynomial splines (Denison et al. 1998). We believe this to be a fertile basis for future investigations, particularly those involving gambles, threshold effects, and reservation prices, all of which have been broadly validated in experimental contexts.

An online supplement to this paper is available on the *Management Science* website (<http://mansci.pubs.informs.org/ecompanion.html>).

6. Electronic Companion

An electronic companion to this paper is available as part of the online version that can be found at <http://mansci.journal.informs.org/>.

Acknowledgments

The authors thank John D. C. Little, Greg Allenby, Peter Lenk, John Liechty, and Michel Wedel for their suggestions.

Appendix. Prior and Posterior Distributions

To conserve space, this appendix gives only a brief summary of the prior distributions for the various parameters in our models. The order of the spline in (5), l_m , is assumed to be given. Because it is unknown, we first consider $l_m = 1$ and increase l_m by one in sequence, determining an appropriate value for l_m by comparing Bayes factors.

To complete the model specification, prior distributions for the two application types are as follows. The individual regression coefficients in both applications are modeled through a hierarchical random-effects specification.

1. Choice-based conjoint application: We assume that $\beta_{h0} \sim N(\mu_0, \sigma_0^2) \forall h$, with $\mu_0 \sim N(g_0, c_0)$ and $\sigma_0^2 \sim IG(a_0, b_0)$, and that $\beta_{h1} \sim N_k(\mu_1, \Sigma_1) \forall h$, with $\mu_1 \sim N_k(g_1, C_1)$ and $\Sigma_1 \sim IW_k(r_1, W_1)$. Here, $N(\mu, \sigma^2)$ denotes a univariate normal distribution with mean μ and variance σ^2 ; $IG(a, b)$ denotes an inverted gamma distribution with shape parameter a and scale parameter b ; and $IW_k(r, W)$ denotes a k -dimensional inverted Wishart distribution with parameters r and W , where $r > 0$ and W is nonsingular.

2. Scanner panel data application: We assume that $\beta_h \sim N_k(\mu_\beta, \Sigma_\beta) \forall h$, with $\mu_\beta \sim N_k(g_\beta, C_\beta)$ and $\Sigma_\beta \sim IW_k(r_\beta, W_\beta)$. The prior distribution for the unknown covariance matrix Σ_u is $IW_f(m, C)$, but Σ_u is restricted to be a correlation matrix.

The parameters related to the spline function are described after (6). Their prior distributions are as follows:

3. We use the same prior distribution for the knot configuration for each individual. Their prior distributions are as follows. The prior distribution for the number of interior knots, q_{hm} , is Poisson with known mean λ_m , truncated at Q_m , the number of all candidate knots in \mathcal{D}_{hm} . In our scanner data applications, we chose $\lambda_1 = 3$. A limited simulation analysis suggested that the results are not really sensitive to departures from this prior distribution.

4. Given the number of interior knots, the prior distribution for the knot locations θ_{hm} is constant.

5. Given the number of interior knots q_{hm} and the knot locations θ_{hm} , the prior distribution for the $(l_m + q_{hm})$ -dimensional spline coefficient vector $\gamma_{hm}^{(q_{hm}, \theta_{hm})} = (\gamma_{hm,1}^{(q_{hm}, \theta_{hm})}, \dots, \gamma_{hm, l_m + q_{hm}}^{(q_{hm}, \theta_{hm})})'$ is such that $\gamma_{hm,i}^{(q_{hm}, \theta_{hm})} \sim N(a_m, b_m)$ for $i = 1, \dots, l_m + q_{hm} \forall h$, possibly subject to constraints like (7).

We use MCMC methods to evaluate the posterior distributions resulting from these prior distributions and the likelihoods for the models described in this paper. More detail on the prior distributions and the MCMC simulation can be obtained in the online appendix and in Kim et al. (2007).

References

- Abe, M. 1998. Measuring consumers' nonlinear alternative choice response to price. *J. Retailing* 74(4) 541–568.
- Allenby, G. M., P. E. Rossi. 1991. Quality perceptions and asymmetric switching between brands. *Marketing Sci.* 10(3) 185–204.
- Andrews, R. L., A. Ansari, I. S. Currim. 2002. Hierarchical Bayes versus finite mixture conjoint analysis models: A comparison of fit, prediction, and partworth recovery. *J. Marketing Res.* 39(1) 87–98.
- Bell, D. R., J. M. Lattin. 2000. Looking for loss aversion in scanner panel data: The confounding effect of price response heterogeneity. *Marketing Sci.* 19(2) 185–200.
- Briesch, R. A., P. K. Chintagunta, R. L. Matzkin. 2002. Semiparametric estimation of brand choice behavior. *J. Amer. Statist. Assoc.* 97 973–982.
- Caves, D. W., L. R. Christensen. 1980. Global properties of flexible functional forms. *Amer. Econom. Rev.* 70(3) 422–432.
- Chib, S., E. Greenberg. 1998. Analysis of multivariate probit models. *Biometrika* 85(2) 347–361.
- Denison, D. G. T., B. K. Mallick, A. F. M. Smith. 1998. Automatic Bayesian curve fitting. *J. Roy. Statist. Soc. Ser. B* 60 333–350.
- Feinberg, F. M., A. Krishna, Z. J. Zhang. 2002. Do we care what others get? A behaviorist approach to targeted promotions. *J. Marketing Res.* 39 277–291.
- Genz, A. 1992. Numerical computation of multivariate normal probabilities. *J. Computational Graphical Statist.* 1 141–149.
- Genz, A. 1993. Comparison of methods for the computation of multivariate normal probabilities. *Comput. Sci. Statist.* 25 400–405.
- Gonzalez, R., G. Wu. 1999. On the shape of the probability weighting function. *Cognitive Psych.* 38 129–166.
- Green, P. J. 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82 711–732.
- Gupta, S., L. G. Cooper. 1992. The discounting of discounts and promotion thresholds (by consumers). *J. Consumer Res.* 19 401–411.
- Hardie, B. G. S., E. J. Johnson, P. S. Fader. 1993. Modeling loss aversion and reference dependence effects on brand choice. *Marketing Sci.* 12 378–394.
- Hastie, T. J., R. J. Tibshirani. 1990. *Generalized Additive Models*. Chapman and Hall, London, UK.

- Kalyanam, K., T. S. Shively. 1998. Estimating irregular pricing effects: A stochastic spline regression approach. *J. Marketing Res.* **35** 16–29.
- Kalyanaram, G., J. D. C. Little. 1989. An empirical analysis of latitude of price acceptance in consumer package goods. *J. Consumer Res.* **21** 408–418.
- Kim, J. G., U. Menzefricke, F. M. Feinberg. 2007. A Bayesian spline approach to capturing heterogeneous utility curves: Theory. Working paper, Ross School of Business, University of Michigan, Ann Arbor, MI.
- Lenk, P. J., W. S. DeSarbo, P. E. Green, M. R. Young. 1996. Hierarchical Bayes conjoint analysis: Recovery of partworth heterogeneity from reduced experimental designs. *Marketing Sci.* **15**(2) 173–191.
- Lindstrom, M. J. 2002. Bayesian estimation of free-knot splines using reversible jumps. *Computational Statist. Data Anal.* **41** 255–269.
- McCulloch, R. E., N. G. Polson, P. E. Rossi. 2000. A Bayesian analysis of the multinomial probit model with fully identified parameters. *J. Econometrics* **99**(1) 173–193.
- Michalek, J. J., F. M. Feinberg, P. Y. Papalambros. 2005. Linking marketing and engineering product design decisions via analytical target cascading. *J. Product Innovation Management* **22**(1) 42–62.
- Montgomery, A. L. 1997. Creating micro-marketing pricing strategies using supermarket scanner data. *Marketing Sci.* **16**(4) 315–337.
- Montgomery, A. L., E. T. Bradlow. 1999. Why analyst overconfidence about the functional form of demand models can lead to overpricing. *Marketing Sci.* **18**(4) 485–503.
- Schumaker, L. L. 1981. *Spline Functions: Basic Theory*. John Wiley & Sons, New York.
- Shively, T. S., G. M. Allenby, R. Kohn. 2000. A nonparametric approach to identifying latent relationships in hierarchical models. *Marketing Sci.* **19**(2) 149–162.
- Wales, T. J. 1977. On the flexibility of flexible functional forms: An empirical approach. *J. Econometrics* **5** 183–193.
- Wegman, E. J., I. W. Wright. 1983. Splines in statistics. *J. Amer. Statist. Assoc.* **78**(382) 351–365.
- Wu, G., R. Gonzalez. 1996. Curvature of the probability weighting function. *Management Sci.* **42** 1676–1690.

Electronic Companion—“Capturing Flexible Heterogeneous Utility Curves: A Bayesian Spline Approach” by Jin Gyo Kim, Ulrich Menzefricke, and Fred M. Feinberg, *Management Science* 2007, 53(2) 340–354.

Online Appendix

This appendix presents the Markov chain Monte Carlo (MCMC) sampler utilized in the main paper. Details on prior literature, set up of the various spline and benchmark models, descriptions of data sets, and all empirical details are included in the main paper.

EC.1. Evaluation of the Posterior Distribution with an MCMC Sampler

Given l_m , the MCMC sampler is designed to estimate individual-specific spline functions with varying knot configuration. Let:

- $\mathbf{y}_h = (y_{h1}, \dots, y_{hT_h})'$, $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_H)'$,
- $\boldsymbol{\beta} = (\boldsymbol{\beta}'_1, \dots, \boldsymbol{\beta}'_H)'$,
- $\mathbf{u}_h = (\mathbf{u}_{h1}, \dots, \mathbf{u}_{hT_h})$ and $\mathbf{u} = (\mathbf{u}_1, \dots, \mathbf{u}_H)$,
- $(\mathbf{q}_h, \boldsymbol{\theta}_h, \boldsymbol{\gamma}_h) = \{(q_{hm}, \theta_{hm}, \gamma_{hm}^{(q_{hm}, \theta_{hm})})\}_{m=1}^M$ and $(\mathbf{q}, \boldsymbol{\theta}, \boldsymbol{\gamma}) = \{(\mathbf{q}_1, \boldsymbol{\theta}_1, \boldsymbol{\gamma}_1), \dots, (\mathbf{q}_H, \boldsymbol{\theta}_H, \boldsymbol{\gamma}_H)\}$,
- $\boldsymbol{\pi}_h^{(\mathbf{q}_h, \boldsymbol{\theta}_h, \boldsymbol{\gamma}_h)} = (\pi_{h1}^{(\mathbf{q}_h, \boldsymbol{\theta}_h, \boldsymbol{\gamma}_h)}, \dots, \pi_{hT_h}^{(\mathbf{q}_h, \boldsymbol{\theta}_h, \boldsymbol{\gamma}_h)})'$ denote effects of $\{v_{hjt, m}\}$ for individual h depending on $\mathbf{q}_h, \boldsymbol{\theta}_h, \boldsymbol{\gamma}_h$, and π_{ht} as in the main paper, and
- $\boldsymbol{\pi}^{(\mathbf{q}, \boldsymbol{\theta}, \boldsymbol{\gamma})} = (\boldsymbol{\pi}'_1^{(\mathbf{q}_1, \boldsymbol{\theta}_1, \boldsymbol{\gamma}_1)}, \dots, \boldsymbol{\pi}'_H^{(\mathbf{q}_H, \boldsymbol{\theta}_H, \boldsymbol{\gamma}_H)})'$.

Then, the full posterior distribution is

$$p(\mathbf{u}, \boldsymbol{\beta}, \boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta, \boldsymbol{\Sigma}_u, \mathbf{q}, \boldsymbol{\theta}, \boldsymbol{\gamma} | \mathbf{y}) \propto p(\mathbf{y} | \mathbf{u})p(\mathbf{u} | \boldsymbol{\beta}, \boldsymbol{\Sigma}_u, \boldsymbol{\pi}^{(\mathbf{q}, \boldsymbol{\theta}, \boldsymbol{\gamma})}) \times p(\boldsymbol{\beta} | \boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta)p(\mathbf{q}, \boldsymbol{\theta}, \boldsymbol{\gamma})p(\boldsymbol{\mu}_\beta)p(\boldsymbol{\Sigma}_\beta)p(\boldsymbol{\Sigma}_u), \quad (\text{EC1})$$

where

$$p(\mathbf{y} | \mathbf{u})p(\mathbf{u} | \boldsymbol{\beta}, \boldsymbol{\Sigma}_u, \boldsymbol{\pi}^{(\mathbf{q}, \boldsymbol{\theta}, \boldsymbol{\gamma})}) \propto \prod_{h=1}^H \prod_{t=1}^{T_h} (N_j(\mathbf{u}_{ht} | \mathbf{x}_{ht}\boldsymbol{\beta}_h + \boldsymbol{\pi}_{ht}^{(\mathbf{q}_h, \boldsymbol{\theta}_h, \boldsymbol{\gamma}_h)}, \boldsymbol{\Sigma}_u)I(\mathbf{u}_{ht} \in A_{ht})),$$

and $A_{ht} = A_{h1t} \times \dots \times A_{hJt}$ is the sample space of \mathbf{u}_{ht} . Furthermore,

$$p(\mathbf{q}, \boldsymbol{\theta}, \boldsymbol{\gamma}) = \prod_{h=1}^H \prod_{m=1}^M p(\gamma_{hm}^{(q_{hm}, \theta_{hm})} | q_{hm}, \theta_{hm})p(\theta_{hm} | q_{hm})p(q_{hm}).$$

To evaluate (EC1), we use Markov chain Monte Carlo methods, sampling all unknown quantities in sequence.

EC.1.1. Sampling from $p(\mathbf{u}, \boldsymbol{\beta}, \boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta, \boldsymbol{\Sigma}_u | \mathbf{q}, \boldsymbol{\theta}, \boldsymbol{\gamma}, \mathbf{y})$

1. Sample \mathbf{u}_{ht} from $p(\mathbf{u}_{ht} | \boldsymbol{\beta}_h, \boldsymbol{\Sigma}_u, \boldsymbol{\pi}_{ht}^{(\mathbf{q}_h, \boldsymbol{\theta}_h, \boldsymbol{\gamma}_h)}, \mathbf{y}) = N_j(\mathbf{u}_{ht} | \mathbf{x}_{ht}\boldsymbol{\beta}_h + \boldsymbol{\pi}_{ht}^{(\mathbf{q}_h, \boldsymbol{\theta}_h, \boldsymbol{\gamma}_h)}, \boldsymbol{\Sigma}_u)I(\mathbf{u}_{ht} \in A_{ht})$ for each individual h and choice occasion t . As J increases, rejection sampling becomes very inefficient. We therefore sample \mathbf{u}_{ht} under the constraint $I(\mathbf{u}_{ht} \in A_{ht})$ by the multivariate slice sampling method (Neal 2003), which allows one to sample multiple quantities simultaneously with only one auxiliary random variable (see Figure 8 on p. 723 of Neal’s paper). In order to implement the multivariate slice sampler, it is important to set the width of slices to be sufficiently large. We set the slice width to be 10.

2. Sample a correlation matrix $\Sigma_u = \{\sigma_{ij}\}$ from $p(\Sigma_u | \mathbf{u}, \boldsymbol{\beta}, \boldsymbol{\pi}^{(q, \theta, \gamma)}) = IW_J(\bar{\mathbf{m}}, \bar{\mathbf{C}})$ by using the slice sampler for off-diagonal elements in sequence given the constraints $\sigma_{ij} = 1$ and $|\sigma_{ij}| < 1$, where $\bar{\mathbf{m}} = \mathbf{m} + \sum_{h=1}^H T_h$ and

$$\bar{\mathbf{C}} = \mathbf{C} + \sum_{h=1}^H \sum_{t=1}^{T_h} (\mathbf{u}_{ht} - \mathbf{x}_{ht} \boldsymbol{\beta}_h - \boldsymbol{\pi}_{ht}^{(q_h, \theta_h, \gamma_h)}) (\mathbf{u}_{ht} - \mathbf{x}_{ht} \boldsymbol{\beta}_h - \boldsymbol{\pi}_{ht}^{(q_h, \theta_h, \gamma_h)})'$$

3. Sample $\boldsymbol{\beta}_h$ from $p(\boldsymbol{\beta}_h | \mathbf{u}_h, \Sigma_u, \boldsymbol{\pi}_{ht}^{(q_h, \theta_h, \gamma_h)}, \boldsymbol{\mu}_\beta, \Sigma_\beta) = N(\bar{\boldsymbol{\mu}}_\beta, \bar{\Sigma}_\beta)$ for each household h , where $\bar{\boldsymbol{\mu}}_\beta = \bar{\Sigma}_\beta \{ \Sigma_\beta^{-1} \boldsymbol{\mu}_\beta + \sum_{t=1}^{T_h} \mathbf{x}'_{ht} \Sigma_u^{-1} (\mathbf{u}_{ht} - \boldsymbol{\pi}_{ht}^{(q_h, \theta_h, \gamma_h)}) \}$ and $\bar{\Sigma}_\beta = (\Sigma_\beta^{-1} + \sum_{t=1}^{T_h} \mathbf{x}'_{ht} \Sigma_u^{-1} \mathbf{x}_{ht})^{-1}$.

4. Sample $\boldsymbol{\mu}_\beta$ from $p(\boldsymbol{\mu}_\beta | \boldsymbol{\beta}, \Sigma_\beta) = N(\bar{\boldsymbol{\mu}}_\beta, \bar{\Sigma}_\beta)$, a multivariate normal density with $\bar{\boldsymbol{\mu}}_\beta = \bar{\Sigma}_\beta (\mathbf{C}_\beta^{-1} \mathbf{m}_\beta + \Sigma_\beta^{-1} \sum_{h=1}^H \boldsymbol{\beta}_h)$ and $\bar{\Sigma}_\beta = (\mathbf{C}_\beta^{-1} + H \Sigma_\beta^{-1})^{-1}$.

5. Sample Σ_β from $p(\Sigma_\beta | \boldsymbol{\beta}, \boldsymbol{\mu}_\beta) = IW(\bar{r}_\beta, \bar{\mathbf{W}}_\beta)$, an inverse Wishart density with $\bar{r}_\beta = r_\beta + H$ and $\bar{\mathbf{W}}_\beta = \mathbf{W}_\beta + \sum_{h=1}^H (\boldsymbol{\beta}_h - \boldsymbol{\mu}_\beta)(\boldsymbol{\beta}_h - \boldsymbol{\mu}_\beta)'$.

EC.1.2. Sampling from $p(\mathbf{q}, \boldsymbol{\theta}, \boldsymbol{\gamma} | \mathbf{u}, \boldsymbol{\beta}, \boldsymbol{\mu}_\beta, \Sigma_\beta, \Sigma_u, \mathbf{y})$

For each household and each spline, we have:

$$\begin{aligned} & p(q_{hm}, \theta_{hm}, \boldsymbol{\gamma}_{hm}^{(q_{hm}, \theta_{hm})} | \mathbf{u}_h, \boldsymbol{\beta}_h, \Sigma_u, \boldsymbol{\pi}_h^{(q_h, \theta_h, \gamma_h)}) \\ & \propto p(\mathbf{u}_h | \boldsymbol{\beta}_h, \Sigma_u, \boldsymbol{\pi}_h^{(q_h, \theta_h, \gamma_h)}) p(\boldsymbol{\gamma}_{hm}^{(q_{hm}, \theta_{hm})} | q_{hm}, \theta_{hm}) \times p(\theta_{hm} | q_{hm}) p(q_{hm}). \end{aligned}$$

It is necessary to allow the dimension of $(q_{hm}, \theta_{hm}, \boldsymbol{\gamma}_{hm})$ to change across iterations. So, as discussed previously, we use the reversible jump MCMC method (Green 1995), which is designed to move around a countable union of subspaces $\Phi = \bigcup_{i=0}^\infty \Phi_i$ by making random transitions between Φ_i and Φ_k . To implement it, we consider four possible transitions: (a) a birth step (addition of a knot), (b) a death step (deletion of a knot), (c) movement of a knot, and (d) update of $\boldsymbol{\gamma}_{hm}$ without changes in q_{hm} and θ_{hm} . Thus, the set of possible moves is $w \in \{U, M, 0, 1, 2, \dots\}$, where U means an update of $\boldsymbol{\gamma}_{hm}$ without changes in q_{hm} , θ_{hm} , M means a random movement of a knot, and $m = 0, 1, 2, \dots$ refers to increasing the number of interior knots from $q_{hm} = m$ to $q_h = m + 1$ or decreasing from $q_{hm} = m + 1$ to $q_{hm} = m$.

We let the probabilities for these four possible transitions be

$$s_{q_{hm}} = a \min\left(1, \frac{p(q_{hm} + 1)}{p(q_{hm})}\right) \quad \text{for a birth step,} \quad (\text{EC2})$$

$$\tau_{q_{hm}} = a \min\left(1, \frac{p(q_{hm} - 1)}{p(q_{hm})}\right) \quad \text{for a death step,} \quad (\text{EC3})$$

$$v_{q_{hm}} = b(1 - s_{q_{hm}} - \tau_{q_{hm}}) \quad \text{for a move step,} \quad \text{and} \quad (\text{EC4})$$

$$\xi_{q_{hm}} = 1 - s_{q_{hm}} - \tau_{q_{hm}} - v_{q_{hm}} \quad \text{for an update step.} \quad (\text{EC5})$$

The positive constant a should be as large as possible subject to $s_{q_h} + \tau_{q_h} \leq 0.9$ for all $q_{hm} = 0, 1, \dots, Q_m$. Note that this has the practical consequence of bounding the move and update step probabilities below by $1 - 2a$. It is easy to find the constant a by examining $(s_{q_h} + \tau_{q_h})$ over all values of $q_{hm} \leq Q_m$. Note that a should fall in the interval $[0, 0.5]$; otherwise, if $a > 0.5$, the sum of the probabilities s_{q_h} and τ_{q_h} could be greater than one for some values of q_{hm} . We take $a = 0.45$, but other values are also valid. Because q_{hm} should be in the interval $[0, Q_m]$, we use $\tau_0 = v_0 = 0$ and $s_{Q_m} = v_{Q_m} = 0$. Note that the ratio $p(q_{hm} \pm 1)/p(q_{hm})$ is not affected by this truncation because $p(q_{hm})I(q_{hm} \leq Q_m) = (1/\sum_{i=0}^{Q_m} \text{Po}(q_{im} | \lambda_m)) \cdot \text{Po}(q_{hm} | \lambda_m)$. In addition, we let $b = 0.5$, but other nonnegative values from the interval $(0, 1)$ are also valid.

EC.1.2.1 Update Step. In the update step, only $\boldsymbol{\gamma}_{hm} = \boldsymbol{\gamma}_{hm}^{(q_{hm}, \theta_{hm})}$ is updated, q_{hm} and θ_{hm} being unchanged. Let the scalar $\tilde{u}_{hjt} = u_{hjt} - \mathbf{x}'_{hjt} \boldsymbol{\beta}_h - \sum_{i=1, i \neq m}^M f_i^h(v_{hjt}, i)$, let $\tilde{\mathbf{u}}_{hjm} = (\tilde{u}_{h1tm}, \dots, \tilde{u}_{hJtm})'$, a

J -dimensional vector, and let $\tilde{\mathbf{u}}_{hm} = (\tilde{\mathbf{u}}_{h1m}, \dots, \tilde{\mathbf{u}}_{hT_h m})$. Furthermore, let the $(l_m + q_{hm})$ -dimensional vector \mathbf{z}_{hjtm} collect the coefficients of γ_{hm} , that is,

$$\mathbf{z}_{hjtm} = \begin{pmatrix} (v_{hjtm} - s_{hm0})_+^1 \\ (v_{hjtm} - s_{hm0})_+^2 \\ \vdots \\ (v_{hjtm} - s_{hm0})_+^{l_m} \\ (v_{hjtm} - s_{hm1})_+^{l_m} \\ \vdots \\ (v_{hjtm} - s_{hmq_{hm}})_+^{l_m} \end{pmatrix},$$

and let $\mathbf{z}_{hmtm} = (\mathbf{z}_{h1tm}, \dots, \mathbf{z}_{hT_h tm})'$ be a $J \times (l_m + q_{hm})$ matrix. Thus, $\tilde{\mathbf{u}}_{hmtm} = \mathbf{z}_{hmtm} \boldsymbol{\gamma}_{hm}^{(q_{hm}, \theta_{hm})} + \boldsymbol{\varepsilon}_{ht}$. We then have

$$p(\boldsymbol{\gamma}_{hm} \mid \tilde{\mathbf{u}}_{hm}, \boldsymbol{\Sigma}_u, \mathbf{q}_h, \boldsymbol{\theta}_h) = \frac{1}{c_{hm}} p(\tilde{\mathbf{u}}_{hm} \mid \boldsymbol{\gamma}_{hm}, \boldsymbol{\Sigma}_u) p(\boldsymbol{\gamma}_{hm} \mid q_h, \boldsymbol{\theta}_{hm}),$$

with normalizing constant

$$c_{hm} = c(q_{hm}, \theta_{hm}) = \frac{\exp\left\{-\frac{1}{2}\left[\sum_{t=1}^{T_h} \tilde{\mathbf{u}}'_{hmtm} \boldsymbol{\Sigma}_u^{-1} \tilde{\mathbf{u}}_{hmtm} + (l_m + q_{hm}) \frac{a_m^2}{b_m} + \mathbf{g}'_{hm} \mathbf{G}_{hm}^{-1} \mathbf{g}_{hm}\right]\right\}}{(2\pi)^{J T_h / 2} |\boldsymbol{\Sigma}_u|^{T_h / 2} \left| \mathbf{I}_{l_m + q_{hm}} + b_m \sum_{t=1}^{T_h} \mathbf{z}'_{hmtm} \boldsymbol{\Sigma}_u^{-1} \mathbf{z}_{hmtm} \right|^{1/2}}, \quad (\text{EC6})$$

where $\mathbf{g}_{hm} = \mathbf{G}_{hm} \{(a_m/b_m) \mathbf{1}_{l_m + q_{hm}} + \sum_{t=1}^{T_h} \mathbf{z}'_{hmtm} \boldsymbol{\Sigma}_u^{-1} \tilde{\mathbf{u}}_{hmtm}\}$, $\mathbf{G}_{hm} = \{(1/b_m) \mathbf{I}_{l_m + q_{hm}} + \sum_{t=1}^{T_h} \mathbf{z}'_{hmtm} \boldsymbol{\Sigma}_u^{-1} \mathbf{z}_{hmtm}\}^{-1}$, $\mathbf{1}_q$ is a vector of q ones, and the prior parameters a_m and b_m are defined as in the main paper, that is, the conditional posterior distribution for $\boldsymbol{\gamma}_{hm}$ given $(\tilde{\mathbf{u}}_{hmtm}, \boldsymbol{\Sigma}_u, \boldsymbol{\pi}_h^{(q_h, \boldsymbol{\theta}_h, \boldsymbol{\gamma}_h)}, \mathbf{q}_h, \boldsymbol{\theta}_h)$ is

$$\boldsymbol{\gamma}_{hm} \sim N_{l_m + q_{hm}}(\mathbf{g}_{hm}, \mathbf{G}_{hm}). \quad (\text{EC7})$$

When a constraint is imposed on the splines, for example, monotonicity, then $\boldsymbol{\gamma}_{hm}$ must be sampled from $N_{l_m + q_{hm}}(\mathbf{g}_{hm}, \mathbf{G}_{hm}) I(\boldsymbol{\gamma}_{hm} \in B)$, where B denotes the space of $\boldsymbol{\gamma}_{hm}$ under the imposed constraint. It is thus necessary to multiply the normalizing constant in (EC6) by a correction factor involving $\phi(q_{hm}, \theta_{hm}) = \int_{\boldsymbol{\gamma}_{hm} \in B} p(\boldsymbol{\gamma}_{hm} \mid \mathbf{g}_{hm}, \mathbf{G}_{hm}) d\boldsymbol{\gamma}_{hm}$ so that the normalizing constant is

$$c_{hm}^* = c^*(q_{hm}, \theta_{hm}) = \frac{1}{\phi(q_{hm}, \theta_{hm})} c(q_{hm}, \theta_{hm}).$$

In this case, we can use the single variable slice sampler (cf., Neal 2003) to sample $\boldsymbol{\gamma}_{hm}$.

EC.1.2.2. Birth, Death, and Move Steps. In the birth, death, and move steps, the transition probabilities in (EC2) to (EC4) yield a proposed new value for $\bar{q}_{hm} = q_{hm} + r$, where $r = -1, 0, 1$. Given \bar{q}_{hm} , we must then propose a new value for θ_{hm} , $\bar{\theta}_{hm}$. To add a new knot or move/delete one of the knots currently present, we have a set of candidate knots, $\mathcal{D}_{hm} = \{D_{hm1}, \dots, D_{hmQ_m}\}$, which can be, for example, a set of prespecified grid points or a subset of $\{v_{hjt, m}\}$ for all j and t . Then, for each iteration, a newly proposed value of $\bar{\theta}_{hm}$, given \bar{q}_{hm} , is generated as follows:

1. Birth step: Add a knot uniformly chosen from one of the $Q_m - q_{hm}$ candidate grid points from \mathcal{D}_{hm} .
2. Death step: Delete a knot uniformly chosen from the q_{hm} knots currently present.
3. Move step: Choose one of the currently present q_h knots uniformly and change its location to a value uniformly chosen from the currently nonpresent knots.

Given the newly proposed knot configuration $(\bar{q}_{hm}, \bar{\theta}_{hm})$, we must finally propose a value for the spline coefficients for each household and each spline, $\bar{\boldsymbol{\gamma}}_{hm} = \bar{\boldsymbol{\gamma}}_{hm}^{(\bar{q}_{hm}, \bar{\theta}_{hm})}$, where we drop the superscript for expository convenience. Our proposal distribution for these spline coefficients, $\bar{\boldsymbol{\gamma}}_{hm}$, is their conditional posterior distribution, that is, $\bar{\boldsymbol{\gamma}}_{hm} \sim N_{l_m + \bar{q}_{hm}}(\bar{\mathbf{g}}_{hm}, \bar{\mathbf{G}}_{hm})$, where $\bar{\mathbf{g}}_{hm}$ and $\bar{\mathbf{G}}_{hm}$ are defined just before (EC7). Note that we have changed the notation from \mathbf{g}_{hm} and \mathbf{G}_{hm} in (EC7) to $\bar{\mathbf{g}}_{hm}$ and $\bar{\mathbf{G}}_{hm}$ to make explicit that these quantities are computed for the newly proposed knot configuration $(\bar{q}_{hm}, \bar{\theta}_{hm})$.

Given the proposal $(\bar{q}_{hm}, \bar{\theta}_{hm}, \bar{\boldsymbol{\gamma}}_{hm}^{(\bar{q}_{hm}, \bar{\theta}_{hm})})$ generated above, we must now decide whether or not to accept it. Using the notation in Green (1995), the acceptance probability for each move type is

$$\alpha = \min\{1, (\text{likelihood ratio}) \times (\text{prior ratio}) \times (\text{proposal ratio}) \times \text{Jacobian}\}.$$

The Jacobian is needed for the steps in which the dimension is changing: It does not matter for the move step, but it does for the birth and death steps. In both these steps, the Jacobian term is one because the new knot configuration, $\bar{\mathbf{q}}_h$ and $\boldsymbol{\theta}_h^{(\bar{\mathbf{q}}_h)}$, and the spline coefficients, $\bar{\boldsymbol{\gamma}}_{hm}^{(\bar{q}_{hm}, \bar{\theta}_{hm})}$, are generated independently of the previous values.

Concentrating first on the pair $(q_{hm}, \theta_{hm}^{(q_{hm})})$ and a birth step, the prior ratio is

$$\frac{p(q_{hm} + 1) \frac{1}{1/\binom{Q_m}{q_{hm}+1}}}{p(q_{hm}) \frac{1}{1/\binom{Q_m}{q_{hm}}}} = \frac{p(q_{hm} + 1) \frac{q_{hm} + 1}{Q_m - q_{hm}}}{p(q_{hm})},$$

and the proposal ratio is

$$\frac{P(\text{death} | q_{hm} + 1) \frac{1}{q_{hm} + 1}}{P(\text{birth} | q_{hm}) \frac{1}{Q_m - q_{hm}}} = \frac{p(q_{hm}) \frac{Q_m - q_{hm}}{q_{hm} + 1}}{p(q_{hm} + 1)},$$

where we have used the fact that $P(\text{death} | q_{hm} + 1) = a \min\{1, p(q_{hm})/p(q_{hm} + 1)\}$ and $P(\text{birth} | q_{hm}) = a \min\{1, p(q_{hm} + 1)/p(q_{hm})\}$. Combining prior and proposal ratios for the pair $(q_{hm}, \theta_{hm}^{(q_{hm})})$, we find the product to be one. For a death step the ratio is the inverse, that is, one, and for a move step, it is also one.

Let us now concentrate on the likelihood ratio, and on the prior ratio and proposal ratio contributed by the spline coefficients. Because the proposal distribution for the spline coefficient vector is its conditional posterior distribution as given in (EC7), the product of likelihood ratio, prior ratio, and proposal ratio is just the corresponding ratio of the normalizing constants in (EC6), $c(\bar{q}_{hm}, \bar{\theta}_{hm})/c(q_{hm}, \theta_{hm})$, where $c(\bar{q}_{hm}, \bar{\theta}_{hm})$ indicates that the normalizing constant is computed for the newly proposed knot configuration.

The acceptance probability is thus

$$\alpha = \min\left\{1, \frac{c(\bar{q}_{hm}, \bar{\theta}_{hm})}{c(q_{hm}, \theta_{hm})}\right\}, \quad (\text{EC8})$$

with

$$\begin{aligned} \frac{c(\bar{q}_{hm}, \bar{\theta}_{hm})}{c(q_{hm}, \theta_{hm})} &= \frac{|\mathbf{I}_{m+q_{hm}} + b_m \sum_{t=1}^{T_h} \mathbf{z}'_{htm} \boldsymbol{\Sigma}_u^{-1} \mathbf{z}_{htm}|^{1/2}}{|\mathbf{I}_{m+\bar{q}_{hm}} + b_m \sum_{t=1}^{T_h} \bar{\mathbf{z}}'_{htm} \boldsymbol{\Sigma}_u^{-1} \bar{\mathbf{z}}_{htm}|^{1/2}} \\ &\quad \times \exp\left\{\frac{1}{2}\left[(q_{hm} - \bar{q}_{hm}) \frac{a_m^2}{b_m} + \mathbf{g}'_{hm} \mathbf{G}_{hm}^{-1} \mathbf{g}_{hm} - \bar{\mathbf{g}}'_{hm} \bar{\mathbf{G}}_{hm}^{-1} \bar{\mathbf{g}}_{hm}\right]\right\}, \end{aligned}$$

where the term $(q_{hm} - \bar{q}_{hm})$ equals 1, 0, and -1 for a death, move, and birth step, respectively.

When there is a constraint on the spline coefficient vector such that $\boldsymbol{\gamma}_{hm} \in B$, each of the normalizing constants in (EC8) must be multiplied by its appropriate correction factor involving $\phi(\bullet)$ as discussed before and we must sample the newly proposed value for the spline coefficient vector, $\bar{\boldsymbol{\gamma}}_{hm}$, from this region. We found rejection sampling to be very inefficient for this purpose, and so used a slice sampler instead.

References

See references list in the main paper.

Neal, R. M. 2003. Slice sampling (with discussion). *Ann. Statist.* 31 705–767.