



## Monotonic Regression Based on Bayesian P-Splines

An Application to Estimating Price Response Functions From Store-Level Scanner Data

Andreas Brezger & Winfried J Steiner

To cite this article: Andreas Brezger & Winfried J Steiner (2008) Monotonic Regression Based on Bayesian P-Splines, Journal of Business & Economic Statistics, 26:1, 90-104, DOI: [10.1198/073500107000000223](https://doi.org/10.1198/073500107000000223)

To link to this article: <https://doi.org/10.1198/073500107000000223>



Published online: 01 Jan 2012.



Submit your article to this journal [↗](#)



Article views: 366



View related articles [↗](#)



Citing articles: 12 View citing articles [↗](#)

# Monotonic Regression Based on Bayesian P–Splines: An Application to Estimating Price Response Functions From Store-Level Scanner Data

**Andreas BREZGER**

Department of Statistics, University of Munich, Munich, Germany

**Winfried J. STEINER**

Department of Marketing, Institute of Management and Economics, Technical University of Clausthal, Clausthal-Zellerfeld, Germany ([winfried.steiner@tu-clausthal.de](mailto:winfried.steiner@tu-clausthal.de))

In many practical situations, it is desirable to restrict the flexibility of nonparametric estimation to accommodate a presumed monotonic relationship between a covariate and the response variable. We follow a Bayesian approach using penalized B-splines and incorporate the assumption of monotonicity in a natural way by an appropriate specification of the respective prior distributions. We illustrate the methodology in an empirical application modeling demand for brands of orange juice and show that imposing monotonicity constraints for own- and cross-item price effects improves the predictive validity of the estimated sales response functions considerably.

**KEY WORDS:** Asymmetric quality tier competition; Generalized additive model; Markov chain Monte Carlo; Own- and cross-item price effects; Sales promotion.

## 1. INTRODUCTION

Generalized additive models (GAMs) are a powerful tool for modeling possibly nonlinear effects of multiple covariates. For continuous covariates, the variety of different approaches for nonlinear modeling comprises, for example, smoothing splines (e.g., Hastie and Tibshirani 1990), regression splines (e.g., Friedman and Silverman 1989; Friedman 1991; Stone, Hansen, Kooperberg, and Truong 1997), local methods (e.g., Fan and Gijbels 1996) as well as P-splines (Eilers and Marx 1996; Marx and Eilers 1998). Bayesian nonparametric approaches make use of adaptive knot selection (e.g., Smith and Kohn 1996; Denison, Mallick, and Smith 1998; Biller 2000; Di Matteo, Genovese, and Kass 2001; Biller and Fahrmeir 2001; Hansen and Kooperberg 2002) or smoothness priors (Hastie and Tibshirani 2000; Fahrmeir and Lang 2001a,b). Lang and Brezger (2004) have adopted the frequentist P-splines of Eilers and Marx (1996) for a Bayesian framework for additive models and Brezger and Lang (2005) have extended their work to GAMs.

Whereas strictly parametric modeling is too restrictive in many cases, the flexibility of non- and semiparametric approaches may lead to implausible results on the other hand. Clearly, the problem of overfitting can be addressed by penalization of too rough functions or by adaptive knot selection. Much less discussed in the literature on nonparametric estimation is, however, the important case when theory and/or empirical evidence strongly suggest a monotonic relationship between a covariate and a response variable. For example, consumers usually will buy less of a brand as its price increases, and therefore one expects a brand's unit sales or market share to decrease monotonically in price. The downward slope of own-price response functions is in accordance with economic theory (e.g., Rao 1993), and there is strong empirical support that own-price elasticities are negative and elastic (e.g., Tellis 1988;

Hanssens, Parsons, and Schultz 2001). Similarly, we generally expect cross-price effects on competitive items (i.e., brand substitutes) to be positive or at least nonnegative, implying that a price cut by one brand may decrease but by no means will increase the unit sales of competitive brands (Sethuraman, Srinivasan, and Kim 1999). Examples for presumed monotonic relationships can also be found in disciplines other than business and economics, as it is the case for many dose-response relationships in medicine. For instance, the concentration of dust and the duration of exposure to it at work is assumed to affect the occurrence of certain lung diseases in a monotonic way (Ulm and Salanti 2003). Monotonic effects are also referred to as isotonic if the respective function is nondecreasing, and antitonic if a function is nonincreasing.

The topic of monotonic regression has already been addressed by Ramsay (1988, 1998), Ulm and Salanti (2003), and Salanti and Ulm (2003) in a frequentist setting. Allenby, Arora, and Ginter (1995) have incorporated prior ordinal information about regression parameters of dummy coded covariates in a Bayesian setting. Specifically, they have proposed a Bayesian model for conjoint analysis with order restrictions (monotonicity constraints) for part-worth utilities. Dunson and Neelon (2003) and Holmes and Heard (2003) have presented further Bayesian approaches to monotonic regression. The former, however, have considered only GLMs and modeling has been based on piecewise constant functions, while the latter have dealt with only a small number of level sets obtained from a categorization of continuous covariates.

In this article we propose to use Bayesian P-splines of an arbitrary degree and enforce monotonicity with an additional

restriction. This restriction may be imposed for either one or an arbitrary number of the additive terms in the model, whereas other terms may be modeled unrestricted. Markov chain Monte Carlo (MCMC) inference involves sampling from multivariate truncated normal distributions. This is accomplished by an “internal” Gibbs sampler in each iteration; that is, we employ a short Gibbs sampler to draw from the proposal density. In the non-Gaussian case, we draw from an iteratively weighted least squares (IWLS) proposal density in a Metropolis–Hastings step. Our methodology is implemented with the public domain software package *BayesX* (Brezger, Kneib, and Lang 2005). It is possible to combine monotonic regression with all types of response distributions supported by *BayesX*. These are the most common one-dimensional distributions such as Gaussian, Binomial, Poisson, Gamma, and Negative Binomial, and multinomial logit and cumulative probit models for multivariate responses. *BayesX* also supports the use of random effects to account for unobserved heterogeneity, Gaussian Markov random field (GMRF) priors for spatial covariates, and varying coefficient terms and surface smoothing for interactions of covariates.

The remainder of the article is organized as follows: Section 2 briefly reviews GAMs and (Bayesian) P-splines, whereas Section 3 provides details on the MCMC techniques employed. In Section 4 we apply the proposed methodology to weekly store-level scanner data to relate unit sales of a particular brand of orange juice in a major supermarket chain to own and competing brands’ promotional instruments. Using a lognormal model and a Gamma model, we illustrate for both Gaussian and non-Gaussian responses that imposing monotonicity constraints on the nonparametric terms for own-item and cross-item price effects improves the predictive validity of the estimated sales response functions considerably. In Section 5 we present some refinements of our methodology, including a relaxation of our MCMC procedure to impose monotonicity as well as a sensitivity analysis on the number and position of knots used for the P-splines. We further apply our methodology to a second brand (the supermarket’s private label brand) to validate our findings obtained in Section 4. We conclude with a summary of the most important contents and key findings in Section 6.

## 2. MODEL ASSUMPTIONS

### 2.1 Generalized Additive Models and P-Splines

Suppose we are given  $N$  observations  $(y_n, x_n, v_n)$ ,  $n = 1, \dots, N$ , where  $y_n$  is a response variable,  $x_n = (x_{n1}, \dots, x_{np})'$  is a vector of continuous covariates, and  $v_n = (v_{n1}, \dots, v_{nq})'$  is a vector of additional covariates. GAMs assume that, given  $x_n$  and  $v_n$ , the response  $y_n$  follows an exponential family distribution (Hastie and Tibshirani 1990; Fahrmeir and Tutz 2001)

$$p(y_n|x_n, v_n) = c(y_n, \phi) \exp \left\{ \frac{y_n \theta_n - b(\theta_n)}{\phi} \right\}, \quad (1)$$

where  $b(\cdot)$ ,  $c(\cdot)$ ,  $\theta_n$ , and  $\phi$  define the respective distribution [e.g., if  $b(\theta_n) = \theta_n^2/2$ ,  $c(y_n, \phi) = \exp(y_n^2/(2\phi))/\sqrt{2\pi\phi}$ ,  $\theta_n = \mu_n$ , and  $\phi = \sigma^2$ , we obtain an  $N(\mu_n, \sigma^2)$  distribution]. GAMs further assume that the mean  $\mu_n = E(y_n|x_n, v_n)$  is linked to a semiparametric additive predictor  $\eta_n$  via a known link function  $g$ :

$$g(\mu_n) = \eta_n, \quad \eta_n = f_1(x_{n1}) + \dots + f_p(x_{np}) + v_n' \gamma. \quad (2)$$

$f_1, \dots, f_p$  are unknown smooth functions of the continuous covariates and  $v_n' \gamma$  represents the parametric part of the predictor.

For modeling the unknown functions  $f_j$ ,  $j = 1, \dots, p$ , we follow Lang and Brezger (2004), who proposed a Bayesian version of the P-splines approach introduced in a frequentist setting by Eilers and Marx (1996). Accordingly, we assume that the unknown functions can be approximated by a polynomial spline of degree  $l$  and with  $k + 1$  equally spaced knots

$$x_{j,\min} = \zeta_{j0} < \zeta_{j1} < \dots < \zeta_{j,k-1} < \zeta_{jk} = x_{j,\max} \quad (3)$$

over the domain of  $x_j$ . The spline can be written in terms of a linear combination of  $\Psi = k + l$  B-spline basis functions (De Boor 2001). Figure 1 gives an illustration of B-spline basis functions of degree 3, which are also referred to as cubic splines. Explicit formulas for the functional form of B-spline basis functions are given in Appendix A. Note that except at the boundaries each basis function overlaps with  $2 \cdot l$  neighboring B-splines (i.e., with six neighboring B-splines in the case of cubic splines; see Fig. 1).

Denoting the  $\psi$ th basis function by  $B_{j\psi}$ , we obtain

$$f_j(x_j) = \sum_{\psi=1}^{\Psi} \beta_{j\psi} B_{j\psi}(x_j). \quad (4)$$

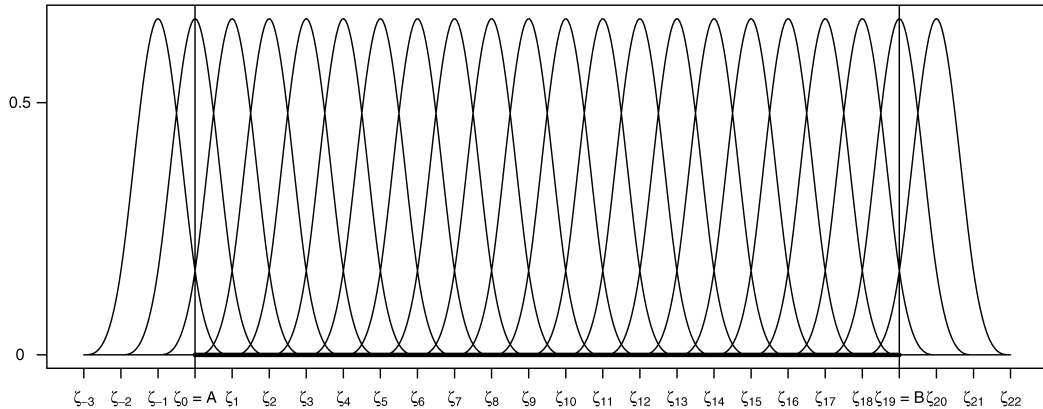


Figure 1. B-spline basis functions of degree 3 covering the interval  $[A, B]$ .

To keep notation simple, we assume an equal number of basis functions  $\Psi$  for all functions  $f_j$ . By defining the  $N \times \Psi$  design matrices  $X_j$ , where the element in row  $n$  and column  $\psi$  is given by  $X_j(n, \psi) = B_{j\psi}(x_{nj})$ , we can rewrite the predictor (2) in matrix notation as

$$\eta = X_1\beta_1 + \cdots + X_p\beta_p + V\gamma. \quad (5)$$

Here  $\beta_j = (\beta_{j1}, \dots, \beta_{j\Psi})'$ ,  $j = 1, \dots, p$ , corresponds to the vector of unknown regression coefficients. The matrix  $V$  is the usual design matrix for parametric effects. To overcome the difficulties in determining the position and the number of the knots involved with regression splines, Eilers and Marx (1996) suggested a relatively large number of knots (usually between 20 and 40) to ensure sufficient flexibility, and to introduce a roughness penalty of first- or second-order differences on adjacent regression coefficients to avoid overfitting. These penalized B-splines have also become known as P-splines. In our Bayesian approach, we replace first- or second-order differences used in this frequentist approach with their stochastic analogues, that is, first- or second-order random walks defined by

$$\begin{aligned} \beta_{j\psi} &= \beta_{j,\psi-1} + u_{j\psi} \quad \text{or} \\ \beta_{j\psi} &= 2\beta_{j,\psi-1} - \beta_{j,\psi-2} + u_{j\psi} \end{aligned} \quad (6)$$

with Gaussian errors  $u_{j\psi} \sim N(0, \tau_j^2)$  and diffuse priors  $\beta_{j1} \propto \text{const}$ , or  $\beta_{j1}$  and  $\beta_{j2} \propto \text{const}$ , for initial values, respectively. The amount of smoothness is controlled by the variance parameter  $\tau_j^2$ , which corresponds to the inverse of the smoothing parameter in the frequentist approach. The amount of smoothness can be estimated simultaneously with the regression coefficients by defining an additional hyperprior for the variance parameters  $\tau_j^2$ .

We assign the conjugate prior for  $\tau_j^2$  (and for the scale parameter  $\sigma^2$  in the Gaussian case) which is an inverse Gamma distribution

$$\tau_j^2 \sim \text{IG}(a_j, b_j) \quad (7)$$

with hyperparameters  $a_j$  and  $b_j$ . A common choice for  $a_j$  and  $b_j$  leading to almost diffuse priors is  $a_j = b_j$ , for example,  $a_j = b_j = .001$ , which is also our default choice. Alternatively, we may set  $a_j = 1$  and  $b_j$  small, for example,  $b_j = .005$  or  $b_j = .0005$ . We estimated all models discussed in this article with alternative settings for the hyperparameters. The results proved to be almost insensitive regarding the specific choice of hyperparameters. All results presented in the remainder of the article are obtained by the default choice.

The random walks defined as priors for the individual basis functions according to (6) can be combined in a global smoothness prior. Taking into account that we use diffuse priors for the initial  $\delta$  parameters (i.e.,  $\delta = 1$  or  $\delta = 2$  corresponding to a first- or second-order random walk, respectively), we can write (6) equivalently as  $D^\delta \beta_j \sim N(0, \tau_j^2 I_{\Psi-\delta})$ , where  $D^\delta$  is a difference matrix of order  $\delta$  and dimension  $(\Psi - \delta) \times \Psi$ , and  $I_{\Psi-\delta}$  is an identity matrix of dimension  $(\Psi - \delta)$ . Hence, by defining the penalty matrix  $K^\delta = D^{\delta'} D^\delta$ , we can formulate the prior for a P-spline term as a joint prior distribution for  $\beta_j$ :

$$p(\beta_j) \propto \exp \left\{ -0.5 \frac{1}{\tau_j^2} \beta_j' D^{\delta'} D^\delta \beta_j \right\} = \exp \left\{ -0.5 \frac{1}{\tau_j^2} \beta_j' K_j^\delta \beta_j \right\}; \quad (8)$$

see Lang and Brezger (2004) for details. For example, for  $\delta = 1$  we get

$$K^1 = \begin{pmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & 2 & -1 & \\ & & \ddots & \ddots & \ddots \\ & & & -1 & 2 & -1 \\ & & & & -1 & 2 & -1 \\ & & & & & -1 & 1 \end{pmatrix} \quad (9)$$

with zero elements outside the first off-diagonals.

## 2.2 Monotonicity Constraints

To obtain monotonicity, that is,  $f_j'(x) \geq 0$  or  $f_j'(x) \leq 0$ , it is sufficient to guarantee that subsequent parameters are ordered, such that

$$\beta_{j1} \leq \cdots \leq \beta_{j\Psi} \quad \text{or} \quad \beta_{j1} \geq \cdots \geq \beta_{j\Psi}, \quad (10)$$

respectively. A proof can be found in Appendix B. In our approach, these constraints are imposed by introducing indicator functions to truncate the prior appropriately to obtain the desired support. This leads to

$$p(\beta_j) = c_1(\beta_j) \exp \left\{ -0.5 \frac{1}{\tau_j^2} \beta_j' K_j^\delta \beta_j \right\} \prod_{\psi=2}^{\Psi} \mathbf{1}(\beta_{j\psi} \geq \beta_{j,\psi-1}) \quad (11)$$

for nondecreasing functions (isotonic case) and

$$p(\beta_j) = c_1(\beta_j) \exp \left\{ -0.5 \frac{1}{\tau_j^2} \beta_j' K_j^\delta \beta_j \right\} \prod_{\psi=2}^{\Psi} \mathbf{1}(\beta_{j\psi} \leq \beta_{j,\psi-1}) \quad (12)$$

for nonincreasing functions (antitonic case), respectively, where  $c_1(\beta_j)$  is a normalizing function depending on  $\beta_j$ .

## 2.3 Extensions

Various extensions regarding the additive predictor (2) are possible. To account for unobserved heterogeneity between different groups or clusters of units, we may add an unstructured group-specific random effect. Suppose we are given a grouping variable that can take values in  $\{1, \dots, G\}$ . Then, we can extend (2) to

$$\eta_n = f_1(x_{n1}) + \cdots + f_p(x_{np}) + v_n' \gamma + f_{\text{random}}(g_n) \quad (13)$$

and assume

$$f_{\text{random}}(g) = b_g \sim N(0, \tau_b^2), \quad g = 1, \dots, G, \quad (14)$$

where  $f_{\text{random}}(g_n) = f_{\text{random}}(g)$  if observation  $n$  belongs to group  $g$ . Using the penalty matrix  $K^b = I$ , we can write (14) in the general form

$$p(b_g | \tau_b^2) \propto \exp \left\{ -\frac{1}{2} b_g' K^b b_g \right\}. \quad (15)$$

If we would presume a spatial correlation between groups, we may additionally introduce a spatial correlated GMRF. Further possible extensions are varying coefficient terms and interactions of covariates (see Lang and Brezger 2004; Brezger and Lang 2005). In the remainder, we focus on models with random effects.

### 3. MCMC INFERENCE

Let  $\alpha$  be the vector of all parameters to be estimated in the model. Bayesian inference is based on the posterior distribution

$$p(\alpha|y) \propto L(y, \beta_1, \dots, \beta_p, \gamma, b_g, \phi) \prod_{j=1}^p (p(\beta_j|\tau_j^2) p(\tau_j^2)) \times p(b_g|\tau_b^2) p(\tau_b^2) p(\gamma) p(\phi), \quad (16)$$

where  $L(\cdot)$  consists of the product of all individual likelihood contributions.  $\phi$  and  $p(\phi)$  have to be omitted for response distributions without a scale parameter. Because (16) is analytically intractable in all but the most simple cases, we employ MCMC techniques to obtain estimates for the parameters of interest. More specifically, we implement a block move, that is, we subsequently draw from the full conditionals  $p(\beta_j|\cdot)$ ,  $j = 1, \dots, p$ ,  $p(\gamma|\cdot)$ , and  $p(b_g|\cdot)$  of the blocks of parameters  $\beta_j$ ,  $j = 1, \dots, p$ ,  $\gamma$ , and  $b_g$ . For Gaussian responses, these blocks can be updated by block-move Gibbs sampling steps. In binary probit and cumulative probit models, we can rely on the same sampling scheme as building block; see Chen and Dey (2000) or Brezger and Lang (2005) for details. In all other cases, we use Metropolis–Hastings steps with IWLS proposals. The variance parameters  $\tau_1^2, \dots, \tau_p^2, \tau_b^2$  (and the scale parameter  $\sigma^2$  in the Gaussian case) are updated by single-move Gibbs sampling steps.

For posterior inference, we discard the draws from an initial burn-in period and take only every  $r$ th draw thereafter to minimize the autocorrelation of the samples. The formulas and algorithms in the following subsections are formulated with respect to isotonic constraints. The adjustments for antitonic constraints are straightforward.

#### 3.1 Gaussian Response

For Gaussian response, the posterior distribution for  $\beta_j$  is given by

$$p(\beta_j|\cdot) \propto c_1(\beta_j) \exp\{-.5(\beta_j - m_j)' P_j(\beta_j - m_j)\} \times \prod_{\psi=2}^{\Psi} \mathbf{1}(\beta_{j\psi} \geq \beta_{j,\psi-1}), \quad (17)$$

and

$$P_j = \frac{1}{\sigma^2} X_j' X_j + \frac{1}{\tau_j^2} K_j^\delta, \quad (18)$$

$$m_j = \frac{1}{\sigma^2} P_j^{-1} X_j' (y - \eta + X_j \beta_j^c),$$

where  $\beta_j^c$  is the current state of  $\beta_j$ .

To sample from this  $\Psi$ -dimensional truncated Gaussian distribution (17), we adopt the method of Robert (1995) and run an extra (short) single-move Gibbs sampler in each MCMC iteration. The algorithm is as follows:

(a) Set  $\beta^{(0)} = \beta_j^c$ .

(b) For  $t = 1, \dots, T$ , successively draw from the one-dimensional truncated Gaussian distributions

$$(1) \beta_1^{(t)} \sim N(\mu_1, \sigma_1^2, -\infty, \beta_2^{(t-1)})$$

$$(2) \beta_2^{(t)} \sim N(\mu_2, \sigma_2^2, \beta_1^{(t)}, \beta_3^{(t-1)})$$

$$(3) \beta_3^{(t)} \sim N(\mu_3, \sigma_3^2, \beta_2^{(t)}, \beta_4^{(t-1)})$$

$\vdots$

$$(\Psi) \beta_\Psi^{(t)} \sim N(\mu_\Psi, \sigma_\Psi^2, \beta_{\Psi-1}^{(t)}, \infty),$$

where  $N(\mu, \sigma^2, \mu_l, \mu_r)$  denotes a Gaussian distribution with mean  $\mu$ , variance  $\sigma^2$  and with left truncation point  $\mu_l$  and right truncation point  $\mu_r$ , respectively. The truncation points in the algorithm above are the current states of the adjacent parameters. Therefore, we have only right truncation for  $\beta_1^{(t)}$  and left truncation for  $\beta_\Psi^{(t)}$ . The parameters  $\mu_\psi$  and  $\sigma_\psi^2$ ,  $\psi = 1, \dots, \Psi$ , are the conditional means and variances of the (nontruncated) posterior (17):

$$\mu_\psi = \frac{1}{p_{\psi\psi}} \left\{ \sum_{\rho < \psi} (\beta_\rho^{(t)} - m_\rho) \cdot p_{\psi\rho} + \sum_{\rho > \psi} (\beta_\rho^{(t-1)} - m_\rho) \cdot p_{\psi\rho} \right\}, \quad (19)$$

$$\sigma_\psi^2 = \frac{1}{p_{\psi\psi}},$$

where  $m_\rho$  is the  $\rho$ th element of  $m_j$  and  $p_{\psi\rho}$  is the element in row  $\psi$  and column  $\rho$  of the precision matrix  $P_j$  in (17). Note that the subscript  $j$  is suppressed in the formulas above.

(c) Take  $\beta^{(T)} = (\beta_1^{(T)}, \dots, \beta_\Psi^{(T)})'$  as a random sample from (17).

Usually, convergence is reached after 10–20 cycles. To reach convergence with considerable certainty, we set  $T = 100$ . Computation is very fast, as the mean  $m_j$  and the precision matrix  $P_j$  have to be computed only once. Moreover,  $m_j$  is obtained by sparse matrix operations exploiting the band structure of  $P_j$ . This involves a Cholesky decomposition and avoids expensive matrix inversions (cf. Rue 2001).

Regarding the parametric effects, we obtain a normal distribution with precision matrix and mean

$$P_\gamma = \frac{1}{\sigma^2} V' V, \quad m_\gamma = (V' V)^{-1} V' (y - \eta + V \gamma) \quad (20)$$

as full conditional, if we use constant priors for the parameters  $\gamma$ . Of course, a highly dispersed Gaussian distribution can be assigned as a conjugate prior as well when a priori information is sparse/lacking.

The full conditionals for the variance parameters  $\tau_j^2$ ,  $j = 1, \dots, p$ ,  $\tau_b^2$ , and the scale parameter  $\sigma^2$  are all inverse Gamma distributions with parameters

$$a'_j = a_j + \frac{\text{rank}(K_j)}{2} \quad \text{and} \quad b'_j = b_j + \frac{1}{2} \beta_j' K_j \beta_j \quad (21)$$

for  $\tau_j^2$ ,  $j = 1, \dots, p$ ,  $\tau_b^2$ , and

$$a'_{\sigma^2} = a_{\sigma^2} + \frac{N}{2} \quad \text{and} \quad b'_{\sigma^2} = b_{\sigma^2} + \frac{1}{2} \epsilon' \epsilon \quad (22)$$

for  $\sigma^2$ , where  $\epsilon$  is the usual vector of residuals.

### 3.2 Non-Gaussian Response

For non-Gaussian response, the posterior  $p(\beta_j|\cdot)$  is

$$p(\beta_j|\cdot) \propto c(y_n, \phi) \exp\left\{\frac{y_n \theta_n - b(\theta_n)}{\phi}\right\} \\ \times c_1(\beta_j) \exp\left\{-.5 \frac{1}{\tau_j^2} \beta_j' K_j^\delta \beta_j\right\} \\ \times \prod_{\psi=2}^{\Psi} \mathbf{1}(\beta_{j\psi} \geq \beta_{j,\psi-1}), \quad (23)$$

which has no longer standard form [e.g., if we consider a Gamma distribution with parameters  $\mu$  and  $\nu$ , as in our empirical application, this implies  $b(\theta_n) = -\log(-\theta_n)$ ,  $c(y_n, \phi) = (\phi^{-1})^{(\phi^{-1})} y_n^{\phi^{-1}-1} / \Gamma(\phi^{-1})$ ,  $\theta_n = -1/\mu$ , and  $\phi = \nu^{-1}$ ]. Thus, we use a Metropolis–Hastings step with an IWLS proposal to update  $\beta_j$ . An IWLS proposal is obtained by a quadratic approximation of the likelihood via Taylor expansion around the current state  $\beta_j^c$  of  $\beta_j$  (cf. Brezger and Lang 2005 or, in the mixed model context, Gamerman 1997). This leads to a truncated multivariate Gaussian proposal

$$q(\beta_j) \propto c_1(\beta_j) \exp\left\{-.5(\beta_j - m(\beta_j^c))' P(\beta_j^c)(\beta_j - m(\beta_j^c))\right\} \\ \times \prod_{\psi=2}^{\Psi} \mathbf{1}(\beta_{j\psi} \geq \beta_{j,\psi-1}) \quad (24)$$

where

$$P(\beta_j^c) = X_j' W(\beta_j^c) X_j + \frac{1}{\tau_j^2} K_j^\delta, \\ m(\beta_j^c) = P(\beta_j^c)^{-1} X_j' W(\beta_j^c) \tilde{y}(\beta_j^c), \\ W(\beta_j^c) = \text{diag}(w_1, \dots, w_N), \\ \tilde{y}(\beta_j^c) = (y - \mu) g'(\mu) + X_j \beta_j^c, \quad (25)$$

with  $w_n^{-1} = b''(\theta_n) \{g'(\mu_n)\}^2$ . Alternatively, we could also use the current mode of  $p(\beta_j|\cdot)$  rather than  $\beta_j^c$  to perform the Taylor expansion, which would simplify the calculation of the acceptance probability; compare Brezger and Lang (2005). The main difference to the Gaussian sampling scheme is that this sample can only be accepted with probability

$$\alpha(\beta_j^c, \beta_j^p) = \min\left\{1, \frac{L(\beta_j^p) p(\beta_j^p) q(\beta_j^p, \beta_j^c)}{L(\beta_j^c) p(\beta_j^c) q(\beta_j^c, \beta_j^p)}\right\} \quad (26)$$

as the new state of  $\beta_j$ . Note that the indicator functions involved (in the prior and in the proposal) cancel out.

The full conditionals for the variance parameters  $\tau_j^2$ ,  $j = 1, \dots, p$ , and  $\tau_b^2$  are again inverse Gamma distributions and therefore updated via Gibbs sampling. Parametric effects are updated by Metropolis–Hastings steps.

## 4. EMPIRICAL APPLICATION: ESTIMATING PRICE RESPONSE FROM STORE-LEVEL SCANNER DATA

### 4.1 Background

It is important for both manufacturers and retailers to know how sales respond to price promotions. For example, if a brand's sales response to own-price cuts shows increasing returns to scale, a firm will run deeper price discounts for the brand than in case of decreasing returns to scale. In the following, we apply the monotonic regression approach to estimating promotional price response functions from store-level scanner data.

It is well documented that sales promotions, especially in the form of temporary price reductions, substantially increase sales of promoted brands (e.g., Wilkinson, Mason, and Paksoy 1982; Blattberg and Neslin 1990; Bemmaor and Mouchoux 1991; Blattberg, Briesch, and Fox 1995). There is also empirical evidence that a temporary price cut by a brand may decrease sales of competitive items significantly (e.g., Blattberg and Wisniewski 1989; Bemmaor and Mouchoux 1991; Mulherne and Leone 1991). Cross-item price effects, however, are usually much lower than own-item price effects; see Hanssens et al. (2001) for an overview of empirical findings. In addition, there is strong empirical support that cross-promotional effects are asymmetric, implying that promoting higher priced/higher quality brands generates more switching from lower priced/lower quality brands than does the reverse (e.g., see Blattberg and Wisniewski 1989; Allenby and Rossi 1991; Blattberg et al. 1995). This phenomenon has also become known as asymmetric quality tier competition (e.g., Sivakumar and Raj 1997). Moreover, a recent meta-analysis of cross-price elasticity estimates revealed strong neighborhood price effects, indicating that brands that are closer to each other in price have larger cross-price effects than brands priced farther apart (Sethuraman et al. 1999).

Despite the wealth of empirical findings on own- and cross-price effects, little was known about the shape of the promotional price response function until recently. Most studies addressing this issue employed strictly parametric functions, and came to different results from model comparisons. For example, Wisniewski and Blattberg (1983) found the own-item price effect curve to be modeled best by an S-shaped function, while Blattberg and Wisniewski (1987) found the curve to show increasing returns with deeper price discounts. The former, however, estimated own-price response functions at the category level rather than the individual brand level, while the latter analyzed a limited range of price discounts. Today, multiplicative (log–log), exponential (semilog), and log-reciprocal functional forms are the most widely used parametric specifications to represent nonlinearities in sales response to promotional instruments (e.g., see Blattberg and Wisniewski 1989; Blattberg and George 1991; Montgomery 1997; Foekens, Leeflang, and Wittink 1999; Kopalle, Mela, and Marsh 1999; van Heerde, Leeflang, and Wittink 2002). These functional forms are inherently monotonic (decreasing for own-price and increasing for cross-price effects) and all use a logarithmic transformation of brand sales to normalize the distribution of the dependent variable, which typically is markedly skewed with promotional data (e.g., Mulherne and Leone 1991). However, there does not

seem to exist a “best” parametric functional form generalizable across product categories or even across brands within a category. Therefore, nonparametric regression methods seem to be highly promising to explore the shape of the promotional price response curve more flexibly.

van Heerde et al. (2001) proposed a kernel-based semiparametric approach in which a brand’s unit sales are modeled as a nonparametric function of own- and cross-item price variables and a parametric function of other predictors. The model can also accommodate flexible interaction effects between price cuts of different brands but may suffer from the curse of dimensionality as the number of competing items increases. van Heerde et al. (2001) obtained superior performance for the semiparametric model in both fit and predictive validity relative to two benchmark parametric models. Their results based on store-level scanner data for three product categories (tuna, beverage, and a third packaged food product) indicate threshold and/or saturation effects for both own- and cross-item price cuts. Threshold effects are present if consumers do not change their purchase intentions unless a promotional price cut exceeds a certain threshold level, say, for example, 15% (Gupta and Cooper 1992). A common argument for the existence of saturation effects is based on the belief that consumers can stockpile and/or consume only limited amounts of a promoted good, for example, due to inventory constraints or perishability (Blattberg et al. 1995; van Heerde et al. 2001). About two-thirds of the nonparametric own-item and cross-item price response curves estimated by van Heerde et al. (2001) showed a (reverse) S-shape reflecting both threshold and saturation effects, with a wide range of different saturation points across brands. Some curves revealed a (reverse) L-shape with a strong kink at a certain level of price cut, while other curves showed neither a threshold nor a saturation effect. These different results across individual brands strongly support the use of nonparametric estimators to let the data determine the shape of price response functions. However, two own-item price response curves indicated a decrease in unit sales as price discounts became very deep. Clearly, this nonmonotonicity is difficult to interpret from an economic point of view. One explanation may be that consumers associate a loss in quality with very deep price cuts, but this argument seems at least questionable with frequently purchased consumer nondurables.

In contrast to van Heerde et al. (2001), Kalyanam and Shively (1998) proposed a stochastic spline regression approach (Wahba 1978) in the context of a hierarchical Bayes model (Wong and Kohn 1996) and found much stronger irregularities in own-price response for some of the brands (tuna, margarine) examined. Especially, although overall downward sloping, the respective curves showed local upturns and downturns with spikes at certain price points resulting in less smooth and nonmonotonic shapes. Kalyanam and Shively (1998) illustrated that these nonmonotonicities may be associated with odd pricing or a complex convolution of odd pricing with other effects like, for example, the existence of segments with distinct reservation prices. Odd pricing refers to the practice of setting prices ending in odd numbers or just below a round number (e.g., .99 cents instead of 1.00 dollar). On the other hand, the curve plots also revealed that the estimates at the very strongest local sales peaks were based only on one or a few data points (see, e.g., the

results for the Starkist brand in Kalyanam and Shively 1998, p. 26). Kalyanam and Shively (1998) themselves pointed out that in case of an insufficient number of data points, the estimated functions may show irregularities where none exist. This problem also applies to another tuna brand (Bumble Bee) where the estimated curve indicated a (monotonic) increase in unit sales with increasing own-price beyond a certain price point (i.e., for higher price levels). This latter irregularity is not in accordance with economic theory and, as a consequence, would suggest an optimal price at infinity.

Besides the problem of inaccurate estimation due to sparse data in some cases, the findings of Kalyanam and Shively (1998) coincided with those of van Heerde et al. (2001) with respect to the existence of threshold effects for several brands, that is, flat own-price response around prices at the upper bound of the range of observed prices. In comparison to a parametric semilog specification, Kalyanam and Shively (1998) obtained a superior fit of their spline model in terms of adjusted  $R^2$  values for each of the brands analyzed. Unfortunately, no model validation results were reported.

The monotonic nonparametric regression approach as proposed in this article is our answer to resolve the problem whether nonmonotonic effects indeed exist when theory and/or empirical experience would rather suggest not. Our perspective is that an unconstrained estimation allowing for nonmonotonicities should be preferred only if it outperforms a constrained estimation in validation samples. Otherwise, nonmonotonic effects are likely to represent an artifact caused by sparse data or merely by too much flexibility of the nonparametric estimator. Importantly, imposing monotonicity constraints does not preclude the estimation of irregular pricing effects like steps and kinks at certain price points or threshold and saturation effects at the extremes of the observed price/price cut ranges.

## 4.2 An Illustration

For illustration, we use weekly store-level scanner data from Dominick’s Finer Foods, a major supermarket chain in the Chicago metropolitan area. The dataset includes unit sales, retail price, and a deal code indicating the use of an in-store display for eight brands of refrigerated orange juice (64 oz). The sample covers individual brand sales in 81 stores ( $s = 1, \dots, 81$ ) of the chain over a time span of 89 weeks ( $t = 1, \dots, 89$ ). Table 1 provides summary statistics pooled across the stores for average weekly prices, market shares, and unit sales of the brands.

As Table 1 reveals, the brands can be classified into three price-quality tiers: the premium brands (made from freshly squeezed oranges), the national brands (reconstituted from frozen orange juice concentrate), and the store brand (Dominick’s private label brand). The differences in quality across the tiers are well represented by higher (lower) average prices for higher (lower) quality tier brands. Average weekly prices and market shares of all brands vary considerably, reflecting the frequent use of promotions.

We now illustrate the usefulness of imposing monotonicity constraints to estimate price response functions considering as example the brand Florida Gold. We focus on two distributional models, namely a lognormal model

$$sales_{st} \sim \text{LN}(\eta_{st}, \sigma^2), \quad (27)$$

Table 1. Descriptive statistics for weekly brand prices, market shares, and unit sales. Refrigerated orange juice category (64 oz)

| Brand                   | Retail price |           |              | Market share |          |             | Unit sales |         |
|-------------------------|--------------|-----------|--------------|--------------|----------|-------------|------------|---------|
|                         | Range (\$)   | Mean (\$) | Std Dev (\$) | Range (%)    | Mean (%) | Std Dev (%) | Minimum    | Maximum |
| <i>Premium brands:</i>  |              |           |              |              |          |             |            |         |
| Tropicana Pure Premium  | [1.60; 3.55] | 2.95      | .53          | [3; 73]      | 15       | 15          | 6,388      | 100,712 |
| Florida Natural Premium | [1.57; 3.16] | 2.86      | .33          | [1; 53]      | 5        | 7           | 1,138      | 56,037  |
| <i>National brands:</i> |              |           |              |              |          |             |            |         |
| Citrus Hill             | [1.09; 2.82] | 2.31      | .31          | [1; 78]      | 8        | 12          | 2,006      | 151,570 |
| Minute Maid             | [1.29; 2.92] | 2.23      | .40          | [3; 87]      | 21       | 22          | 4,805      | 243,711 |
| Tropicana               | [1.49; 2.75] | 2.20      | .35          | [2; 75]      | 21       | 23          | 3,041      | 102,629 |
| Florida Gold            | [.99; 2.83]  | 2.17      | .39          | [1; 63]      | 4        | 8           | 325        | 150,945 |
| Tree Fresh              | [1.07; 2.48] | 2.15      | .27          | [1; 42]      | 4        | 6           | 916        | 39,401  |
| <i>Store brand:</i>     |              |           |              |              |          |             |            |         |
| Dominick's              | [.99; 2.47]  | 1.75      | .4           | [1; 83]      | 22       | 22          | 2,170      | 189,462 |

which can be equivalently written in terms of the assumption of a Gaussian distribution for the natural logarithm of the response as

$$\log(sales_{st}) \sim N(\eta_{st}, \sigma^2), \quad (28)$$

and a Gamma model

$$sales_{st} \sim G(\exp(\eta_{st}), \nu), \quad (29)$$

where  $sales_{st}$  denotes the unit sales of Florida Gold in store  $s$  and week  $t$ . Note that the exponential function is the so-called natural link function for a Gamma model. The scale parameter  $\nu$  is supplied with a Gamma prior with parameters  $a_\nu = .001$ ,  $b_\nu = .001$  and estimated in a Metropolis–Hastings step.

As mentioned above, the use of a lognormal model is the standard approach in marketing to relate brand sales to promotional instruments. The Gamma model, on the other hand, provides high flexibility with respect to the shape of the distribution (e.g., it can take on a highly skewed distribution) and is used to demonstrate the applicability of our method in the non-Gaussian case. Like Kalyanam and Shively (1998) and van Heerde et al. (2001), we choose a semiparametric additive predictor to model sales response, with nonparametric terms for own- and cross-price effects as well as weekly effects, and parametric terms for own and competitive display and store-specific effects. According to economic theory and the empirical findings discussed in Section 3.1, we expect the unit sales of Florida Gold to be an antitonic function in own promotional price and an isotonic function in competitive items' promotional prices rather than to show a nonmonotonic shape, respectively. Specifically, we estimate three variants of the semiparametric additive predictor for both the lognormal and the Gamma model:

$$\begin{aligned} \eta_{st}^{(1)} = & f_{1\_antitonic}^{RW1}(price_{st}) + f_{2\_isotonic}^{RW1}(price\_premium_{st}) \\ & + f_{3\_isotonic}^{RW1}(price\_national_{st}) \\ & + f_{4\_isotonic}^{RW1}(price\_Dominicks_{st}) + f_5^{RW2}(week) \\ & + f_{random}(store) + \gamma_1 display_{st} + \gamma_2 display\_premium_{st} \\ & + \gamma_3 display\_national_{st} + \gamma_4 display\_Dominicks_{st}, \quad (30) \end{aligned}$$

$$\eta_{st}^{(2)} = f_{1\_antitonic}^{RW2}(price_{st}) + f_{2\_isotonic}^{RW2}(price\_premium_{st})$$

$$\begin{aligned} & + f_{3\_isotonic}^{RW2}(price\_national_{st}) \\ & + f_{4\_isotonic}^{RW2}(price\_Dominicks_{st}) + f_5^{RW2}(week) \\ & + f_{random}(store) + \gamma_1 display_{st} + \gamma_2 display\_premium_{st} \\ & + \gamma_3 display\_national_{st} + \gamma_4 display\_Dominicks_{st}, \quad (31) \end{aligned}$$

and

$$\begin{aligned} \eta_{st}^{(3)} = & f_1^{RW2}(price_{st}) + f_2^{RW2}(price\_premium_{st}) \\ & + f_3^{RW2}(price\_national_{st}) \\ & + f_4^{RW2}(price\_Dominicks_{st}) + f_5^{RW2}(week) \\ & + f_{random}(store) + \gamma_1 display_{st} + \gamma_2 display\_premium_{st} \\ & + \gamma_3 display\_national_{st} + \gamma_4 display\_Dominicks_{st}. \quad (32) \end{aligned}$$

The three variants differ in the specification of the unknown smooth functions  $f_1$ – $f_4$  for own- and cross-price effects. These are estimated either by P-splines with monotonicity constraints, with first-order random walk prior ( $\eta^{(1)}$ ) or second-order random walk prior ( $\eta^{(2)}$ ), respectively, or by unconstrained P-splines with second-order random walk prior ( $\eta^{(3)}$ ) as a reference. The choice of the reference specification is based on a study conducted by Lang and Brezger (2004) who reported superior results for P-splines with second-order rather than first-order random walk priors in the unrestricted case. *price* denotes Florida Gold's actual price in store  $s$  and week  $t$ , and *display* is an indicator variable representing the usage (1) or nonusage (0) of an in-store display for Florida Gold in store  $s$  and week  $t$ . Similarly to Blattberg and George (1991), we capture cross-price effects in a more parsimonious way through the use of competitive variables at the tier level rather than the individual brand level: *price\_premium<sub>st</sub>* and *price\_national<sub>st</sub>* indicate the minimum price for competing brands within the premium brand and the national brand tier in store  $s$  and week  $t$ , respectively, whereas *price\_Dominicks<sub>st</sub>* is the actual price of Dominick's private label brand in store  $s$  and week  $t$ . It is important to note that the price of Florida Gold (which itself is a national brand) is excluded from computing *price\_national<sub>st</sub>*. Accordingly, the indicator variables *display\_premium<sub>st</sub>* and *display\_national<sub>st</sub>* take the value '1' if a display is used for at least one brand



within the respective tier in store  $s$  and week  $t$ , and ‘0’ otherwise.  $display\_Dominicks_{st}$  is the display dummy for the private label brand indicating if Dominick’s supports its own brand with a display or not in store  $s$  and week  $t$ .  $\gamma_1$  to  $\gamma_4$  denote the corresponding display effects to be estimated.

The *week* covariate is incorporated to capture seasonal and missing variable (e.g., manufacturer advertising) effects, and the *store* covariate to accommodate differences in base sales of Florida Gold across the stores, for example, due to their spatial location. The effect of *week* is modeled as a P-spline with second-order random walk prior and *store* is incorporated as a random effect. We use cubic splines with 20 knots for all P-spline terms, except for the *week* effect, where we use 40 knots to be able to account for possibly strong time variability. The specification with 40 knots for the time effect, however, is still much less costly in terms of degrees of freedom lost than if we were to use weekly indicator variables. Finally, the hyperparameters  $\sigma^2$  and  $\nu$  are supplied with inverse Gamma priors  $\sigma^2 \sim \text{IG}(.001, .001)$  and  $\nu \sim \text{IG}(.001, .001)$ , respectively, and are estimated simultaneously with the regression parameters. The resulting models are referred to as LN1–LN3 for the lognormal variants and G1–G3 for the Gamma model variants in the following. With regard to the sampling process, we store every 10th sample of a Markov chain of length 10,000 (after the burn-in period) to obtain 1,000 draws for each parameter and take the means as parameter estimates.

#### 4.3 Model Evaluation and Interpretation of Results

We evaluate the different models in terms of the average mean squared error (AMSE) in validation samples (also cf. van Heerde et al. 2001). Specifically, we randomly split the data into nine equally sized subsets and performed ninefold cross-validation. For each subset, we fitted the respective model to the remaining eight subsets making up the estimation sample and calculated the squared prediction errors of the fitted model when applied to the observations in this holdout subset (Efron and Tibshirani 1998). Let  $N$  denote the number of observations of the entire dataset, and  $k(n)$  the holdout subset containing observation  $n$ . Let further  $\widehat{sales}_n^{-k(n)}$  indicate the fitted value of observation  $n$  computed from the estimation sample without subset  $k(n)$ ; then the AMSE of prediction is

$$\text{AMSE} = \frac{1}{N} \sum_{n=1}^N (\text{sales}_n - \widehat{sales}_n^{-k(n)})^2. \quad (33)$$

Because we are interested in unit sales rather than log unit sales of Florida Gold, conditional mean predictions from the estimated lognormal models were obtained as follows (Goldberger 1968; Greene 1997):

$$\widehat{sales}_n^{-k(n)} = \frac{1}{1,000} \sum_{i=1}^{1,000} \exp \left\{ \eta_{ni} + \frac{\sigma_i^2}{2} \right\}, \quad (34)$$

where  $\eta_{ni}$  is the additive predictor for observation  $n$  and stored iteration  $i$ , and  $\sigma_i^2$  denotes the residual variance of the respective lognormal model in iteration  $i$ . For the Gamma model, no correction factor  $\sigma_i^2/2$  is required for the conditional mean predictions.

Table 2. Evaluation of models in terms of AMSE

| Model specification                | Lognormal | Gamma  |
|------------------------------------|-----------|--------|
| $\eta_{st}^{(1)}$ (restricted/RW1) | 49,931    | 49,577 |
| $\eta_{st}^{(2)}$ (restricted/RW2) | 50,046    | 49,370 |
| $\eta_{st}^{(3)}$ (unrestricted)   | 50,800    | 52,200 |

The validation results are displayed in Table 2. Under both the lognormal and the Gamma distribution, the models with monotonicity constraints (LN1, LN2, G1, G2) clearly outperform the respective model without monotonicity constraints (LN3, G3). Interestingly, whereas in the unrestricted case the lognormal model (LN3) yields a smaller AMSE compared to the Gamma model (G3), the restricted Gamma models G1 and G2 provide the highest predictive validity. Furthermore, the differences between restricted models with first-order and second-order random walk priors for the nonparametric terms are virtually negligible. These results indicate that imposing monotonicity constraints on own- and cross-item price effects can substantially improve the predictive validity of a sales response model.

Figures 2 and 3 show the nonparametrically estimated own- and cross-price effects for Florida Gold resulting from the lognormal models (LN1–LN3) and the Gamma models (G1–G3), respectively. Shown are the posterior means as well as 80% and 95% pointwise credible intervals. To ensure identifiability, the functions are centered to have mean zero in every iteration of the MCMC procedure, that is,  $1/\text{range}(x_j) \int f_j(x_j) dx_j = 0$ . The subtracted means are added to an intercept term, which is not displayed here. It is important to note that, because to this centering, it is not necessary to explicitly restrict  $\beta_{1\psi}$  to be negative and  $\beta_{21}, \beta_{31}, \beta_{41}$  to be positive to obtain monotonicity for the own- and cross-item price response curves. As can be seen, the effects are very similar for corresponding model versions (LN1|G1, LN2|G2, and LN3|G3), except for the own-price effect which reveals a stronger increase in unit sales for very low prices under the Gamma distribution. Probably, this difference in own-price response is responsible for the higher predictive validity of the Gamma models. As already indicated by the AMSE values, there is also not much difference in own- and cross-price effects between the restricted Gamma models G1 and G2. We therefore focus in the following on Gamma model G2, the model with the highest predictive validity, for interpretation of results. Importantly, the unrestricted models LN3 and G3, which are inferior in predictive validity, show strong local nonmonotonicities in both own- and cross-price effects, which indicates too much flexibility (strong overfitting) of an unconstrained estimation.

Our results are similar to the findings of van Heerde et al. (2001) with respect to the shape of price response functions. Specifically, the own-price response curve for Florida Gold shows a reverse S-shape with an additional increase in sales for extremely low prices. This strong sales spike can be attributed to an odd pricing effect at 99 cents, the lowest observed price of Florida Gold (cf. Table 1). The cross-price response curve with respect to the premium tier brands reveals a reverse L-shape and a threshold effect for competitive prices over two dollars. In other words, only if one of the premium brands is priced lower than two dollars, unit sales of Florida Gold significantly

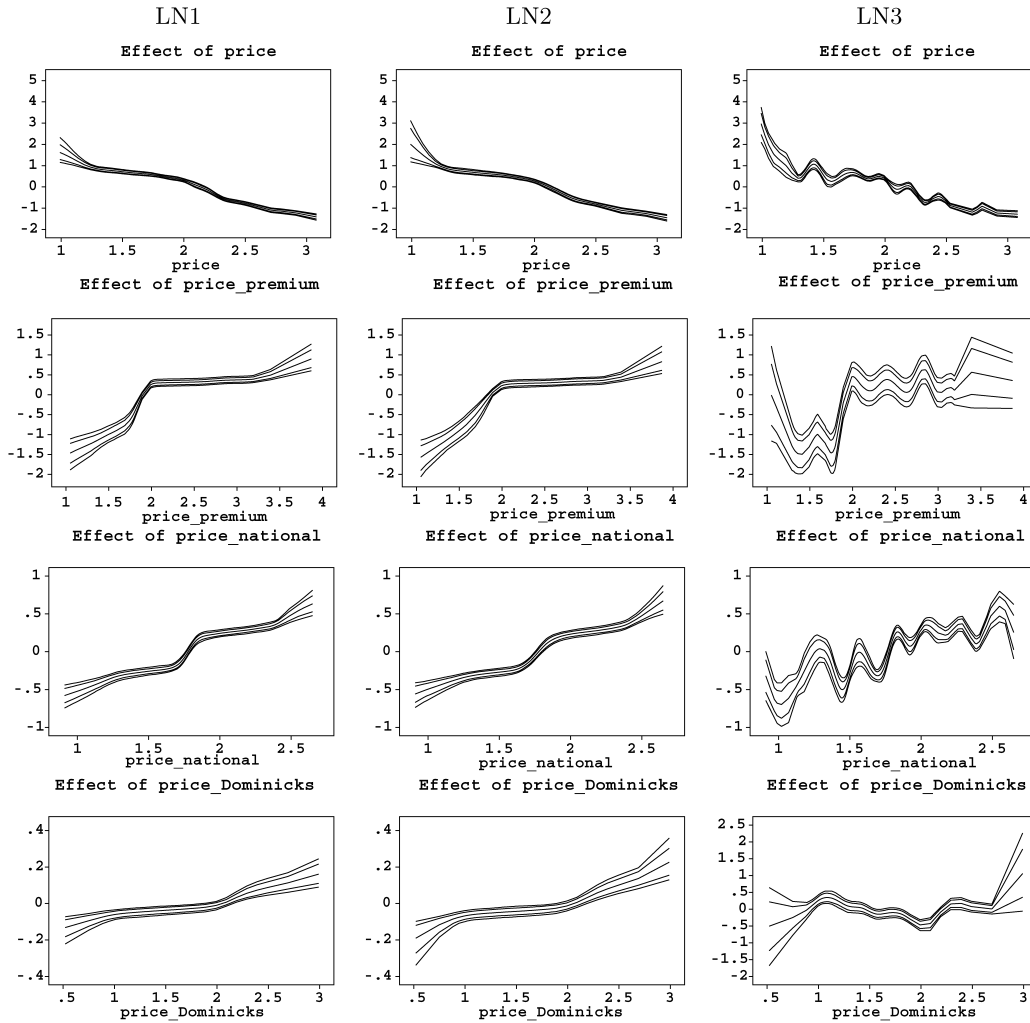


Figure 2. Estimated curves for own-price (*price*) and tier-specific cross-price (*price\_premium*, *price\_national*, *price\_Dominicks*) effects on unit sales of Florida Gold. Columns 1–3 show the effects for the models LN1–LN3. Shown are the posterior means as well as 80% and 95% pointwise credible intervals.

decrease and consumers switch up to the low-priced premium brand. The cross-price effect with respect to the national brand tier (the tier of Florida Gold) is S-shaped but by far less strong than the premium tier effect, which contradicts the hypothesis that brands which are priced closer to each other (like Florida Gold and the other national brands) are more competitive than brands priced farther apart (like Florida Gold and the premium brands). Finally, the cross-price effect of Dominick's private label brand on Florida Gold's sales is almost negligible. Comparing the three cross-price effects in magnitude, our results confirm previous empirical findings of asymmetric quality tier competition. Specifically, a price cut by a premium brand may draw substantial sales from Florida Gold, whereas a price cut by a private label brand does not. As expected, the own-price effect is much stronger than each of the cross-price effects.

Tables 3 and 4 provide parameter estimates for the display effects and the corresponding multiplier effects (Leeftang, Witink, Wedel, and Naert 2000). The multiplier effects are obtained from the transformation

$$\frac{1}{1,000} \sum_{i=1}^{1,000} \exp\{\gamma_{ji}\}, \quad j = 1, \dots, 4. \quad (35)$$

Shown are the posterior means, posterior standard deviations, and the corresponding 2.5% and 97.5% quantiles, respectively. Multipliers with values larger (smaller) than 1 indicate a positive (negative) effect on unit sales of Florida Gold.  $\gamma_{1i}$  denotes the own display effect of Florida Gold, and  $\gamma_{2i}$  to  $\gamma_{4i}$  refer to the tier-specific competitive display effects.  $i$  denotes the  $i$ th stored sample for the respective parameter. Except for the cross-display effect of Dominick's private label brand, the display multipliers show the expected impact. For example, if a display is used for Florida Gold, its unit sales increase on average by a factor of 1.36, whereas a display for a premium brand causes a decrease in Florida Gold's unit sales of about 11% on average. The display effect with respect to the brands in the national tier (except Florida Gold) is not significant. One possible explanation for the positive cross display effect of Dominick's private label could be that promotion activities of Dominick's for its own store brand are especially distinct and stimulate not only own brand sales but also sales of some other brands in the category. As expected, the own display effect is much stronger than competitive display effects.

Finally, Figure 4 shows estimated results for the store-specific random effect. The store effect is portrayed with a spa-

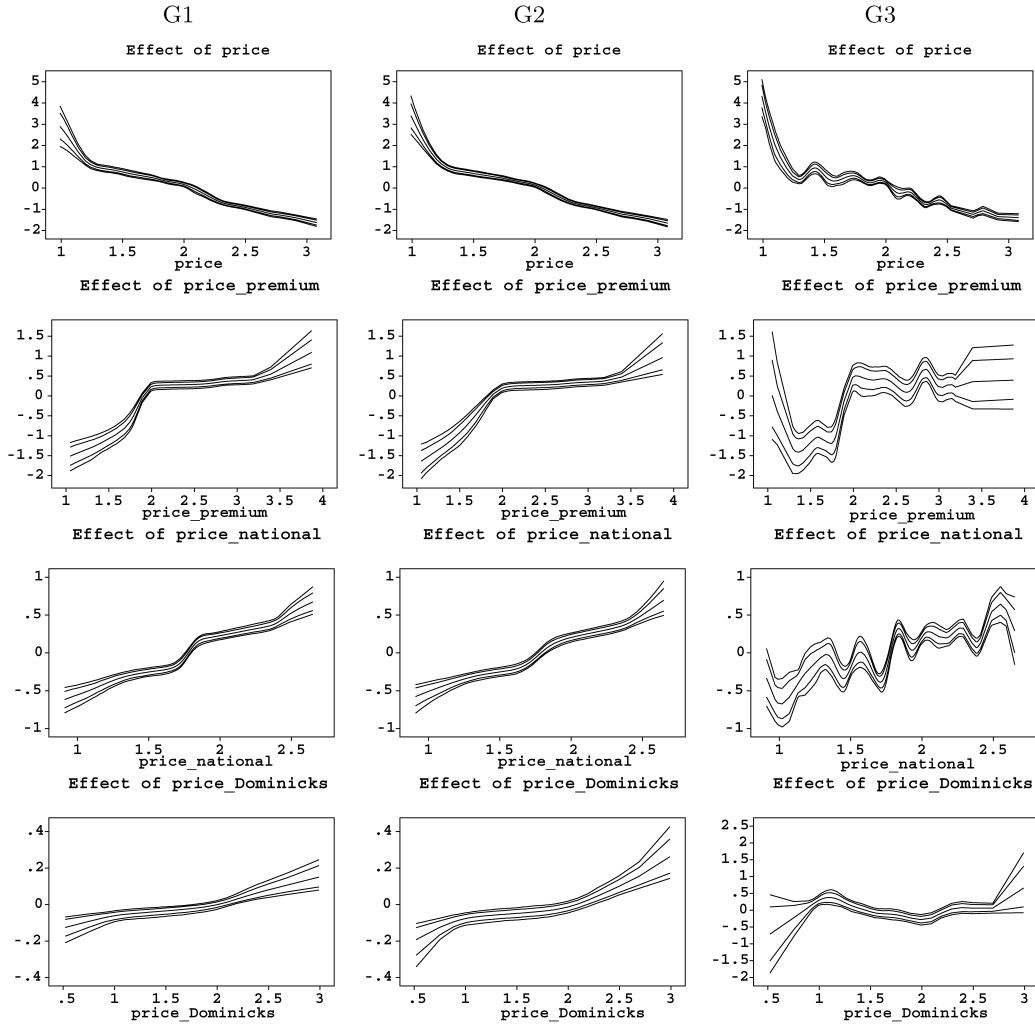


Figure 3. Estimated curves for own-price (*price*) and tier-specific cross-price (*price\_premium*, *price\_national*, *price\_Dominicks*) effects on unit sales of Florida Gold. Columns 1–3 show the effects for the models G1–G3. Shown are the posterior means as well as 80% and 95% pointwise credible intervals.

tial map that represents the store locations of Dominick’s Finer Foods in the Chicago metropolitan area. There is a noticeable difference in base sales across stores, with an apparent drop from the coastline in the east, where we have a high concentration of stores, to the interior region in the west. We found (weak) positive correlations between the store effect and the percentage of the population under age 9 (.28) and the percentage of households with three or more members (.24). Hence, one possible explanation for the east-west drop of base sales may be that more households with little children live in the east part of the Chicago area, and people buy more orange juice

there because they are concerned with their children’s health. We abstain from depicting the estimated effect for the time covariate *week*, because it does not reveal any seasonal pattern nor a trend.

## 5. REFINEMENTS

In this section we present some refinements of our methodology. The refinements are illustrated for the lognormal model, which is the standard approach in marketing to relate brand sales to price and promotional instruments (cf. Sec. 4.2). First,

Table 3. Estimation results (posterior means, standard deviations, and 95% credible intervals) for the display effects (Model G2)

| Effect                         | Posterior mean | 2.5%-quantile | 97.5%-quantile |
|--------------------------------|----------------|---------------|----------------|
| $\gamma_1$ (display)           | .30 (.04)      | .24           | .38            |
| $\gamma_2$ (display_premium)   | −.12 (.04)     | −.19          | −.05           |
| $\gamma_3$ (display_national)  | −.02 (.05)     | −.11          | .08            |
| $\gamma_4$ (display_Dominicks) | .07 (.03)      | .00           | .14            |

Table 4. Estimation results (posterior means, standard deviations, and 95% credible intervals) for the display multiplier effects (Model G2)

| Effect                         | Posterior mean | 2.5%-quantile | 97.5%-quantile |
|--------------------------------|----------------|---------------|----------------|
| $\gamma_1$ (display)           | 1.36 (.05)     | 1.27          | 1.45           |
| $\gamma_2$ (display_premium)   | .89 (.03)      | .83           | .95            |
| $\gamma_3$ (display_national)  | .98 (.05)      | .90           | 1.08           |
| $\gamma_4$ (display_Dominicks) | 1.07 (.04)     | 1.00          | 1.15           |

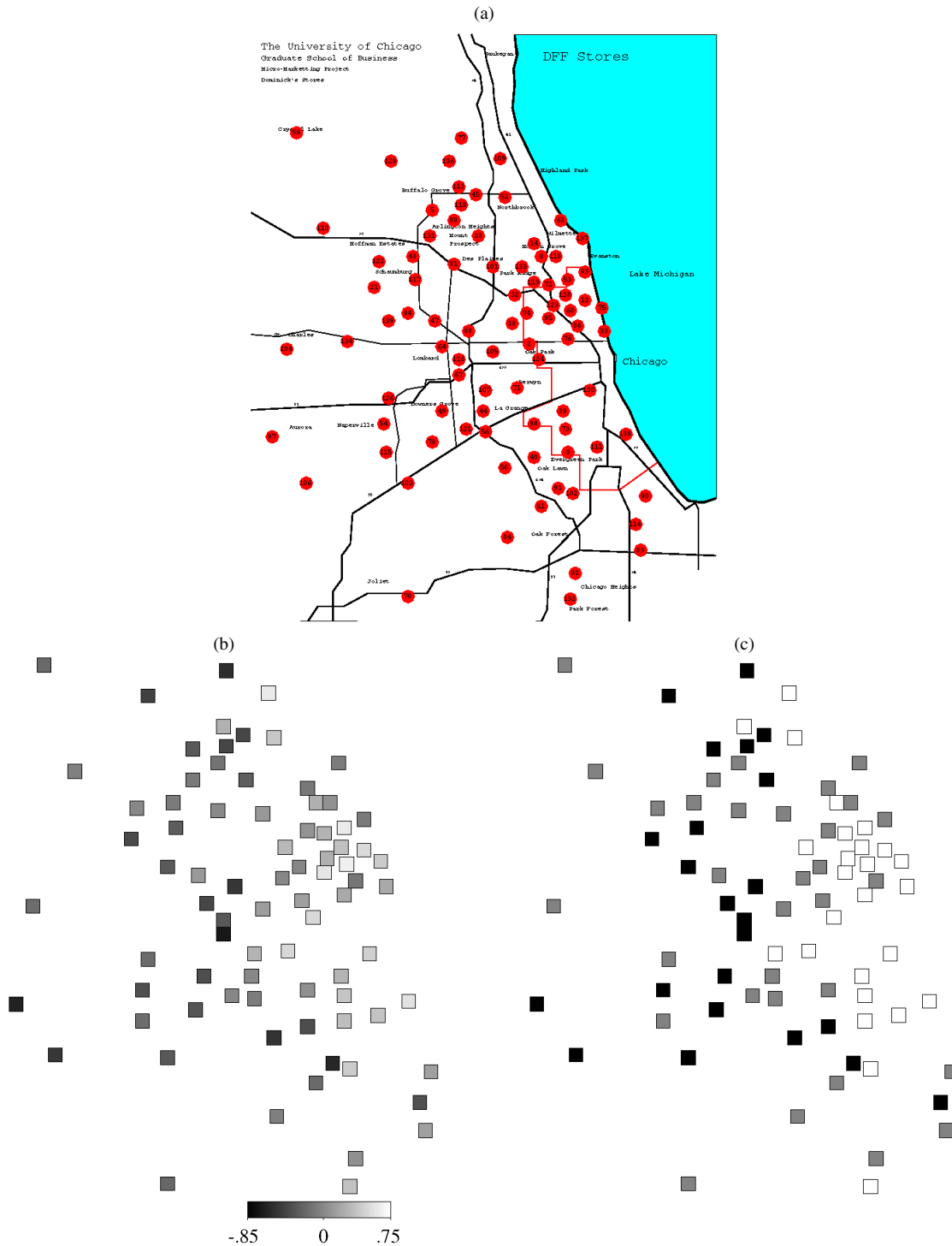


Figure 4. (a) Map of the Chicago metropolitan area with store locations of Dominick's Finer Foods. (b) Estimated random effect of *store* for the Gamma model (G2). (c) Posterior probabilities of *store*. White (black) indicates strictly positive (negative) 95% credible intervals, gray indicates that the 95% credible intervals contain zero.

we provide a relaxation of our procedure to enforce monotonicity on P-splines to encompass a more general class of monotone P-splines. Second, we perform a sensitivity analysis on the number and location of the knots with respect to the nonparametric price terms in our empirical application. Third, we apply our methodology to another brand (Dominick's store brand) to validate our results obtained for Florida Gold. We also compare the predictive power of our semiparametric models to the log-

linear (or exponential) model, which is one of the most widely used parametric specifications to estimate price response.

### 5.1 Generalization of Conditions for Monotonicity

The conditions (10) we suggested in Section 2.2 for imposing monotonicity constraints on nonparametric terms are sufficient, but not necessary. We employed these conditions in each

iteration of our sampler by admitting only posterior draws for a spline coefficient  $\beta_{\psi}^{(t)}$  that fall inside of the interval bounded by the current states of the adjacent spline coefficients  $\beta_{\psi-1}^{(t)}$  and  $\beta_{\psi+1}^{(t-1)}$  (cf. Sec. 3.1). This updating mechanism guarantees that the first derivative of the resulting spline is either positive or negative over its entire domain; hence the estimated curve becomes either isotonic or antitonic. There may be, however, values outside of the interval  $[\beta_{\psi-1}^{(t)}; \beta_{\psi+1}^{(t-1)}]$  that would not have destroyed the monotonicity of the curve (if admitted). In other words, excluding these values is sufficient, but not necessary, for monotonicity. As a consequence, with conditions (10), we covered only the class of monotone P-splines that can be represented with an ordered set of basis function coefficients  $\beta_{j1}, \beta_{j2}, \dots, \beta_{j\Psi}$ .

To get a sense of how restrictive our class of monotone P-splines is, we relaxed our condition to encompass the class of all monotone P-splines. This can be accomplished by admitting a wider range of posterior draws for  $\beta_{\psi}^{(t)}$ ,  $\psi = 1, \dots, \Psi$ ,  $t = 1, \dots, T$ , and rejecting them if and only if we obtain a wrong sign for the derivative of the spline somewhere. For *cubic* P-splines, as considered in this article, checking this new condition is simplified by utilizing the properties that (a) each basis function coefficient impacts function values over a domain of only  $l + 1 = 4$  interknot regions (cf. Fig. 1), (b) the first derivative is continuous, and (c) the second derivative is a piecewise linear curve. Implementation of the new rejection rule is carried out by verifying that the first derivative is positive (negative) at the borders of the four regions and also at the minimum (maximum) of these regions. (We thank an anonymous referee for her/his valuable comments on this topic.)

Based on this relaxation, we estimated two new variants of our semiparametric additive predictor: with first-order random walk prior ( $\eta^{(4)}$ ) and with second-order random walk prior ( $\eta^{(5)}$ ). We included both new variants in the sensitivity analysis whose results are presented below. The new variants are referred to as *restricted/RW1/R* and *restricted/RW2/R* in the following, with *R* indicating the relaxation.

## 5.2 Sensitivity Analysis

To assess how strongly the predictive performance of our models is influenced by the number and location of knots used for the nonparametric price terms, we performed a sensitivity analysis. We estimated the five models  $\eta^{(1)}$ – $\eta^{(5)}$  with (a) 10, 20, 30, and 40 equally spaced (equidistant) knots, (b) 20 and

30 knots positioned at the quantiles of the respective price covariates, as well as (c) knots located at intervals of approximately 10 cents within the price ranges. While alternatives (a) and (b) assume an equal number of knots across own- and cross-item price terms (implying unequal distances between each two knots for differently sized price ranges), alternative (c) instead leads to a different number of knots for price ranges of different span (but to equal distances between adjacent knots).

Table 5 shows the AMSE results of the sensitivity analysis, which can be summarized as follows: Independent of the number and location of knots, the restricted models ( $\eta^{(1)}$ ,  $\eta^{(2)}$ ,  $\eta^{(4)}$ ,  $\eta^{(5)}$ ) are always superior in predictive validity to the respective unrestricted model. Specifically, the improvement in AMSE values from imposing monotonicity constraints is considerable for all four restricted model variants and by far greater than the corresponding differences in AMSE values between the restricted models, respectively. There is a tendency that models with first-order random walk prior ( $\eta^{(1)}$  and  $\eta^{(2)}$ ) perform better than models with second-order random walk prior ( $\eta^{(4)}$  and  $\eta^{(5)}$ ). Furthermore, although the lowest AMSE value across all knot scenarios is associated with model  $\eta^{(4)}$  (48726; see column *30 knots quant.*), models  $\eta^{(4)}$  and  $\eta^{(5)}$  perform worse than their counterparts  $\eta^{(1)}$  and  $\eta^{(2)}$  in almost all other cases, respectively. This is rather unexpected, since  $\eta^{(1)}$  and  $\eta^{(2)}$ , representing the class of monotone P-splines constructable from an ordered set of basis function coefficients, occur as a special case of models  $\eta^{(4)}$  and  $\eta^{(5)}$  representing the class of all monotone P-splines. We therefore conclude that the conditions (10) for imposing monotonicity are not very restrictive, while at the same time requiring much more computational effort (i.e., CPU time).

There is also a tendency that the predictive performance of the restricted models slightly increases with more knots. Importantly, this latter finding did not hold for the Gamma models (neither for the restricted variants nor for the unrestricted model) for which we also conducted a sensitivity analysis. We again, however, obtained a considerable improvement in predictive performance for all Gamma models with monotonicity constraints compared to the unrestricted Gamma model under each of the seven knot scenarios. Based on these insights, the prediction improvement of the restricted models can be attributed to the monotonicity constraint rather than to the number and location of the knots.

Summarizing the main findings from our sensitivity analysis, there is evidence that the monotonicity constraint is the crucial part for the prediction improvement, rather than the number and location of the knots. Of course, more research with respect to

Table 5. Evaluation of lognormal models in terms of AMSE for Florida Gold

|                                      | 10 knots<br>equidist. | 20 knots<br>equidist. | 30 knots<br>equidist. | 40 knots<br>equidist. | 20 knots<br>quant. | 30 knots<br>quant. | 10 cent<br>equidist. |
|--------------------------------------|-----------------------|-----------------------|-----------------------|-----------------------|--------------------|--------------------|----------------------|
| $\eta_{st}^{(3)}$ (unrestricted)     | 55,717                | 50,800                | 50,202                | 52,514                | 52,888             | 51,182             | 53,357               |
| $\eta_{st}^{(1)}$ (restricted/RW1)   | 50,123                | 49,931                | 49,710                | 49,151                | 49,132             | 49,126             | 49,950               |
| $\eta_{st}^{(2)}$ (restricted/RW2)   | 50,042                | 50,046                | 49,877                | 49,726                | 48,996             | 49,130             | 49,294               |
| $\eta_{st}^{(4)}$ (restricted/RW1/R) | 50,255                | 50,058                | 49,600                | 49,247                | 49,369             | 48,726             | 50,058               |
| $\eta_{st}^{(5)}$ (restricted/RW2/R) | 50,246                | 50,234                | 50,039                | 49,994                | 49,493             | 48,970             | 50,327               |

this point could be done, for example, by using a still larger number of knots or by estimating the *optimal* number and location of knots. We leave these issues for future research. The results further indicate that the class of monotone P-splines representable with conditions (10) should not be very restrictive.

Finally, in contrast to our comparison with AMSE, we also computed the DIC upon the suggestion of one reviewer. This statistic favors the unrestricted models. The use of the DIC, however, has been considered critically as far as models with random effects are concerned (e.g., compare the discussion in Spiegelhalter, Best, Carlin, and van der Linde 2002). In our study, we have modeled the store effects by a random effect. We further regard cross-validation as a much stronger criterion to address our central question whether nonmonotonic effects indeed exist or rather represent an artifact: because model evaluation is based on holdout sets not used for parameter estimation. Vehtari and Lampinen (2003) provided a similar argument in favor of cross-validation. They pointed out that information criteria ignore the uncertainty about parameter values.

### 5.3 Illustration for Dominick's Store Brand

To validate our results obtained for Florida Gold, we replicate our application for Dominick's, the retail chain's own private label brand. Contrary to Florida Gold, which has been one of the brands with a rather low average market share (4%), Dominick's store brand has the highest market share (22%) on average in the product category (cf. Table 1). While the sales data indicated an out-of-stock situation for Dominick's in two promotional weeks, we do not exclude these weeks from estimation and validation but instead use an additional dummy to account for this irregularity. Table 6 shows the AMSE results for the three models  $\eta^{(1)}$ ,  $\eta^{(2)}$ , and  $\eta^{(3)}$ , where we use our default choice of 20 equally spaced knots for all nonparametric price terms. Note that because Dominick's is the only private label brand, we have only two cross-price effects included, one for the national brand tier and one for the premium brand tier.

The AMSE results for Dominick's confirm our findings for Florida Gold that imposing monotonicity constraints on own- and cross-item price effects can improve the predictive validity of a sales response model. In particular, both restricted models  $\eta^{(1)}$  and  $\eta^{(2)}$  perform much better than the unrestricted model  $\eta^{(3)}$ , whereas the differences from using first-order or second-order random walk priors are almost negligible. Again, the estimated curves for the unrestricted model showed strong local nonmonotonicities in both own- and cross-price effects, indicating strong overfitting. Also, as expected, the own-price effect of Dominick's turned out to be much stronger than each of the two cross-price effects.

Table 6. Evaluation of models in terms of AMSE for Dominicks

| Model specification                | lognormal |
|------------------------------------|-----------|
| $\eta_{st}^{(1)}$ (restricted/RW1) | 70,620    |
| $\eta_{st}^{(2)}$ (restricted/RW2) | 70,781    |
| $\eta_{st}^{(3)}$ (unrestricted)   | 75,881    |
| log-linear (parametric)            | 80,891    |

To get a sense of how much we can gain from the nonparametric technique in general, we also estimated the log-linear model (or exponential model), which is one of the most widely used parametric models for problems as considered here. However, to provide a fair comparison, we retained the cubic spline for the seasonal effect. The corresponding AMSE value provided in Table 6 indicates that the log-linear model performs much worse compared to the nonparametric models. Whereas the unconstrained nonparametric model suffers from too much flexibility, the log-linear model now suffers from strong underfitting.

## 6. DISCUSSION

We proposed a methodology to incorporate specific prior knowledge of a monotonic relationship between a response variable and one or more continuous covariates into (Bayesian) generalized additive models. Unlike other approaches to monotonic regression, our method offers the possibility of nonparametric monotonic modeling by penalized splines of arbitrary degree. Sampling is accomplished by block updates of nonparametric effects. An internal Gibbs sampler is employed for drawing random numbers from truncated multivariate normal densities. Our approach can also accommodate additional covariates modeled by appropriate other specifications, such as parametric effects, unrestricted P-splines, random effects, or spatial effects as well as varying coefficient terms and interactions of covariates. We illustrated the methodology and its practical relevance in an empirical application estimating sales response for brands of refrigerated orange juice from store-level scanner data. Our results show that imposing monotonicity constraints for own- and cross-item price effects can considerably improve the predictive validity of a sales response model. The methodology is implemented in the public domain software package *BayesX*.

## ACKNOWLEDGMENTS

This research has been financially supported in part by grants from the German Science Foundation (DFG), Sonderforschungsbereich 386 "Statistical Analysis of Discrete Structures." We thank Ludwig Fahrmeir, Harald Hruschka, and Stefan Lang for helpful discussion. The data were provided by the James M. Kilts Center, GSB, University of Chicago.

## APPENDIX A: B-SPLINE BASIS FUNCTIONS

The following results were obtained from De Boor (2001). A B-spline basis function of degree  $l$  with equally spaced knots and distance  $h$  between each two adjacent knots can be written in terms of

$$B_{\psi}^l(x) = (-1)^{l+1} \Delta^{l+1} t(x, \zeta_{\psi}) / (h^l l!), \quad (36)$$

where  $\Delta^{l+1}$  is the difference operator of order  $l+1$  and  $t(x, \zeta_{\psi}) = (x - \zeta_{\psi})^l$  if  $x > \zeta_{\psi}$  and  $t(x, \zeta_{\psi}) = 0$  otherwise.

B-spline basis functions of degree  $l$  with equally spaced knots with distance  $h$  can be derived recursively starting from the corresponding B-spline basis function of degree 0:

$$B_{\psi}^0(x) = \begin{cases} 1, & \text{for } \zeta_{\psi-1} \leq x < \zeta_{\psi} \\ 0, & \text{otherwise,} \end{cases} \quad (37)$$

$$B_{\psi}^l(x) = \frac{x - \zeta_{\psi-l-1}}{lh} B_{\psi}^{l-1}(x) + \frac{\zeta_{\psi} - x}{lh} B_{\psi+1}^{l-1}(x) \quad \text{for } l \geq 1. \quad (38)$$

Explicit formulas can also be derived. For example, for a cubic B-spline basis function (i.e.,  $l = 3$ ), we obtain

$$B_{\psi}^3(x) = \frac{1}{6h^3} \begin{cases} (x - \zeta_{\psi-4})^3, & \text{for } \zeta_{\psi-4} \leq x < \zeta_{\psi-3} \\ h^3 + 3h^2(x - \zeta_{\psi-3}) \\ + 3h(x - \zeta_{\psi-3})^2 \\ - 3(x - \zeta_{\psi-3})^3, & \text{for } \zeta_{\psi-3} \leq x < \zeta_{\psi-2} \\ h^3 + 3h^2(\zeta_{\psi-1} - x) \\ + 3h(\zeta_{\psi-1} - x)^2 \\ - 3(\zeta_{\psi-1} - x)^3, & \text{for } \zeta_{\psi-2} \leq x < \zeta_{\psi-1} \\ (\zeta_{\psi} - x)^3, & \text{for } \zeta_{\psi-1} \leq x < \zeta_{\psi}. \end{cases} \quad (39)$$

One fundamental property of a B-spline basis function is  $B_{\psi}^l(x) > 0$  for  $x \in ]\zeta_{\psi-l-1}, \zeta_{\psi}[$  and  $B_{\psi}^l(x) = 0$  otherwise. A corresponding recursion formula for first-order derivatives is given by  $B_{\psi}^l(x) = \frac{1}{h}(B_{\psi}^{l-1}(x) - B_{\psi+1}^{l-1}(x))$ .

## APPENDIX B: CONDITIONS FOR MONOTONICITY

Let  $f(x) = \sum_{\psi=1}^{\Psi} \beta_{\psi} B_{\psi}(x)$ , where  $B_{\psi}(x)$ ,  $\psi = 1, \dots, \Psi$ , denotes B-spline basis functions of a certain degree. To ensure that  $f'(x) \geq 0$  or  $f'(x) \leq 0$ , it is sufficient to guarantee that subsequent parameters are ordered, such that

$$\beta_1 \leq \dots \leq \beta_{\Psi} \quad \text{or} \quad \beta_1 \geq \dots \geq \beta_{\Psi}, \quad (40)$$

respectively.

*Proof.* We exploit the recursion formula for derivatives of B-splines and the fact that B-spline functions are nonnegative on their domain (see App. A).

Letting the superscript  $l-1$  denote basis functions of degree  $l-1$ , we can write  $f'(x)$  in terms of

$$\begin{aligned} f'(x) &= \frac{1}{h} \sum_{\psi=1}^{\Psi} \beta_{\psi} (B_{\psi}^{l-1}(x) - B_{\psi+1}^{l-1}(x)) \\ &= \frac{1}{h} \sum_{\psi=2}^{\Psi} (\beta_{\psi} - \beta_{\psi-1}) B_{\psi}^{l-1}(x), \end{aligned} \quad (41)$$

where  $h$  denotes the distance between two adjacent knots. The second equivalence in (41) holds, because  $B_1^{l-1}(x) = 0$  for  $x \in [x_{\min}, x_{\max}]$ , and  $B_{\Psi+1}^{l-1}(x) = 0$  for  $x \in [x_{\min}, x_{\max}]$  (cf. Fig. 1). Because  $h > 0$  and  $B_{\psi}^{l-1}(x) \geq 0$ , it follows that  $f'(x) \geq 0$  if  $\beta_{\psi} - \beta_{\psi-1} \geq 0$  for all  $\psi \in \{2, \dots, \Psi\}$ . Correspondingly, from  $\beta_{\psi} - \beta_{\psi-1} \leq 0$  for all  $\psi \in \{2, \dots, \Psi\}$  it follows that  $f'(x) \leq 0$ .

## REFERENCES

- Allenby, G. M., Arora, N., and Ginter, J. L. (1995), "Incorporating Prior Knowledge Into the Analysis of Conjoint Studies," *Journal of Marketing Research*, 32, 152–162.
- Allenby, G. M., and Rossi, P. E. (1991), "Quality Perceptions and Asymmetric Switching Between Brands," *Marketing Science*, 10, 185–204.
- Bemmaor, A. C., and Mouchoux, D. (1991), "Measuring the Short-Term Effect of In-Store Promotion and Retail Advertising on Brand Sales: A Factorial Experiment," *Journal of Marketing Research*, 28, 202–214.
- Berger, J. O., and Pericchi, L. R. (2001), "Objective Bayesian Methods for Model Selection: Introduction and Comparison," in *Model Selection, Institute of Mathematical Statistics Lecture Notes—Monograph Series*, Vol. 38, ed. P. Lahiri, pp. 135–207.
- Billier, C. (2000), "Adaptive Bayesian Regression Splines in Semiparametric Generalized Linear Models," *Journal of Computational and Graphical Statistics*, 9, 122–140.
- Billier, C., and Fahrmeir, L. (2001), "Bayesian Varying-Coefficient Models Using Adaptive Regression Splines," *Statistical Modeling*, 2, 195–211.
- Blattberg, R. C., Briesch, R., and Fox, E. J. (1995), "How Promotions Work," *Marketing Science*, 14 (Part 2), G122–G132.
- Blattberg, R. C., and George, E. I. (1991), "Shrinkage Estimation of Price and Promotional Elasticities," *Journal of the American Statistical Association*, 86, 304–315.
- Blattberg, R. C., and Neslin, S. A. (1990), *Sales Promotion: Concepts, Methods, and Strategies*, Englewood Cliffs, NJ: Prentice-Hall.
- Blattberg, R. C., and Wisniewski, K. J. (1987), "How Retail Price Promotions Work," Marketing Working Paper 42, University of Chicago.
- (1989), "Price-Induced Patterns of Competition," *Marketing Science*, 8, 291–309.
- Brezger, A., Kneib, T., and Lang, S. (2005), "BayesX: Analysing Bayesian Structured Additive Regression Models," *Journal of Statistical Software*, 14(11). Open domain software available from <http://www.stat.uni-muenchen.de/~bayesx/>.
- Brezger, A., and Lang, S. (2005), "Generalized Structured Additive Regression Based on Bayesian P-Splines," *Computational Statistics and Data Analysis*, 50, 967–991.
- Chen, M. H., and Dey, D. K. (2000), "Bayesian Analysis for Correlated Ordinal Data Models," in *Generalized Linear Models: A Bayesian Perspective*, eds. D. K. Dey, S. K. Ghosh, and B. K. Mallick, New York: Marcel Dekker, pp. 133–158.
- De Boor, C. (2001), *A Practical Guide to Splines* (rev. ed.), New York: Springer.
- Denison, D. G. T., Mallick, B. K., and Smith, A. F. M. (1998), "Automatic Bayesian Curve Fitting," *Journal of the Royal Statistical Society, Ser. B*, 60, 333–350.
- Di Matteo, I., Genovese, C. R., and Kass, R. E. (2001), "Bayesian Curve-Fitting With Free-Knot Splines," *Biometrika*, 88, 1055–1071.
- Dunson, D. B., and Neelon, B. (2003), "Bayesian Inference on Order Constrained Parameters in Generalized Linear Models," *Biometrics*, 59, 286–295.
- Efron, B., and Tibshirani, R. J. (1998), *An Introduction to the Bootstrap*, Boca Raton, FL: Chapman and Hall/CRC.
- Eilers, P. H. C., and Marx, B. D. (1996), "Flexible Smoothing Using B-Splines and Penalized Likelihood" (with comments and rejoinder), *Statistical Science*, 11, 89–121.
- (2004), "Splines, Knots and Penalties," technical report, available at <http://www.stat.lsu.edu/bmarx/>.
- Fahrmeir, L., and Lang, S. (2001a), "Bayesian Inference for Generalized Additive Mixed Models Based on Markov Random Field Priors," *Journal of the Royal Statistical Society, Ser. C*, 50, 201–220.
- (2001b), "Bayesian Semiparametric Regression Analysis of Multicategorical Time-Space Data," *Annals of the Institute of Statistical Mathematics*, 53, 10–30.
- Fahrmeir, L., and Tutz, G. (2001), *Multivariate Statistical Modelling Based on Generalized Linear Models*, New York: Springer.
- Fan, J., and Gijbels, I. (1996), *Local Polynomial Modelling and Its Applications*, London: Chapman and Hall.
- Foekens, E. W., Leeftang, P. S. H., and Wittink, D. R. (1999), "Varying Parameter Models to Accommodate Dynamic Promotion Effects," *Journal of Econometrics*, 89, 249–268.
- Friedman, J. H. (1991), "Multivariate Adaptive Regression Splines" (with discussion), *The Annals of Statistics*, 19, 1–141.
- Friedman, J. H., and Silverman, B. L. (1989), "Flexible Parsimonious Smoothing and Additive Modeling" (with discussion), *Technometrics*, 31, 3–39.
- Gamerman, D. (1997), "Efficient Sampling From the Posterior Distribution in Generalized Linear Models," *Statistics and Computing*, 7, 57–68.

- Geweke, J. (1991), "Efficient Simulation From the Multivariate Normal and Student-*t* Distribution Subject to Linear Constraints," in *Computing Science and Statistics: Proceedings of the Twenty-Third Symposium on the Interface*, pp. 571–578.
- Goldberger, A. (1968), "The Interpretation and Estimation of Cobb–Douglas Functions," *Econometrica*, 35, 464–472.
- Greene, W. (1997), *Econometric Analysis*, Englewood Cliffs, NJ: Prentice-Hall.
- Gupta, S., and Cooper, L. (1992), "The Discounting of Discounts and Promotion Thresholds," *Journal of Consumer Research*, 19, 401–411.
- Hansen, M. H., and Kooperberg, C. (2002), "Spline Adaptation in Extended Linear Models," *Statistical Science*, 17, 2–51.
- Hanssens, D. M., Parsons, L. J., and Schultz, R. L. (2001), *Market Response Models: Econometric and Time Series Analysis*, London: Chapman and Hall.
- Hastie, T., and Tibshirani, R. (1990), *Generalized Additive Models*, London: Chapman and Hall.
- (2000), "Bayesian Backfitting," *Statistical Science*, 15, 193–223.
- Holmes, C. C., and Heard, N. A. (2003), "Generalized Monotonic Regression Using Random Change Points," *Statistics in Medicine*, 22, 623–638.
- Kalyanam, K., and Shively, T. S. (1998), "Estimating Irregular Pricing Effects: A Stochastic Spline Regression Approach," *Journal of Marketing Research*, 35, 16–29.
- Kopalle, P. K., Mela, C. F., and Marsh, L. (1999), "The Dynamic Effect of Discounting on Sales: Empirical Analysis and Normative Pricing Implications," *Marketing Science*, 18, 317–332.
- Lang, S., and Brezger, A. (2004), "Bayesian P-Splines," *Journal of Computational and Graphical Statistics*, 13, 183–212.
- Leeflang, P. S. H., Wittink, D. R., Wedel, M., and Naert, P. A. (2000), *Building Models for Marketing Decisions*, Boston: Kluwer Academic Publishers.
- Marx, B. D., and Eilers, P. H. C. (1998), "Direct Generalized Additive Modeling With Penalized Likelihood," *Computational Statistics and Data Analysis*, 28, 193–209.
- Montgomery, A. L. (1997), "Creating Micro-Marketing Pricing Strategies Using Supermarket Scanner Data," *Marketing Science*, 16, 315–337.
- Mulherne, F. J., and Leone, R. P. (1991), "Implicit Price Bundling of Retail Products: A Multiproduct Approach to Maximizing Store Profitability," *Journal of Marketing*, 55, 63–76.
- Ramsay, J. O. (1988), "Monotone Regression Splines in Action," *Statistical Science*, 3, 425–441.
- (1998), "Estimating Smooth Monotone Functions," *Journal of the Royal Statistical Society, Ser. B*, 60, 365–375.
- Rao, V. (1993), "Pricing Models in Marketing," in *Handbooks in Operations Research and Management Science*, Vol. 5, eds. J. Eliashberg and G. L. Lilien, Amsterdam: Elsevier, pp. 517–552.
- Rue, H. (2001), "Fast Sampling of Gaussian Markov Random Fields With Applications," *Journal of the Royal Statistical Society, Ser. B*, 63, 325–338.
- Robert, C. P. (1995), "Simulation of Truncated Normal Variables," *Statistics and Computing*, 5, 121–125.
- Salanti, G., and Ulm, K. (2003), "The Analysis of Dose-Response Relationship for Binary Data Using Monotonic Regression," *American Journal of Epidemiology*, 157, 273–291.
- Sethuraman, R., Srinivasan, V., and Kim, D. (1999), "Asymmetric and Neighborhood Cross-Price Effects: Some Empirical Generalizations," *Marketing Science*, 18, 23–41.
- Sivakumar, K., and Raj, S. P. (1997), "Quality Tier Competition: How Price Change Influences Brand Choice and Category Choice," *Journal of Marketing*, 61, 71–84.
- Smith, M., and Kohn, R. (1996), "Nonparametric Regression Using Bayesian Variable Selection," *Journal of Econometrics*, 75, 317–343.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002), "Bayesian Measures of Model Complexity and Fit" (with discussion), *Journal of the Royal Statistical Society, Ser. B*, 64, 583–639.
- Stone, C. J., Hansen, M., Kooperberg, C., and Truong, Y. K. (1997), "Polynomial Splines and Their Tensor Products in Extended Linear Modeling" (with discussion), *The Annals of Statistics*, 25, 1371–1470.
- Tellis, G. J. (1988), "The Price-Elasticity of Selective Demand," *Journal of Marketing Research*, 25, 331–341.
- Ulm, K., and Salanti, G. (2003), "Estimation of General Threshold Limit Values for Dust," *International Archives of Occupational Environmental Health*, 76, 233–240.
- van Heerde, H. J., Leeflang, P. S. H., and Wittink, D. R. (2001), "Semiparametric Analysis to Estimate the Deal Effect Curve," *Journal of Marketing Research*, 38, 197–215.
- (2002), "How Promotions Work: SCAN\*PRO-Based Evolutionary Model Building," *Schmalenbach Business Review*, 54, 198–220.
- Vehtari, A., and Lampinen, J. (2003), "Expected Utility Estimation via Cross-Validation," in *Bayesian Statistics*, Vol. 7, eds. J. M. Bernardo, M. J. Bayarri, et al., New York: Oxford University Press, pp. 701–710.
- Wahba, G. (1978), "Improper Priors, Spline Smoothing and the Problem of Guarding Against Model Errors in Regression," *Journal of the Royal Statistical Society, Ser. B*, 40, 364–372.
- Wilkinson, J. B., Mason, J. B., and Paksoy, C. H. (1982), "Assessing the Impact of Short-Term Supermarket Strategy Variables," *Journal of Marketing Research*, 19, 72–86.
- Wisniewski, K. J., and Blattberg, R. C. (1983), "Response Function Estimation Using UPC Scanner Data," in *Proceedings of ORSA-TIMS Marketing Science Conference*, ed. F. S. Zufryden, pp. 300–311.
- Wong, C. M., and Kohn, R. (1996), "A Bayesian Approach to Additive Semiparametric Regression," *Journal of Econometrics*, 74, 209–235.