

Brezger, Andreas; Steiner, Winfried J.

**Working Paper**

## Monotonic regression based on Bayesian P-splines: an application to estimating price response functions from store-level scanner data

Discussion Paper, No. 331

**Provided in Cooperation with:**

Collaborative Research Center (SFB) 386: Statistical Analysis of discrete structures -  
Applications in Biometrics and Econometrics, University of Munich (LMU)

*Suggested Citation:* Brezger, Andreas; Steiner, Winfried J. (2003) : Monotonic regression based on Bayesian P-splines: an application to estimating price response functions from store-level scanner data, Discussion Paper, No. 331, Ludwig-Maximilians-Universität München, Sonderforschungsbereich 386 - Statistische Analyse diskreter Strukturen, München, <https://doi.org/10.5282/ubm/epub.1709>

This Version is available at:

<https://hdl.handle.net/10419/31088>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# 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  
Ludwigstr. 33  
80539 Munich  
Germany

Winfried J. Steiner  
Department of Marketing  
University of Regensburg  
Universitätsstr. 31  
93053 Regensburg  
Germany

## ABSTRACT

Generalized additive models have become a widely used instrument for flexible regression analysis. In many practical situations, however, it is desirable to restrict the flexibility of nonparametric estimation in order to accommodate a presumed monotonic relationship between a covariate and the response variable. For example, consumers usually will buy less of a brand if its price increases, and therefore one expects a brand's unit sales to be a decreasing function in own price. 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 a brand 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 function considerably.

Key words: Generalized Additive Model, Markov Chain Monte Carlo, Sales Promotion, Own- and Cross-Item Price Effects, Asymmetric Quality Tier Competition

## 1. INTRODUCTION

Generalized additive models (GAM) 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 (2003) have extended their work to GAMs.

While 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 a 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 exposition to it at working places 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 in Ulm and Salanti (2003) and Salanti and Ulm (2003) in a frequentist setting. Dunson and Neelon (2003) and Holmes and Heard (2003) have presented 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 paper, we propose to use Bayesian P-splines of an arbitrary degree and enforce monotonicity in a straightforward way by an additional restriction of the prior distribution via indicator functions. This restriction may be imposed either for one or an arbitrary number of the additive terms in the model, whereas other terms may be modeled unrestricted. MCMC inference involves sampling from multivariate truncated normal distributions. This is accomplished by an "internal" Gibbs sampler in each iteration, i.e., we employ a short Gibbs sampler in order to draw from the proposal density. In the non-Gaussian case, this procedure is used to draw from an iteratively weighted least squares (IWLS) proposal density in a Metropolis-Hastings step. Our methodology is implemented in the public domain software package *BayesX* (Brezger, Lang and Kneib 2003) and it is possible to combine monotonic regression with all types of response distributions supported by *BayesX*. These are the most common one dimensional distributions like 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, varying coefficient terms and surface smoothing for interactions of covariates.

The remainder of the paper 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 log-normal 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. We conclude with a summary of the most important contents and key findings in section 5.

## 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)

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

and 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  $g$ :

$$g(\mu_n) = \eta_n, \quad \eta_n = f_1(x_{n1}) + \dots + f_p(x_{np}) + v_n' \gamma. \quad (1)$$

$f_1, \dots, f_p$  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

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

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 1978). Figure 1 gives an illustration of B-spline basis functions of degree three, which are also referred to as cubic splines. Note that except at the boundaries each basis function overlaps with  $2 \cdot l$  neighboring B-splines. Denoting the  $\psi$ -th basis function by  $B_{j\psi}$ , we obtain

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

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 (1) in matrix notation as

$$\eta = X_1 \beta_1 + \dots + X_p \beta_p + V \gamma. \quad (2)$$

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 fixed effects. To overcome the

difficulties in determining the position and the number of the knots involved with regression splines, Eilers and Marx (1996) suggest a relatively large number of knots (usually between 20 to 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, i.e., first or second order random walks defined by

$$\beta_{j\psi} = \beta_{j,\psi-1} + u_{j\psi}, \quad \text{or} \quad \beta_{j\psi} = 2\beta_{j,\psi-1} - \beta_{j,\psi-2} + u_{j\psi} \quad (3)$$

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

$$\tau_j^2 \sim IG(a_j, b_j)$$

with hyperparameters  $a_j$  and  $b_j$ . A common choice for  $a_j$  and  $b_j$  leading to almost diffuse priors is  $a_j = b_j$ , e.g.  $a_j = b_j = 0.001$ , which is also our default choice. Alternatively, we may set  $a_j = 1$  and  $b_j$  small, e.g.  $b_j = 0.005$  or  $b_j = 0.0005$ . We estimated all models discussed in this paper 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 paper are obtained by the default choice.

Defining a penalty matrix  $K^\delta$  corresponding to a random walk of order  $\delta$  enables us to formulate the prior for a P-spline term as a joint prior distribution for  $\beta_j$ :

$$p(\beta_j) \propto \exp \left\{ -0.5 \frac{1}{\tau_j^2} \beta_j' K_j^\delta \beta_j \right\}, \quad (4)$$

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

$$K^1 = \begin{pmatrix} 1 & -1 & & & & \\ -1 & 2 & -1 & & & \\ & -1 & 2 & -1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & -1 & 2 & -1 \\ & & & & -1 & 2 & -1 \\ & & & & & -1 & 1 \end{pmatrix}$$

with zero elements outside the first off-diagonals.

## 2.2 Monotonicity constraints

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

$$\beta_{j1} \leq \dots \leq \beta_{j\Psi} \quad \text{or} \quad \beta_{j1} \geq \dots \geq \beta_{j\Psi},$$

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

$$p(\beta_j) = c_1(\beta_j) \exp \left\{ -0.5 \frac{1}{\tau_j^2} \beta_j' K_j^\delta \beta_j \right\} \prod_{\psi=2}^{\Psi} \mathbf{1}(\beta_{j\psi} \geq \beta_{j\psi-1}) \quad (5)$$

for nondecreasing functions (isotonic case) and

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

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 (1) are possible. In order 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 (1) to

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

and assume

$$f_{\text{random}}(g) = b_g \sim N(0, \tau_b^2), \quad g = 1, \dots, G, \quad (6)$$

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 (6) in the general form

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

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 Brezger and Lang 2003; Lang and Brezger 2004). 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

$$\begin{aligned} p(\alpha | y) &\propto L(y, \beta_1, \dots, \beta_p, \gamma, b_g, \phi) \prod_{j=1}^p (p(\beta_j | \tau_j^2) p(\tau_j^2)) \\ &\quad p(b_g | \tau_b^2) p(\tau_b^2) p(\gamma) p(\phi) \end{aligned} \quad (7)$$

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 (7) is analytically intractable in all but the most simple cases, we employ Markov Chain Monte Carlo (MCMC) techniques to obtain estimates for the parameters of interest. More specifically, we implement a block move, i.e. 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 (2003) for details. In all other cases, we use Metropolis-Hastings steps with iteratively weighted least squares (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 in order 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

$$p(\beta_j|\cdot) \propto c_1(\beta_j) \exp\{-0.5(\beta_j - m_j)' P_j(\beta_j - m_j)\} \prod_{\psi=2}^{\Psi} \mathbf{1}(\beta_{j\psi} \geq \beta_{j,\psi-1}), \quad (8)$$

and

$$\begin{aligned} P_j &= \frac{1}{\sigma^2} X_j' X_j + \frac{1}{\tau_j^2} K_j^\delta \\ m_j &= \frac{1}{\sigma^2} P_j^{-1} X_j' (y - \eta + X_j \beta_j^c) \end{aligned}$$

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

In order to sample from this  $\Psi$ -dimensional truncated Gaussian distribution (8), 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:

- (i) Set  $\beta^{(0)} = \beta_j^c$ .
- (ii) 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 (8):

$$\begin{aligned}\mu_\psi &= \frac{1}{p_{\psi\psi}} \left\{ \sum_{\rho < \psi} (\beta_\rho^{(t)} - m_\rho) \cdot p_{\psi\rho} + \sum_{\rho > \psi} (\beta_\rho^{(t-1)} - m_\rho) \cdot p_{\psi\rho} \right\} \\ \sigma_\psi^2 &= \frac{1}{p_{\psi\psi}}\end{aligned}$$

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 (8). Note that the subscript  $j$  is suppressed in the formulae above.

(iii) Take  $\beta^{(T)} = (\beta_1^{(T)}, \dots, \beta_\Psi^{(T)})'$  as a random sample from (8).

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 (compare Rue 2001).

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

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

as full conditional.

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

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

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

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

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

$$p(\beta_j | \cdot) \propto c(y_n, \theta_n) \exp \left\{ \frac{y_n \theta_n - b(\theta_n)}{\phi} \right\} c_1(\beta_j) \exp \left\{ -0.5 \frac{1}{\tau_j^2} \beta_j' K_j^\delta \beta_j \right\} \prod_{\psi=2}^{\Psi} \mathbf{1}(\beta_{j\psi} \geq \beta_{j,\psi-1}),$$



which has no longer standard form. 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  $\beta_j$ , compare Brezger and Lang (2003) or, in the mixed model context, Gamerman (1997). This leads to a truncated multivariate Gaussian proposal

$$q(\beta_j) \propto c_1(\beta_j) \exp\{-0.5(\beta_j - m(\beta_j^c))' P(\beta_j^c)(\beta_j - m(\beta_j^c))\} \prod_{\psi=2}^{\Psi} \mathbf{1}(\beta_{j\psi} \geq \beta_{j\psi-1}) \quad (9)$$

where

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

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 (2003). Generating a proposed value  $\beta_j^p$  from (9) is again accomplished by an extra Gibbs sampler as described in subsection 3.1. It has been our experience that convergence in the non-Gaussian case is slower than in the Gaussian case. We therefore set the number of iterations for the single move Gibbs sampler to  $T = 250$  (as opposed to  $T = 100$  for Gaussian response) to ensure convergence, which implies that we take the 250th sample as a random sample from (9). The main difference to the Gaussian sampling scheme is that this sample can only be accepted with probability

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

as the new state of  $\beta_j$ . Note that the normalizing functions  $c_1(\cdot)$  cancel out and we have the same acceptance probability as in the unrestricted case.

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. Fixed 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., Mulherne and Leone 1991; Blattberg and Wisniewski 1989, Bemmaor and Mouchoux 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 modelled 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 (semi-log) 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; Kopalle, Mela and Marsh 1999; Foekens, Leeftang and Wittink 1999, van Heerde, Leeftang 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 is modelled 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, e.g., 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, e.g., due to inventory constraints or perishability (Blattberg et al. 1995; van Heerde, Leeflang and Wittink 2001). About two-third 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 do not show 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 become 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 show 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, e.g., 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., 0.99 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 point 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) agree with those of van Heerde et al. (2001) with respect to the existence of threshold effects for several brands, i.e., 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 paper 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 artefact 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 data set includes unit sales, retail price and a deal code indicating the use of an in-store display for 11 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 log-normal model

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

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

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

and a Gamma model

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

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  $a_\nu = 0.001$ ,  $b_\nu = 0.001$  and estimated in a Metropolis-Hastings step.

As mentioned above, the use of a log-normal 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 log-normal and the Gamma model:

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

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

and

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

The three variants differ in the specification of the unknown smooth functions  $f_1$  to  $f_4$  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 report 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  $t$ . Similar 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<sub>st</sub>* is the corresponding fixed effect for the private label brand.

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, e.g., due to their spatial location. The effect of *week* is modelled 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  $\sigma^2 \sim IG(0.001, 0.001)$  and  $\nu \sim IG(0.001, 0.001)$ , respectively, and are estimated simultaneously with the regression parameters. The resulting models are referred to as LN1-LN3 for the log-normal 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 compare van Heerde et al. 2001). Specifically, we randomly split the data into nine equally-sized subsets and performed nine-fold 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 data set, 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:

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

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

$$\widehat{sales}_{st}^{-k(n)} = \frac{1}{1000} \sum_{i=1}^{1000} \exp\{\eta_{sti} + \sigma_i^2/2\}, \quad (10)$$

where  $\eta_{sti}$  is the additive predictor for store  $s$ , week  $t$  and stored iteration  $i$  and  $\sigma_i^2$  denotes the residual variance of the respective log-normal model in iteration  $i$ . For the Gamma model, no correction factor  $\sigma_i^2/2$  is required for the conditional mean predictions.

The validation results are displayed in table 2. Under both the log-normal 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 log-normal model (LN3) yields a smaller AMSE compared to the Gamma model (G3), the restricted Gamma models G1 and G2 provided 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 log-normal 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, i.e.  $1/\text{range}(x_j) \int f_j(x_j) dx_j = 0$ . The subtracted means are added to an intercept term, which is not displayed here. 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 (compare 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 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 (Leeflang, Wittink, Wedel and Naert 2000). The multiplier effects are obtained from the transformation

$$\frac{1}{1000} \sum_{i=1}^{1000} \exp\{\gamma_{ji}\}, \quad j = 1, \dots, 4.$$

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 not only stimulate 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 spatial map which 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 nine (0.28) and the percentage of households with three or more members (0.24). 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 childrens 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. 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. Convergence of the internal Gibbs sampler is fast in the Gaussian case, but might be improved for other response distributions. Our approach can also accommodate additional covariates modelled by appropriate other specifications, like fixed 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 a brand 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*.

### Acknowledgement:

This research has been partly financially supported 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 is provided by the James M. Kilts Center, GSB, University of Chicago.



## APPENDIX: CONDITIONS FOR MONOTONICITY

To ensure that  $f'_j(x) \geq 0$  or  $f'_j(x) \leq 0$ , it is sufficient to guarantee that subsequent parameters are ordered, such that

$$\beta_{j1} \leq \cdots \leq \beta_{j\Psi} \quad \text{or} \quad \beta_{j1} \geq \cdots \geq \beta_{j\Psi}, \quad (11)$$

respectively.

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

$$\begin{aligned} f'_j(x) &= \frac{1}{h} \sum_{\psi=1}^{\Psi} \beta_{j\psi} (B_{j\psi}^{l-1}(x) - B_{j,\psi+1}^{l-1}(x)) \\ &= \frac{1}{h} \sum_{\psi=2}^{\Psi} (\beta_{j\psi} - \beta_{j,\psi-1}) B_{j\psi}^{l-1}(x), \end{aligned} \quad (12)$$

where  $h$  denotes the distance between two adjacent knots. The second equivalence in (12) holds, because  $B_{j1}^{l-1}(x) = 0$  and  $B_{j,\Psi+1}^{l-1}(x) = 0$  for  $x \in [x_{j,min}, x_{j,max}]$ . Since  $h > 0$  and  $B_{j\psi}^{l-1}(x) \geq 0$ , it follows that  $f'_j(x) \geq 0$  if  $\beta_{j\psi} - \beta_{j,\psi-1} \geq 0$  for all  $\psi \in \{2, \dots, \Psi\}$ . Correspondingly, from  $\beta_{j\psi} - \beta_{j,\psi-1} \leq 0$  for all  $\psi \in \{2, \dots, \Psi\}$  it follows that  $f'_j(x) \leq 0$ .

## References

- Allenby, G.M. and Rossi, P.E. (1991), "Quality Perceptions and Asymmetric Switching Between Brands," *Marketing Science*, 10(3), 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(2), 202–214.
- Biller, C. (2000), "Adaptive Bayesian Regression Splines in Semiparametric Generalized Linear Models," *Journal of Computational and Graphical Statistics*, 9, 122–140.
- Biller, C. and Fahrmeir, L. (2001), "Bayesian Varying-coefficient Models using Adaptive Regression Splines," *Statistical Modeling*, 2, 195–211.
- Blattberg, R.C. and Neslin, S.A. (1990), "Sales Promotion: Concepts, Methods, and Strategies", *Englewood Cliffs*, New Jersey.
- Blattberg, R.C. and George, E.I. (1991), "Shrinkage Estimation of Price and Promotional Elasticities," *Journal of the American Statistical Association*, 86(414), 304–315.
- Blattberg, R.C., Briesch, R. and Fox, E.J. (1995), "How Promotions Work," *Marketing Science*, 14(3)(Part 2), G122–G132.
- 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(4), 291–309.

- Brezger, A. and Lang, S. (2003), "Generalized structured additive regression based on Bayesian P-splines," SFB 386 Discussion paper 321, Department of Statistics, University of Munich.
- Brezger, A., Lang, S. and Kneib, T. (2003), "BayesX: Analysing Bayesian Semiparametric Regression Models," SFB 386 Discussion paper 321, Department of Statistics, University of Munich.
- Chen, M. H. and Dey, D. K. (2000), "Bayesian Analysis for Correlated Ordinal Data Models," in: *Generalized linear models: A Bayesian perspective*, eds. Dey, D. K., Ghosh, S. K. and Mallick, B. K., New York: Marcel Dekker.
- De Boor, C. (1978), *A Practical Guide to Splines*, Springer, New York.
- Denison, D.G.T., Mallick, B.K. and Smith, A.F.M. (1998), "Automatic Bayesian curve fitting," *Journal of the Royal Statistical Society 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*, (submitted).
- Efron, B. and Tibshirani, R.J. (1998), *An Introduction to the Bootstrap*, Chapman and Hall/CRC, Boca Raton.
- Eilers, P.H.C. and Marx, B.D. (1996), "Flexible smoothing using B-splines and penalized likelihood (with comments and rejoinder)," *Statistical Science*, 11 (2), 89–121.
- 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 C (Applied Statistics)*, 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*, Springer, New York.
- Fan, J. and Gijbels, I. (1996), *Local Polynomial Modelling and Its Applications*, Chapman and Hall, London.
- Foekens, E.W., Leeftang, P.S.H., Wittink, D.R. (1999), "Varying Parameter Models to Accomodate Dynamic Promotion Effects," *Journal of Econometrics*, 89, 249–268.
- Friedman, J. H. (1991), "Multivariate Adaptive Regression Splines (with discussion)," *Annals of Statistics*, 19, 1–141.
- Friedman, J. H. and Silverman, B. L. (1989), "Flexible Parsimonious Smoothing and Additive Modeling (with discussion)," *Technometrics*, 1989, 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*, 571–578, Alexandria.
- Goldberger, A. (1968), "The Interpretation and Estimation of Cobb-Douglas Functions," *Econometrica*, 35, 464–472.
- Greene, W. (1997), *Econometric Analysis*, Prentice Hall, New Jersey.
- Gupta, S., Cooper, L. (1992), "The Discounting of Discounts and Promotion Thresholds," *Journal of Consumer Research*, 19, 401–411.
- Hansen, M. H., 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*, Chapman and Hall, London.
- Hastie, T. and Tibshirani, R. (1990), *Generalized Additive Models*, Chapman and Hall, London.
- (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., Shively, T.S. (1998), "Estimating Irregular Pricing Effects: A Stochastic Spline Regression Approach," *Journal of Marketing Research*, 35(1), 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(3), 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*, Kluwer Academic Publishers, Boston.
- 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(4), 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(4), 63–76.
- Rao, V. (1993), "Pricing Models in Marketing," in: *Handbooks in Operations Research and Management Science*, Vol. 5, Marketing, 517–552, eds. Eliashberg, J. and Lilien, G.L., Elsevier Science Publishers B.V., Amsterdam.
- Rue, H. (2001), "Fast Sampling of Gaussian Markov Random Fields with Applications," *Journal of the Royal Statistical Society, Series 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(1), 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(3), 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," *Journal of the Royal Statistical Society B*, 65, 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)," *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," *Int Arch Occup Environ 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(2), 197–215.
- (2002), "How Promotions Work: SCAN\*PRO-Based Evolutionary Model Building," *Schmalenbach Business Review*, 54, 198–220.
- Wahba, G. (1978), "Improper Priors, Spline Smoothing and the Problem of Guarding Against Model Errors in Regression," *Journal of the Royal Statistical Society, Series 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(1), 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*, 300–311, ed. F.S. Zufryden.
- Wong, C.M., Kohn, R. (1996) "A Bayesian Approach to Additive Semiparametric Regression," *Journal of Econometrics*, 74, 209–235.

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	0.53	[3;73]	15	15	6,388	100,712	
Florida Natural Premium	[1.57; 3.16]	2.86	0.33	[1;53]	5	7	1,138	56,037	
<i>National Brands:</i>									
Citrus Hill	[1.09; 2.82]	2.31	0.31	[1;78]	8	12	2,006	151,570	
Minute Maid	[1.29; 2.92]	2.23	0.40	[3;87]	21	22	4,805	243,711	
Tropicana	[1.49; 2.75]	2.20	0.35	[2;75]	21	23	3,041	102,629	
Florida Gold	[0.99; 2.83]	2.17	0.39	[1;63]	4	8	325	150,945	
Tree Fresh	[1.07; 2.48]	2.15	0.27	[1;42]	4	6	916	39,401	
<i>Store Brand:</i>									
Dominick's	[0.99; 2.47]	1.75	0.4	[1;83]	22	22	2,170	189,462	

Table 2: Evaluation of models in terms of AMSE.

Model specification	log-normal	Gamma
$\eta_{st}^{(1)}$ (restricted/RW1)	6239.05	6194.83
$\eta_{st}^{(2)}$ (restricted/RW2)	6253.44	6169.00
$\eta_{st}^{(3)}$ (unrestricted)	6347.70	6522.69

Table 3: Estimation results for the display effects (Model G2).

effect	posterior mean		2.5%-quantile	97.5%-quantile
$\gamma_1$ (display)	0.30	(0.04)	0.24	0.38
$\gamma_2$ (display_premium)	-0.12	(0.04)	-0.19	-0.05
$\gamma_3$ (display_national)	-0.02	(0.05)	-0.11	0.08
$\gamma_4$ (display_Dominicks)	0.07	(0.03)	0.00	0.14

Table 4: Estimation results for the display multiplier effects (Model G2).

effect	posterior mean		2.5%-quantile	97.5%-quantile
$\gamma_1$ (display)	1.36	(0.05)	1.27	1.45
$\gamma_2$ (display_premium)	0.89	(0.03)	0.83	0.95
$\gamma_3$ (display_national)	0.98	(0.05)	0.90	1.08
$\gamma_4$ (display_Dominicks)	1.07	(0.04)	1.00	1.15

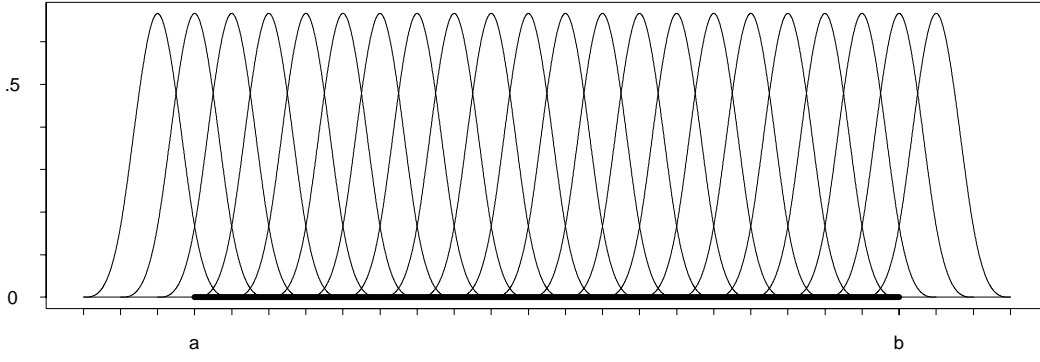


Figure 1: B-spline basis functions of degree three covering the interval  $[a, b]$ .

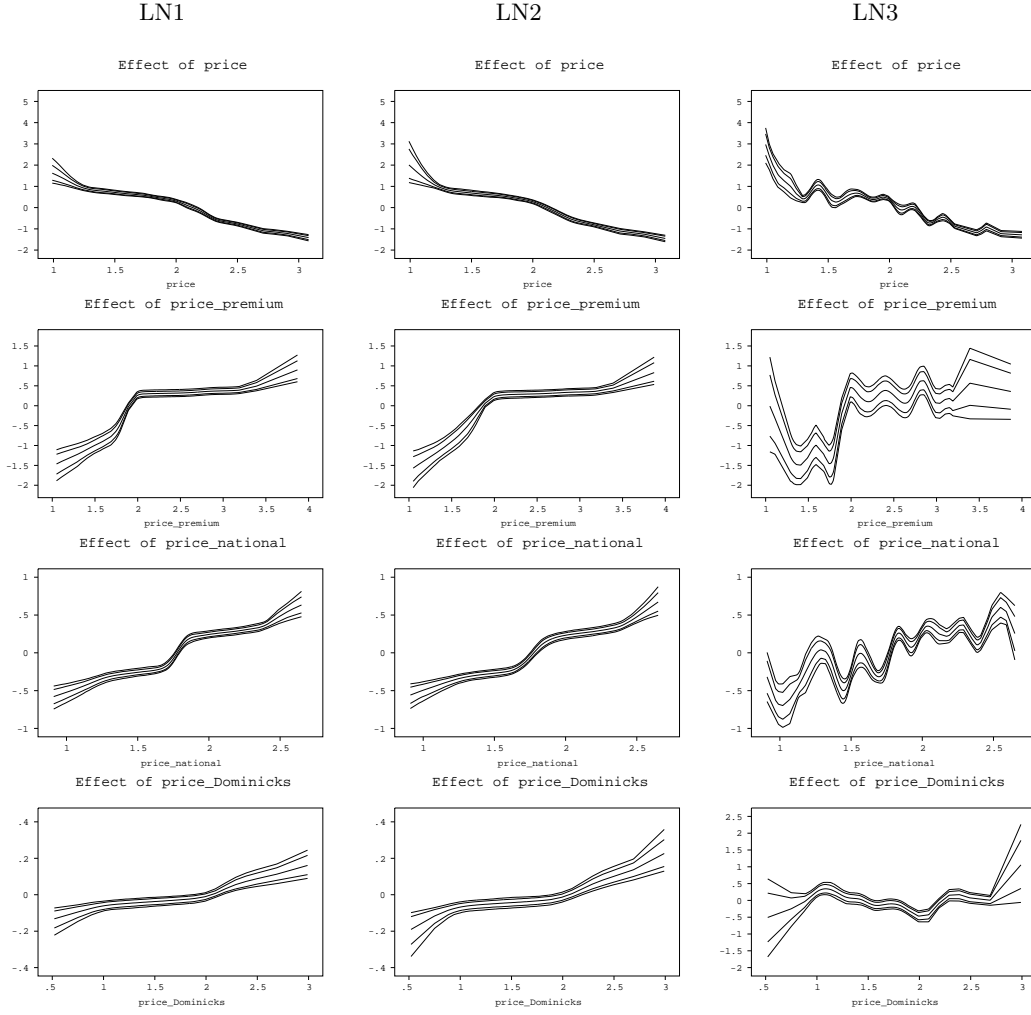


Figure 2: Estimated curves for own-price (*price*) and tier-specific cross-price (*price\_premium*, *price\_national*, *price\_Dominicks*) effects on unit sales of Florida Gold. Columns 1-3 show the effects for the models LN1-LN3. Shown are the posterior means as well as 80% and 95% pointwise credible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

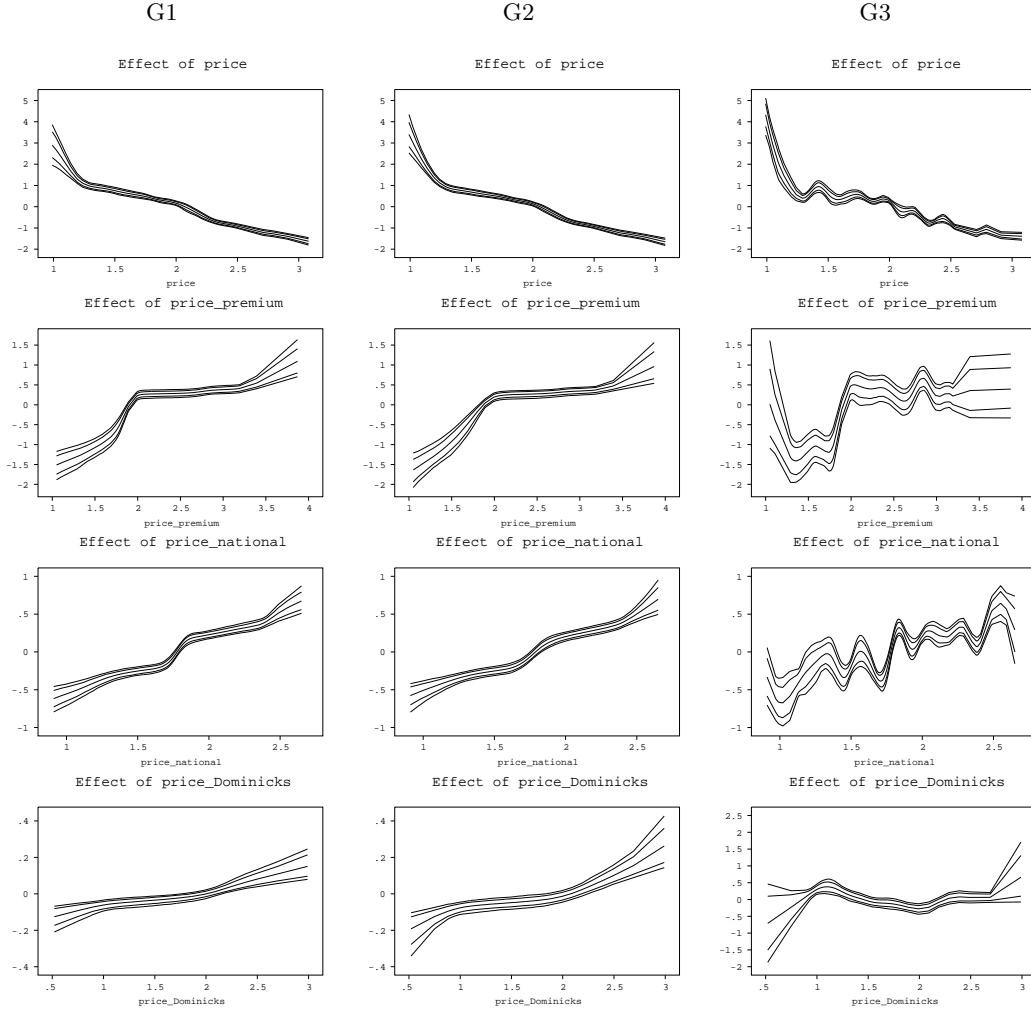


Figure 3: Estimated curves for own-price ( $price$ ) and tier-specific cross-price ( $price\_premium$ ,  $price\_national$ ,  $price\_Dominicks$ ) effects on unit sales of Florida Gold. Columns 1-3 show the effects for the models G1-G3. Shown are the posterior means as well as 80% and 95% pointwise credible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



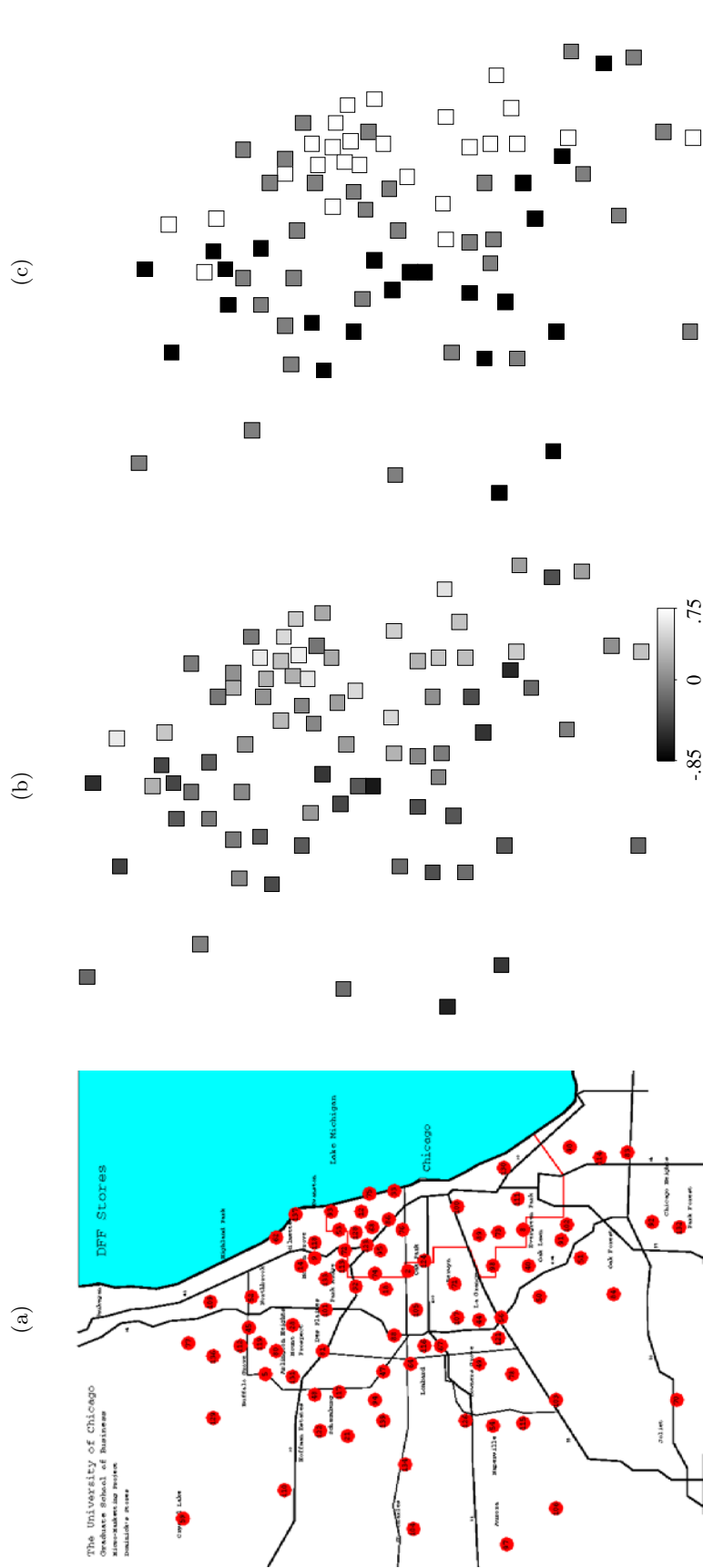


Figure 4: (a) Map of the Chicago metropolitan area with store locations of Dominick’s Finer Foods. (b) Estimated random effect of *store* for the Gamma model (G2). (c) Posterior probabilities of *store*. White (black) indicates strictly positive (negative) 95% credible intervals, gray indicates that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