AUSTRIAN JOURNAL OF STATISTICS
Volume 37 (2008), Number 3&4, 255–270

Additive Models with Random Scaling Factors: Applications to Modeling Price Response Functions

Peter Wechselberger[1], Stefan Lang[1] and Winfried J. Steiner[2]

[1] University of Innsbruck, Austria
[2] Technical University of Clausthal, Germany

Abstract: We discuss inference for additive models with random scaling factors. The additive effects are of the form $(1+\gamma)f(z)$ where f is a nonlinear function of the continuous covariate z modeled by P(enalized)-splines and $1+\gamma$ is a random scaling factor. Additionally, monotonicity constraints on the nonlinear functions are possible.

Our work is motivated by the situation of a retailer analyzing the impact of price changes on a brand's sales in its orange juice product category. Relating sales to a brand's own price as well as to the prices of competing brands in the category, we estimate own- and cross-item price response functions flexibly to represent nonlinearities and irregular pricing effects in sales response. Monotonicity constraints are imposed so that a brand's own price is inversely related and the prices of competing brands are directly related to the number of items sold, as suggested by economic theory. Unobserved store-specific heterogeneity is accounted for by allowing the price response curves to vary between different stores.

Zusammenfassung: Wir behandeln additive Modelle mit zufälligen Skalierungsfaktoren. Die additiven Effekte haben die Form $(1+\gamma)f(z)$. f ist eine nichtlineare Funktion der stetigen Kovariable z, modelliert mittels P(enalized)-splines und $1+\gamma$ ist ein zufälliger Skalierungsfaktor. Den nichtlinearen Funktionen können zusätzlich Monotonierestriktionen auferlegt werden.

Den Ausgangspunkt unserer Arbeit bildet die Situation eines Einzelhändlers, der den Einfluss von Preisänderungen auf den Absatz einer Orangensaftmarke in seinem Sortiment analysieren möchte. Eine entsprechende Absatzreaktionsfunktion lässt sich schätzen, indem der Absatz der betrachteten Marke als nichtlineare Funktion des eigenen Preises sowie der Preise der Konkurrenzmarken modelliert wird. Monotonierestriktionen für die Preiseffekte gewährleisten darüber hinaus einen inversen Verlauf des Absatzes bezüglich des eigenen Preises sowie eine direkte Beziehung des Absatzes zu Konkurrenzpreisen, wie es in Anlehnung an die ökonomische Preistheorie zu erwarten ist. Unbeobachtete Heterogenität wird berücksichtigt, indem die Preiseffekte über die einzelnen Geschäfte des Händlers zufällig variieren können.

Keywords: P-Splines, Monotonicity Constraints, Multiplicative Random Effects, Price Response, Own- and Cross-item Price Effects.

1 Introduction

This paper is motivated by a frequently encountered application problem in marketing: estimating price response from sales data. Specifically, we are interested in modeling the dependence of a brand's unit sales from its own price and the prices of competing brands. Most previous studies have employed strictly parametric functions to represent nonlinearities in sales response to price changes (e.g., Blattberg und Wisniewski, 1989, Montgomery, 1997, and Heerde, Leeflang, und Wittink, 2002). It is important to note that all those parametric functional forms have been inherently monotonic, i.e. decreasing for own-price effects and increasing for cross-price effects, which is in accordance with economic theory (e.g., see Hanssens, Parsons, und Schultz, 2001). Kalyanam and Shively (1998) and Heerde, Leeflang, and Wittink (2001) suggested stochastic spline regression and kernel regression, respectively, to explore the shape of price response curves more flexibly, and both obtained superior performance for their models compared to strictly parametric models. Recently, Brezger and Steiner (2007) demonstrated that imposing monotonicity constraints on nonparametrically estimated own- and cross-price effects not only preserves a reasonable economic interpretation but can also improve the predictive validity of sales response functions considerably. However, all flexible approaches mentioned above do not account for possible heterogeneity of price effects across different stores of a retail chain. In this paper we add to the body of knowledge by addressing nonlinearity *and* heterogeneity in estimating price response. For illustration we use weekly store-level scanner data from Dominick's Finer Foods, a major supermarket chain in the Chicago metropolitan area.

Figure 1 shows the relationship between the unit sales of a certain orange juice brand and its own price for two different stores of the retail chain. The figure exhibits two characteristic features: (1) The unit sales of the brand depend on the own price in a nonlinear way. (2) The level and presumably also the scaling of the underlying price response function differ across stores.

The purpose of the paper is to provide statistical methodology for modeling and estimating those two features: nonlinearity *and* heterogeneity. Specifically, we propose additive models (Hastie und Tibshirani, 1990) with random scaling factors. The additive effects are of the form $(1 + \gamma)f(z)$ where f is a nonlinear function of the continuous covariate z modeled by P(enalized)-splines and $1 + \gamma$ is a cluster specific random scaling factor. With respect to our marketing application, z is the own price of the brand while the cluster variable refers to the outlet index. To ensure economically plausible results, additional monotonicity constraints on the nonlinear functions are imposed as discussed above.

P-splines have been introduced by Eilers and Marx (1996), see also Lang and Brezger (2004) for a Bayesian version. Monotonicity constraints in the context of P-spline modeling are discussed in Bollaerts, Eilers, and Mechelen (2006) and Brezger and Steiner (2007). An overview concerning additive models and extensions is given in Hastie and Tibshirani (1990), Fahrmeir, Kneib, and Lang (2007) and Wood (2006). Additive models with random scaling factors are in the spirit of regression models for functional data, see, e.g., Ramsay and Silverman (2005) and Ramsay and Silverman (2002). Statistical inference is based on a recent paper by Belitz and Lang (2007) that allows for simultaneous

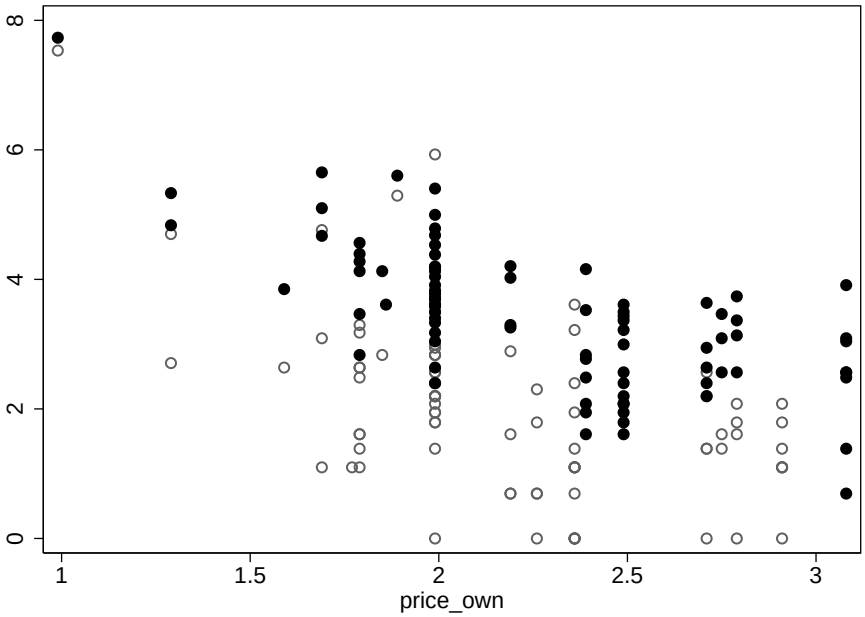

Figure 1: Scatterplot of unit sales versus own price for a brand of orange juice in two stores.

estimation of unknown regression and smoothing parameters as well as selection of relevant model terms. The approach is also able to discriminate between linear and nonlinear relationships.

The remainder of this article is structured as follows. In the next section, we introduce additive models with random scaling factors and discuss algorithms for estimation. In Section 3, the results of some simulations are presented, followed by an application to marketing data in Section 4. The final section 5 concludes.

2 Methodology

2.1 Additive and Varying Coefficient Models Based on P-Splines

Suppose that observations $(y_i, \mathbf{z}_i)$, $i = 1, \ldots, n$, are given, where y_i is a continuous response variable and $\mathbf{z}_i = (z_{i1}, \ldots, z_{iq})'$ is a vector of continuous covariates to be modeled nonlinearly. We assume an additive decomposition of the effects of z_{ij} and obtain the additive model

$$y_i = f_1(z_{i1}) + \cdots + f_q(z_{iq}) + \varepsilon_i \,, \tag{1}$$

where f_1 to f_q are nonlinear functions of the covariates z_{ij}. The error terms ε_i are assumed to be mutually independent Gaussian with mean 0 and variance σ^2, i.e. $\varepsilon_i \sim N(0, \sigma^2)$.

The nonlinear functions f_j are modeled by P(enalized)-splines introduced by Eilers and Marx (1996). The approach assumes that a particular unknown function f of a covariate z can be approximated by a polynomial spline of degree l and with equally spaced knots

$$z_{min} = \zeta_0 < \zeta_1 < \cdots < \zeta_{m-1} < \zeta_m = z_{max}$$

over the domain of z. The spline can be written in terms of a linear combination of $K = m + l$ B-spline basis functions (De Boor, 2001)

$$f(z) = \sum_{k=1}^{K} \beta_k B_k(z)\,,$$

where $B_k(z)$ are (known) B-spline basis functions evaluated at covariate value z and the β_k are unknown regression coefficients to be estimated. Defining the $n \times K$ design matrix $\mathbf{Z}$ with elements $\mathbf{Z}_{i,k} = B_k(z_i)$, the vector $\mathbf{f} = (f(z_1), \ldots, f(z_n))'$ of function evaluations can be written in matrix notation as $\mathbf{f} = \mathbf{Z}\boldsymbol{\beta}$. Accordingly, for model (1) we obtain

$$\mathbf{y} = \boldsymbol{\eta} + \boldsymbol{\varepsilon} = \mathbf{Z}_1\boldsymbol{\beta}_1 + \cdots + \mathbf{Z}_q\boldsymbol{\beta}_q + \boldsymbol{\varepsilon}\,, \tag{2}$$

where $\mathbf{y}$ is the vector of observations, $\boldsymbol{\eta}$ is the additive predictor and $\boldsymbol{\varepsilon}$ is the error vector.

The additive modeling framework can be easily extended to cover varying coefficients models as introduced in Hastie and Tibshirani (1993). Varying coefficient models contain terms of the form

$$y_i = \cdots + f(z_i) \cdot u_i + \cdots\,,$$

where the effect of the additional covariate u varies smoothly over the course of the continuous covariate z. Covariate z is called the effect modifier of u. Terms of this type are incorporated into the structure of (2) through a modification of the design matrix $\mathbf{Z}$. The matrix of B-spline basis functions must be multiplied row-wise with the observations u_i of the additional covariate u. Hence, for a varying coefficients term the elements of $\mathbf{Z}$ are given by $\mathbf{Z}_{i,k} = u_i B_k(z_i)$.

In a simple regression spline approach the unknown regression coefficients $\boldsymbol{\beta}_j$ are estimated using standard algorithms and software. The crucial point is the choice of the number (and position) of knots. For a small number of knots, the resulting spline may be not flexible enough to capture the variability of the data. For a large number of knots, estimated curves tend to overfit the data and, as a result, too rough functions are obtained. To overcome the difficulties involved with regression splines, Eilers and Marx (1996) suggest a relatively large number of knots (usually between 20 to 40) to ensure enough flexibility, and to introduce a roughness penalty on adjacent regression coefficients based on squared r-th order differences, i.e.

$$\lambda \sum_{k=r+1}^{K} (\Delta^r \beta_k)^2 = \lambda \boldsymbol{\beta}' \mathbf{K} \boldsymbol{\beta}\,.$$

The penalty matrix is given by $\mathbf{K} = \mathbf{D}_r'\mathbf{D}_r$ where $\mathbf{D}_r$ is a r-th order difference matrix. Typically, second or third order differences are used.

The approach can be extended to impose monotonicity or more general shape constraints. We follow an approach proposed by Bollaerts et al. (2006). A sufficient condition for a decreasing spline is given by $\beta_k \leq \beta_{k-1}$, i.e. a parameter β_k is less than its predecessor β_{k-1}. The simple but powerful idea of Bollaerts et al. (2006) is to impose the required constraint by expanding the penalty by an additional term. More specifically they propose the penalty

$$\lambda \boldsymbol{\beta}' \mathbf{K} \boldsymbol{\beta} + \mu \boldsymbol{\beta}' \mathbf{L} \boldsymbol{\beta}\,,$$

where the additional penalty matrix $\mathbf{L}$ is a diagonal matrix with entries 1 whenever the condition $\beta_k \leq \beta_{k-1}$ fails and 0 otherwise. For increasing functions $\mathbf{L}$ has to be adapted accordingly. The parameter μ is not estimated but set large enough to produce monotonic functions.

2.2 Simultaneous Selection of Variables and Smoothing Parameters

A main building block of the estimation algorithms for additive and varying coefficients models are smoothers of the form

$$S(\mathbf{y}, \lambda) = \mathbf{Z}\hat{\boldsymbol{\beta}} \qquad \hat{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{Z} + \lambda\mathbf{K} + \mu\mathbf{L})^{-1}\mathbf{Z}'\mathbf{y}\,. \tag{3}$$

Consecutively applying smoothers S_j corresponding to the j-th function f_j in (1) to the current partial residual reveals the well-known backfitting algorithm to minimize the overall PLS-criterion

$$PLS = (\mathbf{y} - \boldsymbol{\eta})'(\mathbf{y} - \boldsymbol{\eta}) + \sum_{j=1}^{q}(\lambda_j\boldsymbol{\beta}_j'\mathbf{K}_j\boldsymbol{\beta}_j + \mu_j\boldsymbol{\beta}_j'\mathbf{L}_j\boldsymbol{\beta}_j)\,.$$

The complexity of the fit may be determined by the equivalent degrees of freedom df as a measure of the effective number of parameters. The equivalent degrees of freedom df are typically approximated by the sum of the degrees of freedom of individual smoothers, i.e.

$$df = \sum_{j=1}^{q} df_j + p\,,$$

where df_j is computed as

$$df_j = \text{trace}(\mathbf{Z}_j(\mathbf{Z}_j'\mathbf{Z}_j + \lambda_j\mathbf{K}_j + \mu_j\mathbf{L}_j)^{-1}\mathbf{Z}_j') - 1\,. \tag{4}$$

Now the approach for simultaneous selection of variables and smoothing parameters can be described follows:

1. **Initialization**
   Define for every possible nonlinear term f_j, $j = 1, \ldots, q$, a discrete number M_j of decreasing smoothing parameters $\lambda_{j1} > \cdots > \lambda_{jM_j}$. To include a linear fit $\lambda_{j1} = \infty$ is always specified.

2. **Start model**
   Choose a start model with current predictor

   $$\hat{\boldsymbol{\eta}} = \hat{\mathbf{f}}_1 + \cdots + \hat{\mathbf{f}}_q\,,$$

   where $\hat{\mathbf{f}}_j$ is the vector of function evaluations at the observations. Choose a goodness of fit criterion G (e.g. AIC, BIC, Cross validation, etc.).

3. **Iteration** For $j = 1, \ldots, q$:

(a) For $m = 0, \ldots, M_j$:
Compute the fits

$$\hat{\mathbf{f}}_{jm} := \begin{cases} \mathbf{0} & m = 0 \\ S_j(\mathbf{y} - \hat{\boldsymbol{\eta}}_{[j]}, \lambda_{jm}) & m = 1, \ldots, M_j \end{cases}$$
$$= \begin{cases} \mathbf{0} & m = 0 \\ (\mathbf{Z}_j'\mathbf{Z}_j + \lambda_{jm}\mathbf{K}_j + \mu_j\mathbf{L}_j)^{-1}\mathbf{Z}_j'(\mathbf{y} - \hat{\boldsymbol{\eta}}_{[j]}) & m = 1, \ldots, M_j \end{cases}$$

and the corresponding predictors $\hat{\boldsymbol{\eta}}_{jm} := \hat{\boldsymbol{\eta}}_{[j]} + \hat{\mathbf{f}}_{jm}$. Here, $\hat{\boldsymbol{\eta}}_{[j]}$ is the current predictor with the j-th fit $\hat{\mathbf{f}}_j$ removed.

Compute the updated estimate

$$\hat{\mathbf{f}}_j = \text{argmin } G(\hat{\mathbf{f}}_{jm}) \,,$$

i.e. among the fits $\hat{\mathbf{f}}_{jm}$ for the j-th component, choose the one that minimizes the goodness of fit criterion G.

4. **Termination**
The iteration cycle in 3. is repeated until the model, regression and smoothing parameters do not change anymore.

Note that when updating the function estimates $\hat{\mathbf{f}}_j$ the other terms in the model are *not* re-estimated as in a backfitting procedure. However, the algorithm automatically collapses to backfitting as soon as the variables and smoothing parameters included in the model do not change anymore. Avoiding backfitting in step 2 dramatically reduces computing time without loss of estimation accuracy. More details on the estimation algorithm can be found in Belitz and Lang (2007).

2.3 Multiplicative Random Effects

As described in the introduction, the super market scanner data, that motivated our work, are clustered in 81 stores (we use the terms cluster, store and outlet interchangeably). It is usually not justified to assume homogeneous price response functions across stores. Heterogeneity in price response may come from different levels of consumer price sensitivity in different geographic regions of the supermarket chain's trading area, and/or from differences in interbrand competition across the stores of the chain. Drivers of these store-specific differences, in turn, may be locally varying demographic and competitive characteristics of a store's neighborhood, related to income and family size structure or the number of and distances to competing retail stores (e.g. Montgomery, 1997). In the following, we therefore allow for cluster specific (random) scaling factors for every nonlinear function f_j as well as a cluster specific random intercept. This leads to the model

$$y_i = \gamma_{0c_i} + (1 + \gamma_{1c_i}) f_1(z_{i1}) + \cdots + (1 + \gamma_{qc_i}) f_q(z_{iq}) + \varepsilon_i \,, \tag{5}$$

where $c_i \in \{1, \ldots, C\}$ is the cluster index of the i-th observation and the γ_{jc}, $j = 0, \ldots, q$ are normally distributed random effects with mean 0 and variance τ_j^2, i.e.

$$\gamma_{jc} \sim N(0, \tau_j^2) \,, \quad c = 1, \ldots, C \,.$$

The result of the multiplicative effects is the scaling up of the nonlinear function by increasing its steepness if $(1 + \gamma_{jc}) > 1$ and scaling down by decreasing its steepness if $(1 + \gamma_{jc}) < 1$.

In order to estimate the model we assume for the moment that the nonlinear functions f_j are known. Rearranging model terms yields

$$\begin{aligned} y_i &= \gamma_{0c_i} + \gamma_{1c_i} f_1(z_{i1}) + \cdots + \gamma_{qc_i} f_q(z_{iq}) + f_1(z_{i1}) + \cdots + f_q(z_{iq}) + \varepsilon_i \\ &= \gamma_{0c_i} + \gamma_{1c_i} x_{i1} + \cdots + \gamma_{qc_i} x_{iq} + o_i + \varepsilon_i \,. \end{aligned} \tag{6}$$

For known functions f_j equation (6) is a simple random effects model with random intercept γ_{0c}, random slopes γ_{jc}, $j = 1, \ldots, q$ of the pseudo or transformed covariates $x_{ij} := f_j(z_i)$ and with an additional offset

$$o_i = f_1(z_{i1}) + \cdots + f_q(z_{iq}) \,.$$

Model (6) has formally the same structure as the additive or more precisely the varying coefficient model described in section 2.1. The role of the effect modifier is taken by the cluster index c and the interacting variables are the pseudo covariates $x_j = f_j(z_j)$. In matrix notation we may write

$$\mathbf{y} = \tilde{\mathbf{Z}}_0 \boldsymbol{\gamma}_0 + \tilde{\mathbf{Z}}_1 \boldsymbol{\gamma}_1 + \cdots + \tilde{\mathbf{Z}}_q \boldsymbol{\gamma}_q + \mathbf{o} + \boldsymbol{\varepsilon}$$

with design matrices $\tilde{\mathbf{Z}}_0 = \mathbf{C}$ and $\tilde{\mathbf{Z}}_j = \text{diag}(x_{1j}, \ldots, x_{nj})\mathbf{C}$. The $n \times C$ matrix $\mathbf{C}$ is a 0/1 incidence matrix whose entry in the i-th row and k-th column is 1 if the i-th observation belongs to the k-th cluster and 0 otherwise.

For given variance parameters $\tau_0^2, \ldots, \tau_q^2$ the random effects γ_{jc}, $j = 0, \ldots, q$, $c = 1, \ldots, C$ may be estimated by minimizing the following penalized least squares criterion

$$PLS = (\mathbf{y} - \boldsymbol{\eta})'(\mathbf{y} - \boldsymbol{\eta}) + \sum_{j=0}^{q} \tilde{\lambda}_j \boldsymbol{\gamma}_j' \boldsymbol{\gamma}_j \,,$$

where the "smoothing parameters" are given by $\tilde{\lambda}_j = \sigma^2/\tau_j^2$ and $\boldsymbol{\gamma}_j = (\gamma_{j1}, \ldots, \gamma_{jC})'$ are vectors of random effects coefficients. The vectors $\boldsymbol{\gamma}_j$ may be estimated via backfitting analogous to the parameters $\boldsymbol{\beta}_j$ of the nonlinear functions f_j in the preceding subsection. The corresponding smoothers are given by

$$\tilde{S}_j(\mathbf{y}, \tilde{\lambda}) = \tilde{\mathbf{Z}}_j \hat{\boldsymbol{\gamma}}_j \qquad \hat{\boldsymbol{\gamma}}_j = (\tilde{\mathbf{Z}}_j' \tilde{\mathbf{Z}}_j + \tilde{\lambda}_j \mathbf{I}_j)^{-1} \tilde{\mathbf{Z}}_j' \mathbf{y} \,.$$

Simultaneous selection of smoothing parameters and relevant random effects can be done in the same way as described in Subsection 2.2.

We are now prepared to describe estimation of our model (5). Estimation is carried out in the following two steps which may be iterated:

1. In a first step we assume homogeneous functions f_j in model (5), i.e. the random effects coefficients γ_{jc} are assumed to be identical to zero. Using the algorithms of section 2.2 estimates $\hat{f}_j$ of the nonlinear functions are obtained. Because of the built-in model selection some of the functions may be linear or identical to zero.
2. In the second step we estimate the random effects coefficients as described above by keeping fixed the estimated functions $\hat{f}_j$ from the first stage.

3 Simulation

3.1 Setup

The true model consists of the response y and three covariates x_1, x_2, and x_3. They exert a nonlinear influence which is modified by multiplicative effects as given in equation (7),

$$y = (1+\gamma_1)f_1(x_1) + (1+\gamma_2)f_2(x_2) + (1+\gamma_3)f_3(x_3) + \varepsilon\,, \tag{7}$$

where f_1 is the decreasing part of the sine-function in the interval $[1.58; 4.71]$, f_2 is the natural logarithm in the interval $[5; 15]$, f_3 is the value of the cumulative distribution function of the standard normal distribution evaluated in the interval $[-2.5; 2.5]$ and ε is the usual i.i.d. Gaussian error term with some variance σ^2. The γ_j are normally distributed and centered about 0, which means that the multiplicative effects $(1+\gamma_j)$ are centered about 1. We choose different levels of variance for γ_j in order to assess the behavior of the estimation technique for different strengths of the multiplicative effects. In particular, we set the variances equal to 0.4^2, 0.2^2 and 0.1^2 for γ_1, γ_2 and γ_3, respectively:

$$\gamma_1 \sim N(0; 0.4^2) \quad \gamma_2 \sim N(0; 0.2^2) \quad \gamma_3 \sim N(0; 0.1^2)\,.$$

In the simulation setup, we use 100 clusters, each having 50 observations. The effects $(1+\gamma_j)f_j(X_j)$ are shown in Figure 2.

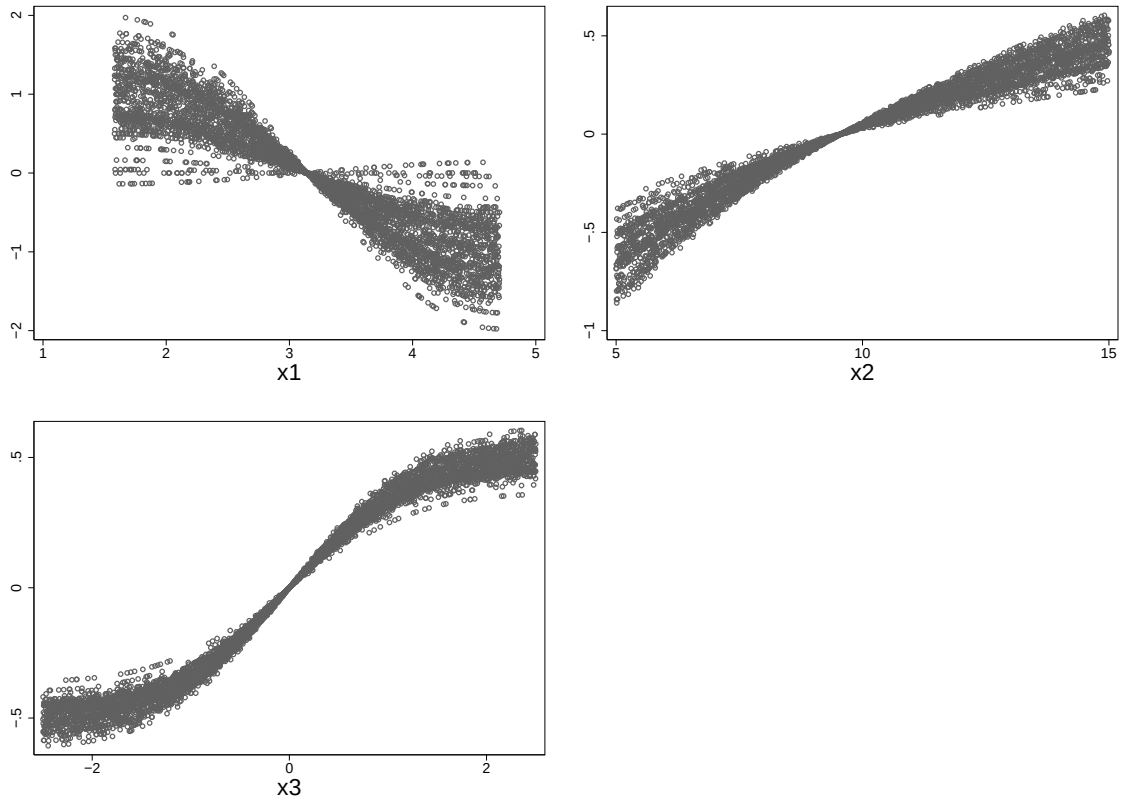

Figure 2: The three nonlinear functions, multiplied with their respective random effects.

Furthermore, we study the three different signal-to-noise ratios 3, 2 and 1. The signal-to-noise ratio is computed as the ratio of the standard deviation of the predictor η to the standard deviation of the error term ε:

$$\text{SNR} = \frac{\sigma_\eta}{\sigma_\varepsilon}\,.$$

A ratio of 3 corresponds to a moderate signal, 2 means a weak signal and 1 implies a very weak signal.

We calculate 250 replications of our model and carry out the estimation procedure described in the previous section. The results are shown in figure 3. From the 100 cluster-specific effects $(1+\gamma_j)f_j(x_j)$ three are picked out for closer inspection. The first is the cluster corresponding to the 5%-, the second to the 50%- and the third to the 95%-quantile of the random effects γ_j. The arithmetic mean from the 250 replications is computed and displayed in the graph (solid). In order to facilitate comparison, the true effects are plotted, too (dashed). Clearly, the average from 250 estimates is more biased for smaller signal-to-noise ratios. Furthermore, the estimate is more biased the weaker the random effect is. These two characteristics are a well-known feature of random effects estimators, see for example Gelman and Hill (2007).

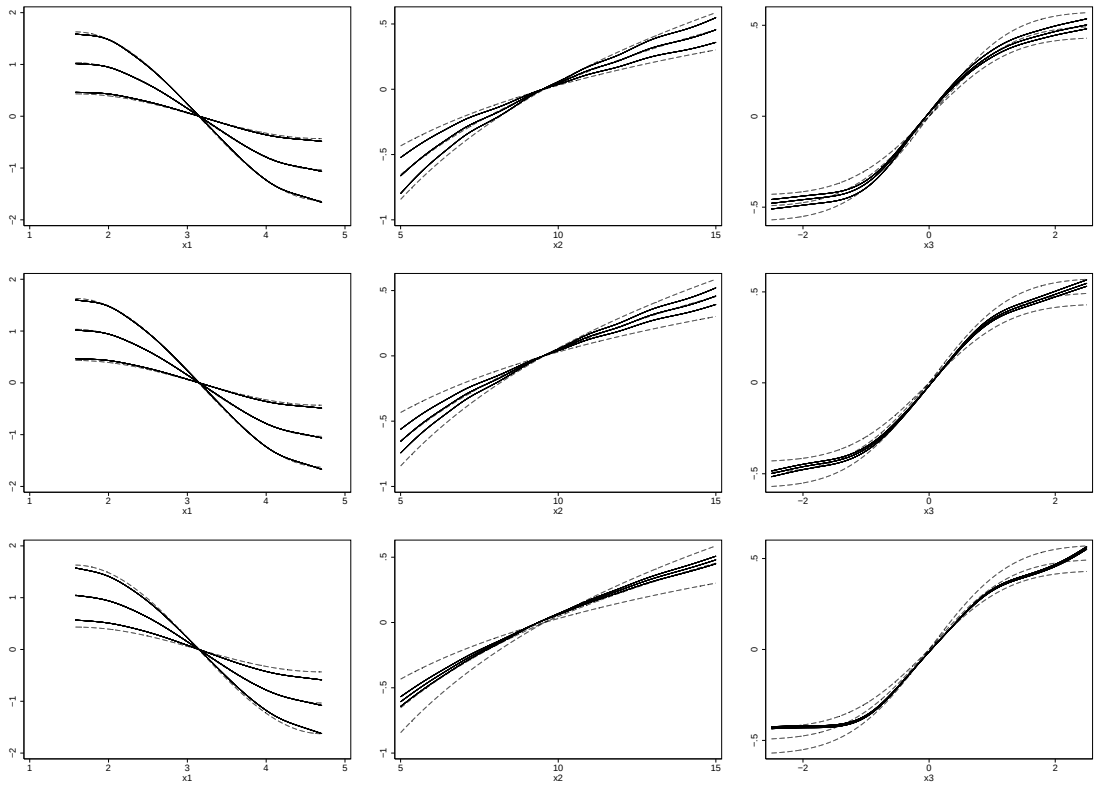

Figure 3: The average estimates of 250 replications (solid) and the true effects (dashed) $(1+\gamma_j)f_j(X_j)$ for the signal-to-noise ratio 3, 2 and 1 in the first, second and third row, respectively.

It remains to be investigated how many errors the estimation procedure has produced, i.e. how often have covariates been eliminated although they actually should have been

Table 1: Number of false exclusions for different signal-to-noise ratios.

Signal-to-noise ratio	γ_1	γ_2	γ_3
3:1	0	0	0
2:1	0	0	28
1:1	0	57	111

included. In step one, regardless of the covariate and the signal-to-noise ratio no errors have occurred. Contrary to that, errors have appeared in the second step. Table 1 points out that the multiplicative effects of f_1 have never been excluded. The weak multiplicative effect of f_3 is excluded most often, with a signal-to-noise ratio of 1 nearly half of the time.

We repeat the analysis using the same 250 replications of the model in equation (7), but now we calculate the model using a fully Bayesian approach with MCMC techniques (see Lang und Brezger, 2004 and Brezger und Lang, 2006 for details). This approach, however, is not able to perform model selection. The results are visually indistinguishable from those in figure 3, which is why we omit the depiction.

We compare the two estimation procedures in terms of their efficiency. For this reason we compute the mean squared error MSE $= \mathrm{E}(\hat{y}-y)^2$ for the 250 replications of the three clusters considered in figure 3. The results are reported in Figure 4. The upper left panel depicts a boxplot of the MSE for a signal-to-noise ratio of 3, the upper right of 2 and the lower of 1. We can see that generally the error increases the weaker the signal is. Next to the MSE of the estimation procedure described in section 2 we also report the MSE of Bayesian techniques, indicated by $MCMC$. It does not, however, differ to a notable extent.

Summarizing the above we can state that the estimation procedure performs reasonably well. Even for a moderate strength of the signal the results are not too much biased. We have seen, though, that there are limits. In particular, with a very low signal-to-noise ratio of 1 there seems to be too much noise to estimate the effects precisely.

4 Application

We apply our methodology to data from "Dominick's Finer Foods", a major supermarket chain operating in the Chicago metropolitan area. The data include weekly unit sales and corresponding retail prices for different brands of orange juice (premium, national and store brands) in 81 stores of the chain over a time period of 89 weeks. In the following, we illustrate our methodology for one of the national brands, the brand "Florida Gold".

To account for multicollinearity and for the fact that cross-item price effects are usually much lower than own-item price effects (see, e.g., Hanssens et al., 2001), we capture cross-price effects at the tier level rather than the individual brand level: we define $price\_premium_{it}$ ($price\_national_{it}$) as the minimum price for a premium brand (national brand) in store *i* and week *t*, while ($price\_dominicks_{it}$) denotes the price of the only private label brand, Dominick's store brand, in store *i* and week *t*. It is important to note that price activities of the national brand Florida Gold are excluded from the computation of $price\_national_{it}$.

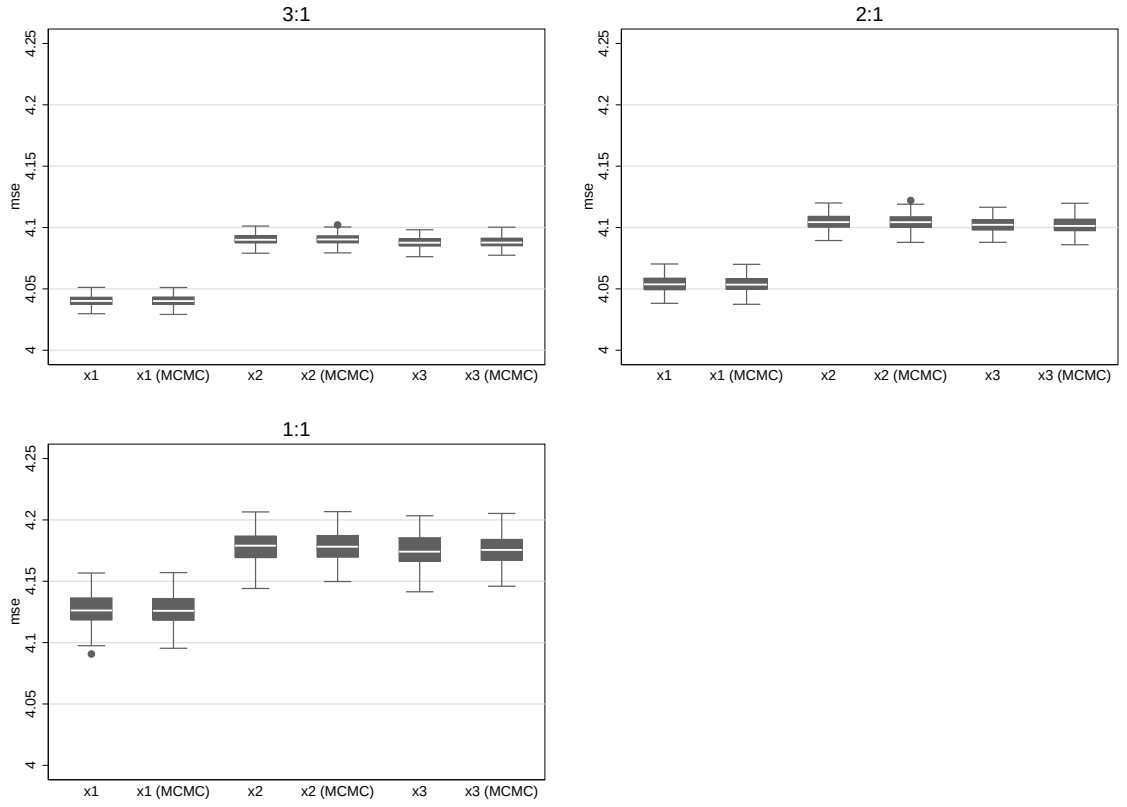

Figure 4: The MSE for a signal-to-noise ratio of 3 (upper left panel), 2 (upper right panel) and 1 (lower panel). For each effect the MSE of the estimation procedure as described in section 2 as well as a Bayesian procedure (indicated with $MCMC$) is reported.

A scatter plot of log unit sales and the own-item price of the brand Florida Gold is shown in the upper left panel of Figure 5, indicating the expected inverse relationship between unit sales and own-item price. The situation is less clear-cut for the impact of competitive prices on the sales of Florida Gold: it is hard to discern the expected direct relationships in the scatter plots since there is much noise in the data.

We apply the estimation procedure described in section 2 and estimate the model

$$\ln q_{i,t} = \gamma_{0i} + (1+\gamma_{1i})f_1 + (1+\gamma_{2i})f_2 + (1+\gamma_{3i})f_3 + (1+\gamma_{4i})f_4 + \varepsilon_{i,t}\,, \qquad (8)$$

where $q_{i,t}$ denotes unit sales of Florida Gold in store i and week t, γ_{0i} is a random intercept accounting for heterogeneity in baseline sales of Florida Gold across different stores, f_1 is a nonlinear function of the price of "Florida Gold", f_2 to f_4 are nonlinear functions of the competitive prices w.r.t. the premium brand tier, the national brand tier and the store brand, respectively, and $\varepsilon_{i,t} \sim N(0, \sigma_\varepsilon^2)$ is the usual i.i.d. Gaussian error term. The nonlinear functions are modeled using P-splines with 20 knots and a second order difference penalty. Note that for ease of notation we have omitted the arguments of the nonlinear functions.

Selection of penalty parameters and relevant terms is carried out using the algorithm described in subsection 2.3. The algorithm deletes the random effect of the premium

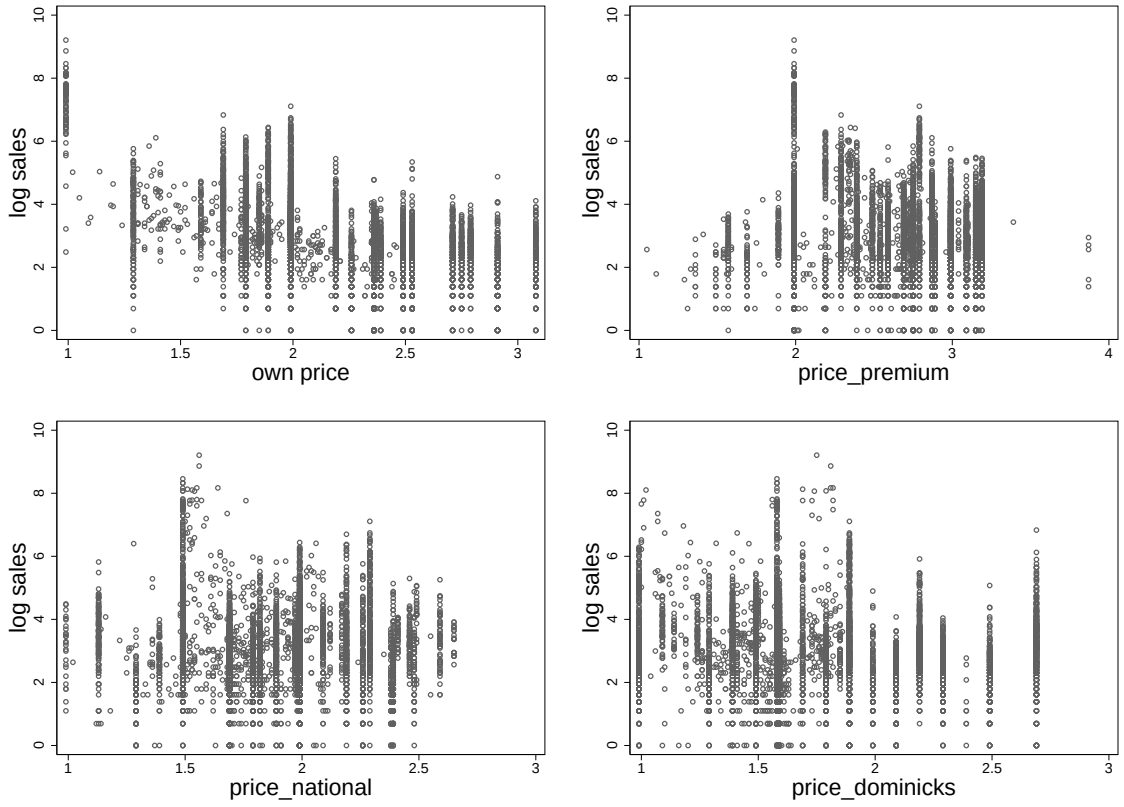

Figure 5: Scatter plots between the log number of sold packages and the price of "Florida Gold" (upper left panel) as well as prices of competitive products in three quality tiers.

brand tier from the model, yet the fixed effect is still incorporated. The resulting model therefore is

$$\log q_{i,t} = \gamma_{0i} + (1+\gamma_{1i})f_1 + f_2 + (1+\gamma_{3i})f_3 + (1+\gamma_{4i})f_4 + \varepsilon_{i,t}\,, \tag{9}$$

where f_2 is homogeneous across outlets. The minimum and maximum random marginal effects, holding all other covariates constant at the mean in the dataset and after transforming log unit sales to unit sales, are shown in figure 6.

The estimated price effects exhibit a highly nonlinear behavior with steps and kinks at certain price points indicating threshold and saturation effects. For example, unit sales of Florida Gold do not increase until the own price falls below about 2.25 dollars. The cross-price response curve with respect to the premium brands shows an inverse L-shape and a strong kink at a price of about 2.25 dollars, below which the unit sales of Florida Gold rapidly decrease. And, unit sales of Florida Gold rapidly increase if the lowest price for one of the competing national brand exceeds 2.5 dollars. We further observe strong heterogeneity across stores for the cross-price effect of Dominick's own orange juice brand.

We replicated the same analysis using MCMC techniques (Lang und Brezger, 2004 and Brezger und Lang, 2006). Figure 7 displays the estimation results. A comparison with Figure 6 reveals only minor differences between the estimation procedures regarding

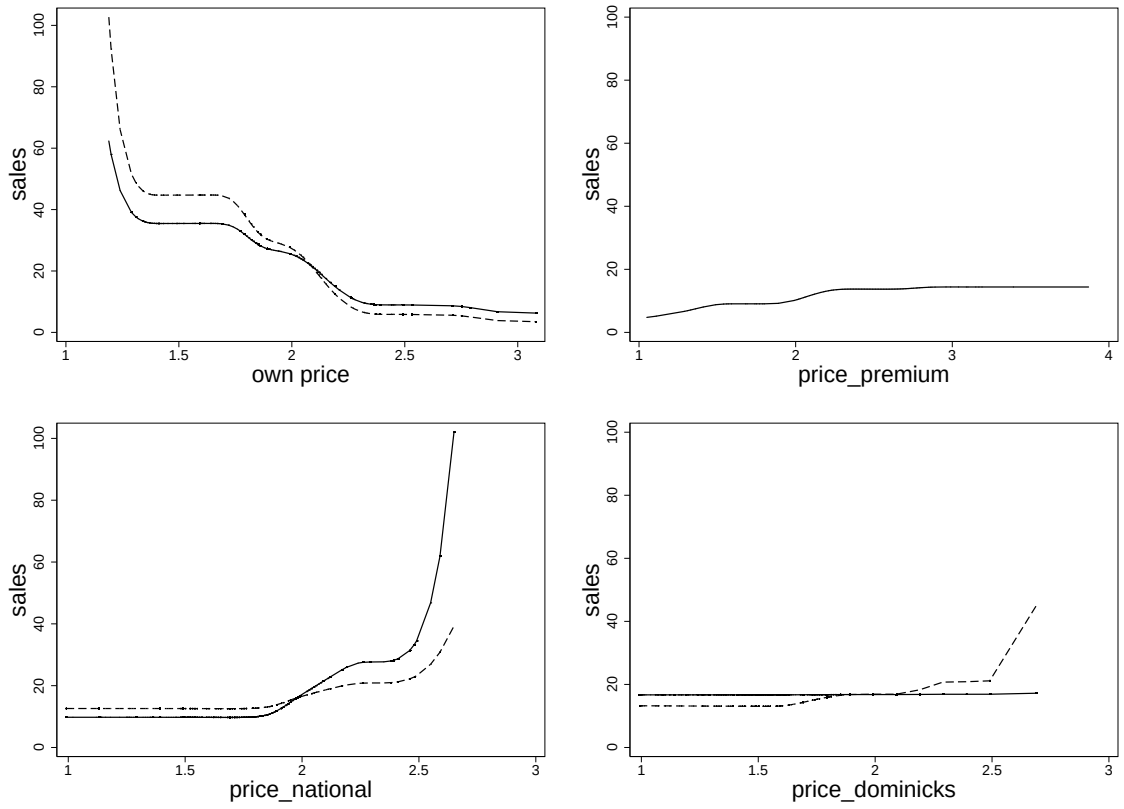

Figure 6: Maximum and minimum random marginal price effects for the three price variables that were selected. The price w.r.t. the premium brand tier (upper right panel) was selected not to vary across outlets.

the shapes of the price effects, except for the premium tier effect which is now much smoother. Furthermore, since this approach is incapable of variable selection, the random effect of *price_premium* is included in addition to the fixed effect (though revealing only little variation).

We compared the model performance in terms of predictive validity to the model

$$\log q_{i,t} = \gamma_0 + f_1 + f_2 + f_3 + f_4 + \varepsilon_{i,t} \tag{10}$$

which contains no random effects at all, and to the model

$$\log q_{i,t} = \gamma_{0i} + f_1 + f_2 + f_3 + f_4 + \varepsilon_{i,t} \tag{11}$$

which only includes the random intercept but not the random scaling factors.

In particular, we randomly split the data into five and ten equally-sized subsets and performed five-fold and ten-fold cross-validation, respectively. The results displayed in Table 2 indicate a considerable improvement in predictive performance when allowing the nonlinear functions to vary across stores.

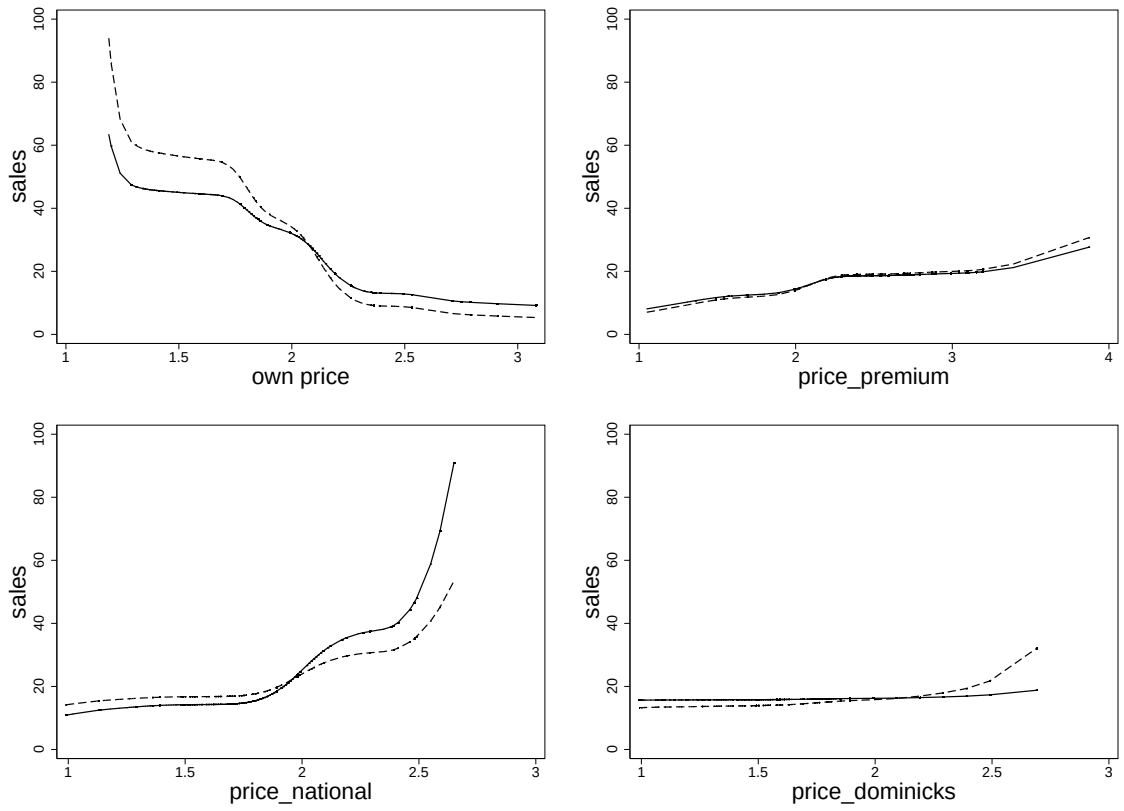

Figure 7: Maximum and minimum random marginal price effects for the four price variables estimated using MCMC.

Table 2: Cross-validation criteria for different model specifications. Smaller values correspond to better model fits.

Model	(8)	(10)	(11)
CV5	0.7243	0.8408	0.8201
CV10	0.7242	0.8406	0.8175

5 Conclusion

The paper provides inference for additive models with random scaling factors and presents an empirical application concerning the estimation of own- and cross item price effects from retail sales data.

Several directions for future research are conceivable. First, the methodology could be extended to non-Gaussian responses. Second, we plan a fully Bayesian version based on MCMC simulation techniques. Third, since the stores of the retail chain exhibit a spatial structure, we plan to introduce spatially correlated random scaling factors rather than uncorrelated factors, as employed in the current paper.

Acknowledgement

The data is provided by the James M. Kilts Center, GSB, University of Chicago. The authors furthermore wish to thank two anonymous referees for helpful suggestions.

References

Belitz, C., und Lang, S. (2007). Simultaneous selection of variables and smoothing parameters in structured additive regression models. *Technical report, submitted for publication*.

Blattberg, R., und Wisniewski, K. (1989). Price-induced patterns of competition. *Marketing Science*, *8*, 291-309.

Bollaerts, K., Eilers, P. H. C., und Mechelen, I. van. (2006). Simple and multiple P-splines regression with shape constraints. *British Journal of Mathematical and Statistical Psychology*, *59*, 451-469.

Brezger, A., und Lang, S. (2006). Generalized structured additive regression based on Bayesian P-splines. *Computational Statistics and Data Analysis*, *50*, 947-991.

Brezger, A., und Steiner, W. (2007). Monotonic regression based on Bayesian P-splines: An application to estimating price response functions from store-level scanner data. *Journal of Business and Economic Statistics*. (In press)

De Boor, C. (2001). *A Practical Guide to Splines*. Springer, New York.

Eilers, P. H. C., und Marx, B. D. (1996). Flexible smoothing using B-splines and penalized likelihood (with comments and rejoinder). *Statistical Science*, *11*, 89-121.

Fahrmeir, L., Kneib, T., und Lang, S. (2007). *Regression. Modelle, Methoden und Anwendungen*. Berlin: Springer.

Gelman, A., und Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press.

Hanssens, D., Parsons, L., und Schultz, R. (2001). *Market Response Models - Econometric and Time Series Analysis*. London: Chapman and Hall.

Hastie, T., und Tibshirani, R. (1990). *Generalized Additive Models*. London: Chapman and Hall.

Hastie, T., und Tibshirani, R. (1993). Varying–Coefficient Models. *Journal of the Royal Statistical Society B*, *55*, 757-796.

Heerde, H. van, Leeflang, P., und Wittink, D. (2001). Semiparametric analysis to estimate the deal effect curve. *Journal of Marketing Research*, *38*, 197-215.

Heerde, H. van, Leeflang, P., und Wittink, D. (2002). How promotions work: SCAN*PRO-based evolutionary model building. *Schmalenbach Business Review*, *54*, 198-220.

Kalyanam, K., und Shively, T. (1998). Estimating irregular pricing effects: a stochastic spline regression approach. *Journal of Marketing Research*, *35*, 19-29.

Lang, S., und Brezger, A. (2004). Bayesian P-splines. *Journal of Computational and Graphical Statistics*, *13*, 183-212.

Montgomery, A. (1997). Creating micro-marketing pricing strategies using supermarket scanner data. *Marketing Science*, *16*, 345-337.

Ramsay, J., und Silverman, B. (2002). *Applied Functional Data Analysis*. Berlin: Springer.

Ramsay, J., und Silverman, B. (2005). *Functional Data Analysis*. Berlin: Springer.

Wood, S. N. (2006). *Generalized Additive Models: An Introduction with R*. Chapman and Hall.

Authors' Addresses:

Peter Wechselberger, Stefan Lang
Department of Statistics
University of Innsbruck
Universitätsstraße 15
6020 Innsbruck
Austria

Winfried Steiner
Department of Marketing
Technial University of Clausthal
Julius-Albert-Straße 2
38678 Clausthal-Zellerfeld
Germany

E-Mail: peter.wechselberger@uibk.ac.at