Semiparametric Stepwise Regression to Estimate Sales Promotion Effects

Winfried J. Steiner[1], Christiane Belitz[2], and Stefan Lang[3]

[1] Department of Marketing,
University of Regensburg, 93040 Regensburg, Germany
[2] Department of Statistics,
University of Munich, 80539 Munich, Germany
[3] Institute of Empirical Economic Research,
University of Leipzig, 04109 Leipzig, Germany

Abstract. Kalyanam and Shively (1998) and van Heerde et al. (2001) have proposed semiparametric models to estimate the influence of price promotions on brand sales, and both obtained superior performance for their models compared to strictly parametric modeling. Following these researchers, we suggest another semiparametric framework which is based on penalized B-splines to analyze sales promotion effects flexibly. Unlike these researchers, we introduce a stepwise procedure with simultaneous smoothing parameter choice for variable selection. Applying this stepwise routine enables us to deal with product categories with many competitive items without imposing restrictions on the competitive market structure in advance. We illustrate the new methodology in an empirical application using weekly store-level scanner data.

1 Introduction

Kalyanam and Shively (1998) and van Heerde et al. (2001) have proposed nonparametric techniques (a kernel-based and a stochastic spline regression approach, respectively) to estimate promotional price effects. In both studies, the authors obtained superior performance for their semiparametric models compared to strictly parametric modeling. The empirical results of these two studies indicate that own- and cross-promotional price effects may show complex nonlinearities which are difficult or not at all to capture by parametric models. Moreover, no unique patterns for own- and cross-promotional price response curves generalizable across or even within product categories could be identified. These findings strongly support the use of nonparametric techniques to let the data determine the shape of promotional price response functions. A recent empirical comparison of parametric and seminonparametric sales response models (the latter specified as multilayer perceptrons) conducted by Hruschka (2004) also provides superior results for the more flexible neural net approach.

We follow Kalyanam and Shively (1998) and van Heerde et al. (2001) and propose a semiparametric model based on penalized B-splines to estimate

sales promotion effects flexibly. We add to the body of knowledge by suggesting a stepwise regression procedure with simultaneous smoothing parameter choice for variable selection. Applying this stepwise routine enables us to deal with product categories with many competing brands and to resolve the problem of identifying relevant cross-promotional effects between brands without imposing restrictions on the competitive market structure in advance. Since cross-item price effects are usually much lower in magnitude than own-item price effects (e.g., Hanssens et al. (2001)), and frequently not all competing brands in a product category are close substitutes to each other (e.g., Foekens (1995)), a stepwise selection to reduce the number of predictors in a sales response model seems very promising. Many previous approaches to analyze sales response to promotional activities have tackled this problem by imposing restrictions on the competitive market structure, e.g., by capturing competitive promotional effects in a highly parsimonious way through the use of a single competitive variable (e.g., Blattberg and George (1991), Kopalle et al. (1999)) or by focusing only on a limited number of major brands in a product category (e.g., Kalyanam and Shively (1998), van Heerde et al. (2001)).
The paper is organized as follows: in section 2, we propose the semiparametric model to estimate promotional effects and provide details about the P-splines approach we use to model the unknown smooth functions for own- and cross-promotional price effects; in section 3, we introduce the stepwise routine which includes a simultaneous smoothing parameter selection for the continuous price variables; in section 4, we illustrate the new methodology in an empirical application using weekly store-level scanner data for coffee brands; section 5 summarizes the contents of the paper.

2 A semiparametric approach to analyze promotional data

To estimate sales promotion effects, we model a brand's unit sales as (1) a nonparametric function of own- and cross-item price variables using penalized B-splines (e.g., Eilers and Marx (1996), Lang and Brezger (2004)) and (2) a parametric function of other promotional instruments:

$$\begin{aligned} ln(Q_{is,t}) = &\sum_s \alpha_{is} O_s + \sum_j f_{ij}(P_{js,t}) + f_{ii}(P_{is,t-1}) + \\ &\sum_j \sum_k \gamma_{ijk} D_{jks,t} + \sum_q \delta_{iq} W_{q,t} + \varepsilon_{is,t}; \quad \varepsilon \sim N(0, \sigma^2), \end{aligned} \tag{1}$$

where

$Q_{is,t}$: unit sales of item i (brand i) in store s and week t;
O_s: store dummy to capture heterogeneity in baseline sales of brand i across different stores;

$f_{ij}(P_{js,t})$: unknown smooth functions for the effect of own-item price ($j = i$) and prices of competing items ($j \neq i$) on unit sales of brand i;
$P_{js,t}$: actual price of item j in store s and week t;
$P_{is,t-1}$: lagged price of item i in store s and week t;
$D_{jks,t}$: indicator variables capturing usage (= 1) or nonusage (= 0) of non-price promotional instrument k (e.g., display, feature) for brand j in store s and week t;
$W_{q,t}$: seasonal dummy indicating if public holiday q falls in week t (= 1) or not (= 0);
α_{is}, δ_{iq}: store intercept for item i and store s, effect of holiday q on unit sales of brand i;
γ_{ijk}: effect of non-price promotional instrument k of item j on unit sales of brand i, representing own ($j = i$) and cross ($j \neq i$) promotional effects;

As common in commercially applied sales response models, we pool the data across stores and focus on one brand at a time (e.g., Wittink et al. (1988), van Heerde et al. (2002)). We use log unit sales ($ln(Q_{is,t})$) instead of unit sales to normalize the distribution of the criterion variable which is typically markedly skewed with promotional data. We further include indicator variables ($W_{q,t}$) to account for "seasonal" fluctuations in a brand's unit sales due to holidays (e.g., Christmas, Easter). We also include a lagged variable for own price ($P_{is,t-1}$) to accomodate the fact that promotions often accelerate sales of a brand during the promotional period leading to a trough after the promotional period (e.g., Blattberg and Neslin (1990), Blattberg and George (1991)).

To model the unknown smooth functions for own- and cross-price effects, we adopt the P-splines approach proposed by Eilers and Marx (1996). This approach can be characterized by three properties: (a) It is assumed that the unknown functions f_{ij} (or f_{ii}) can be approximated by a spline of degree l with equally spaced knots within the range of the respective price P_j. We use cubic splines and, hence, assume degree 3. Suppressing brand index i, store index s and time index t, we can write such a spline in terms of a linear combination of M_j cubic B-spline basis functions B_{jm}, $m = 1, \ldots, M_j$:

$$f_j(P_j) = \sum_{m=1}^{M_j} \beta_{jm} B_{jm}(P_j), \tag{2}$$

where

B_{jm}: m-th B-spline basis function;
β_{jm}: regression coefficient for the m-th B-spline basis function.

It would be beyond the scope of the paper to go into the details of B-splines. We refer to De Boor (1978) as a key reference. (b) Eilers and Marx (1996) suggest to use a moderately large number of knots to ensure enough flexibility

for the unknown functions. For simplicity, we use 20 knots for every price response curve, i.e., $M_j = M = 20$. (c) To guarantee sufficient smoothness of the fitted curves, a roughness penalty based on squared differences (of order k) of adjacent B-spline coefficients is specified.
Let v_n denote the vector of all parametric effects of the model for the nth observation and let index j, $j = 1, \ldots, J+1$ cover all smooth functions for own- and competitive price effects (including the lagged own-price effect as the $(J+1)$th price effect), this leads to the penalized least-squares criterion

$$\sum_{n=1}^{N}\left(y_n - \sum_{j=1}^{J+1} f_j(P_{jn}) - \mathrm{v}_n'\zeta\right)^2 + \sum_{j=1}^{J+1}\lambda_j \sum_{l=k+1}^{M}(\Delta^k\beta_{j,l})^2, \tag{3}$$

where

N : sample size as product of number of stores and number of weeks;
Δ^k: differences of order k between adjacent regression coefficients;
λ_j: smoothing parameter for function f_j.

In the following, we restrict ourselves to penalties based on second order differences, i.e., $\Delta^k\beta_{j,l} = \beta_{j,l} - 2\beta_{j,l-1} + \beta_{j,l-2}$. The penalized sum of squared residuals (3) is minimized with respect to the unknown regression coefficients β_{jm} (compare equation (2)) and ζ. The trade off between flexibility and smoothness is controlled by the smoothing parameters λ_j, $j = 1, \ldots, J+1$, which are determined within the stepwise routine (see section 3). Estimation of the semiparametric model (1) given the smoothing parameters is carried out with backfitting (Hastie and Tibshirani (1990)).
To give a benchmark for the performance of the semiparametric model (1), we compare it in our empirical application presented in section 4 to the exponential model (4), which is one of the most widely used parametric models to analyze sales response (e.g., Montgomery (1997), Kalyanam and Shively (1999)):

$$\begin{aligned} ln(Q_{is,t}) = {} & \sum_s \alpha_{is}' O_s + \sum_j \beta_{ij} P_{js,t} + \nu_i P_{is,t-1} + \\ & \sum_j \sum_k \gamma_{ijk}' D_{jks,t} + \sum_q \delta_{iq}' W_{q,t} + \varepsilon_{is,t}'; \quad \varepsilon' \sim N(0,\sigma^2). \end{aligned} \tag{4}$$

Model (4) differs from model (1) only with respect to own- and cross-price effects which are specified linearly (parametrically).

3 Stepwise routine with simultaneous smoothing parameter selection

Based on the P-splines approach outlined above, we suggest a stepwise regression procedure for markets with many competing brands and promotional instruments. This procedure does not only allow for variable selection but also enables to determine the degree of smoothness of effects which can be modeled nonparametrically. The objective of using the stepwise routine is to select relevant predictors (and especially relevant cross-promotional effects) for the unit sales of a brand under consideration, while at the same time not losing much explanatory power by excluding other variables from the model. Importantly, by obtaining a parsimonious sales response model that way, overspecification effects arising from the inclusion of all possible but not necessarily important cross effects (typically reflected by unreliable coefficients resulting from overparametrization, wrong signs and unexpected magnitudes of coefficients due to multicollinearity) can be avoided.

The stepwise procedure works as follows: For each independent variable, we consider a hierarchy of specification alternatives defined in terms of equivalent degrees of freedom *df*. It is well-known that the equivalent degrees of freedom *df* of a smooth function can be calculated from the trace of the corresponding smoother matrix (which in turn depends on the smoothing parameter value), and it is common practice to choose the value of a smoothing parameter simply by specifying the *df* for the smooth (Hastie and Tibshirani 1990). Clearly, there are only two possible specifications for indicator variables (like display, feature or seasonal dummy variables): excluded from the model ($df = 0$) or included in the model ($df = 1$). For the continuous price variables, however, we allow for a much broader interval of possible values for degrees of freedom ranging at integer increments from $[0; 10]$. Setting $df = 0$ implies that the respective price variable is excluded from the model. For $df = 1$, the effect is included linearly. With increasing *df* (i.e., decreasing smoothing parameter), the penalty term in expression (3) becomes less important and the estimated function gets more and more rough.

Variable selection starts from the linear model, which includes all independent variables at $df = 1$ (i.e., parametrically). In each iteration, a set of new models is estimated by passing through the independent variables successively: (a) For each independent variable, the number of *df* is increased and decreased by one (where feasible) and the respective models are estimated leaving the number of *df* with respect to all other independent variables unchanged; (b) From the pool of new models estimated, the best model is then determined according to the BIC criterion:

$$BIC = N \cdot ln(\hat{\sigma}^2) + ln(N) \cdot df_{total} \tag{5}$$

where

$\hat{\sigma}^2$: estimated variance for the error term $\varepsilon_{is,t}$;
df_{total} : overall degrees of freedom

It is convenient to approximate df_{total} by adding up the degrees of freedom used for the individual functions/terms included in the model (Hastie and Tibshirani (1990)). (c) If the BIC of the best model selected is less (i.e., better) than the BIC of the start model, GO TO (b) and use the selected model as the new start model; otherwise STOP.

4 Empirical study

In this section, we present results from an empirical application of our semiparametric framework to weekly store-level scanner data for nine brands of coffee offered in five German supermarkets. The data were provided by MADAKOM GmbH (50825 Cologne, Germany) and include unit sales, retail prices and deal codes indicating the use of non-price promotional instruments (display, feature, other advertising activities) for the nine brands over a time span of 104 weeks. Table 1 shows summary statistics pooled across the stores for average, minimum and maximum market shares as well as price ranges of the individual brands. The weekly market shares of all brands vary considerably reflecting the frequent use of price promotions.

brand	average market share	minimum market share	maximum market share	price range
1	24.94	6.67	62.16	[5.99; 8.49]
2	8.85	0.88	57.13	[4.99; 7.49]
3	20.00	2.26	74.41	[4.99; 7.49]
4	16.83	1.91	50.30	[6.99; 8.99]
5	2.80	0.20	17.89	[5.99; 8.99]
6	5.00	0.27	30.21	[5.99; 7.99]
7	5.15	0.45	25.75	[5.99; 7.99]
8	8.14	0.64	33.24	[5.99; 7.99]
9	8.29	1.42	45.32	[5.99; 9.49]

Table 1. Market shares and price ranges

Table 2 shows the BIC values for the best semiparametric and exponential models selected by the stepwise routine, respectively. Importantly, the stepwise procedure based on the exponential model (4) was only allowed to select own- and cross-price effects parametrically (i.e., at $df = 1$), as opposed to the stepwise selection with simultaneous smoothing parameter choice based on the semiparametric model (1). In addition, the overall degrees of freedom for the start model $df_{total/start}$, which includes *all* effects parametrically, versus the overall degrees of freedom for the best semiparametric model $df_{total/semipar}$ are reported as a kind of benchmark for model improvement through variable selection. With the exception of brands 3, 7 and 8, the semiparametric approach clearly outperforms the strictly parametric approach, with the most dramatic improvement occurring for brand 9. The improvement from nonparametric modeling is only slightly for brand 7, while no differences between the final models occur with respect to brands 3 and 8. The latter implies that nonparametric modeling of price effects does not matter for these two brands, and that the semiparametric model here actually degenerates into the exponential model. A comparison between the number of degrees of freedom used in the start model to those used in the best semiparametric model demonstrates the usefulness of the stepwise routine in providing very parsimonious sales response models.

brand	semiparametric model	exponential model	$df_{total/start}$	$df_{total/semipar}$
1	-1001.13	- 969.15	66	23
2	- 319.18	- 311.13	64	19
3	- 873.38	- 873.38	60	13
4	-1201.98	-1182.00	62	18
5	- 556.69	- 536.91	64	16
6	- 542.31	- 512.17	61	12
7	- 521.62	- 519.85	60	15
8	- 502.43	- 502.43	61	14
9	- 616.63	- 543.38	62	15

Table 2. BIC values and overall degrees of freedom

The following results refer to the brands for which the improvement in BIC values for the semiparametric approach is substantial (i.e., not brands 3, 7 and 8): (a) For five out of six brands, exactly the same price variables were selected in the semiparametric model (1) and the exponential model (4). This implies that the greater flexibility in nonlinear effects for the price variables provided by the semiparametric approach is the reason for the BIC improvement relative to the exponential model. (b) For each brand, the current own-price effect is included nonparametrically. For three brands, however, the BIC improvement can also be attributed to nonparametrically selected cross-price effects which show strong nonlinearities. (c) Out of 72 possible cross-price effects (8 per brand), only 21 were selected across brands. This confirms previous empirical findings that only some of the brands in a product category may be close substitutes to each other. (d) Nearly all selected non-price promotional instruments (referring to the use of display, feature and other advertising activities) have signs in the expected direction, i.e., positive for own-promotional effects, negative for cross-promotional effects. Figure 1 illustrates the differences between the semiparametric model and the strict parametric one, considering the estimated own-item price effect for brand 9 and the cross-item price effect of brand 4 on the unit sales of brand 6 as two examples. Although the nonparametric and parametric own price response curves for brand 9 are shaped rather similar, the differences in predicted sales are substantial. In particular, the parametric model dramatically understates the effect for low prices (up to a difference of 800 units at 5.99), and it overstates the effect for medium prices. With respect to the cross-promotional price effect of brand 6 on brand 4, the parametric model understates the sales effect for low and high prices and overstates the sales effect for medium prices. Importantly, the nonparametric curve reveals a threshold effect at 6.99, up to which the unit sales of brand 4 are insensitive to price changes of brand 6.

5 Conclusions

We presented a semiparametric regression model including a stepwise procedure for variable selection to analyze promotional data. While the semipara-

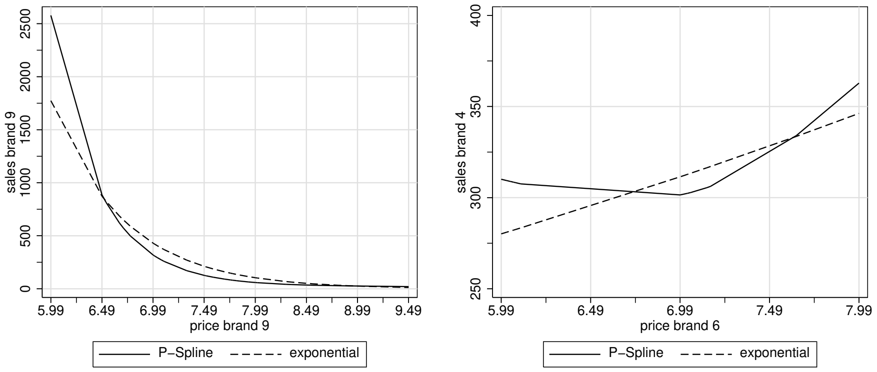

Fig. 1. Nonparametrically estimated own-/cross-promotional price effects

metric model provides high flexibility in modeling nonlinear effects for the continuous price variables, the stepwise routine is used to identify the relevant predictors in markets with many competing items and many promotional instruments. The new approach is illustrated in an empirical application using weekly store-level scanner data.

References

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. and NESLIN, S.A. (1990): *Sales Promotion: Concepts, Methods, and Strategies.* Englewood Cliffs, New Jersey.

DE BOOR, C. (1978): *A Practical Guide to Splines.* Springer, New York.

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.*

FOEKENS, E.W. (1995): *Scanner Data Based Marketing Modelling: Empirical Applications.* Labyrinth Publications, The Netherlands.

HANSSENS, D.M., PARSONS L.J. and SCHULTZ, R.L. (2001): *Market Response Models: Econometric and Time Series Analysis.* Chapman & Hall, London.

HASTIE, T. and TIBSHIRANI, R. (1990): *Generalized Additive Models.* Chapman & Hall, London.

HRUSCHKA, H. (2004): Relevance of Functional Flexibility for Heterogeneous Sales Response Models. A Comparison of Parametric and Seminonparametric Models. *Discussion Paper 394, Faculty of Economics, University of Regensburg.*

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.*

MONTGOMERY, A.L. (1997): Creating Micro-Marketing Pricing Strategies Using Supermarket Scanner Data. *Marketing Science, 16(4), 315–337.*

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.*

VAN HEERDE, H.J., LEEFLANG, P.S.H. and WITTINK, D.R. (2002): How Promotions Work: SCAN*PRO-Based Evolutionary Model Building. *Schmalenbach Business Review, 54(3), 198–220.*

WITTINK, D.R., ADDONA, W.J. HAWKES and PORTER, J.C. (1988): SCAN*PRO: The Estimation, Validation and Use of Promotional Effects Based on Scanner Data. *Internal Paper, Cornell University.*