

Semiparametric Analysis to Estimate the Deal Effect Curve

Harald J. van Heerde Peter S.H. Leeflang Dick R. Wittink *

SOM-theme B Marketing and interactions between firms

Also available (and downloadable) in electronic version: <http://www.eco.rug.nl/SOM/>

* Harald J. van Heerde is Assistant Professor of Marketing and Marketing Research, Department of Business Administration and Management Sciences, Faculty of Economics, University of Groningen, P.O. Box 800, 9700 AV Groningen, The Netherlands. Peter S.H. Leeflang is Professor of Marketing at the same department. Dick R. Wittink is the General George Rogers Clark Professor of Management and Marketing at the Yale School of Management, New Haven, CT, and Professor of Marketing and Marketing Research, at the University of Groningen. The first author can be contacted by telephone (+31 50 3637093), fax (+31 50 3637202), or by email (H.J.VAN.HEERDE@ECO.RUG.NL). All authors thank ACNielsen (USA and The Netherlands) and an anonymous third party for providing the data, and Makoto Abe, Gary Erickson, David Gautschi, Robert Jacobson, Bertrand Melenberg, Ton Steerneman, Roland Rust, Mike Waldman for useful suggestions.

Abstract

The marketing literature suggests several phenomena that may contribute to the shape of the relationship between sales and price discounts. These phenomena can produce severe nonlinearities and interactions in the curves, and we argue that those are best captured with a flexible approach. Since a fully nonparametric regression model suffers from the “curse of dimensionality,” we propose a semiparametric regression model. Store-level sales over time is modeled as a nonparametric function of own- and cross-item price discounts, and a parametric function of other predictors (all indicator variables).

We compare the predictive validity of the semiparametric model with that of two parametric benchmark models and obtain better performance on average. The results for three product categories indicate a.o. threshold- and saturation effects for both own- and cross-item temporary price cuts. We also show how the own-item curve depends on other items’ price discounts (flexible interaction effects). In a separate analysis, we show how the shape of the deal effect curve depends on own-item promotion signals. Our results indicate that prevailing methods for the estimation of deal effects on sales are inadequate.

1. Introduction

One of the key issues with limited empirical results in sales promotion research is the shape of the deal effect curve (Blattberg and Neslin 1990, pp. 350-351; Blattberg, Briesch, and Fox 1995). The deal effect curve shows how sales responds to temporary price discounts. Little is known about the shape of the deal effect curve. For example, Blattberg, Briesch, and Fox (1995, p. G127) ask whether the curve is linear, concave, convex, or S-shaped. The shape is important because if it is convex, the firm will run deeper deals than if the effect is concave.

Wisniewski and Blattberg (1983) found the deal shape fits an S-shaped function. However, their result is based on an analysis of promotion effects across multiple product categories. Consequently this finding may be confounded by category differences. Blattberg and Wisniewski (1987) compared alternative specifications of deal effects within product categories and found the the curve to be convex. However, since their data covered a limited range of deal discounts, we cannot be sure about the absence of a saturation point.

Given the multitude of effects that can shape the deal effect curve it is appropriate to explore this shape flexibly. Therefore, we propose a semiparametric regression model for the deal effect curve, based on store-level scanner data. Our model is semiparametric in the sense that there is both a nonparametric and a parametric part. An important advantage of a nonparametric model is the flexibility in modeling both main- and interaction effects. However, a fully nonparametric model suffers from the curse of dimensionality: as the number of dimensions (number of predictors) increases, the required number of observations explodes. Given the available sample sizes and the number of predictors, we cannot include all predictors nonparametrically. We therefore focus on the flexible estimation of deal effects only: sales is modeled as a nonparametric function of own- and cross-item price discounts. We include the remaining predictors (indicator variables for all but one data set) parametrically.

The nonparametric part of the model allows us to explore the shape of the deal effect curve separately for each of the own-item effects as well as for cross-item effects. The flexibility in interaction effects allows us to determine how the own-effect curve changes at different discount levels for other items. Importantly, the nonparametric part accommodates three-dimensional functions, compared to only two-dimensional ones in the current literature. Empirically, the curves have a considerable amount of similarity across product categories and across countries (USA and the Netherlands). This is notable because the promotional intensity in the Netherlands is far less than it is in the USA.

This paper is organized as follows. In section 2 we review the literature on the deal effect curve. In section 3 we propose our modeling approach. We describe the data in section 4, provide empirical results in section 5, and present our conclusions in section 6.

2. The deal effect curve

The literature suggests the following four phenomena relevant to the shape of the deal effect curve:

- Threshold effects (Gupta and Cooper 1992);
- Saturation effects (Blattberg, Briesch, and Fox 1995; Gupta and Cooper 1992);
- Promotion signal effects (Inman, McAlister, and Hoyer 1990); Mayhew and Winer 1992);
- Cross-item price effects (Sethuraman 1996).

2.1 Threshold effects

Although many managers believe that the minimum value of a price promotion required to change consumers' purchases is 15 percent (Della Bitta and Monroe 1980), there is little empirical support for this claim. An exception is an experiment by Gupta and Cooper (1992) which shows that consumers do not change their intentions unless the promotional discount exceeds a threshold level of 15 percent. Bucklin and Gupta (2000) conclude that price thresholds remain one of the key questions in pricing research.

2.2 Saturation effects

Two mechanisms may cause saturation effects in the deal effect curve. One is a limit on the amount consumers can stockpile (Blattberg, Briesch, and Fox 1995) and/or consume in response to a price discount. The other is a "discounting of discounts" effect. Consumers evaluate and encode information provided to them, and it is their perception of the information and not the information itself that affects their behavior. It has been shown that consumers' perceptions of discounts are typically less than the real discounts (see, e.g., Blair and Landon 1981; Mobley, Bearden, and Teel 1988). Gupta and Cooper (1992) suggest that consumers discount the discounts. Based on experimental data they estimated saturation levels to occur at a 20-30 percent discount level for aerobic shoes.

2.3 Promotion signal effects

Inman, McAlister, and Hoyer (1990) and Mayhew and Winer (1992) suggest that “low need for cognition” consumers react to the simple presence of a promotion signal whether or not the price of the promoted item is reduced. We consider this issue separately, when we examine the deal effect curve shape across promotion signals.

2.4 Cross-item deal effects

Cross-item deal effects may show nonlinearities as well. For example, Sethuraman (1996) analyzed the effect of price discounts by high-priced items on the sales of low-priced items. Graphically, his model implies a linearly increasing cross-item deal effect with a step at the price of the focal item (Sethuraman 1996, Figure 2).

2.5 Contributions of this research

Given the multitude of effects suggested by the four phenomena described above, it is important to explore the form of the deal effect curve flexibly. Consequently, we employ *nonparametric* techniques. The term nonparametric refers to the flexible functional form of the regression curve (Härdle 1990), but this flexibility pertains also to interaction effects.

This research makes the following contributions:

- *Substantively*, previous studies considered the effects of only one or two of the phenomena on the shape of the deal effect curve. Gupta and Cooper (1992) studied promotion threshold effects and discounting of discounts effects, whereas Sethuraman (1996) examined cross-item effects. Our research explores the nature of the deal effect curve more generally: we let the data determine the shape of the entire response curve;
- *Methodologically*, we provide an application of nonparametric regression techniques. These techniques have seen few applications in marketing because of the “curse of dimensionality”. To manage this problem we use a *semiparametric* (semilinear) regression model. To our knowledge this is the first application of a semiparametric *regression* model in marketing (i.e., the criterion variable has an interval scale at minimum). Both Abe (1995) and Bult and Wansbeek (1995) applied a semiparametric *discrete choice* model in a marketing context, and Hruschka (1998) presents a semiparametric regression (market share) model, but his model was developed later than ours;
- We estimate all deal effect curves with *store-level scanner data*, whereas previous research used experimental data to estimate own-item deal effect

curves (Gupta and Cooper 1992), or used store-level scanner data to estimate cross-item deal effect curves only (Sethuraman 1996). Gupta and Cooper (1992) mention that “finding promotion threshold- and saturation effects in choice data (e.g., scanner data) will therefore be a useful next step”. A basic data requirement for estimating the deal effect curve is that the deal sizes have sufficient variation. We employ three store-level scanner data sets with different amounts of deal size variation;

- We estimate both *own-item* and *cross-item deal effect curves*. Sethuraman (1996) only focuses on cross-item deal effects, while Gupta and Cooper (1992) consider just own-item deal effects.
- We explore *interaction effects* between own- and cross-item price discounts, and in a separate analysis between own-item price discounts and own-item promotion signals;
- We *compare* the semiparametric model with two parametric benchmark models based on fit and predictive validity, and contrast the flexible deal effect curves with standard ones.

It is also useful to contrast our study with Kalyanam and Shively (1998) who use a spline regression approach to relate sales and price. We use a more flexible Kernel regression approach (explained in section 2.3.3) which also differs from theirs in the following respects. One, they do not distinguish between regular and promotional prices. We isolate deal effects which have been reported to be different from regular price effects (Blattberg, Briesch, and Fox 1995). Two, the flexibility of our approach allows us to model own-brand, cross-brand, as well as interaction effects. Three, we demonstrate superior predictive validation for our semiparametric approach relative to two traditional models. Kalyanam and Shively do not allow for flexible interaction effects, and they do not consider predictive validities.¹

3. Models

3.1 Rationale for regression-type approach

We use a regression-type approach for the deal effect curve based on store-level scanner data. Household-level purchase data are potentially richer for describing consumer purchase behavior. As we have already indicated in section 1.6, we think

¹ Wedel and Leeflang (1998) present a study that is related to Kalyanam and Shively (1998). Whereas Wedel and Leeflang propose a spline model for psychologic price effects based on *survey* data, the model by Kalyanam and Shively is calibrated with *scanner* data.

it is important for researchers to also develop and use methods for the estimation of models based on store data.

Since we cannot relate the empirical shapes of the deal effect curves to individual consumer behavior, our research deals with the question “what is the shape of the deal effect curve?” and not with the question “why is the shape of the deal effect curve the way it is?” Bucklin and Gupta (2000) indicate that the regression approach applied to (pooled) store-level scanner data is the typical approach used in practice (IRI, AC Nielsen) for estimating constant deal elasticities. We relax the assumption of constant elasticities.

3.2 Parametric approach

Studies of promotional effects typically use parametric regression models that relate sales to own- and other items’ promotional instruments. Blattberg and Wisniewski (1989) use a semilog functional form, whereas Wittink et al. (1988) propose a log-log model. The parametric modeling strategy has a number of optimality properties (Powell 1994, pp. 2445-6). However, the model specification is subject to uncertainty such that the estimates may be biased and/or inconsistent. One can consider alternative parametric specifications and perform standard specification tests, but there is no guarantee that any of the parametric specifications considered will be the true one. In fact, there is no guarantee that the true functions belong to any parametric family (Briesch, Chintagunta, and Matzkin 1997). Thus, the cost of imposing the strong restrictions required for parametric estimation can be considerable (Härdle and Linton 1994).

3.3 Nonparametric approach

A nonparametric regression model does not project the observed data into a Procrustean bed² of a fixed parameterization (Härdle 1990). Nonparametric modeling imposes few restrictions on the form of the joint distribution of the data, so there is little room for (functional form) misspecification, and consistency of the estimator of the regression curve is established under much more general conditions than is the case for parametric modeling. Rust (1988) introduced *nonparametric* regression models to marketing research. He emphasizes that nonlinearity, non-normal errors, and heteroscedasticity are automatically accommodated. We believe that the primary benefits consist of the relaxation of functional form constraints and the allowance for flexible interactions.

² Procrustes is a figure in Greek mythology. He was a robber who took his victims home to cut or stretch them so that they fit in his bed.

Greater flexibility does impose a cost: the convergence rate of a nonparametric estimator is usually slower³ than it is for parametric estimators (Powell 1994, p. 2448). Thus, the precise estimation of a nonparametric multidimensional regression surface requires very large sample sizes. Yet, as is true for parametric models, the statistical precision is a function of the variances and covariances of the predictors, and not of the sample size.⁴ Nevertheless, Silverman (1986, p. 94) and Scott (1992, pp. 198-200) show how the “required” sample size increases with the number of predictors. For example, Scott’s formulas suggest that if the required sample size is 10 for a one-predictor variable model, it is 1,000+ for a model with three predictors. However, for a model of one item’s sales as a function of three continuous variables for each of three items, the required sample size would be more than 7.5 million! The reason for this “curse of dimensionality” is the flexible accommodation of all possible interaction effects.

A widely used nonparametric regression estimator is the Kernel method. It is straightforward to implement and it is easy to understand (Härdle 1990). The intuition behind the Kernel method is that it computes a local weighted average of the criterion variable given the values of the predictors. The weights are inversely related to the “distances” between (a) the values of the predictor variables for which one wants to estimate the criterion variable and (b) the values of the same variables that lie in the proximity. The dependence of the weight on these distances is described by the Kernel function which tends to be symmetric around zero where it reaches its maximum. The Kernel method also includes a bandwidth parameter which controls the peakedness of this function. The smaller this parameter, the more peaked the function, and the more weight is placed on the nearest observations. A bandwidth parameter of (almost) zero leads to a response curve that connects the observations, resulting in no bias but maximal variance. An infinite bandwidth parameter leads to a horizontal response curve: maximal bias and minimal variance. Thus, the choice of the bandwidth parameter involves a trade-off between bias and variance. Most bandwidth selection techniques minimize a mean squared error criterion. In Appendix 1 we describe nonparametric regression in general terms and we provide formulas for the Kernel method.

Härdle (1990), Hastie and Tibshirani (1990), and Scott (1992) provide other

3 For further discussion of parametric versus nonparametric regression models, see Härdle (1990, pp. 3-6), and Powell (1994, pp. 2444-7).

4 In parametric estimation, one desires maximum variance and minimum covariances for the predictors, under the implicit assumption that the functional form is known. For nonparametric estimation, assuming that one is equally interested in the effects of all possible combinations of values, one desires uniformly distributed values for the predictors, not maximum variances.

nonparametric estimators, such as the k -nearest neighbor estimator, cubic smoothing splines and orthogonal series estimators. Härdle (1990, pp. 77-88) shows that based on a finite-sample simulation, different methods yield essentially the same curves. Importantly, the semiparametric model we propose below requires the Kernel method.

3.4 Semiparametric approach

Our primary interest is in own- and cross-item deal effects, so we use a semiparametric approach that has the advantage of nonparametric regression (flexibility) for the deal variables, and the advantage of parametric regression (efficiency) for all other variables. We employ the semiparametric (semilinear) model proposed by Robinson (1988). We express sales as the sum of a nonparametric function of discount variables and a linear function of other predictors (almost all indicator variables). The indicator variables represent feature and display activities, special pack promotions, weekly effects and store effects.

The specification of the semiparametric regression model for the estimation of the deal effect curve is based on the well-known Scan*Pro model (Blattberg and Neslin 1990, pp. 369-70; Foekens, Leeflang, and Wittink 1994). In Scan*Pro, the criterion variable is log unit sales (instead of unit sales) to accommodate multiplicative effects for the predictor variables (which applies to the semiparametric model in its parametric part). The price variables are log price indices. The price index is the ratio of actual price for an item relative to its regular price. Price indices less than one represent temporary price cuts. The usage of price indices allows us to isolate the deal effects and it enhances the comparability of (deal) effects across items and product categories.

We do not include regular prices as predictors in our model. The regular prices tend to be collinear with store indicator variables, because they vary especially across stores. Over time, almost all changes in regular prices occur in the same week(s) in all stores within each chain we use for analysis. Therefore, the weekly indicator variables capture such effects adequately, and there is no missing variable bias⁵ due to the exclusion of regular prices.

5 To study this conjecture we also included regular price effects in all three models for all three product categories. The estimation results showed that there were no noticeable changes in the shapes of the deal effect curves. The fit- and predictive validity statistics improved slightly, but the relative performances of the models stayed the same. The estimated regular price elasticities, however, had wrong signs and unexpected magnitudes in some cases as a result of a high degree of multicollinearity.

For item k , $k = 1, \dots, J$, the specification for the semiparametric model is:

$$\ln S_{ik,t} = m(\ln(\text{PI}_{i1,t}), \ln(\text{PI}_{i2,t}), \dots, \ln(\text{PI}_{iJ,t})) + \sum_j \sum_l \gamma_{jkl} I_{ijl,t} \alpha_{ik} S_i + \xi_{tk} W_t + u_{ik,t} \quad (1)$$

where: $S_{ik,t}$ is unit sales of item k in store i , $i = 1, \dots, n$, in week t , $t = 1, \dots, T$;
 $m(\cdot)$ is a nonparametric function;
 $\text{PI}_{ij,t}$ is the the price index (ratio of actual to regular price) of item j in store i in week t ; $j = 1, \dots, J$;
 $I_{ijl,t}$ is an indicator variable capturing usage (=1) or nonusage (=0) of feature-only ($l=1$), display-only ($l=2$), feature and display ($l=3$) and bonus packs ($l=4$) of item j in store i in week t , if applicable ;
 S_i is a store indicator variable; = 1 if observation is from store i , = 0 otherwise;
 W_t is a weekly indicator variable; = 1 if observation is from week t , = 0 otherwise;
 γ_{jkl} is the effect of promotion l for item j on the sales of item k ;
 α_{ik} and ξ_{tk} are store intercepts for store i , item k , and week intercepts for week t , item k , respectively;
 $u_{ik,t}$ is a disturbance term for item k in store i in week t .

Equation (1) is fully flexible as far as the main effects of the predictors are concerned, since the continuous predictors (price indices) are modeled nonparametrically. It also includes flexible interaction effects between the price index variables of different items. We discuss the three-step estimation of model (1) in Appendix 2. We describe the specific choices of the Kernel and the bandwidth parameter in Appendix 3, where we also explain the construction of confidence intervals for the estimated deal effect curves.

3.5 Benchmark parametric models

One parametric benchmark we use is AC Nielsen's Scan*Pro model for promotional effects. Wittink et al. (1988) developed this model for commercial purposes, and it and its variants have been used by AC Nielsen in over 2000 different commercial

applications in North America, Europe and Asia. We use the following specification:

$$\ln S_{ik,t} = \sum_{j=1}^J \beta_{jk} \ln(\text{PI}_{ij,t}) + \sum_j \sum_l \gamma'_{jkl} I_{ijl,t} + \alpha'_{ik} S_i + \xi'_{ik} W_t + u'_{ik,t} \quad (2)$$

for $t = 1, \dots, T$ and $i = 1, \dots, n$. The variables have the same definitions as in (1). The only difference from model (1) is the specification of deal effects. In (2) β_{jk} is the (constant) elasticity of the sales of item k with respect to a price discount of item j . For further details see Blattberg and Neslin (1990, pp. 369-70); Foekens, Leeflang, and Wittink (1994); and Gupta et al. (1996).

The second parametric benchmark model we use is based on Blattberg and Wisniewski's (1989, equation (5.1)) model. We adapt their model by excluding regular prices. The Blattberg and Wisniewski model (B&W model) differs from Scan*Pro in the type of nonlinearities assumed for own- and cross-brand deal effects:

$$\ln S_{ik,t} = \eta_{kk}(1 - \text{PI}_{kk,t}) + \sum_{j \neq k}^J \eta_{jk} / \text{PI}_{ij,t} + \sum_j \sum_l \gamma''_{jkl} I_{ijl,t} + \alpha''_{ik} S_i + \xi''_{ik} W_t + u''_{ik,t}. \quad (3)$$

We estimate equations (2) and (3) by ordinary least squares.

4. Data description

We apply the models to two American and one Dutch weekly store-level scanner data sets. The data sets are described in Tables 4.1-4.3. All data sets are from stores belonging to one chain. It is worth noting that the percent of observations with a price promotion is very high in the American data (more than 29 percent for each item). The Dutch data show no price promotions for two items, and less than ten percent of the observations for the other two items.

We pool data across stores of a given chain, and use the first half of the weekly observations for estimation, leaving an equal number for validation. Importantly, promotional and other market conditions can change systematically over time, and such changes allow the predictive validities to show which model truly captures the deal effect curves best.

The first American data set, which is from AC Nielsen, contains the three largest

national brands in the 6.5 oz. canned tuna fish product category (Table 4.1). All items have a substantial amount of variation in the price discount levels. We have 104 weeks of data for each of the 28 stores of one supermarket chain in a metropolitan area which generates an estimation sample of 1456 observations. Each item is discounted in more than 400 observations in the estimation samples.

Table 4.1: Description of American tuna data set

Number of stores	28		
Weeks per store	104		
Pooled sample size	2912		
Estimation sample size	1456		
Validation sample size	1456		
Unit measure	6.5 oz.		
Percentage of total category	74%		
Item number	1	2	3
Brand name	<i>A</i>	<i>B</i>	<i>C</i>
Average regular unit price (US\$)	0.86	0.87	0.86
Purchase share (%)	41	37	22
Price promotion obs. (%)	42	31	31
Average price discount (%)	26	25	26

The second American data set is from an unnamed source (third party). We have data on three beverage items for each of the largest two brands (Table 4.2). To reduce the curse of dimensionality we specify for each item's sales the three own-item effects nonparametrically, and the three cross-item effects parametrically.⁶ We have 117 weekly observations at maximum for 151 stores. Since some items were unavailable in the first part of the data, we use only the store-week observations in which all six items were available. The minimum number of times an item is discounted in estimation is 800.

The Dutch data set (from AC Nielsen) is on the four largest items of a packaged food product (Table 4.3). Only items 3 and 4 (both brand *B*) have been offered at a discount. Therefore, only the price indices of these two items are included in the nonparametric part of the semiparametric model for all four items. For this product category we have 144 weekly observations for almost all 48 stores in a national sample from one large supermarket chain. The estimation sample contains 3440 observations, including 274 with a discount for item 3, and 238 with a discount for item 4.

⁶ This split is based on a preliminary parametric analysis which showed that the average cross-deal elasticity is far greater within than across brands.

Table 4.2: Description of American beverage product data set

Number of stores	151					
Weeks per store	117					
Pooled sample size	5718					
Estimation sample size	2890					
Validation sample size	2828					
Unit measure	12 oz.					
Percentage of total category	46%					
Item number	1	2	3	4	5	6
Brand name	A	A	A	B	B	B
Average regular unit price (US\$)	0.60	0.58	0.61	0.59	0.58	0.61
Purchase share (%)	14	26	12	13	23	13
Price promotion obs. (%)	37	29	30	30	36	35
Average price discount (%)	12	10	9	13	9	9

We focus on one item at a time using both cross sectional and time series variation. To see whether the time series variation is similar across stores, we checked the variation in magnitudes of price discounts *within* stores. For every item in each product category, we split the observations with own-item price indices less than one into three equally sized intervals: lowest, medium, and highest. For each of these intervals, we computed the percentage of stores that offered price discounts. For tuna the average percentage across intervals and across items is 95 percent. For the American beverage product it is 76 percent. For the Dutch food category all stores offered price discounts in all three intervals for items 3 and 4 (100 percent). Thus, the nonparametric estimation of deal effect curves will not just reflect average differences in deal magnitudes between stores.

5. Results

In this section we first report fit- and predictive validity statistics for the semiparametric model (1) and the parametric models (2) and (3).⁷ We then show the nature of the deal effect curves.

⁷ A complete overview of the estimation results is available from the first author on request.

Table 4.3: Description of Dutch food product data set

Number of stores	48			
Weeks per store	144			
Pooled sample size	6884			
Estimation sample size	3440			
Validation sample size	3444			
Unit measure	1 kg.			
Percentage of total category	72%			
Item number	1 ^a	2 ^a	3	4
Brand name	A	A	B	B
Average regular unit price (Dfl.)	7.15	7.07	6.26	6.13
Purchase share (%)	27	26	21	26
Price promotion obs. (%)	0.0	0.0	7.1	6.8
Average price discount (%)	—	—	13	18

^a This item's deal effect is not modeled at all, because of the absence of discounts for this item.

Comparison of fit and predictive validity

To obtain fit- and predictive validity statistics, we estimate unit sales, in the estimation sample, for model (1) as follows:

$$\hat{S}_{ik,t} = \exp(\hat{m}(\ln(\text{PI}_{i1,t}), \ln(\text{PI}_{i2,t}), \dots, \ln(\text{PI}_{iJ,t}))) + \sum_j \sum_l \hat{\gamma}_{jkl} I_{ijl,t} + \hat{\alpha}_{ik} S_i + \hat{\xi}_{tk} W_t + \frac{1}{2} \hat{\sigma}_{ik,t}^2) \quad (4)$$

$t = 1, \dots, T$ and $i = 1, \dots, n$.

The term $\frac{1}{2} \hat{\sigma}_{ik,t}^2$, where $\hat{\sigma}_{ik,t}^2$ is the residual variance, is included to minimize the bias in the conditional mean predictions due to estimation in the log space (Goldberger 1968). We obtain $\hat{\sigma}_{ik,t}^2$ by the nonparametric estimate for $E((\ln S)_{ik,t}^2 | X_{ik,t}) - \{E(\ln S_{ik,t} | X_{ik,t})\}^2$, where $X_{ik,t}$ captures the values of the predictor variables relevant for item k in store i in week t . The unit sales estimates for models (2) and (3) are obtained analogously, except that we use an OLS-based homoscedastic $\hat{\sigma}_k^2$ for the bias reduction. The validation sample estimates are obtained in a similar way.

5.1 Fit

We report the fit of the three models in the estimation sample, as measured by the weighted average Mean Squared Error in sales (MSE), in Tables 5.1-5.3. For tuna, the weighted (by squared purchase share) average MSE is 14.5 percent lower for the semiparametric model than for Scan*Pro, and 12.6 percent lower than the B&W model (Table 5.1). The beverage product has average gains of 17.6 and 18.6 percent

(Table 5.2), while the Dutch food product shows average gains of 3.4 and 3.5 percent respectively (Table 5.3).

Table 5.1: Tuna data: fit and predictive validity^a statistics for semiparametric and parametric models

Fit statistic	Semiparametric model (1)	Difference from	
		Scan*Pro model (2)	B&W model (3)
<i>Estimation sample</i>			
MSE all observations	27694	-14.5%	-12.6%
MSE own-price cut observations	65544	-17.8%	-16.7%
<i>Validation sample</i>			
MSE all observations	55654	-47.9%	-19.0%
Squared bias all observations	398	-82.6%	-72.9%
Variance all observations	55256	-47.2%	-17.9%
<i>Validation sample</i>			
MSE own-price cut observations	112974	-51.9%	-22.2%
Squared bias own-price cut observations	890	-88.8%	-80.6%
Variance own-price cut observations	112084	-50.6%	-20.3%

^a The validation sample results for all models are negatively affected by the omission of the weekly indicator variable effects in the predicted values. We ignored these variables because seasonal effects can change from one week to another over time (e.g., Easter may occur in a differently sequenced weeks) and missing variable effects are very unlikely to apply to the same week (e.g., regular price changes) in multiple years. This omission should not, however, affect the relative performances of alternative models.

Table 5.2: Beverage data: fit and predictive validity^a statistics for semiparametric and parametric models

Fit statistic	Semiparametric model (1)	Difference from	
		Scan*Pro model (2)	B&W model (3)
<i>Estimation sample</i>			
MSE all observations	1764.0	-17.6%	-18.6%
MSE own-price cut observations	4722.1	-20.4%	-22.0%
<i>Validation sample</i>			
MSE all observations	4967.5	-9.1%	-9.8%
Squared bias all observations	536.0	-20.8%	-23.8%
Variance all observations	4431.6	-7.4%	-7.8%
<i>Validation sample</i>			
MSE own-price cut observations	7527.8	-11.6%	-10.6%
Squared bias own-price cut observations	661.7	-17.8%	-24.8%
Variance own-price cut observations	6866.1	-11.0%	-9.0%

^a See footnote a Table 5.1.

Table 5.3: Dutch food data: fit and predictive validity^a statistics for semiparametric and parametric models

Fit statistic	Semiparametric model (1)	Difference from	
		Scan*Pro model (2)	B&W model (3)
<i>Estimation sample</i>			
MSE all observations	28.5	-3.4%	-3.5%
MSE own-price cut observations	148.5	-8.4%	-8.7%
<i>Validation sample</i>			
MSE all observations	54.0	-11.1%	-12.0%
Squared bias all observations	7.1	-9.0%	-10.3%
Variance all observations	47.0	-11.4%	-12.2%
<i>Validation sample</i>			
MSE own-price cut observations	378.2	-21.0%	-22.4%
Squared bias own-price cut observations	50.0	-40.7%	-44.2%
Variance own-price cut observations	328.2	-16.8%	-17.5%

^a See footnote a Table 5.1.

These results indicate that the greater flexibility in nonlinear- and interaction effects (for the deal variables) that is the hallmark of nonparametric estimation improves the fit to the sample data relative to both parametric models. The smaller percentage improvements for the Dutch food product are consistent with the Dutch data having fewer observations with temporary discounts.

The benefit of the semiparametric model should be concentrated in observations with discounts. To examine this, we select the data with own-item discounts. The unit sales of these observations fitted by the semiparametric model reflect both the flexible nonlinearity of the own-item deal effect and the interaction with cross-item deals when other items are discounted at the same time. Indeed, we obtain stronger relative gains in fit in all three categories for these observations. For tuna the improvement in MSE for the semiparametric model relative to Scan*Pro is 17.8 percent and relative to B&W it is 16.7 percent (Table 5.1). For the beverage product the relative improvements are 20.4 and 22.0 percent (Table 5.2), and for the Dutch food category these are 8.4 and 8.7 percent (Table 5.3), relative to Scan*Pro and B&W respectively. We note that the average MSE is much higher for these observations than for the entire sample. This reflects the much greater variance in sales under promoted conditions than under non-promoted conditions.

5.2 Predictive validity

Since improvement in fit is virtually guaranteed with more flexible methods, a real test consists of a comparison of performances in the validation samples. We expect a superior description of deal effects from the semiparametric model but these effects will have greater variance, relative to parametric models. Thus, there is no guarantee that the reduction in bias (due to greater flexibility) is greater than the possible increase in variance (due to reduced efficiency). We show MSE results for the validation sample in the middle panels of Tables 5.1, 5.2, and 5.3. In all cases, these values are much higher in the validation than in the estimation samples (in part because the weekly indicator variables' effects are not used). Importantly, the weighted improvement for the semiparametric model over the parametric models is a higher percentage in validation than in estimation for two of the categories. The most dramatic difference occurs for tuna, relative to Scan*Pro (47.9 percent), and with a respectable improvement of 19.0 percent relative to B&W (Table 5.1). The relative differences are 9.1 and 9.8 percent in Table 5.2 and 11.1 and 12.0 percent in Table 5.3. The improvements in weighted average validation sample MSE values for the semiparametric model are significant, based on paired t-tests ($p < .01$ for all categories).

To examine the trade-off between bias and variance directly, we decomposed the validation sample mean squared error (MSE) into squared bias and variance. For tuna, the reduction in MSE relative to the Scan*Pro model consists of a 82.6 percent reduction in squared bias and a 47.2 percent reduction in variance (middle panel Table 5.1). Relative to the B&W model the reductions in squared bias and variance are 72.9 and 17.9 percent. The beverage data also show larger relative improvements in squared bias than in variance: 20.8 and 7.4 percent versus Scan*Pro and 23.8 and 7.8 percent relative to the B&W model (middle panel Table 5.2). However, in the Dutch data the reduction relative to Scan*Pro consists of only a 9.0 percent decrease in squared bias and a higher, 11.4 percent, decrease in variance (middle panel Table 5.3). The percent reductions relative to the B&W model are 10.3 and 12.2.

We obtain even more striking reductions in MSE, squared bias and variance when we isolate the own-item price cut observations in the validation samples (lower panels of Tables 5.1, 5.2, and 5.3). The reductions in weighted average MSE for the semiparametric model relative to the Scan*Pro and B&W model are 51.9 and 22.2 percent for tuna, 11.6 and 10.6 percent for the beverage data, and 21.0 and 22.4 percent for the Dutch food category. The decompositions into squared bias and variance for these MSE's show in all cases that the squared bias is reduced to a more substantial extent than the variance is.

From these results it is clear that the semiparametric model outperforms both

parametric benchmarks in all three product categories. As expected the improvement is especially noticeable in the observations with own-item discounts. In addition, the relative improvement tends to be greater in the bias than in the variance component of MSE. This is consistent with the notion that the semiparametric model captures complex patterns better. However, since it does this with lower efficiency, one could imagine that the variance component favors the parametric benchmarks. The reason this is not the case is that the bias is computed across all price discount observations and not separately for, say, discount intervals. Thus, the variance component also reflects some of the advantages in the semiparametric model of capturing complex deal effect curves. We use graphical representations of the estimated deal effect curves to demonstrate the nature of systematic differences.

5.3 Deal effect curves

We show deal effects from the semiparametric model, and contrast these curves with the corresponding ones for the best performing parametric benchmark.

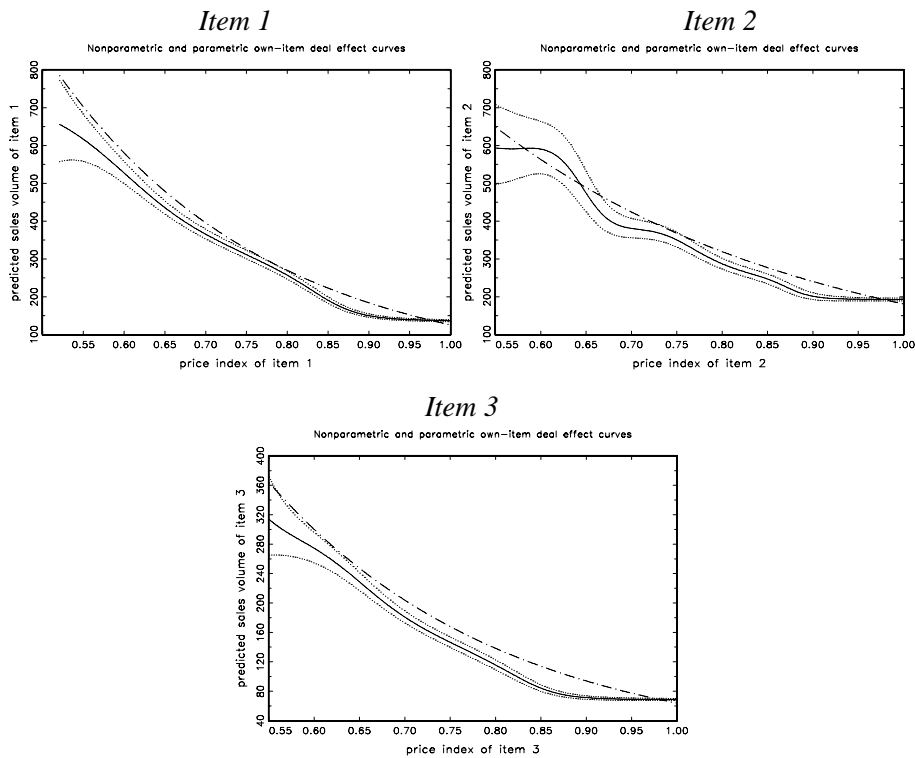
5.3.1 Own-item deal effect curves

We show nonparametrically estimated own-item deal effect curves in Figures 5.1-5.3, with own-item price indices on the x-axis and predicted sales on the y-axis. For the x-axis we generated 300 values, equally spaced between the lowest and highest price indices observed in the estimation sample. For each of these 300 price indices we estimated the log of the sales (for an average week and an average store) as outlined in Appendix 2, while fixing the other items' price indices at one. The transformation to sales included the error variance adjustment shown in (4).

We show the own-item deal effect curves for the three tuna items in Figure 5.1, for the six beverage items in Figure 5.2, and for Dutch food items 3 and 4 in Figure 5.3. The solid lines in these figures represent the nonparametric deal effect curves from model (1). The dotted lines around the solid line represent the lower and upper 95 percent pointwise confidence bounds (see equation (20) in Appendix 3). The dashed lines represent deal effect curves for the parametric benchmark model (2) or (3). For tuna we used the B&W model (see Table 5.1), and we used the the Scan*Pro model in the other two cases (Tables 5.2 and 5.3)

The differences between the nonparametric- and parametric deal curves in Figures 5.1-5.3 are noteworthy in all product categories. A reverse-S shape is visible in almost all nonparametric own-item deal effect curves, indicating threshold- and saturation levels, whereas the parametric model yields exponential deal effect curves. The parametric curves often overstate the effect for small discounts (by omitting threshold

Figure 5.1: Tuna data: own-item deal effect curves^a



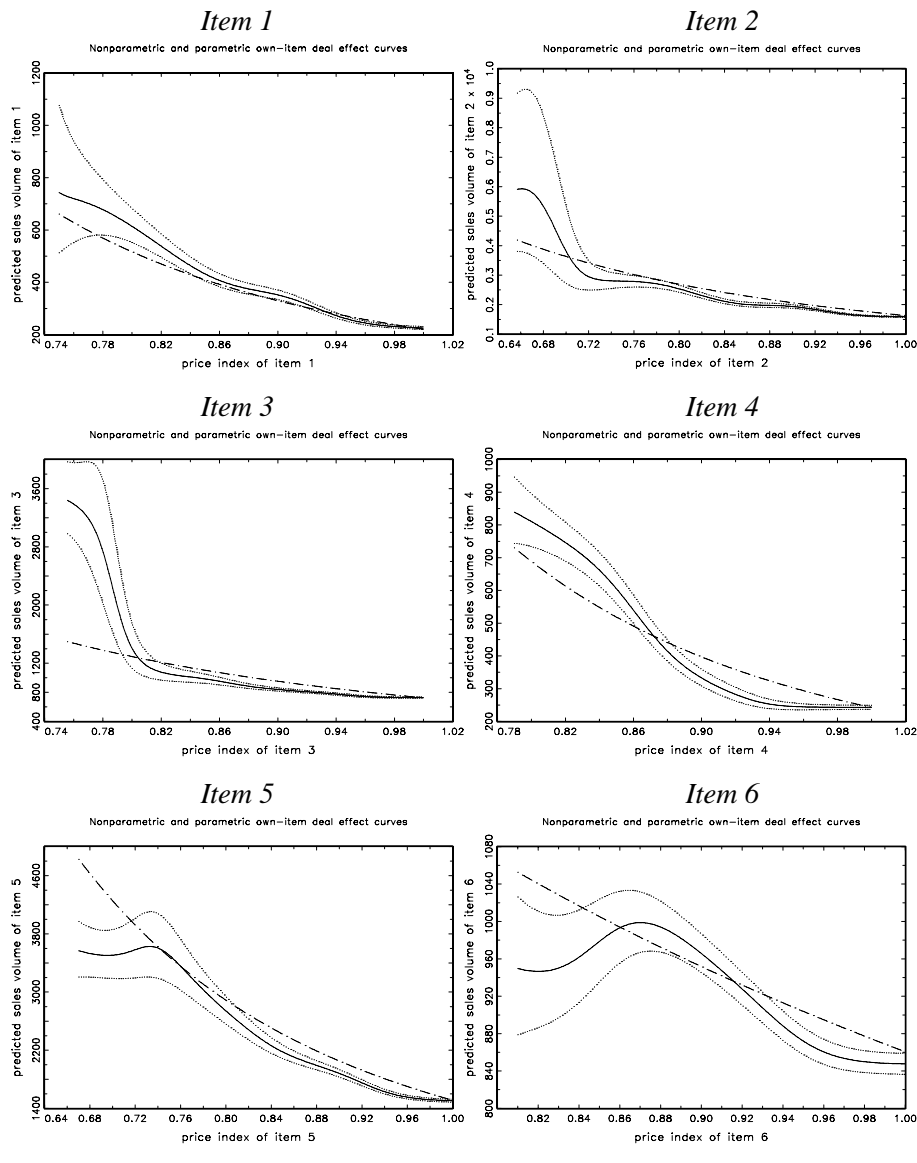
^a solid line = nonparametric estimate, dashed line = parametric B&W estimate, dotted lines = nonparametric confidence bounds (95%)

effects), understate the effect for intermediate discounts, and overstate the effect for the largest discounts (by omitting saturation effects). Such differences are best illustrated for item 3 in the Dutch food category. We note that the deal effect curves for beverage item 6 indicate a slight decrease in sales toward the lower limit of the price index range (Figure 5.3). This nonmonotonicity may be due to chance, however, as indicated by sufficiently wide confidence intervals.

5.3.2 Cross-item deal effect curves

We show selected cross-item deal effect curves in Figures 5.4-5.6. We base our selection on (1) the representativeness of the selected curve for all cross-item deal effect curves and (2) the significance of the corresponding cross-item effect in the parametric benchmark model used in the figures. The cross-item deal effect curves

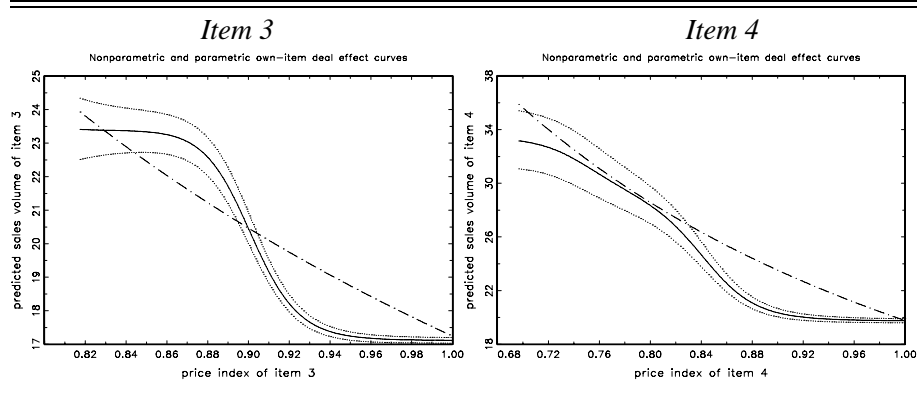
Figure 5.2: Beverage data: own-item deal effect curves^a



^a solid line = nonparametric estimate, dashed line = parametric B&W estimate, dotted lines = nonparametric confidence bounds (95%)

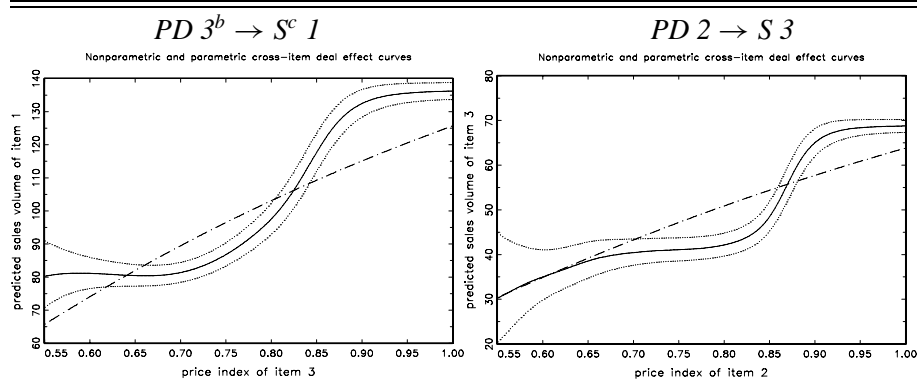
were obtained by varying the values of one cross-price index, while setting the own-item and the other cross-item price indices equal to one.

Figure 5.3: Dutch food data: own-item deal effect curves^a



^a solid line = nonparametric estimate, dashed line = parametric B&W estimate, dotted lines = nonparametric confidence bounds (95%)

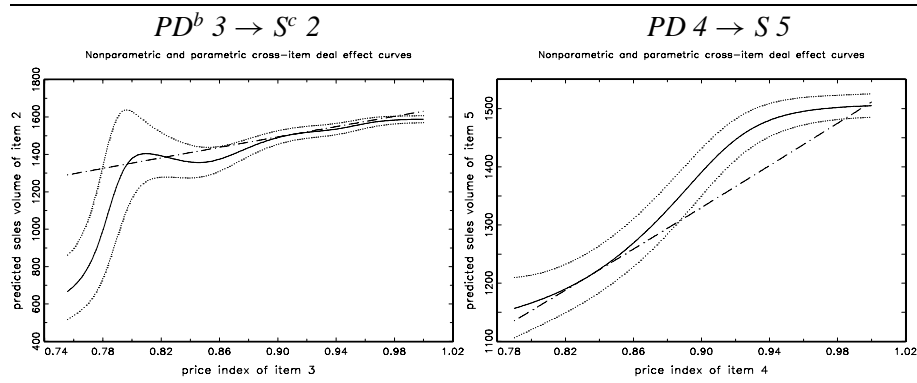
Figure 5.4: Tuna data: cross-item deal effect curves^a



^a solid line = nonparametric estimate, dashed line = parametric B&W estimate, dotted lines = nonparametric confidence bounds (95%)

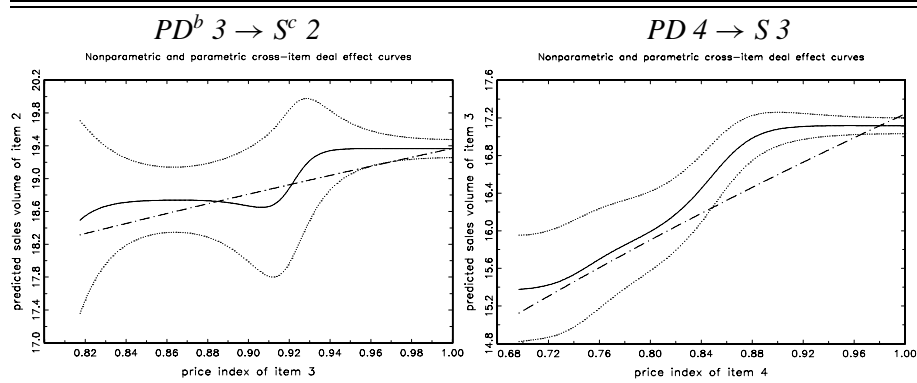
^{b,c} PD = Price Discount, S = Sales

Figure 5.5: Beverage data: cross-item deal effect curves^a



^a solid line = nonparametric estimate, dashed line = parametric B&W estimate, dotted lines = nonparametric confidence bounds (95%)
^{b,c} PD = Price Discount, S = Sales

Figure 5.6: Dutch food data: cross-item deal effect curves^a



^a solid line = nonparametric estimate, dashed line = parametric B&W estimate, dotted lines = nonparametric confidence bounds (95%)
^{b,c} PD = Price Discount, S = Sales

The cross-item deal effects tend to show S-shapes in all three product categories. For most cross effects shown in Figures 5.4-5.6, the nonparametric curve is above the parametric one for small and large discounts. Interestingly, the large *increase* in sales of beverage item 3 with own-item price indices of 0.80 and lower (Figure 5.2), corresponds to a large *decrease* in sales of item 2 for item 3 price indices of 0.80 and lower (Figure 5.5).⁸

⁸ We could not test Sethuraman's (1996) theory in our data sets. For instance, the prices are almost

5.3.3 Key features

We summarize key characteristics of the nonparametric own- and cross-item deal effects, for which the parametric model has significant effects, in Tables 5.4-5.6. For tuna fish, the semiparametric model shows a threshold level for all own- and cross-item effects at about 10 percent (Table 5.4). The saturation level of the own-item effect is around 40 percent (although there is little evidence of saturation effects for items 1 and 3). The cross-item effects show saturation levels varying from 20 to 45 percent. The beverage items show a mixture of (reverse) S-shaped- and (reverse-) L-shaped own-item- and cross-item deal effect curves (Table 5.5). For the Dutch food items, the own-item effect thresholds are 5 and 10 percent, and the saturation levels are 15 and 25 percent (Table 5.6). The cross-effects show multiple shapes.

Table 5.4: Tuna data: description of deal effect curves^a

	Shape	Threshold level ^b	Saturation level ^b
<i>own-item deal effect curves</i>			
PD 1 → S 1	convex / (reverse S)	10%	(45%) ^c
PD 2 → S 2	reverse-S	10%	40%
PD 3 → S 3	convex / (reverse S)	10%	(40%)
<i>cross-item deal effect curves</i>			
PD 2 → S 1	S + additional decrease from 40%	10%	25%
PD 3 → S 1	S	10%	30%
PD 1 → S 2	S + additional decrease from 40%	10%	35%
PD 3 → S 2	S	10%	40%
PD 1 → S 3	S	10%	45%
PD 2 → S 3	S + additional decrease from 40%	10%	20%

^a only those with significant parametric deal effects

^b rounded to the nearest multiple of 5%

^c numbers in parentheses indicate uncertainty about the existence of a saturation level

5.3.4 Confidence bounds

The width of the confidence bounds varies considerably between items and categories. This width is determined by at least three factors: (1) the number of price discount observations, (2) the distribution of price discounts, and (3) the dimensionality of the nonparametric estimation problem. To illustrate, consider the own-item deal effect curves for tuna fish items 1 and 3 (Figure 5.1). For these items, the average discount percent is the same (see Table 4.1), and so is the dimensionality.

equal across items in two of the three data sets (tuna and beverage). In the third data set only the two items with lower regular prices engage in discounting (Dutch food).

Table 5.5: Beverage data: description of deal effect curves^a

	Shape	Threshold level ^b	Saturation level ^b
<i>own-item deal effect curves</i>			
PD 1 → S 1	linear	none	none
PD 2 → S 2	L (kink at 30%)	none	35%
PD 3 → S 3	L (kink at 20%)	none	25%
PD 4 → S 4	reverse S	5%	25%
PD 5 → S 5	reverse S	5%	25%
PD 6 → S 6	reverse S	5%	15%
<i>cross-item deal effect curves</i>			
PD 2 → S 1	S	5%	25%
PD 1 → S 2	S/linear	(5%) ^c	20%) ^c
PD 3 → S 2	reverse-L (kink at 20%)	none	none
PD 1 → S 3	S	none	15%
PD 2 → S 3	S	5%	25%
PD 4 → S 5	S	5%	20%
PD 6 → S 5	reverse-L (kink at 15%)	none	none
PD 4 → S 6	reverse-L (kink at 10%)	none	none

^a only those with significant parametric deal effects

^b rounded to the nearest multiple of 5%

^c numbers in parentheses indicate uncertainty about the existence of a threshold- or saturation level

Item 1, however, has a discount 42 percent of the time versus 31 percent for item 3 (about 610 versus 450 observations). Thus, item 1 has somewhat tighter confidence bounds for the own-item deal effect curve. The impact of the second factor is apparent within a given curve. For example, tuna fish item 1 has a widening of the bounds for its deal effect when the price index goes below 0.70. This reflects the reduced number of observations at lower price indices for this item in the sample. We do not have sufficient variation in the dimensionality to show how it affects the confidence bounds.

5.4 Interaction effects

We illustrate two types of interaction effects for tuna only: between price discounts of different items, and between own-item price discount and own-item promotion signals (feature-only, display-only, and feature plus display). For the latter interaction effects, we use separate analyses.

5.4.1 Interaction effects between price discounts of different items

The nonparametric part of equation (1) accommodates flexible interaction effects between price discounts of different items, since $m(\cdot)$ is a function of each item's

Table 5.6: Dutch food data: description of deal effect curves^a

	Shape	Threshold level ^b	Saturation level ^b
<i>own-item deal effect curves</i>			
PD 3 → S 3	reverse S	5%	15%
PD 4 → S 4	reverse S	10%	25%
<i>cross-item deal effect curves</i>			
PD 3 → S 1	linear	none	none
PD 3 → S 2	S	5%	10%
PD 4 → S 3	concave	10%	30%
PD 3 → S 4	S	5%	10%

^a only those with significant parametric deal effects

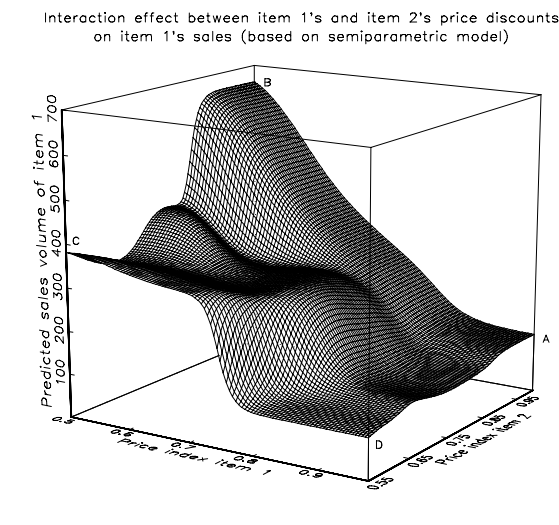
^b rounded to the nearest multiple of 5%

price index. Analogous to the two-dimensional own-item and cross-item deal effect curves (Figures 5.1-5.6), we can construct a three-dimensional deal effect surface. To obtain a reliable surface we need observations with a price discount for two (or more) items simultaneously. Out of the 1456 tuna observations, 114 represent situations in which items 1 and 2 both have price discounts.

We show two deal effect surfaces in Figure 5.7-5.8. In Figure 5.7 the vertical axis represents the predicted sales volume of item 1. The other two axes represent the price indices of items 1 and item 2 respectively. The top part of the three-dimensional surface (the curve A-B) is item 1's own-item deal effect, also shown in Figure 5.1, when the other two items both have a price index of one. The figure shows that this curve tends to decrease in magnitude with a decrease in item 2's price index. But, while there is little evidence of saturation when item 2's index equals one, saturation is very clear when item 2's index is about 0.85 or less. Thus, the interaction effect is highly nonlinear which would be very difficult to model parametrically. Substantively, if item 1 is promoted with a deep discount (when its price index is ≤ 0.75), item 2 can reduce item 1's sales gain considerably if it also has a price index is ≤ 0.75 .

We show the interaction effect for the same price indices on item 2's sales in Figure 5.8. Here, the curve A-B represents item 2's own-item deal effect, when items 1 and 3 have a price index of one (see Figure 5.1). This curve also tends to decrease in magnitude when the price index of item 1 decreases, but the decrease is not as dramatic as it is in Figure 5.7. The own-item deal effect for item 2 appears to be less sensitive to item 1's discounting than vice versa. Thus, the interaction effects appear to be asymmetric.

Figure 5.7: 3D tuna deal effect surface for item 1



5.4.2 Interaction effects between price discounts and promotion signals

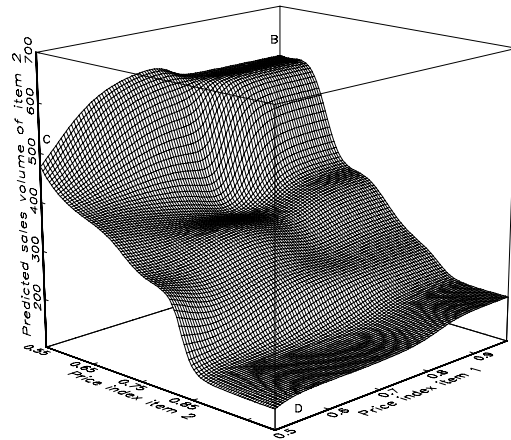
Promotion signals such as feature and display have strong effects on item sales (Blattberg, Briesch, and Fox 1995, p. G125). However, we know little about the synergies between feature advertising, displays, and price discounts. Inman, McAlister, and Hoyer (1990) and Mayhew and Winer (1992) suggest that “low need for cognition” consumers react to the simple presence of a promotion signal whether or not the price of the promotion is reduced. If these consumers form a large group, the deal effect curve for price discounts accompanied by a promotion signal will be shifted upward from the deal effect curve for price discounts without a promotion signal, but show less price sensitivity.

The semiparametric model (1) allows for flexible interaction effects between own- and cross-item price indices (such as those illustrated in Figures 5.7 and 5.8) but not between own-item price index and own-item promotion signals. However, the methodology can be adapted to study these interactions flexibly as well. To accomplish this, we split the sample of 1456 tuna observations into four subsamples (Lee 1996, p. 158). We report interaction effects between price discounts and promotion signals for tuna item 1 only, for which the subsample sizes are: feature-only 73, display-only 103, combined use of feature and display 263, and neither feature nor display 1017.

We estimate a separate own-item deal effect curve for each subsample, analogous

Figure 5.8: 3D tuna deal effect surface for item 2

Interaction effect between item 1's and item 2's price discounts on item 2's sales (based on semiparametric model)

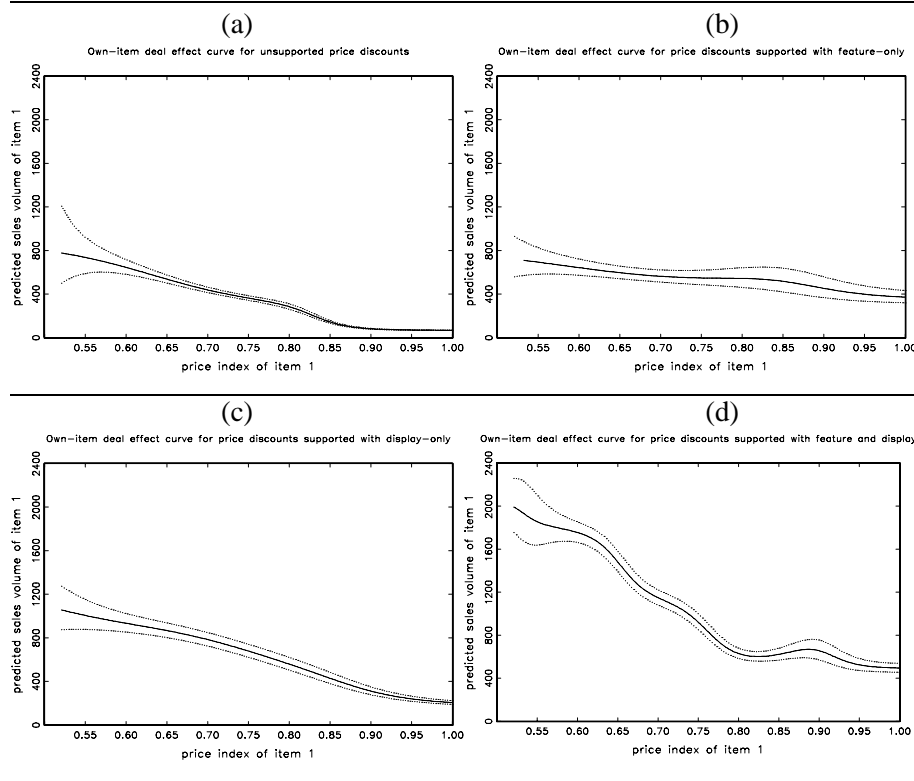


to model (1), except for two differences: (1) the instrument used to construct a subsample is excluded from the parametric part, and (2) all weekly indicators are omitted to reduce collinearity.

We show in Figure 5.9 the own-item deal effect curves for item 1 under four conditions. Figure 5.9.a shows the deal effect when there is no promotion signal (neither feature, nor display). The shape of this curve is similar to the one in Figure 5.1. A comparison of Figures 5.9.b and 5.9.c suggests that at the rightmost point (no discount) the effect of feature-only is greater than that of display-only. However, the curves show greater deal sensitivity with display than with feature. For example, the two curves intersect at a discount of 20 percent, after which the deal effect with display exceeds the effect with feature. Thus, feature is more effective (than display) at no/low discount, while display is more effective at high discount. Such a cross-over interaction is very difficult to anticipate and almost impossible to diagnose from the residuals of parametric models.

The “feature-only” deal effect curve (Figure 5.9.b) is rather flat. This pattern is consistent with the idea that the response to feature is especially by “need for cognition” consumers (Inman, McAlister, and Hoyer 1990; Mayhew and Winer 1992) who do not pay much attention to the actual price. Both in an absolute sense and in a relative sense all three other deal curves show (much) greater sensitivity to discount

Figure 5.9: Interaction effects between price discounts and promotion signals for tuna item 1



levels.⁹ For example, Figure 5.9.b shows that average sales is almost 400 when the item is featured and the price index equals 1.0. At an index less than 0.55, average sales is less than 800. On the other hand, Figure 5.9.c has average sales of 200 when the item is displayed and the price index is 1.0. But it is about 1100 when the

⁹ This pattern suggests also that the semiparametric model (1) for this item is incomplete. For example, here the display-only sales multiplier estimate is 1.79, the feature-only multiplier is 2.19, and the multiplier for the combined use of feature and display is 3.53 (not shown). This suggests that feature provides a greater sales impact than display at all discount levels. Thus, the parametric results for the multipliers obtained from semiparametric model (1) only represent (weighted) average effects. Figure 5.9 suggests the display-only effect is often larger than the feature-only effect. The reason why feature has a higher constant multiplier in equation (1) is due to the distribution of the price indices for both promotion signals. Of the 103 display-only observations, 44 concern price indices equal to one, whereas of the 73 feature-only observations, just eight concern price indices equal to one. Figure 5.9 shows that for a price index of one, the curve for display-only is lower than the curve for feature-only, and this point is weighted far more heavily for display-only in the estimation of model (1).

index is below 0.55. Thus the display-only deal effect shows both a larger relative increase (5.5 versus 2.0) and a larger absolute increase (900 versus 400). Figure 5.9 shows there is a substantively meaningful cross-over interaction effect between these variables. Even Figure 5.9.a shows more price sensitivity than 4b; with neither feature, nor display, sales increase from 100 (at a price index of 1.0) to 800 (index below 0.55), for a relative increase of 8.0, and an absolute increase of 700.

6. Conclusions

The key contributions of this research are the following. We provide empirical support for threshold levels and saturation points: almost all own- and cross-item deal effect curves indicate the existence of threshold effects, and most curves indicate saturation effects as well. We obtain this support with a methodology that can be adapted to flexibly estimate promotion signal effects. Indeed we obtain a cross-over interaction in the deal effect between display and feature advertising. For the tuna item, feature has a greater effect at low discounts, while display has the stronger effect at high discounts.

Our semiparametric model accommodates full flexibility for the deal effect curves while it incorporates the effects of other predictors parametrically. This approach addresses Rust's (1988) call for methods to address the dimensionality issue. It is easy to use, and it provides users with a multiple-regression-like analysis that accommodates nonlinearity and interactions in a flexible manner in the nonparametric part. We accommodate interaction effects, both between price discounts of different items and between price discounts of one item and the separate or combined use of feature and display for the same item. The latter interactions were estimated through a separate data partitioning scheme.

The comparison of our semiparametric model with benchmark parametric models showed that (1) the parametric model generally overstates the effects of the smallest and largest discounts and understates the effects of intermediate discounts; (2) the semiparametric model has a better fit for all items and better average forecast performance in each of the three categories; and (3) the bias reduction due to the semiparametric model is relatively more substantial than the variance reduction.

Once a flexible deal effect curve has been identified with our method one could try to accommodate the specific nonlinearities found by a nonlinear parametric model. With new data one could also perform standard specification tests. However, this can be done only if the true (deal) curves belong to a parametric family (Briesch, Chintagunta, and Matzkin 1997). Our results also show complex interaction effects

which further decrease the potential for a parametric specification to fit the data. In addition, as data become more numerous, the flexibility afforded by nonparametric and semiparametric estimation will become more valuable.

Further research is needed to determine, given the curse of dimensionality, whether the flexibility should be concentrated in own- and cross-deal effects, as in model (1), or in separate own-deal effects for alternative promotion signals, as in Figure 5.9. Another fruitful area, we believe, is the decomposition of the promotional sales effect separately for different price discount levels. Incremental volume for an item at a given store due to a promotion may come from at least four sources: (1) other retailers (store switching), (2) other items (item or brand switching), (3) purchase and/or quantity acceleration (stockpiling), and (4) increased consumption (Blattberg and Neslin 1990, p. 353). The percentage contribution of each of these sources to a sales effect may well depend on the discount level. Indeed, the shape of the deal effect curve should depend on the nature and amount of the contribution of each of these sources at different price cut levels.

Appendix 1: Kernel method

¹⁰Consider $q = p + 1$ economic variables (Y, X') where Y is the criterion variable and X is a $p \times 1$ vector of predictors. These q variables are completely determined by their unknown joint density $f(y, x_{(1)}, x_{(2)}, \dots, x_{(p)}) = f(y, x)$ at the points y, x . Then the nonparametric regression model can be expressed as:

$$Y = m(x) + U, \quad (5)$$

where $m(x) = E(Y|X = x)$ provided $E|Y| < \infty$, and U represent errors. The term nonparametric is used because $m(x)$ has no parametric functional form.

If we have data (y_i, x'_i) ($i = 1, \dots, N$), where x_i is a $p \times 1$ vector of variables and y_i is a scalar, then from (5)

$$y_i = m(x_i) + u_i, \quad (6)$$

where the error term u_i has the properties $E(u_i|x_i) = 0$ and $E(u_i^2|x_i) = \sigma^2(x_i)$. We also assume that the u_i 's are independent.

The aim of nonparametric estimation of $m(x)$ in (6) is to approximate $m(x)$ arbitrarily close, given a large enough sample. This can be accomplished by noting that the estimation of $m(x)$ implies the estimation of the population mean of Y when $X = x$. This is given by a general class of linear nonparametric estimators, essentially the weighted mean of y_i , as

$$\hat{m}(x) = \sum_{i=1}^N w_i(x) y_i, \quad (7)$$

where $w_i = w(x_i, x)$ represents the weight assigned to the i -th observation y_i , which depends on the distance of x_i from the point x . Usually, the weight is high if the distance is small and low if the distance is large.

A large class of nonparametric estimators contains special cases of (7). They differ mainly with respect to the choice of w_i . For the Nadaraya (1964) and Watson (1964) estimator the choice of w_i in $\hat{m}(\cdot)$ is

$$w_i(x) = K\left(\frac{x_i - x}{h}\right) / \sum_{j=1}^N K\left(\frac{x_j - x}{h}\right), \quad (8)$$

where $K(\cdot)$ is a "Kernel function", and h the "bandwidth parameter". The Kernel $K(\cdot)$ is a continuous, bounded function which integrates to one, $\int K(u)du = 1$. Bandwidth h controls the amount of smoothing imposed on the data: the larger h the

10 This appendix is based on Ullah and Vinod (1993).

smoother the curve.

Appendix 2: Estimation procedure for semiparametric model

The estimation procedure we use is described in Lee (1996). The semiparametric model (1) can be simplified as follows:

$$y = m(z) + x'\beta. \quad (9)$$

Taking $E(\cdot|z)$ on (9), we obtain:

$$E(y|z) = m(z) + E(x|z)'\beta. \quad (10)$$

Subtracting (10) from (9) gives:

$$y - E(y|z) = \{x - E(x|z)\}'\beta + u \Leftrightarrow y = E(y|z) + \{x - E(x|z)\}'\beta + u, \quad (11)$$

which excludes $m(z)$. Here the deterministic component of y is decomposed into two parts: one is the effect of z on y , and the other is the effect of x on y net of z . The three-step estimation procedure first eliminates the influence of x on y and determines next the influence of z on y :

Step 1 Estimate $E(y|z)$ and $E(x|z)$, respectively, as:

$$E_N(y|z_i) \equiv \sum_{j \neq i}^N K((z_j - z_i)/h_{y,1}) \cdot y_j / \sum_{j \neq i}^N K((z_j - z_i)/h_{y,1}), \quad (12)$$

and

$$E_N(x|z_i) \equiv \sum_{j \neq i}^N K((z_j - z_i)/h_{x,1}) \cdot x_j / \sum_{j \neq i}^N K((z_j - z_i)/h_{x,1}). \quad (13)$$

Step 2 Substitute (12) and (13) into (11) to define a new criterion variable $y - E_N(y|z)$ and predictors $x - E_N(x|z)$. Apply Least Squares Estimation (LSE) to the new model $y - E_N(y|z) \cong \{x - E_N(x|z)\}'\beta + u$ to get:

$$b_N = \left[\sum_i \{x_i - E_N(x|z_i)\} \cdot \{x_i - E_N(x|z_i)\}' \right]^{-1} \cdot \left[\sum_i \{x_i - E_N(x|z_i)\} \cdot \{y_i - E_N(y|z_i)\} \right] \quad (14)$$

The least squares estimate b_N has the following asymptotic distribution (which is used to compute standard errors):

$$\sqrt{N}(b_N - \beta) \stackrel{d}{=} N(0, A^{-1}BA^{-1}),$$

$$A = {}^p A_N \equiv \frac{1}{N} \sum_i \{x_i - E_N(x|z_i)\} \cdot \{x_i - E_N(x|z_i)\}',$$

$$B = {}^p B_N \equiv \frac{1}{N} \sum_i \{x_i - E_N(x|z_i)\} \cdot \{x_i - E_N(x|z_i)\}' \cdot v_i^2,$$

$$v_i \equiv \{y_i - E_N(y|z_i)\} - \{x_i - E_N(x|z_i)\}'b_N.$$

The usual intercept term in LSE disappears due to the mean subtraction. Thus the intercept is not estimable with (14) and is absorbed in $m(\cdot)$.¹¹

Step 3 The estimation of $m(z)$ is achieved by nonparametric regression of $\tilde{y} = y - x'b_N$ on z , for $E(y - x'\beta|z) = m(z)$ in (10):

$$\hat{m}(z) = \frac{\sum_{i=1}^N K\left(\frac{z_i - z}{h_{y,3}}\right) \tilde{y}_i}{\sum_{j=1}^N K\left(\frac{x_j - x}{h_{y,3}}\right)}. \quad (15)$$

The computation time of the model varies, given bandwidth parameters, from 11 seconds (tuna), 39 seconds (Dutch food product), to 52 seconds (beverage) on a Pentium II 266 Mhz. personal computer. The Gauss code is available from the first author.

11 Kernel estimators allow for heteroscedastic errors (Lee 1996, p. 151), i.e. the variance of the error may depend on the predictors. The Kernel method (like the standard regression model) does not allow for errors to depend on the predictors due to endogeneity problems.

Appendix 3: Model operationalizations for semiparametric model

We address three model operationalizations: the choice of the Kernel, the choice of the bandwidth parameter, and the construction of confidence intervals.

Kernel choice

For a nonparametric model the Kernel function $K(z)$ must be specified. It turns out that for large samples it makes very little difference what the Kernel function is (Silverman 1986). Often it can be chosen on the basis of tractability or convenience. Like Rust (1988) and Abe (1995) we use the multivariate standard normal distribution (with independent components):

$$K(z) = K(z_1, \dots, z_k) = (2\pi)^{-\frac{1}{2}k} \prod_{i=1}^k e^{-\frac{1}{2}z_i^2}. \quad (16)$$

Bandwidth parameter choice

The bandwidth parameter h (see equation (8)) determines how fast the value of $K(\cdot)$ falls as the absolute value of the argument increases (Abe 1995). It affects the smoothness of the resulting response function, and its value involves the tradeoff between bias and variance of the estimator. The value of h can be determined on the basis of a criterion such as the minimization of the mean squared error.¹²

We distinguish three different bandwidth selection techniques:

- *Lee's rule of thumb.* An approximately mean integrated square error minimizing h (see Lee 1996, p. 154) is given by $h = N^{-1/(k+4)}$.
- *Least squares cross validation (LSXV).* Define a criterion variable y and a predictor ($k \times 1$) variable x , and the nonparametric regression function $y = m(x) + u$. The "leave-one-out" Kernel estimator for $m(x_j)$ (j -th observation of x) is computed for all $j = 1, \dots, N$ by:

$$\hat{m}_j(x_j) \equiv \sum_{i \neq j}^N K((x_i - x_j)/h) y_i / \sum_{i \neq j}^N K((x_i - x_j)/h). \quad (17)$$

Then h is chosen to minimize the cross-validation (CV) criterion (see Härdle and Marron 1985):

$$(1/N) \sum_{j=1}^N \{y_j - \hat{m}_j(x_j)\}^2. \quad (18)$$

¹² See Abe (1995, p. 306-7) for a more elaborate discussion about the choice of h .

Note that the “cross-validation sample” here is the estimation sample excluding observation j .

- *Subjective approach.* No single value of h can be universally regarded as “optimal” (Abe 1995). The problem is akin to the researcher’s task of specifying a model structure in parametric modeling. In nonparametric regression the need to pick the bandwidth parameter involves some arbitrariness. If the dimension of x is small, the best way to choose h seems to be trial and error (Lee 1996, p. 156).

We note that the estimation procedure requires the choice of h in two different steps. First in the nonparametric estimation of $E_N(y|z_i)$ (h defined as $h_{y,1}$, see (12)) and $E_N(x|z_i)$ (h defined as $h_{x,1}$, see (13)) in Step 1, and next in the nonparametric regression of $y - x'b_N$ on z (h defined as $h_{y,3}$), i.e. the estimation of $m(z)$ in Step 3.

For both first-step bandwidth parameters we employed automated techniques, since the model results appeared to be relatively insensitive to this bandwidth parameter. For $h_{y,1}$ we applied the LSXV criterion. For $h_{x,1}$ we used Lee’s rule of thumb for all regressors in x .

We varied the third-step bandwidth parameter $h_{y,3}$ to determine its influence on the deal effect curve. We first based this bandwidth parameter on the LSXV-criterion. For some items, the resulting own-item and cross-item deal effect curves had too much variation. To smooth those curves, we multiplied the LSXV-parameter by 1.5.

In practice, one may use k different bandwidths, say h_1, \dots, h_k , because the k regressors in z have different scales. Although the use of regressor-specific bandwidths is more advantageous in principle, the task of choosing bandwidths becomes quite complex (Lee 1996, p. 154). Instead, we standardized the regressors and used one common bandwidth h for each regressor in z .¹³

Construction of confidence intervals

The asymptotic distribution of the estimate of the nonparametric regression function at some point x_0 ($\hat{m}(x_0)$) is given by:

$$(Nh^k)^{0.5}\{\hat{m}(x_0)-m(x_0)\} \stackrel{d}{=} N(\text{Asym.Bias}, V(u|x_0) \int K(z)^2 dz / f(x_0)), (19)$$

where Asym. Bias is the asymptotic bias of $\hat{m}(x_0)$ (see Lee 1996, p. 151), $V(u|x_0)$ is the (heteroscedastic) error variance, and $f(x_0)$ is the multivariate density of the regressors at point x_0 . As $f(x_0)$ increases, the variance decreases, since more data

13 We also standardized the x predictors for all models.

are available for $\hat{m}(x_0)$. The estimated 95 percent (pointwise) confidence interval for $\hat{m}(x_0)$ is given by:

$$\begin{aligned} & [\hat{m}(x_0) - 1.96\sqrt{(Nh^k)^{-1}\hat{V}(u|x_0)\int\widehat{K(z)^2dz/\hat{f}(x_0)}}, \\ & \hat{m}(x_0) + 1.96\sqrt{(Nh^k)^{-1}\hat{V}(u|x_0)\int\widehat{K(z)^2dz/\hat{f}(x_0)}}]. \end{aligned} \quad (20)$$

In (20), $\hat{V}(u|x_0) = V(y|x_0)$ is estimated by $E_N(y^2|x_0) - \{E_N(y|x_0)\}^2$, $\int\widehat{K(z)^2dz}$ is approximated by Monte Carlo integration, and $\hat{f}(x_0)$ is obtained by nonparametric density estimation (see Lee 1996, p. 152).

The confidence interval (20) for $\hat{m}(x_0)$ is asymptotically valid both for the one-step estimator of nonparametric regression models and for the three-step estimator of our semiparametric regression model. The convergence rate of step three is much slower than the convergence rate of steps one and two. Therefore, from an asymptotic point of view we can consider the parameter estimate for β as fixed in step three. As a result, we do not account for the variance of $\hat{\beta}$ in the confidence interval (20). Specifically, steps one and two of the semiparametric estimation sequence yield a $N^{\frac{1}{2}}$ consistent estimator of parameter vector β (Robinson 1988). Step three yields a $(Nh^k)^{\frac{1}{2}}$ consistent estimator for $m(x_0)$. If we use Lee (1996)'s rule of thumb (p. 154) for h , $h = N^{-1/(k+4)}$, the convergence rate is $N^{\frac{1}{2} \frac{4}{4+k}}$. For all k , this convergence rate dominates the convergence rate of $N^{\frac{1}{2}}$ for $\hat{\beta}$.

References

- Abe, Makoto (1995), "A Nonparametric Density Estimation Method for Brand Choice Using Scanner Data," *Marketing Science* 14 (issue 3, part 1 of 2), 300-325.
- Blair, Edward A. and E. Laird Landon jr. (1981), "The Effects of Reference Prices in Retail Advertising," *Journal of Marketing* 45 (Spring), 61-69.
- Blattberg, Robert C., Richard Briesch, and Edward J. Fox (1995), "How Promotions Work," *Marketing Science* 14 (issue 3, part 2 of 2), g122-g132.
- Blattberg, Robert C. and Scott A. Neslin (1990), *Sales Promotion, Concepts, Methods and Strategies*. Englewood Cliffs, New Jersey: Prentice Hall.
- Blattberg, Robert C. and Kenneth J. Wisniewski (1987), "How Retail Price Promotions Work: Empirical Results," Marketing working paper no. 42, Graduate School of Business, University of Chicago.
- Blattberg, Robert C. and Kenneth J. Wisniewski (1989), "Price-Induced Patterns of Competition," *Marketing Science* 8 (Fall), 291-309.
- Briesch, Richard A., Pradeep K. Chintagunta, and Rosa L. Matzkin (1997), "Nonparametric and Semiparametric Models of Brand Choice Behavior," Working Paper, Department of Marketing, University of Texas at Austin.
- Bucklin, Randolph E. and Sunil Gupta (2000), "Commercial Adoption of Advances in the Analysis of Scanner Data," *Marketing Science*, forthcoming.
- Bult, Jan Roelf and Tom Wansbeek (1995), "Optimal Selection for Direct Mail," *Marketing Science* 14, 378-94.
- Della Bitta, Albert J. and Kent B. Monroe (1980), "A Multivariate Analysis of the Perception of Value from Retail Price Advertisements," in *Advances in Consumer Research* 8, Monroe, Kent B. and Ann Arbor eds. MI: Association for Consumer Research.
- Foekens, Eijte W., Peter S.H. Leeflang, and Dick R. Wittink (1994), "A Comparison and an Exploration of the Forecasting Accuracy of a Nonlinear Model at Different Levels of Aggregation," *International Journal of Forecasting* 10, 245-261.
- Goldberger, Arthur S. (1968), "The Interpretation and Estimation of Cobb-Douglas Functions," *Econometrica* 35 (July-October), 464-472.

- Gupta, Sachin, Pradeep K. Chintagunta, Anil Kaul, and Dick R. Wittink (1996), "Do Household Scanner Data Provide Representative Inferences From Brand Choices: A Comparison With Store Data," *Journal of Marketing Research* 33 (November), 383-398.
- Gupta, Sunil and Lee G. Cooper (1992), "The Discounting of Discounts and Promotion Thresholds," *Journal of Consumer Research* 19 (December), 401-411.
- Härdle, Wolfgang (1990), *Applied Nonparametric Regression*. (Econometric Society Monographs, No. 19.) Cambridge: Cambridge University Press.
- Härdle, Wolfgang and Oliver Linton (1994), "Applied Nonparametric Methods," in *Handbook of Econometrics. Vol. 4*, Engle, R.F. and D.L. McFadden eds. Amsterdam: Elsevier Science B.V.
- Härdle, Wolfgang and J.S. Marron (1985), "Optimal Bandwidth Selection in Nonparametric Regression Function Estimation," *Annals of Statistics* 13, 1465-1481.
- Hastie, T., and R. Tibshirani (1990), *Generalized Additive Models*. New York: Chapman and Hall.
- Hruschka, Harald (1998), "Market Share Models with Semi-Parametric Brand Attractions," *Working Paper* nr. 316, University of Regensburg.
- Inman, Jeffrey, Leigh McAlister, and Wayne D. Hoyer (1990), "Promotion Signal: Proxy for Price Cut?," *Journal of Consumer Research* 17 (June), 74-81.
- Judge, George G., W.E. Griffiths, R. Carter Hill, Helmut Lütkepohl, and Tsoung-Chao Lee (1985), *The Theory and Practice of Econometrics*, 2nd edition. Singapore: John Wiley and Sons.
- Kalyanam, Kirithi and Thomas S. Shively (1998), "Estimating Irregular Pricing Effects: A Stochastic Spline Regression Approach," *Journal of Marketing Research* 35 (February), 16-29.
- Lee, Myong-jae (1996), *Methods of Moments and Semiparametric Econometrics for Limited Dependent Variable Models*. New York: Springer-Verlag.
- Leeflang, Peter S.H. and Alex J. Olivier (1985), "Bias in Consumer Panel and Store Audit Data," *International Journal of Research in Marketing* 2, 27-41.
- Mayhew, Glenn E. and Russell S. Winer (1992), "An Empirical Analysis of Internal and External Reference Prices Using Scanner Data," *Journal of Consumer Research* 19 (June), 62-70.

- Mobley, Mary F., William O. Bearden, and Jesse E. Teel (1988), "An Investigation of Individual Responses to Tensile Price Claims," *Journal of Consumer Research* 15 (September), 273-279.
- Nadaraya, E.A. (1964), "On Estimating Regression," *Theory Probability Applications* 9, 141-142.
- Powell, James L. (1994), "Estimation of Semiparametric Models," in *Handbook of Econometrics Vol. 4*, Engle, R.F. and D.L. McFadden eds. Amsterdam: Elsevier Science B.V.
- Robinson, P.M. (1988), "Root-N-Consistent Semiparametric Regression," *Econometrica* 56 (July), 931-954.
- Rust, Roland T. (1988), "Flexible Regression," *Journal of Marketing Research* 25 (February), 10-24.
- Scott, David W. (1992), *Multivariate Density Estimation: Theory, Practice and Visualization*. John Wiley & Sons.
- Sethuraman, Raj (1996), "A Model of How Discounting High-Priced Brands Affects the Sales of Low-Price Brands," *Journal of Marketing Research* 33 (November), 399-409.
- Silverman, B.W. (1986), *Density Estimation for Statistics and Data Analysis*. (Monographs on Statistics and Applied Probability 26) London: Chapman & Hall.
- Ullah, A. and H.D. Vinod (1993), "General Nonparametric Regression Estimation and Testing in Econometrics," in *Handbook of Statistics, Vol. 11*, Maddala, G.S., C.R. Rao, and H.D. Vinod eds. Amsterdam: Elsevier Science Publishers B.V.
- Watson, G.S. (1964), "Smooth Regression Analysis," *Sankhyā Ser. A* 26, 359-372.
- Wedel, Michel and Peter Leeflang (1998), "A Model for the Effects of Psychological Pricing in Gabor-Granger Price Studies," *Journal of Economic Psychology* 19, 237-60.
- Wisniewski, Kenneth J. and Robert C. Blattberg (1983), "Response Function Estimation Using UPC Scanner Data," *Proceedings of ORSA-TIMS Marketing Science Conference*, Zufryden, Fred S. ed.
- Wittink, Dick R., Michael J. Addona, William J. Hawkes, and John C. Porter (1988), "The Estimation, Validation, and Use of Promotional Effects Based on Scanner Data," Working Paper, Johnson Graduate School of Management, Cornell University, Ithaca.